Enzyme-Catalysed Carbohydrate Biosynthesis and Modification

A thesis submitted for the degree of

Doctor of Philosophy

Alexandra B. Merkel

November 2004

Contents

I. Abbreviations

Amino Acids:

A – Ala	Alanine	M – Met	Methionine
C – Cys	Cysteine	N – Asp	Asparagine
D – Asp	Aspartate	P – Pro	Proline
E – Glu	Glutamate	Q – Gln	Glutamine
F – Phe	Phenylalanine	R – Arg	Arginine
G – Gly	Glycine	S – Ser	Serine
H – His	Histidine	T – Thr	Threonine
I – Ile	Isoleucine	V – Val	Valine
K – Lys	Lysine	W – Trp	Tyrptophan
L – Leu	Leucine	Y – Tyr	Tyrosine

Nucleic acids:

A	Adenosine
C	Cytosine
G	Guanine
T	Thymidine
U	Uracil

ΔG	Change in Gibbs Free Energy
ACB	Amylose column buffer
ADSC	Area Detector Systems Corporation
AMU	atomic mass unit
ATP	Adenosine triphosphate
au	Asymmetric unit
BCA	Bicinchoninic Acid
BSA	Bovine Serum Albumin
BOG	n-Octyl-β-D-glucopyranoside

CCD	Charge coupled device
CCP4	Collaborative Computing Project Number 4
CD	Circular Dichroism (Spectroscopy)
CDC	United States Centres for Disease Control and Prevention
CNS	Crystallography and NMR system
CV	Column Volume
Da	Dalton
DCM	Dichloromethane
DHAP	Dihydroxyacetone phosphate
DEAE	Diethylaminoethyl
DERA	2-deoxyribose-5-phosphate Aldolase
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease I
dTDP	deoxythymidine 5'-diphosphate
dTMP	deoxythymidine 5′-monophosphate
DTT	Dithiothreitol
EC	Enzyme Commission
ED	Entner-Doudoroff pathway
EDTA	Ethylenediaminetetra acetic acid
EMP	Embden-Meyerhof-Parnas pathway
ESI	Electrospray Ionisation (-TOF) mass spectrometry
ESP	ESPript 2.2 (http://prodes.toulouse.inra.fr/ESPript/cgi-bin/ESPript.cgi)
ESRF	European Synchrotron Radiation Facility
FBP	Fructose-1,6-bisphosphate
FMN	Flavin mononucleotide
G3P	Glyceraldehyde-3-phosphate
G6P	Glucose-6-phosphate
GC MS	Gas chromatography/Mass spectrometry
GlcNAc	*N*-acetyl glucosamine
H-bond	Hydrogen bond

HPLC	High Performance Liquid Chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
K_{cat}	Enzyme turnover number
KDPG	2-keto-3-deoxy-6-phosphogluconate
KHG	2-keto-4-hydroxyglutarate
K_m	Michaelis Menton constant (Substrate affinity at ½ V_{max})
LB	Luria Bertani (growth medium)
MALDI	Matrix-assisted laser desorption ionisation (mass spectrometry)
MCS	Multiple coning site
MIC	Minimum inhibitory concentration
MPD	2-methyl-2,4-pentadiol
MRSA	Methicillin-resistant *Staphylococcus aureus*
MS	Mass Spectrometry
MW	Molecular weight
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NEB	New England BioLabs
NTP	Nucleotide triphosphate
OD	Optical Density
OPOE	n-octylpolyoxyethylene
PAGE	Polyacrylamide gel electrophoresis
PBP	Penicillin-binding protein
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
Pfu	*Pyrococcus furiosus* (DNA polymerase)
PMSF	Phenylmethylsulfonyl flouride
r1	*Chi1* rotamer
Rf	R_{factor}
Rf/sig	R_{factor} / sigma
Rms	Root mean square

Rmsd	Root mean square deviation
RNase	Ribonucease I
SAM	S-adenosyl methionine
SDR	Short chain dehydrogenase / reductase enzyme
SDS	Sodium dodecylsulphate
SeMet	Selenomethionine
TB	Terrific Broth or Tuberculosis (depending on context)
TCA	Tricarboxylic acid cycle
TDP	Thymidine-diphosphate
TEV	Nuclear Inclusion from Tobacco etch virus
TFA	Triflouroacetic acid
TIR	Translation initiation regions
TLS	Translation, libration and screw-rotation
TMP	Thymidine monophosphate
TOF	Time of flight
TYE	Tryptone yeast extract
VISA	Vancomycin-insusceptible *S. aureus*
V_m	Matthews coefficient
V_{max}	Maximum velocity of an enzymatic reaction following Michaelis Menton kinetics
VRE	Vancomycin-resistant Enterococci
VRSA	Vancomycin-resistant *S. aureus*
WHO	World Health Organisation

II. List of Figures

III. List of Tables

IV. Declarations

I, Alexandra Brigitta Merkel, herby certify that this thesis, which is approximately, 50,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date.17/12/04... Signature of candidate. .....

I was admitted as a research student in October 2001 as a candidate for the degree of Ph.D. in October 2002; the higher study for which this is a record was carried out in the University of St. Andrews between 2001 and 2004.

Date.17/12/04... Signature of candidate. .....

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D. in the University of St. Andrews and the candidate is qualified to submit this thesis in application for that degree.

Date.17/12/04. Signature of supervisor.....

In submitting this thesis to the University of St. Andrews In understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research workers.

Date.17/12/04... Signature of candidate.. .....

V. Abstract

Glycopeptides are regarded as the last line of defence antibiotics. Since their peptide backbone has no biological activity, these compounds depend on attached carbohydrates for their antibacterial potency, as well as for their selectivity and specificity.

In *Amycolatopsis orientalis*, five enzymes are responsible for the synthesis of the L-epivancosamine deoxysugar from dTDP-glucose in the chloroeremomycin biosynthesis.

This thesis discusses the structural characterisation of two of these enzymes: EvaD, a 5-epimerase; and EvaA, a 2,3-dehydratase. Structural data on EvaD crystals have been obtained to 1.5Å resolution for apo- and to 1.4Å resolution for a dTMP co-complex. The M131F/L135A mutant of EvaD was also crystallised and data to 2.1Å collected. Kinetic analysis and deuterium incorporation studies with wild-type and mutant EvaD allowed the identification of all residues involved in catalysis and showed that this enzyme is naturally a mono-epimerase, despite being structurally similar to RmlC-type epimerases. The main reason for this anomaly is the orientation of the catalytic acid, Tyr133.

EvaA and several homologues from other Actinomycetales have been cloned, expressed and purified; and the activity of EvaA has been confirmed. Initial crystallisation conditions have been identified for EvaA, but no diffraction has been obtained.

A second part of this thesis deals with the investigations into the reaction mechanism of KDPG aldolase from *E. coli*. This enzyme is useful for the enzymatic synthesis of C-C bonds; however, its substrate profile is too limited for chemical synthesis. To engineer this enzyme to accept non-natural substrates, the reaction mechanism and basis for selectivity have to be understood. Prior to this study, the aldolase had been crystallised, but citrate contained in these conditions bound to the active site preventing co-crystallisation and soaking experiments. Several new crystal forms without citrate were identified facilitating substrate soaks; however, no clear results were obtained.

VI. Acknowledgements

First of all I need to thank my supervisor Jim Naismith for his guidance and support throughout these years in St. Andrews. Thanks for his encouragement through some tough times and for sharing his enjoyment over my little victories in the lab…

Thank you to everyone who has helped me throughout my PhD: Myriam Asunction was the first person who taught me the difference between BL21 and pET22; Huanting Liu and Louise Major had to carry on from that; James Errey for his illegible scribbles – and then answering all my questions. Thanks to all the chemists, especially Steven Patterson for investing far too much time in my chemistry education. Thanks to the mass spectrometry people, Catherine Botting, Robin Antrobus and Alex Houston, to Sharon Kelly and Nic Price at Glasgow for their CD Spectroscopy work and to Petr Beir in St. Andrews for helping me with the GC-MS analysis.

Gracias to Francesco and the remaining staff at Costa coffee for their hospitality. A lot of people have been involved in proof-reading and commenting on sections of this dissertation – and apparently they did not fall asleep (some of the time): Adrian Merkel, Charlie Ivamy, Charlotte Ryan, Jane Potter, Louise Major and Vicki Bamford, I owe you at least a couple of drinks – and let's blame Microsoft for the remaining typos.

A big thank you also to my parents for their support throughout everything and putting up with not seeing me - ever…

Thanks also to Charlie, Charlotte, Christina, Colleen and Kaitlyn for being extremely understanding as well as tough with me under trying conditions; it is hard to express, what I owe you guys.

Thanks to everybody else I have spent large parts of my spare time with and who have put up with me, especially Allison, Charlotte R., Charlotte S., Jocelyn, Vicki, Lou, Andy and Walter and everybody else I have not mentioned by name.

Thanks to everyone in the Taylor and Naismith groups for being such a good bunch and everyone in the corridor for supplying the necessary distraction to get me through the day...

Wenn das Gehirn des Menschen so einfach wäre, dass wir es verstehen könnten, dann wären wir so dumm, dass wir es doch nicht verstehen würden.

If the human mind was simple enough to understand, we'd be too simple to understand it.

Emerson Pugh

A man who is nobel does not pretend to be nobel, any more than an eloquent man feigns eloquence. When a man exaggerates his qualities it is because something is lacking in himself.

al-Jazir

To my parents Peter and Brigitte…

Chapter 1

Introduction

1.1 Carbohydrate Use in Nature and Industry

Carbohydrates are aldehyde or ketone compounds with multiple hydroxyl groups attached of the general formula $C_m(H_2O)n$. They make up most of the organic matter on earth (Stryer, 1995a) and play multiple roles in all forms of life. They serve as energy stores, fuels and metabolic intermediates; they form part of the structural framework of the nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA); they are part of the structural elements in the cell walls of bacteria and plants and in the exoskeletons of arthropods, and they are linked to many proteins and lipids, where they regulate cell-recognition processes.

Cellular respiration is the sum total of the biological reactions in which organic material is oxidised to simpler compounds in the living organism (McIlroy, 1967). In aerobic respiration the substrate acted upon is often completely oxidised to carbon dioxide and water, and most frequently the primary substrates are either glucose or fructose. Initially, the carbohydrate is broken down to pyruvate, which can then be further oxidised to carbon dioxide and water. Glycolysis for the generation of energy is a nearly universal pathway in biological systems. It is defined as 'the sequence of reaction steps that converts glucose into pyruvate with the concomitant production of a relatively small amount of adenosine triphosphate (ATP)' (Stryer, 1995b).

In animal and plant tissue, the major glycolysis pathway is the Embden-Meyerhof-Parnas pathway (EMP) (Figure 1.1). Elucidation of the full pathway took place in the first half of the twentieth century. In 1897 Hans Buchner and Eduard Buchner concluded that fermentation could occur outside living cells, when they observed that cell-free yeast extract was capable of fermenting added sucrose. This disproved the conviction of that time that fermentation could only occur in living cells as had been postulated by Pasteur. In 1905 Arthur Harden and William Young established that organic phosphate was crucial to glycolysis, and they went on to isolate fructose-1,6-bisphosphate as one of the pathway intermediates. The pioneering work, carried out by Gustav Emden, Otto Meyerhof, Carl Neuberg, Jacob Parnas, Otto Warburg, Gerty Cori and Carl Cori, meant that the full EMP pathway was elucidated by 1940. Bacteria preferentially use the Entner-Doudoroff pathway (ED) (Figure 1.2), which branches off at the glucose-6-phosphate intermediate of the EMP pathway, as this is converted to glucono-δ-lactone-6-phosphate, rather than glucose-1-phosphate.

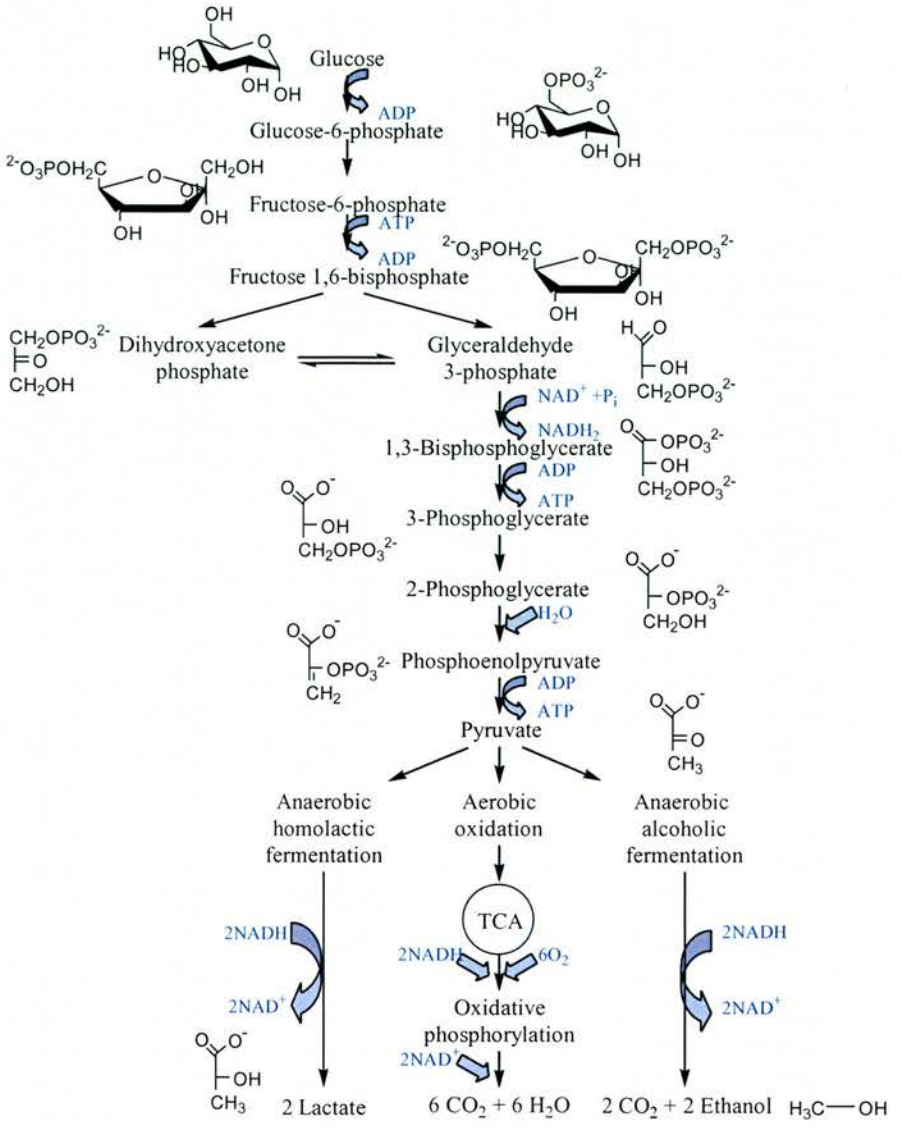


Figure 1.1 – Embden-Meyerhof-Parnas Pathway: (glycolysis) for the breakdown of carbohydrates (adapted from Stryer, 1995).

Carbohydrate metabolism is very important to cells, as it converts complex sugars from the surrounding environment to simpler products, which serve them as the main carbon source for biosynthesis, as well as, the main energy source. The most abundant monosaccharide in nature is D(+)-glucose (=dextrose), which occurs free in plant juices, blood and cerebrospinal fluid, and as a component of oligosaccharides, such as sucrose, which is a combination of D-glucose and D-fructose.

The transformation of one molecule of glucose to two molecules of pyruvate yields a net synthesis of three molecules of ATP and a net transfer of two pairs of protons to the respiratory chain. Aerobic respiration refers to the fate of pyruvate after glycolysis: in this case it is metabolised by the enzymes of the tricarboxylic acid cycle (TCA cycle). These catalyse several decarboxylation and oxidation steps to yield much

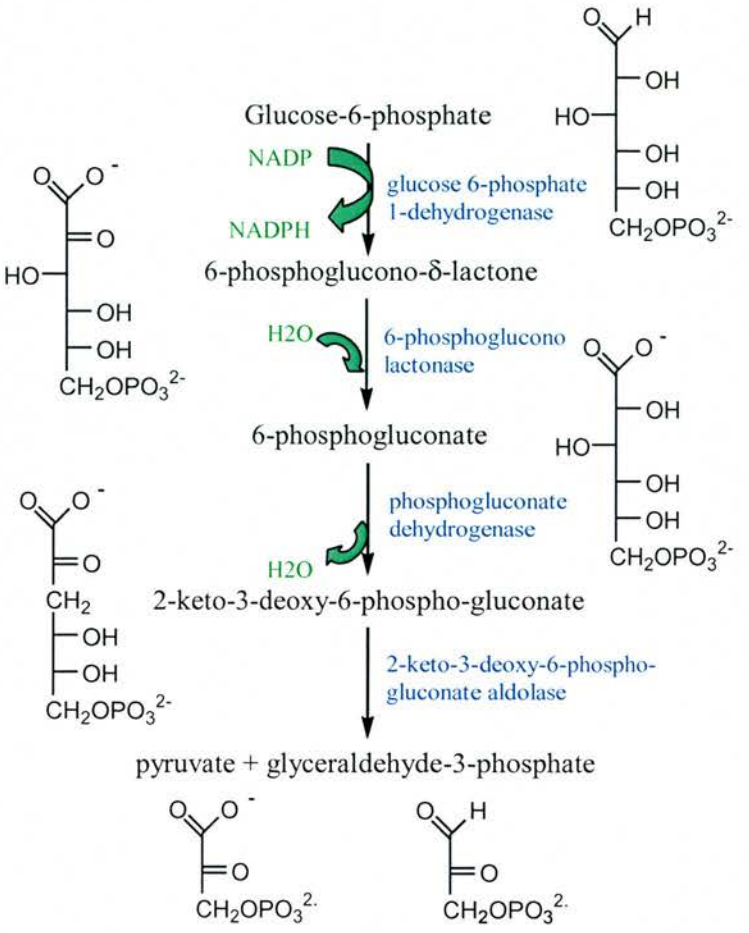


Figure 1.2 - Entner-Doudoroff Pathway: for the breakdown of hexoses, as found in many bacteria (adapted from Stryer, 1995).

more energy than during the anaerobic breakdown of pyruvate. Anaerobically, the pyruvate may be fermented to ethanol and carbon dioxide or to lactic acid. In the former pathway, pyruvate is decarboxylated to acetaldehyde and then reduced to ethanol. In yeast species and under oxygen-limited conditions, such as working muscles, the reduction of pyruvate to lactic acid is catalysed by lactate dehydrogenase. An alternative metabolic route to the EMP and ED pathways is the pentose phosphate pathway (also known as monophosphate shunt or phosphogluconate pathway) (Figure 1.3). It may function in conjunction with the other pathways, depending on tissue type or organism and its biological state. For example, mammary tissue oxidises glucose preferentially by way of the pentose phosphate pathway during lactation. Certain bacteria use this method rather than the ED pathway for their energy production during the generation of secondary metabolites. The net result of six turns of this cycle is the complete oxidation of one glucose molecule to carbon dioxide and the reduction of twelve molecules of $NADP^+$ (McIlroy, 1967).

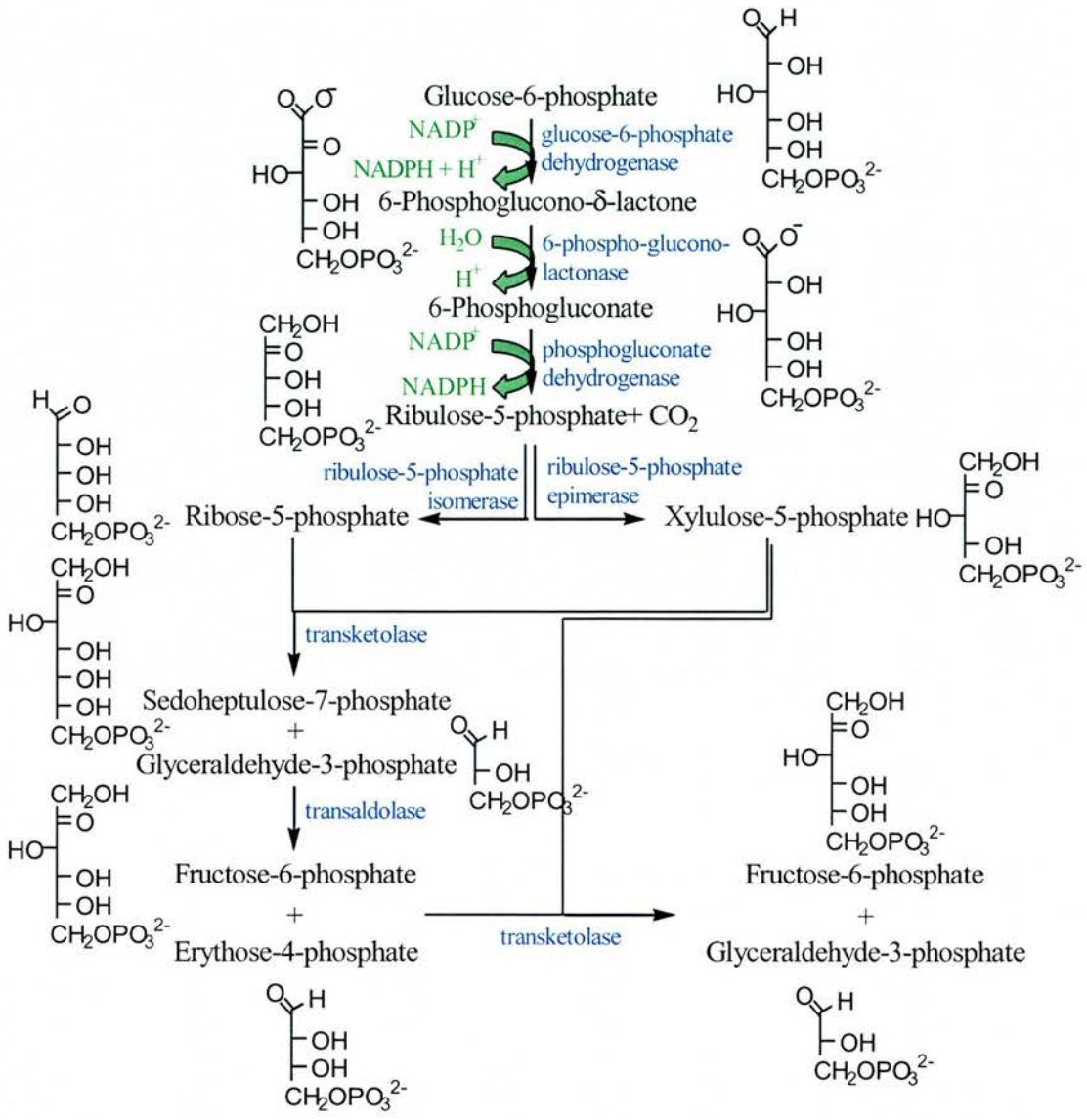


Figure 1.3 - Pentose Phosphate Pathway: also known as hexose monophosphate shunt or phosphogluconate pathway; it generates NADPH by the oxidation of G6P via an alterntive pathway to glycolysis (adapted from Stryer, 1995).

The laboratory synthesis of carbohydrate material was first achieved by Emil Fischer in 1889 (Fischer, 1890). When Fischer treated glycerol with nitric acid or bromine, or allowed mild alkali to act upon formaldehyde, he obtained optically inactive reducing substances, which he called 'acroses'. From these mixtures he isolated glucosazone and, using this product as the starting material, he carried out a series of syntheses, which are illustrated in Figure 1.4. α-acrose might have been DL-fructose, DL-glucose, or DL-mannose, since all potentially yield the same osazone and regenerate fructose. The identity was established by yeast fermentation since this yielded a dextro (+) rotary solution. Had the original solution been DL-glucose or DL-mannose, the D(+) would have fermented giving a laevo (-) rotary solution from the unchanged L(-) form, whereas DL-fructose would give a dextro-rotatory solution because only laevo rotary D(-) fructose is fermentable (McIlroy, 1967).

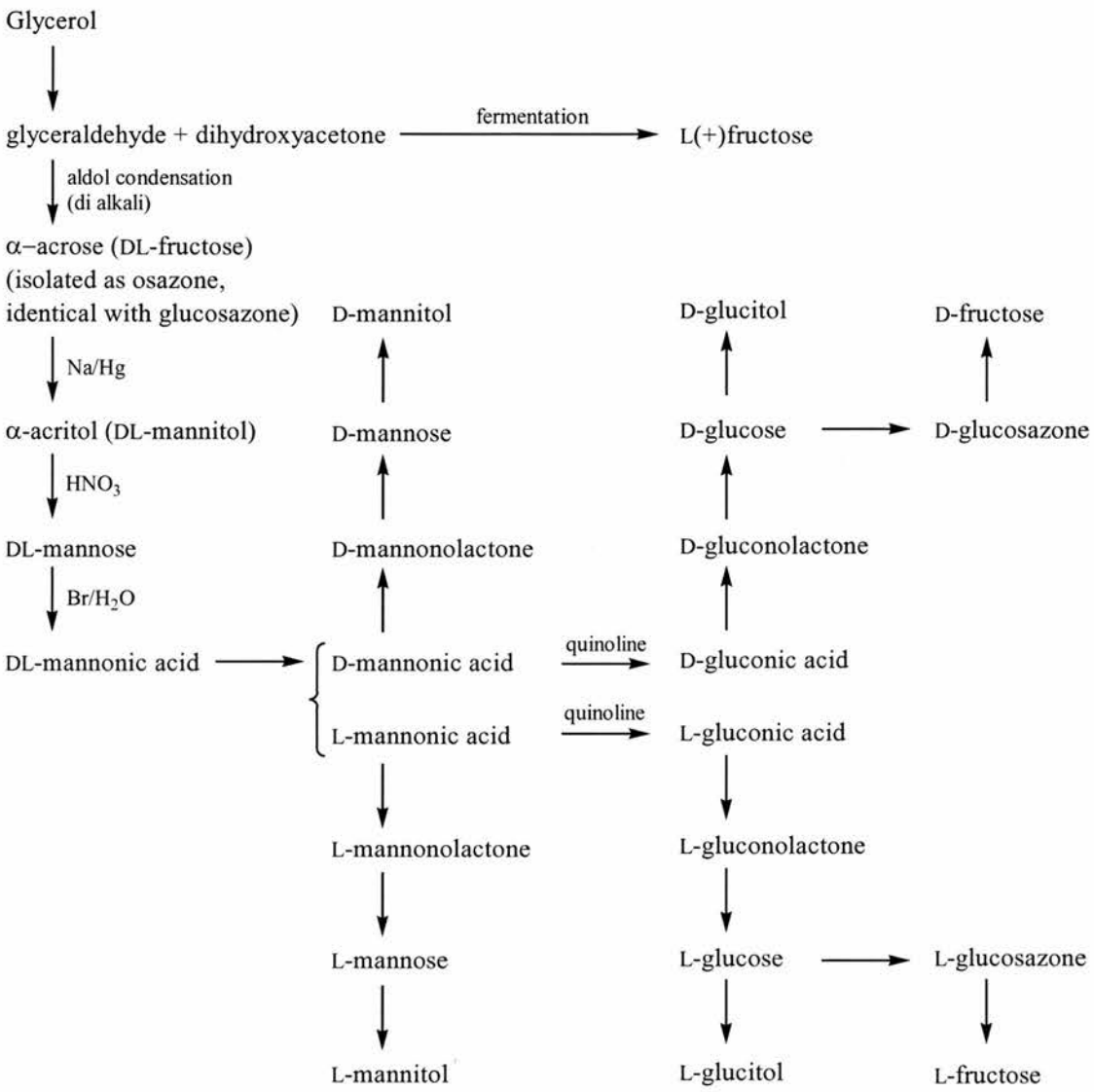


Figure 1.4 - Series of Carbohydrate Syntheses: from the isolated starting material glucosazone, as published by Fischer, 1890.

Prior to 1942, the only method known for extending the chain length of a monosaccharide was the cyanohydrin synthesis, which had been discovered by Kiliani, 1895. As Figure 1.5 shows, a new asymmetric carbon is introduced and two isomers are formed in this synthetic pathway. The aldonic acid products can be isolated as they readily form lactones; Fischer reduced these lactones with sodium amalgam in acid solution and obtained the corresponding aldoses. In this way he prepared glucoheptoses, nonoses and a decose.

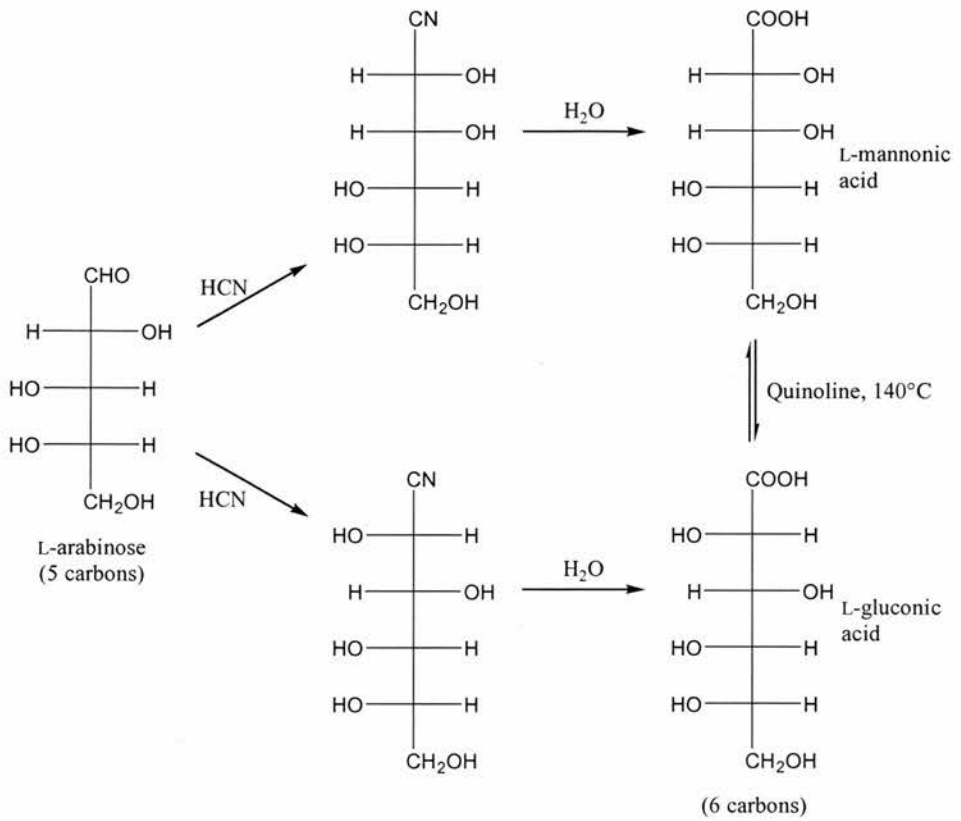


Figure 1.5 - Ascent of the Sugar Series: an example of the cyanohydrin synthesis as discovered by Kiliani in 1895.

Later on, other methods for extending carbohydrate chain lengths were established, such as the nitromethane synthesis by Sowden and Fischer, 1947 and the diazomethane synthesis.

Obviously, carbon-carbon bond formation is at the heart of organic synthesis. For complex multifunctional target structures, that are difficult to prepare by conventional means, asymmetric framework construction by enzymatic catalysis is an attractive alternative to standard chemical methods. It uses mild reaction conditions, does not require the protection of sensitive or reactive functional groups and it normally displays an uncompromised stereo-chemical fidelity, which greatly facilitates the purification of intermediates and products (Schmid *et al.*, 2001; Wymer and Toone, 2000). Enzymes that naturally form carbon-carbon bonds or convert carbohydrates into other products are readily found in all living cells.

For this reason, many of the enzymes involved in the carbohydrate processing pathways in microorganisms, as well as, in higher organisms, have useful or potentially useful applications in synthetic organic chemistry. For example, the class I lyase *N*-Acetylneuraminic acid aldolase (EC 4.1.3.3) catalyses the reversible addition of

pyruvate to *N*-acetyl-D-mannosamine forming the parent sialic acid. This product has great therapeutic potential as sialic acids are important to a variety of recognition events, such as bacterial cell-cell recognition during development and differentiation, viral or microbial infections, tumour metastasis and leukocyte adhesion during inflammation. Neuraminic acid has been studied extensively, as the structure of zanamivir, an inhibitor of viral sialidases, is designed on this model. This product will be marketed for the treatment of influenza infection in the near future and is, therefore, a prime example of the application of industrially-developed aldolase reactions. Process re-design for the equilibrium-controlled bioproduction of neuraminic acid allowed an expansion of operations to the multi-ton scale and optimised downstream processing of the subsequent product (Blayer *et al.*, 1996; Mahmoudian *et al.*, 1997).

It has been shown that the knowledge gained from studies on lyases *in vitro* may readily be applied to *in vivo* systems: a good example is the use of the enzyme cocktail found in human cell lines to synthesise and process neuraminic acid analogues for cell-surface engineering. Acceptance by the cellular machinery of non-natural *N*-levulinoyl-D-mannosamine (Figure 1.6), which contains a ketone moiety in the *N*-acyl group, produced cell-surface oligosaccharides modified in the terminal neuraminic acid residues. Reactive ketone groups are displayed on the cell surface and could then be used in a highly versatile fashion for the covalent redecoration of cells under physiological conditions, for example, by nucleophiles. The simplicity and apparent generality of this approach opens an avenue to a myriad of potential biological and medicinal applications (Mahal *et al.*, 1997).

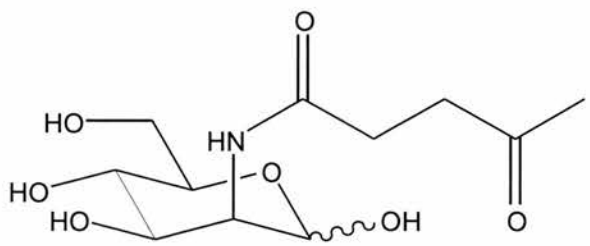


Figure 1.6 - *N*-levulinoyl D-mannosamine
(figure redrawn from Mahal, et al. 1997)

The fructose-1,6-bisphosphate aldolase, which belongs to the class II - metal-dependant aldolases, has been used extensively in the synthesis of C6-substituted (Page *et al.*, 1996) and unsubstituted sugars (Kim and Lim, 1996). This enzyme is particularly useful as it is very stable, with a half-life of two hours at 90°C, and it works

at the low optimum pH of 6.5 at which the base-catalysed decomposition of dihydroxyacetone phosphate (DHAP) is minimised (De Montigny and Sygusch, 1996).

Another important class of enzymes are hydroxynitrilases as their products are cyanohydrins, chiral carboxylic acid derivatives. In their enantiopure form they can yield 2-amino-1-arylethanols, which are important adrenergic drugs, such as bronchodilators, which can be prepared from (R)-terbutaline and (R)-salbutamol via suitable cyanohydrins (Effenberger and Jäger, 1997). Preparations from *Prunus amygdalus* (almond) and *Sorghum bicolor* (millet) display (R)- and (S)-specificity respectively. The plant hydroxynitrilases of (R)- and (S)-specificity have very different properties with regard to their primary and quaternary structures. The proposed mechanism for the hydroxynitrile formation involves nucleophilic attack of cyanide directly to the aldehyde carbonyl (Figure 1.7), which is promoted by hydrogen bond co-ordination of the carbonyl oxygen to surface residues. This is fundamentally different from the uncatalysed solution chemistry and involves a higher transition state for the C-C bond formation (Hasslacher *et al.*, 1997). However, as the yet unpublished work with enzyme-substrate complexes by J. Zuegg observed the addition of cyanide to a carbonyl group, which is far more conventional (as quoted by Fessner, 1998).

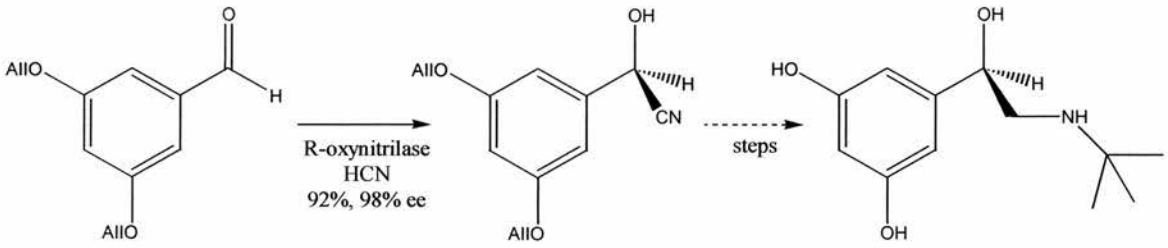


Figure 1.7 – HCN Addition to Benzaldehyde: the oxynitrilase-catalysed, highly enantioselective HCN addition to benzaldehyde is the keystep to adrenergic pharmaceuticals such as terbutaline (Figure copied from Fessner, 1998)

For industrial process design and development involving enzymes for the asymmetric synthesis of complex multifunctional molecules, an intimate knowledge base is required on subjects such as substrate tolerance for donor and acceptor components, and the influence of substrate structure on enantioselectivity and diastereoselectivity of C-C binding. The range of potential applications is broadening on all frontiers with increasing complexity of target structures, whilst the inherent limitations of the enzymes are being taken into account. This understanding is fuelled by the almost explosive growth of genome sequences and enzyme structures available to researchers in the academic environment as well as industry (Schmid *et al.*, 2001).

1.2 Aldolases

The retro-aldol reaction is defined as the condensation of the carbon at the position α to the ketogroup of a carbonyl, carbonate, carbonate ester or carbonate amide to the ketogroup of an aldehyde or ketone. The products of these reactions are known as aldols, β-hydroxycarbonyl compounds (Stryer, 1995c). In organic chemistry, the base catalysed aldol addition (Figure 1.8) is completely reversible, the Gibbs free energy of the reaction (ΔG) is close to zero and it does not normally proceed stereoselectively (Brückner, 1996).

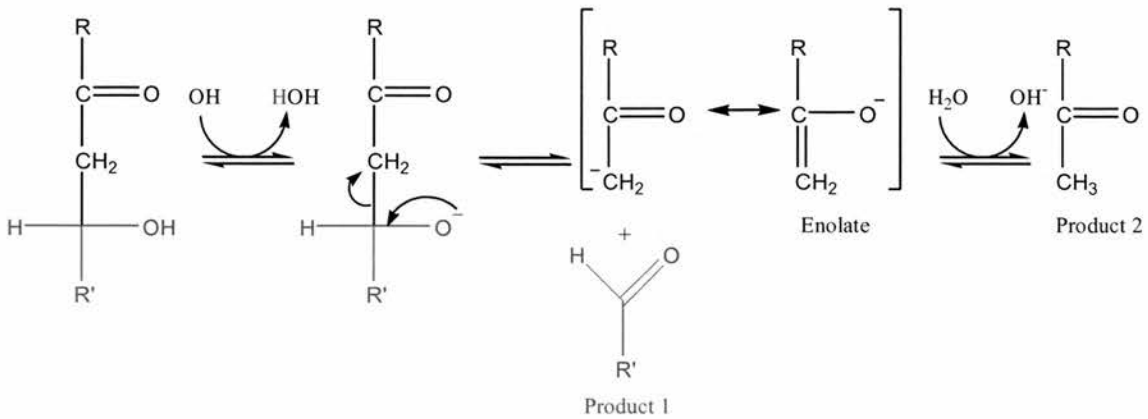


Figure 1.8 - Reaction Scheme for the Base-catalysed Aldol Cleavage: the aldol condensation (retro-aldol reaction) occurs by the reverse mechanism (figure adapted from Stryer, 1995).

In nature, these reactions are readily carried out by aldolase enzymes, which are a subgroup of the lyase enzyme family. They display almost perfect stereoselectivity and do not require the presence of protective groups. Over 30 different aldolases have been identified, forming two distinct functional categories. Both catalyse the aldol condensation/cleavage by activating the substrate nucleophile; but the class I aldolases activate it by forming a Schiff base/enamine intermediate with the substrate (Schurmann and Sprenger, 2001), and the class II metal dependant aldolases use a divalent metal ion and an active site base to increase the activity of the nucleophile (Hall *et al.*, 2002).

Class I aldolases are found in all forms of life (Verlinde and Quigley, 1999) Humans possess three aldolase isoenzymes (A, B and C), each of which has a unique activity profile and a tissue-specific distribution (Voet and Voet, 1995).The Class II aldolases are present in lower forms of life such as fungi, algae and some bacteria, and may therefore serve as targets for antibacterial or antifungal therapies (Machajewski and Wong, 2000).

Both enzyme classes can be divided into four sub-groups according to their reaction nucleophiles:

1. dihydroxyacetone phosphate (DHAP) dependant aldolases
2. glycine dependant aldolases
3. 2-deoxyribose-5-phosphate (DERA) dependant aldolases
4. pyruvate/phosphoenylpyruvate dependant aldolases

An example of a class II aldolase reaction (Figure 1.9) is the one catalysed by the Fructose-1,6-bisphosphate (FBP) aldolase from *Escherichia coli*. Dihydroxy-acetone-phosphate chelates with the catalytic zinc at the active site and is deprotonated by an active site base producing a carbanion. Glyceraldehyde-3-phosphate (G3P) binds parallel to the ene-diolate nucleophile, then the carbon-carbon bond is formed and the product FBP is released (Hall *et al.*, 1999).

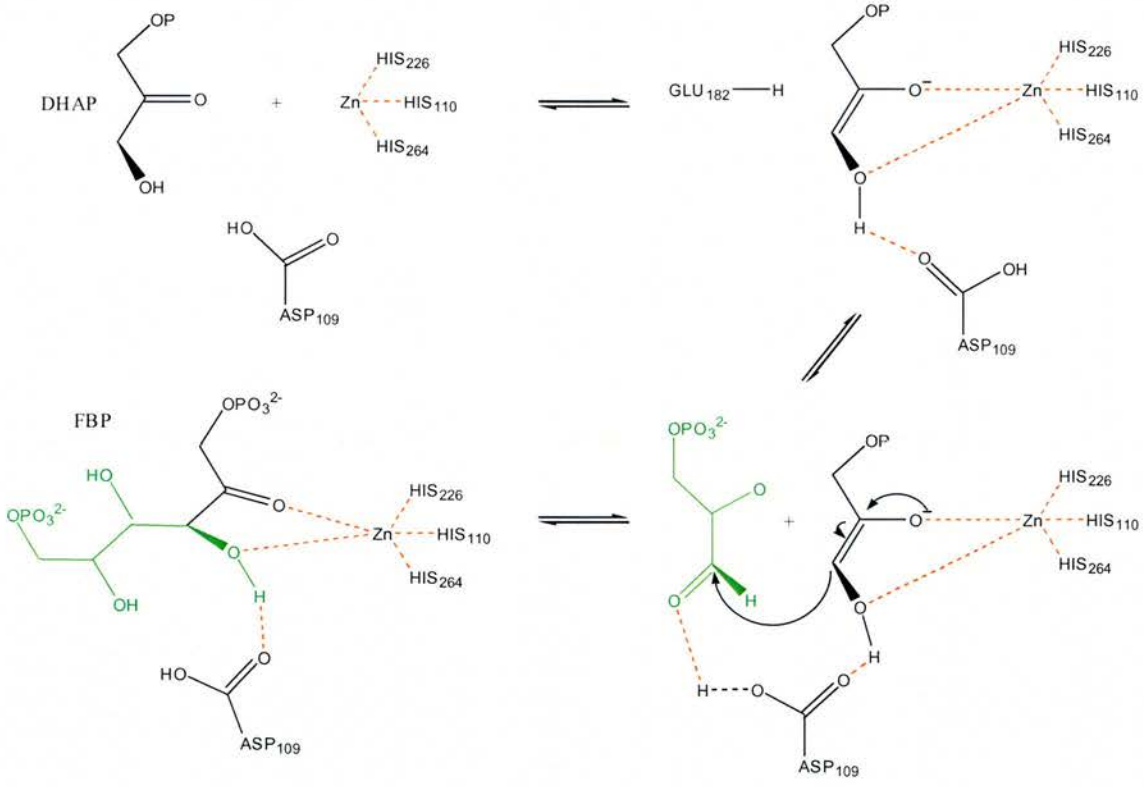


Figure 1.9 - Proposed Reaction Mechanism for the Class II FBP Aldolase from *E. coli*: taken from (Hall *et al.*, 1999).

In humans, the same reaction is catalysed by the class I FBP aldolase; it catalyses the reaction using the DHAP substrate as the catalytic nucleophile, which condenses with G3P to yield FBP. Figure 1.10 gives an overview of the reaction mechanism. A nucleophile attacks the carbonyl group to form a tetrahedral

intermediate, which then dehydrates. The resulting protonated Schiff base of DHAP, together with the ε-amino group of a conserved active site Lys, plays a critical role in catalysis. It promotes the formation of the enolate anion of dihydroxyacetone phosphate, serving as an electron acceptor. The enolate ion then adds to the aldehyde group of G3P to form a protonated Schiff base of FBP. The Schiff base is deprotonated and hydrolysed to yield FBP and the regenerated enzyme. In glycolysis the reverse of the above reaction is a key step; it is catalysed by the same enzymes (Dalby *et al.*, 1999).

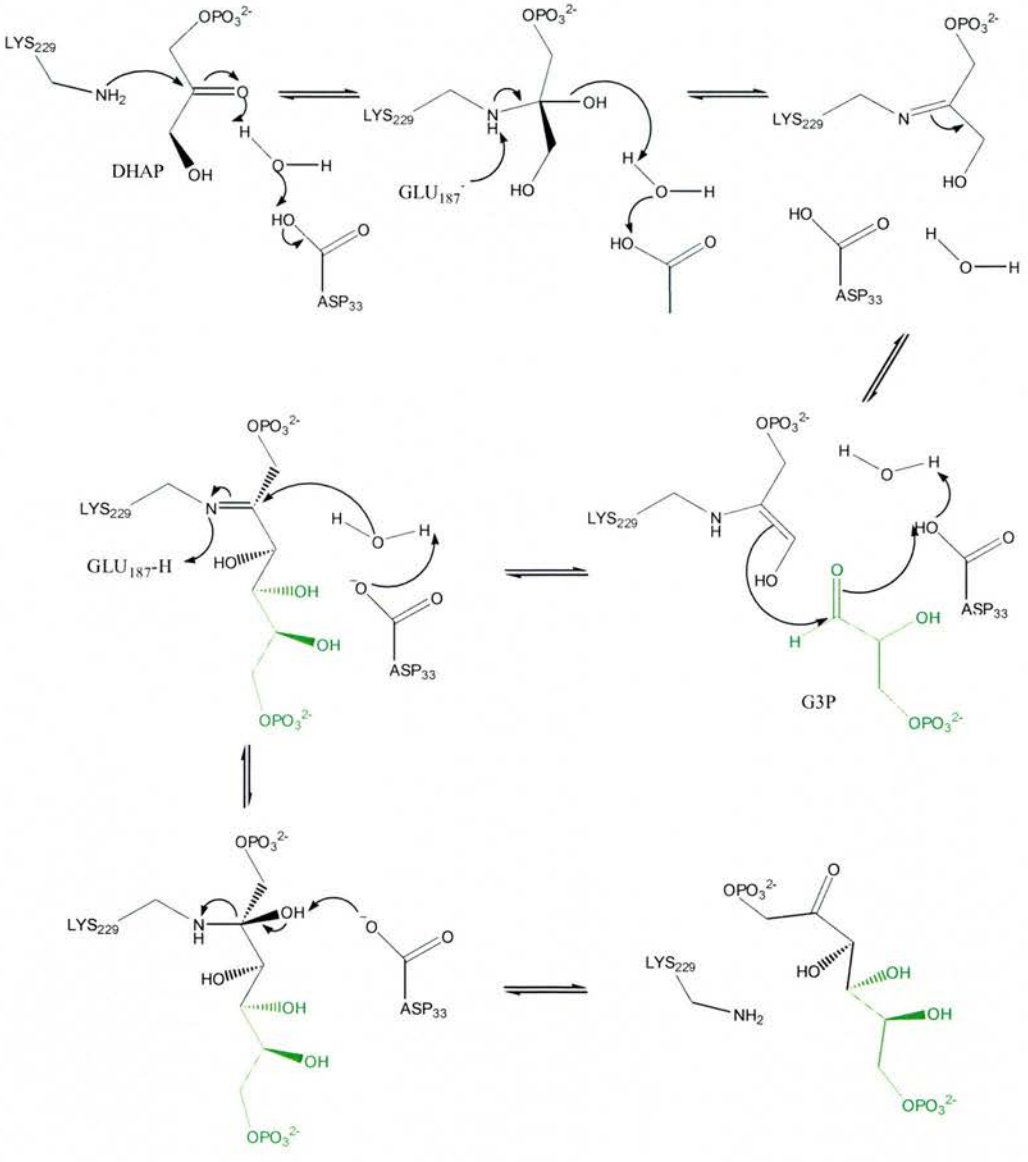


Figure 1.10 – Class I Human FBP Aldolase: mechanism for the condensation of DHAP and G3P by the human class I FBP aldolase (figure taken from Dalby *et al.*, 1999).

These aldolases are capable of forming all four diastereoisomers of the 1,3,4-trihydroxy-2-butanone skeleton. They have been used extensively in the preparation of amino sugars and, to some degree, as the precursors of cyclitols (Chou *et al.*, 1995). The

aldolase-catalysed reaction of DHAP and pentose or hexose provides easy access to novel long-chain sugars, which are difficult to obtain from either natural sources or chemical synthesis (Fessner and Walter, 1996; Gijsen *et al.*, 1996).

Glycine dependant aldolases catalyse the reversible condensation of glycine and a ketone, producing an aldol. Mainly alpha amino acids are produced in this fashion and some of the aldolases in this category accept a variety of different substrates, for example, L-threonine aldolase catalyses the production of erythro- and threo-amino acids in the presence of the co-factor Pyridoxyl-5′-phosphate. This ability has led to the use of L-threonine aldolase in the production of novel compounds, such as the immunosuppressant Mycestericin D (Shibata *et al.*, 1996)

The DERA aldolase is capable of catalysing the condensation of two aldehydes (Figure 1.11). The high-resolution structure of the DERA aldolase from *E. coli* gave an insight into its catalytic mechanism. A Lys in the active site (Lys167) has been identified as the Schiff-base forming residue. After the enamine is formed, the system is

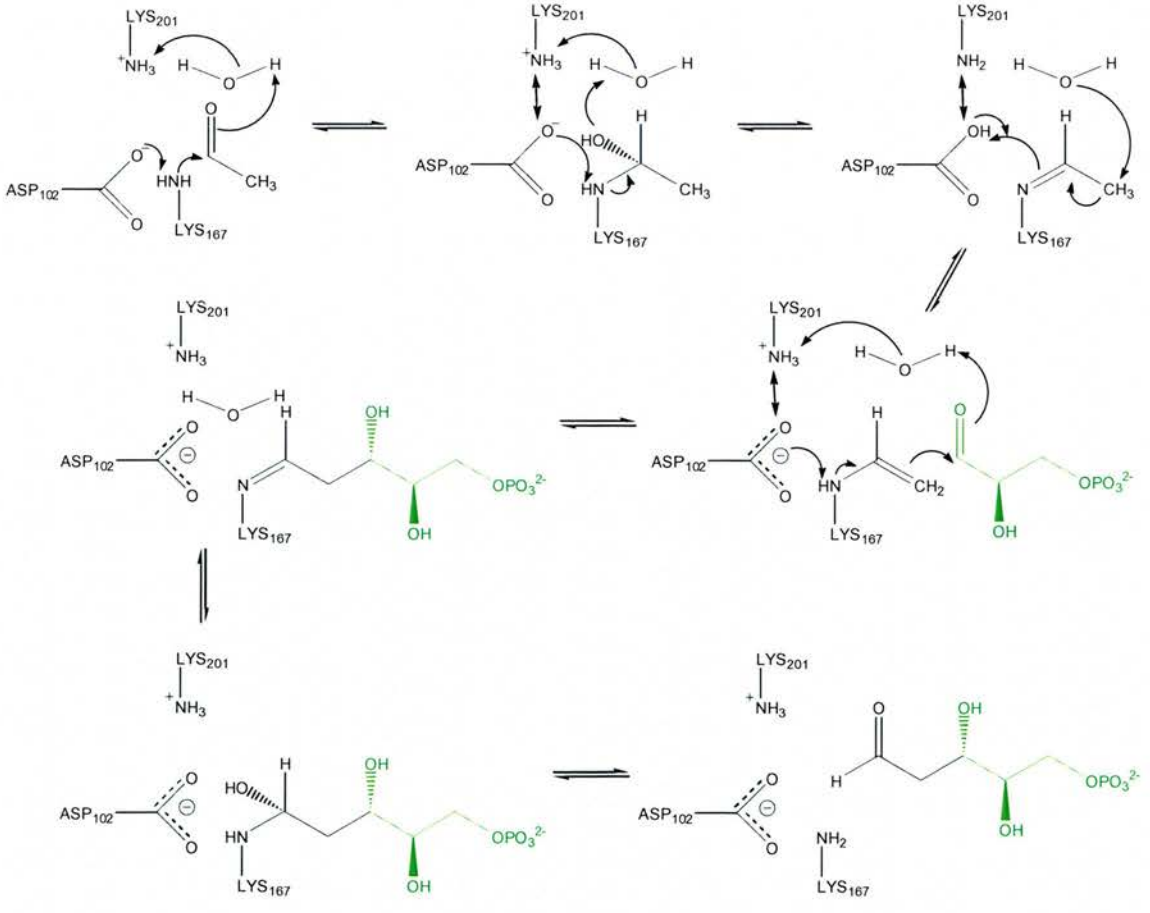


Figure 1.11 - Catalytic Mechanism of the DERA Aldolase from *E. coli*: catalysing the condensation of acetaldehyde and G3P (Heine *et al.*, 2001).

poised for nucleophilic attack on the acceptor aldehyde G3P, requiring the protonation of its carbonyl group. The proton is provided via a water molecule through a proton relay system that is composed of an Asp (Asp102), a Lys (Lys201) and the active site water, which is responsible for shuffling the proton between C2′ of the acetaldehyde imine and enamine and subsequent C3′ hydroxyl protonation (Heine *et al.*, 2001).

1.3 KDPG Aldolase

An example of a pyruvate/phosphoenolpyruvate dependant aldolase is the 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase (EC 4.1.2.14). It is best known for its role in the ED pathway in bacteria (Figures 1.2 and 1.12), which carries out the gluconate metabolism in these organisms. This enzyme catalyses the reversible cleavage

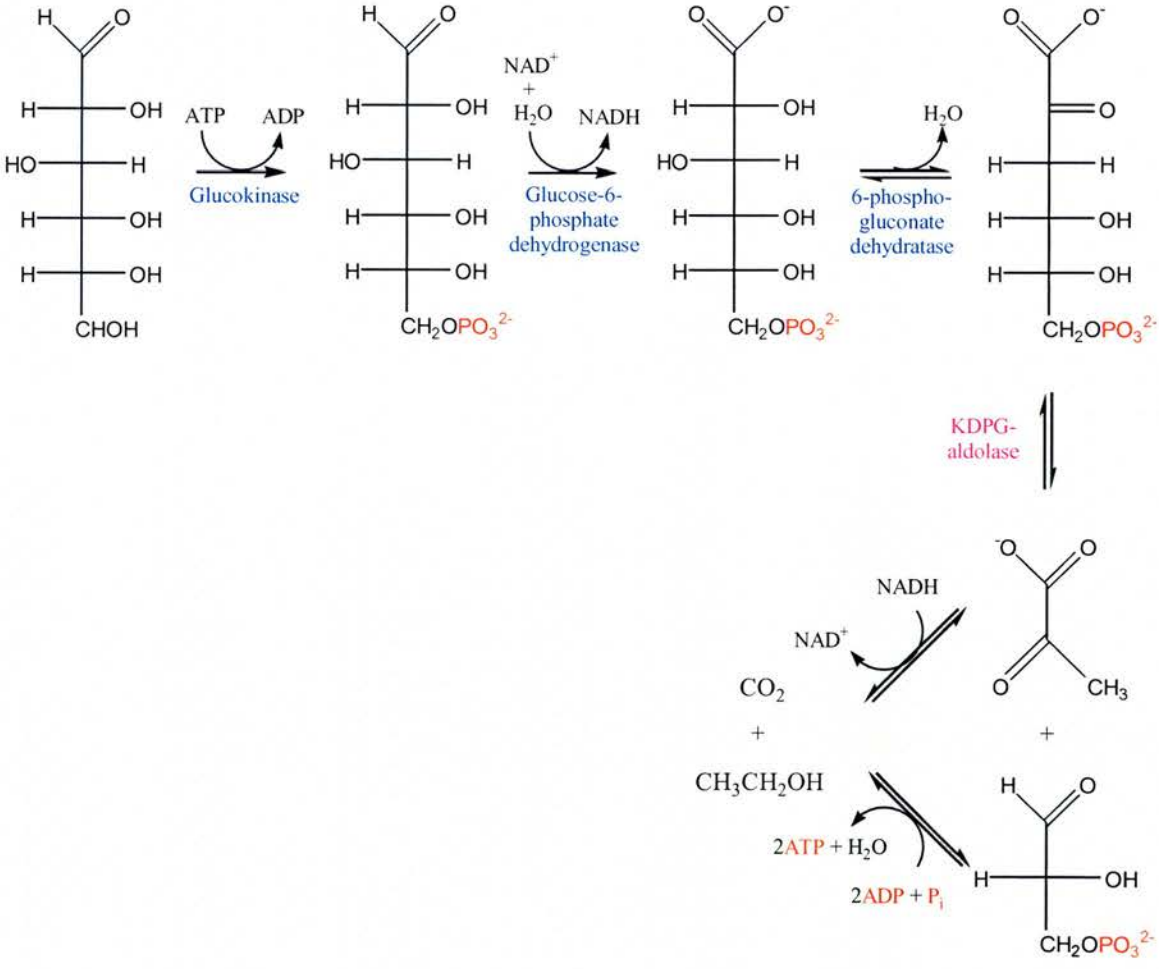


Figure 1.12 - Entner-Doudoroff Pathway for Glucose Breakdown: the KDPG aldolase catalyses the breakdown of 2-keto-3-deoxy-6-phosphogluconate into pyruvate and glyceraldehyde-3-phosphate. (taken from (Voet and Voet, 1995).

of KDPG into pyruvate and G3P. Gene sequencing of *E. coli*, carried out by Egan *et al.*, 1992, showed that the KDPG aldolase is similar in sequence to the 2-keto-4-hydroxyglutarate (KHG) aldolase, which shows a specificity toward glyoxylate, converting it to KHG in the presence of pyruvate. This enzyme is capable of regulating the glyoxylate levels in bacterial cells during growth on acetate carbon sources.

In mammals, the KHG aldolase (EC 4.1.3.16) catalyses a central reaction in the catabolism of hydroxyproline; additionally, this enzyme catalyses the β-decarboxylation of oxaloacetate (Kobes and Dekker, 1971).

The reaction mechanism, as proposed by Allard *et al.*, 2001, for the aldol condensation of pyruvate and G3P to yield KDPG as catalysed by the KDPG aldolase, is depicted in Figure 1.13. The Schiff base formation with pyruvate is thought to occur through the transient formation of a dipolar carbinolamine with the keto group resulting in the neutral carbinolamine shown. This neutral species is then dehydrated to form the imine form of the Schiff base. A reactive enamine is formed by proton abstraction, which then yields a second Schiff base with KDPG on condensation with G3P. Hydrolysis causes the formation of a neutral carbinolamine species from the Schiff base, and on decomposition of the dipolar carbinolamine, KDPG is released from the active site, thereby regenerating the enzyme. In bacteria this enzyme operates under kinetic control (Meloche *et al.*, 1975) to provide stereochemically pure products (Shelton *et al.*, 1996).

This reaction mechanism is scientifically interesting as it is one of nature's ways of forming carbon-carbon bonds. However, in addition to the academic curiosity, chemical industry values this enzyme as it displays a high substrate tolerance; it efficiently catalyses the stereospecific aldol addition of pyruvate to a wide range of unnatural electrophilic substrates, at rates that are practical for large-scale organic synthesis (Fessner and Walter, 1996). Additionally, directed evolution for the *E. coli* KDPG aldolase has been reported: altering key amino acids in the vicinity of the active site allows an alteration of substrate specificity, increasing the production potential of a wide variety of stereogenic compounds (Fong *et al.*, 2000; Wymer *et al.*, 2001).

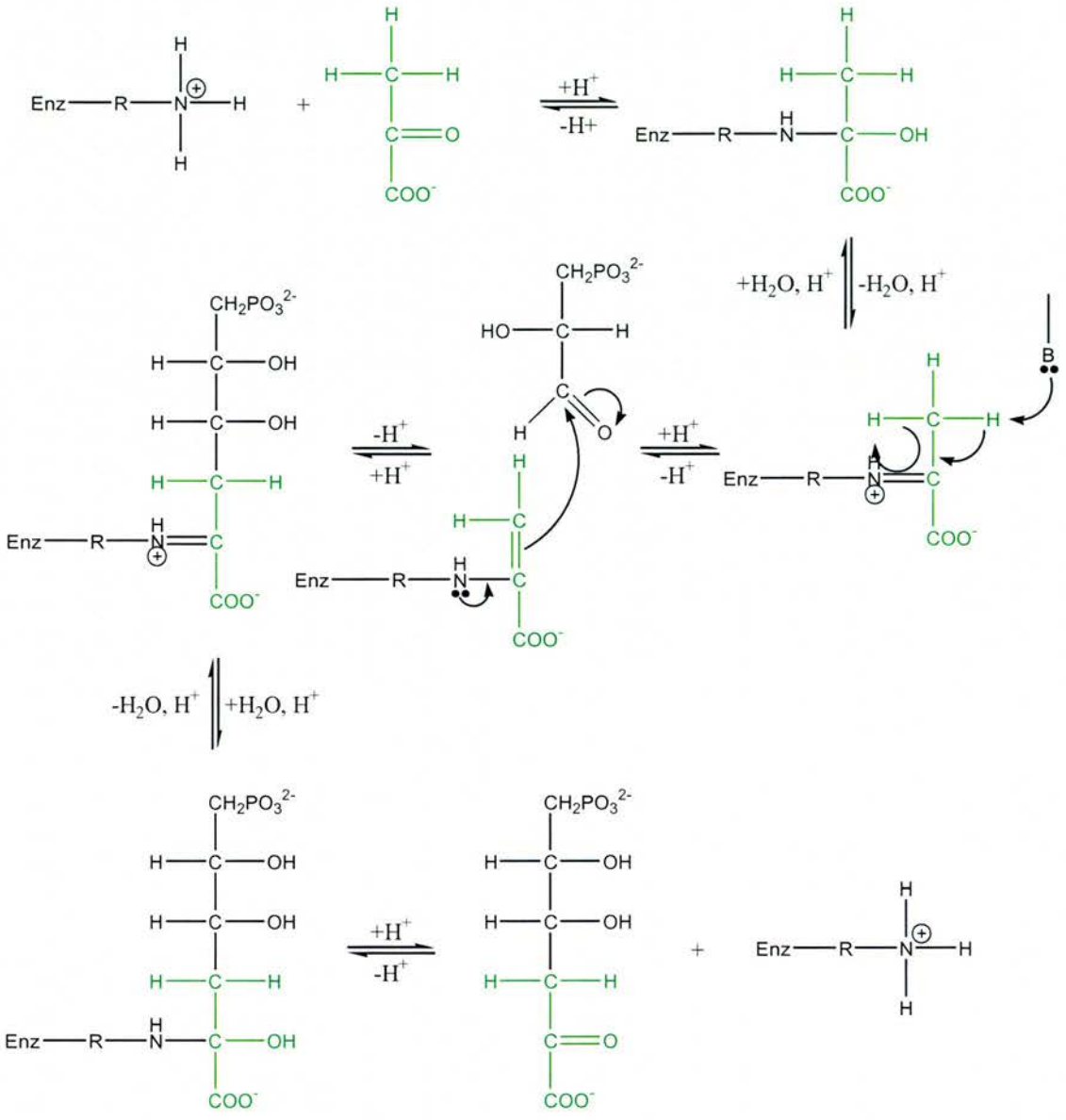


Figure 1.13 - Reaction Mechanism for the Aldol Condensation: between pyruvate and G3P. This is catalysed by the KDPG aldolase. The mechanism was proposed by Allard *et al.* in 2001 (figure reproduced from Allard *et al.*, 2001)

1.4 Antibiotic Resistance

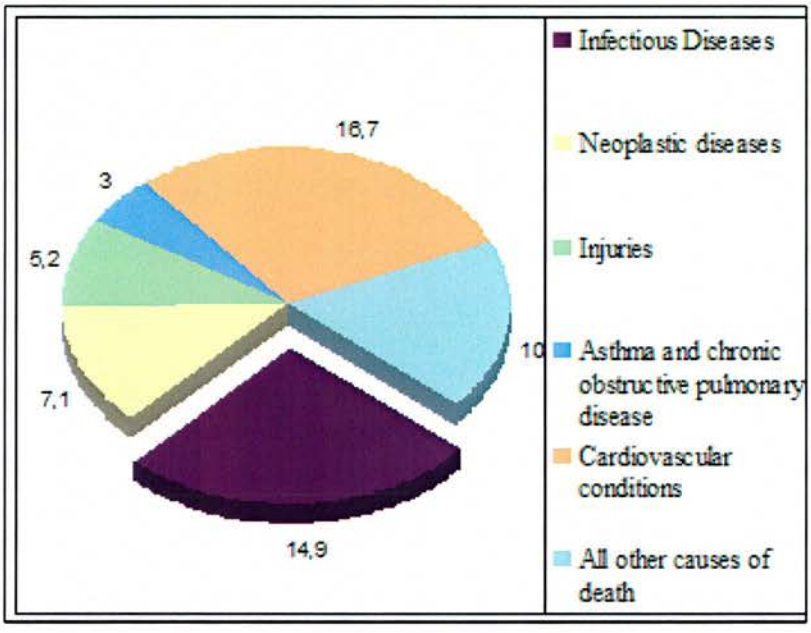
Since Alexander Fleming discovered the activity of penicillin in 1928, physicians have used it and other antibiotics to effectively treat bacterial infections from pink eye to pneumonia. He originally noticed lysis of bacterial colonies occurring next to the contaminant bread mould *Penicillium notatum* in an old petri dish, and then hypothesised, and later proved, that the mould produced a substance that perfused through the medium, killing the cells (Fleming, 1929). This was to be the first of a number of 'magical bullets' that selectively kill bacteria, leaving the host largely unharmed. But just four years after drug companies had launched the mass-production of penicillin in 1943, microbes began surfacing that could resist it. In an article in the New York Times, Fleming warned as early as 1945 of the potential risks associated with the improper use and misuse of antibiotics: 'The greatest possibility of evil in self-medication is the use of too small doses so that instead of clearing the infection, the microbes are educated to resist penicillin and a host of penicillin-fast organisms is bred out which can be passed to other individuals and from them to others until they reach someone who gets a septicaemia or pneumonia which penicillin cannot save.' (Fleming, 1945)

As Fleming warned correctly, misuse of antibiotics in human and veterinary medicine has generated selective pressures that favour the spread of resistant bacterial strains far more quickly than would have occurred through spontaneous mutations (Guillemot, 1999; Morris *et al.*, 1998). In addition, the prescription of antibiotics, such as Linezolid, for viral illnesses, including the common cold, is still widely practised even though these drugs specifically target bacteria, and are therefore completely ineffective against viruses. In 1999, it was reported that more than 20% of antibiotic prescription to adults in the developed world was for viral respiratory diseases (Guillemot, 1999). Another important problem is the increased used of antibiotics under circumstances where they are not necessarily needed, such as in household cleaning products (Marwick, 1999; Stephenson, 2001; Stephenson, 2002; Vastag, 2002).

The issue is crucial to public health, and officials at the World Health Organisation (WHO), as well as the United States Centres for Disease Control and Prevention (CDC), have recently drawn focus to the problems of drug-resistant diseases.

Representatives from both agencies have called upon scientists and funding organisations to take seriously the threat that the world could soon be plunged back into a 'pre-antibiotic era' as microorganisms become increasingly immune to our current disease-fighting arsenal of drugs.

For example, deaths from septicaemia have risen from 1 per 100,000 in 1960 to 10 per 100,000 US population in 2001, and 100,000 persons each year are hospitalised with methicillin-resistant *Staphylococcus aureus* (MRSA) infections in the United States alone. It was estimated that infections with pathogenic bacteria and parasites accounted for almost 20% of all deaths worldwide in 2001 (Figure 1.14) (WHO, 2003).



Infectious diseases	Annual deaths (million)
Respiratory infections	3.96
HIV/AIDS	2.77
Diarrhoeal diseases	1.80
Tuberculosis	1.56
Vaccine-preventable childhood diseases	1.12
Malaria	1.27
STDs (other than HIV)	0.18
Meningitis	0.17
Hepatitis B and C	0.16
Tropical parasitic diseases	0.13
Dengue	0.02
Other infectious diseases	1.76

Figure 1.14 – Mortality Statistics for 2001: WHO report of 2003 emphasises that more than 25% (15 million out of 57 million) of all deaths worldwide in 2001 were due to infectious diseases.

Under normal circumstances *S. aureus* is non-pathogenic; however, it can cause severe disease in immuno-compromised individuals. In these it may cause skin and bone infections, pneumonia and severe septicaemia. In the UK, in the year 2002, it caused 800 deaths and another 5,000 known cases of blood poisoning, which is a significant increase over the past 10 years when only 51 people were reported to have died from MRSA infections. The organism is spread by close contact: almost always by direct physical contact but also through indirect contact and sometimes even through droplets. Statistics show (Figure 1.15) that almost 70% of the *S. aureus* organisms isolated from nosocomial infections in the United States are resistant to methicillin. This is more than double the proportion that had been found in 1989 (CDC, 2004).

The growing number of deaths and severe illness attributable to bacterial infectious disease is thought to be caused mainly by the widespread dissemination of antibiotic resistance. Large-scale studies in US hospitals also agreed with the above findings as no hospitals so far have been successful at preventing or halting the spread of antimicrobial resistance (Diekema *et al.*, 2004).

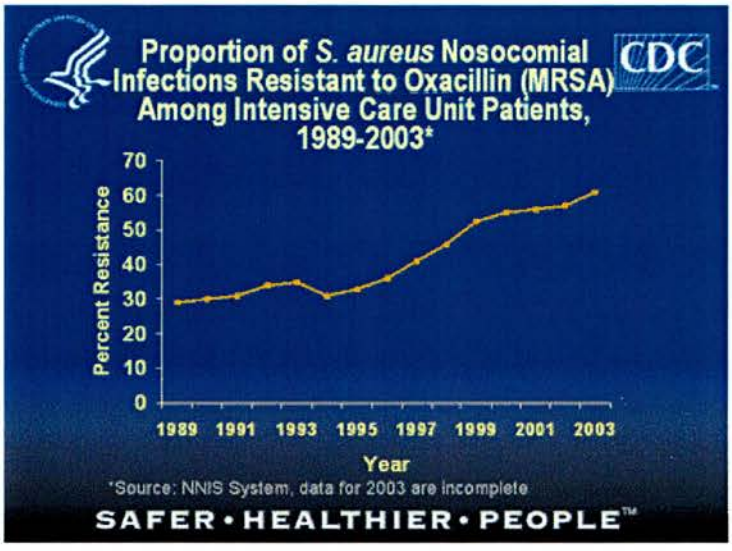


Figure 1.15 - MRSA Infections in Isolates from ICU Patients (figure taken from the CDC web-page http://www.cdc.gov/ncidod/hip/ARESIST/mrsa.htm).

Over the past few decades, only two new chemical entities have entered clinical practice as antibacterial agents, and only one whose target is in a new biochemical class (Miesel *et al.*, 2003; Nathan, 2004; Projan, 2003; Wenzel, 2004). The driving force for the development of new antibiotics has to come from industry, since governmental funding into this area of research has decreased ever since the United States Surgeon General Council pronounced, in its 1969 testimony to Congress, that infectious diseases were under control, and it was only a matter of time before they were eradicated completely. Economic pressures forced industry to go for less risky options, such as increasing the efficacy and bioavailability of existing antibiotics, rather than developing drugs against new targets (Low and Scheld, 1998; Trias and Gordon, 1997). Additionally, since the late 1980s and early 1990s, the search for better antibiotics has given way to the development of drugs to treat non-infectious diseases, such as high cholesterol, asthma, cancer and central nervous system disorders, and of so called 'life-style drugs' such as Viagra and anti-obesity drugs (Barrett CT and Barret JF, 2003). Also, there is a considerable lack of incentive for industry to develop new antibiotics: not much money can be made from selling antibiotics in poor countries, and the call for appropriate use of antibiotics is seen to have a negative impact on sales, prompting

many companies to shut down that particular branch of their drug development programs altogether (WHO, 2001). Table 1.1 summarises the resistance found in common pathogenic bacteria as published by the WHO, 2003.

Table 1.1 - List of Infectious Bacteria: bacteria which were previously susceptible to antibiotics and are now resistant (as summarised by the WHO, 2003)

Disease caused	Bacterium	Resistance
Pneumonia	*Strep. Pneumoniae*	Penicillin
Dysentry	*Shigella dysenteriae*	Muti-resistant
Typhoid	*Salmonella typhi*	Multi-resistant
Gonorrhoea	*Neisseria gonnorrhoeae*	Penicillin & Tetracycline
Tuberculosis	*Mycobacterium tuberculosis*	Rifamicillin & Inah
Nosocomial Infections caused by:		
	Staphylococcus aureus	Methicillin (& vancomycin)
	Enterococcus spp.	Vancomycin
	Klebsiella	Multi-resistant
	Pseudomonas	Multi-resistant

Antibiotic resistance spreads through the accumulation of adaptive mutations that give resistant strains a selective advantage over sensitive ones. The spread of resistance is mediated by the acquisition of resistance genes, which can be gained by three mechanisms (Morris *et al.*, 1998):

- Conjugation - the transfer of genetic material through cell to cell contact.
- Transduction – the transfer of resistance gene carrying bacteriophages from one bacterium to another making use of sex pili.
- Transformation – direct acquisition and absorption of DNA from the surrounding environment by competent bacteria.

The genes themselves may be carried in the bacterial cells on circular DNA, such as plasmids or parasitic stretches of DNA, such as transposons and integrons. Both plasmids and transposons may spread resistance determinants by conjugation. Transposons are particularly efficient at this, as they are self-transmissible encoding factors that allow them to excise themselves and integrate into plasmids or chromosomal DNA (Schaberg *et al.*, 1981). Integrons are normally involved in the generation of large resistance genes (Fluit and Schmitz, 2004).

There are four main categories of antibiotic resistance mechanisms in pathogenic bacteria: modification of the drug, efflux mediated antibiotic resistance, mutation of the drug target and cell wall impermeability.

Chemical modification of the drug can be divided into two main mechanisms: aminoglycoside-modifying enzymes and beta-lactamase action. The enzyme beta-lactamase confers resistance to beta-lactam (penicillin-type) antibiotics. Beta-lactams bind to penicillin-binding proteins (PBPs) in the cell wall of Gram positive bacteria, inactivating them by forming covalent bonds with the active site serine. This causes steric hindrance and prevents the PBPs from carrying out the transpepidation reaction during peptidoglycan synthesis. Beta-lactamase forms the same acyl-enzyme complex with beta-lactam, but this bond formation is followed by de-acylation, resulting in loss of the ring conformation of the beta-lactam antibiotic. This compound is then unable to carry out the initial nucleophilic attack at the acyl carbon of PBPs and is, therefore, ineffective at inhibiting cell wall synthesis (Massova and Mobashery, 1998).

Efflux mediated antibiotic resistance is found in Gram-positive, as well as, Gram-negative bacteria: resistant bacteria express transmembrane proteins that mediate the expulsion of antibiotics from the cell. For example, *Streptococcus pneumoniae* expresses the efflux associated genes Mef(A) and Mef(E), which confer an almost 64-fold increase in resistance to macrolide antibiotics (Sutcliffe *et al.*, 1996). *E. coli* also possess drug efflux pumps, which act as antiporters that use the proton-motive force generated during respiration to exclude antibiotics and other compounds from the cytoplasm in exchange for protons (Borges-Walmsley and Walmsley, 2001).

The drug's target may get mutated to allow bacterial resistance to a particular antibacterial agent. An example of this is the resistance of some *E. coli* strains to erythromycin A. This resistance is due to a change in only three amino acids (MKR) in the ribosomal L22 protein. Presumably, this mutation prevents the recognition or binding of the aminoglycoside to its target protein (Chittum and Champney, 1994).

The change in cell wall permeability of bacterial cells may be due to different proteins being expressed in altered amounts preventing antibacterials from passing through the membrane or cell wall, thus conferring resistance.

1.5 Actinomycetes

Gram positive bacteria are divided into two major classes based on the composition of their genomes. The low G+C organism class contains bacterial species, such as Bacillus, Clostridium, Staphylococcus and Streptococcus, whereas the high G+C organisms are referred to as Actinomycetes. The G+C content of the latter group's DNA is about 70 to 74%. The gross composition of their chromosomal DNA and that of their plasmids is very similar. The strains belonging to this group used to be assigned to their genera on morphology alone, principally on the basis of the tendency of their vegetative hyphae to fragment into small, squarish bits referred to as 'fragments'. However, as this characteristic is wide spread amongst Actinomycetes belonging to different genera, they had to be reclassified on chemical characteristics, such as their cell wall amino acids, fatty acid and lipid patterns and whole cell sugars (Kieser *et al.*, 2000a). Of course, with the availability now of DNA and RNA sequences, the most important tool used by taxonomists is the sequence analysis of the 16S ribosomal RNA (Hackl, *et al.*, 2004). This reclassification has led to some confusion in the literature over taxonomy. For example, *Streptomyces violaceoruber* was formerly classified as *Streptomyces lividans* and both names are still in use. Another case is that of Nocardiae: some species in this bacterial class were found to lack mycolic acids, and were therefore reclassified into the genera *Amycolatopsis* and *Amycolata*. The commercially important species in these genera include *Amycolatopsis* (*Streptomyces*) *orientalis*, *Amycolatopsis (Nocardia/Streptomyces) mediterranei*, *Nocardia rugosa*, *Nocardia lurida* and *Nocardia sulphurea*.

Actinomycetes produce two-thirds of all antibiotics that are synthesised by micro-organisms; amongst them, 80% originate from the genus Streptomyces. It is not the overall chemical versatility of the Actinomycetes that is unparalleled – fungi and higher eukaryotes produce far more versatile chemicals – but the proportion of their output that has a known biological effect, in other words, they are more bioactive compared to other especially higher organisms (Eckwall and Schottel, 1997). In their natural environment, there is only circumstantial evidence of antibiotic production by Actinomycetes, as it is difficult to isolate and extract antibiotics from soil. However, the potato pathogen *Streptomyces scabies* has been found to be specifically controlled by an antibiotic produced by another Streptomycete that coexists in the same surroundings.

Actinomycetes colonise soil and their growth is facilitated by their vegetative hyphal mass, which can differentiate into spores, which assist in spread and persistence (Cochrane, 1961; Mayfield *et al.*, 1972; Morita, 1985). They impart resistance to low nutrient levels and drought, whereas the mycelial stage of the bacteria's life cycle is sensitive to both events (Lloyd, 1969). Most Streptomyces species in soil form spores most of the time, as germination needs exogenous nutrients, water and calcium ions (Ensign, 1978). Spores put into non-sterile soil show a very low growth rate due to competition with endogenous micro-organisms (Lloyd, 1969).

Rigorous model experiments on agar (Wiener, 1996) have indicated that the antibiotic production by Streptomycete colonies can play a significant role in preventing invasion by competing *Bacillus subtilis*. However, it does not seem to improve the ability of the antibiotic producer to invade an established population of antibiotic-sensitive *B. subtilis*, nor help competition between established Streptomycete and bacillus populations. This fits the idea that the major role for antibiotics is to defend the colony from competitors at the time of development of the aerial mycelium in the lysing vegetative state (Chater and Merrick, 1979; Karagouni *et al.*, 1993).

The production of antibiotics is under genetic control, and most antibiotics are the product of complex biosynthetic pathways, with a cluster of genes dedicated to the synthesis of any one compound. These gene clusters usually contain pathway-specific regulatory genes, which act as transcriptional activators, which in turn, may themselves be subject to control by pleiotropic regulatory genes. The onset of antibiotic biosynthesis is determined and influenced by a variety of physiological and environmental factors (Chater and Bibb, 1997; Demain, 1992; Demain and Fang, 1995; Hood *et al.*, 1992; Horinouchi and Beppu, 1994; Yang *et al.*, 1995).

The genetics of Actinomycetes has been studied extensively and is quite well understood (Keller *et al.*, 1985). Nearly all plasmids are self-transmissible fertility factors; both linear and circular plasmids exist, and like the chromosome, they have terminal inverted repeats and 5' ends. The circular plasmids reside in the chromosomes but, after they have transferred independently by conjugation to strains lacking that particular plasmid, they become autonomously replicating plasmids with moderate copy numbers. Actinomycetes also have non-integrating, conventional plasmids; the large ones have a low copy number and replicate by the 'theta' mode, bidirectionally from their origin to generate new double-stranded plasmid copies. The smaller ones have a

high copy number and replicate by a rolling circle mechanism ('sigma' mode), via circular, single-stranded replicating intermediates. Replication is initiated by a characteristic, plasmid-encoded replication site (Kieser *et al.*, 2000b). For several decades, industrial processes used the natural strains which were improved for antibiotic titre using chemical or radiation induced random mutagenesis (Baltz, 1986). Additionally, use of natural mating or protoplase fusion to combine desirable genes from divergent lines of a mutation-screening pedigree has been successful on many occasions. Due to commercial secrecy, this is not well documented and neither is the presence of genetic engineering for titre improvement (Chater and Merrick, 1979). It is, however, well documented that genetic engineering has been used to enhance the expression of specific rate-limiting steps in the biosynthetic pathway for commercial antibiotic production, to enhance end-product titre (Baltz *et al.*, 1997; Chater, 1990). Attempts to increase yield by genetic engineering may well have been successful but this is not known for certain.

New hybrids are the visible results of genetic engineering; few useful compounds have resulted from it, but it is definitely possible, and it does possess great potential for the construction of high-yielding strains and the production of novel protein pharmaceuticals (Hutchinson and Fujii, 1995; Tsoi and Khosla, 1995). The data yielded by genome sequencing projects especially will further genetic engineering of bacteria. An example of successful engineering is that of the glucose isomerase enzyme: this is typically an intracellular enzyme, but it has been engineered to gain an increased thermal and alkaline stability for commercial production (Baltz, 1990; Binnie *et al.*, 1997).

In order to successfully engineer antibiotics which are made naturally by Actinomycetes, it is necessary to fully understand every step of the biosynthetic pathway of that particular compound (Baltz and Seno, 1988).

1.6 Glycopeptides

Glycopeptide antibiotics are often regarded as the drugs of last resort in the fight against infections caused by multi-drug-resistant bacteria. They are antibacterial agents that consist of a multi-peptide backbone to which a number of carbohydrate moieties are attached (Pootoolal *et al.*, 2002). The first person to isolate a compound from this class

of antibiotics was the microbiologist Selman Abraham Waksman, who received the Nobel Prize in Physiology or Medicine in 1952 for his discovery of streptomycin (Figure 1.16). This antibiotic was first isolated in September 1943, and first publically announced in January 1944. Before the end of that year, streptomycin was already being submitted to clinical trials and was then used widely during the closing stages of World War II (Waksman, 1952). In addition to its being active against Gram negative bacteria, it was found to be highly effective against the tuberculosis-causing pathogen *Mycobacterium tuberculosis in vitro*, as well as, *in vivo*. Its efficacy in the treatment of tuberculosis - first in animal models and later in humans - was fully established by Dr. W.H. Feldman and Dr. H.C. Hinshaw at the Mayo Clinic (Feldman and Hinshaw, 1944; Feldman and Hinshaw, 1945). Within a very short period of time, Streptomycin production grew into a large industry with a monthly output of 25,000 kilograms in the United States alone, and many more similar antibiotics were to follow. The next compound to be isolated was vancomycin (Figure 1.17a), which was found in 1956 in soil samples from India and Indonesia. Nowadays, vancomycin often denotes the whole class of glycopeptide antibiotics as it is one of two compounds that are approved for clinical use in humans, the other one being teicoplanin (Figure 1.17b) (Foldes *et al.*, 1983; Hubbard and Walsh, 2003).

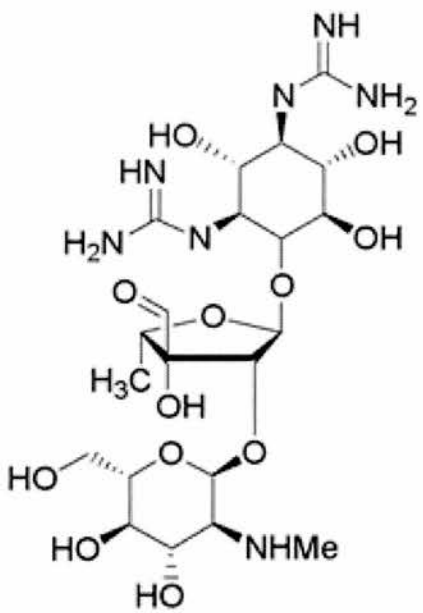
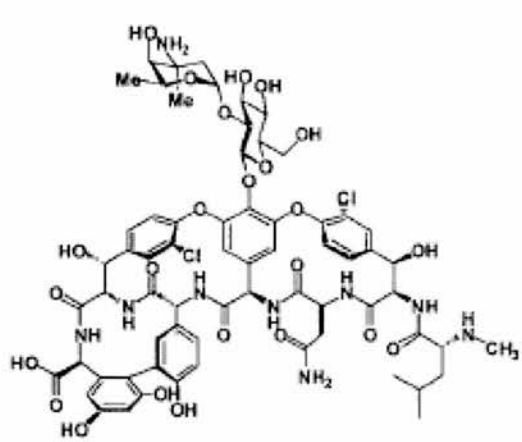


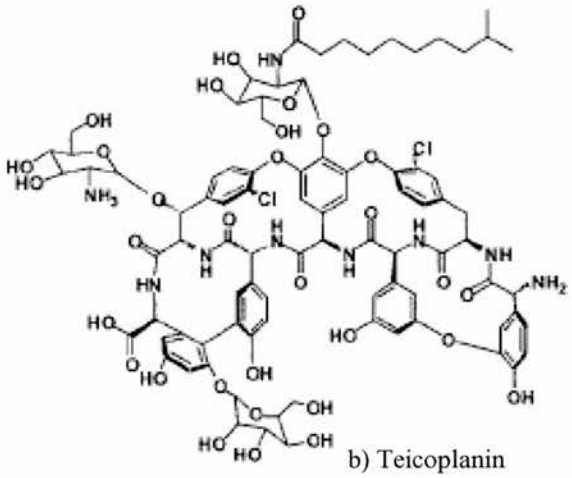
Figure 1.16 - Structure of Streptomycin (taken from http://www.alanwood.net/pesticides/structures/streptomycin.gif)

These two compounds can also be used to exemplify the two sub-classes that exist in this antibiotic family. Each is made up of a heptapeptide moiety that has undergone extensive cross-linking of the aryl side chains of the amino acid residues to

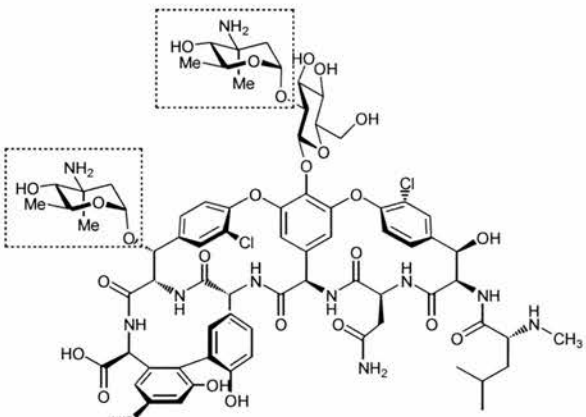
impart a rigid, dome-shaped architecture to the peptide scaffold. Several of these amino acids are nonproteinogenic: residues 4 and 5 in vancomycin, as well as 1, 4 and 5 in teicoplanin, are 4-hydroxyphenylglycines; residue 7 in vancomycin, and 3 and 7 in teicoplanin, are 3,5-dihydroxyphenylglycines. The specific order of D and L amino acids throughout the peptide is thought to be crucial to the formation of the oxidative cross links in the aryl side chains. Vancomycin contains three of these between residues 2 and 4, 4 and 6, and 5 and 7, whilst teicoplanin has an additional link between side chains at residues 1 and 3. This means that all side chains of teicoplanin, but not of vancomycin, are interconnected.



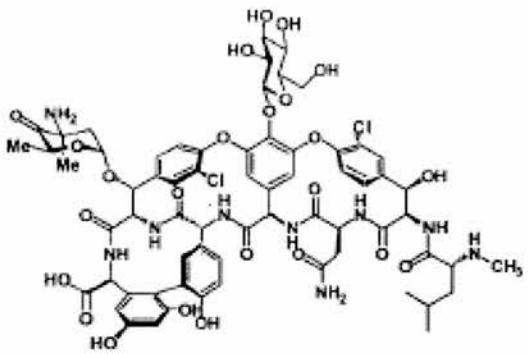
a) Vancomycin



b) Teicoplanin



c) Chloroeremomycin



d) Balhimycin

Figure 1.17 –Structures of Several Glycopeptide Antibiotics
figure taken from Hubbard and Walsh, 2003 and Merkel *et al.*, 2004)

Not only are there differences in the heptapeptide backbone between these two sub-classes in the glycopeptide antibiotic super-family, but also in the sugars attached to the aglycone moieties. A vancosaminyl-1,2-glucosyl moiety is attached to the phenolic oxygen of 4-OH-PheGly$_4$ in vancomycin. The epivancosamine-type sugars are part of a

number of vancomycin-type antibiotics. For example, the 4-epi isomer of L-vancosamine (3-amino-2,3,6-trideoxy-3*C*-methyl-L-*arabino*-hexopyranose) is made specifically as a constituent of chlororeremomycin (Figure 1.17c) by *Amycolatopsis orientalis*. It is attached to the benzylic oxygen of β-OH-Tyr_6 as well as the O2 of the glucose, which in turn is bound to the 4-OH-$PheGly_4$. From here onwards, this compound will be referred to as L-epivancosamine. Balhimycin (Figure 1.17d) contains the 4-oxo-L-vancosamine in the latter of the above positions, where vancomycin itself contains the L-vancosamine sugar.

No comparable amino-deoxysugars are found in the teicoplanin group of glycopeptides, rather they contain the more conventional sugars, 2-*N*-acetyl-D-glucose (Glc*N*Ac) and D-mannose, which are also found in bacteria's primary metabolism. Glc*N*Ac is introduced at the comparable phenolic oxygen atom (4-OH-$PheGly_4$) and has the short chain acetyl replaced by a branched C10′ fatty acyl group. Additionally, residues 6 and 7 have *N*-acyl-D-glucosamine and a common Glc*N*Ac attached respectively. However, there are teicoplanin family members without any sugar substituents, such as A47934, produced by *Streptomyces toyocaensis*, which contains only a sulfate group on residue 3. Table 1.2 gives an overview of some glycopeptide antibiotic family members and their attached sugar moieties.

Table 1.2 - Selection of Sugars Found in the Glycopeptide Antibiotic Family

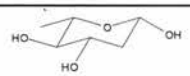
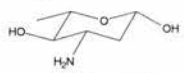
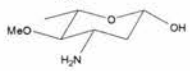
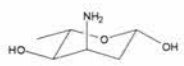
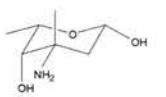
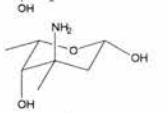
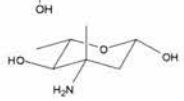
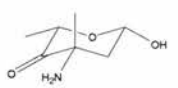
Sugar	Structure	Antibiotic
L-olivose		Orienticin B
L-rhodinose		UK69542
L-acosamine		Actinoidin A, MM47767, MM55256
L-actinosamine		Actinoidin, MM47767, MM55256
L-ristosamine		Actaplanin, avoparcin, chloropolysporin A-C, galacardin, symnonicin A-C, helvecardin, ristocetin
L-vancosamine		dechlorovancomycin, demethylvancomycin, vancomycin CDP-I, vancomycin
3-epi-L-vancosamine		A35512B
L-eremonsamine		Chloroorienticin, dechloroeremomycin, eremomycin, MM47761, MM49721, orenticin
4-oxovancosamine		A83850, balhimycin V, dechlorobalhimycin V, deglucobalhimycin, demethylbalhimycin, methylbalhimycin

Table reproduced from Hubbard and Walsh, 2003

The glycosylation of the aglycone is the last stage of antibiotic maturation; it occurs after the cross-linking of the backbone. The sugars are produced as dTDP sugars and then attached to the aglycone by a glycosyl transferase.

Even though the complete organic synthesis of vancomycin has been reported by two separate groups (McAtee *et al.*, 2002; Nicolaou *et al.*, 1999), the low yield and long production time of the complete organic synthesis make this route economically unfeasible on a commercial level. Instead, vancomycin-type antibiotics are produced by fermentation of the bacteria, that naturally produce them. Manufacture of higher affinity analogues to overcome or manage bacterial resistance will, therefore, most likely arise via synthetic modification of the fermented product or genetic engineering of the organism to produce altered products. The latter route is particularly attractive as it may allow considerable structural diversity and will utilise existing technology. Biological means of incorporating modified carbohydrates into antibiotics have been achieved in the syntheses of novobiocin (Albermann *et al.*, 2003; Freel Meyers *et al.*, 2003; Meyers *et al.*, 2003), spinosyn (Gaisser *et al.*, 2002), premithramycin (Trefzer *et al.*, 2002), oleandomycin (Rodriguez *et al.*, 2002) and elloramycin analogues (Rodriguez *et al.*, 2000).

A rational genetic engineering approach requires detailed knowledge of both the structural basis of substrate recognition and the biosynthetic reaction at each step of the biosynthetic pathway.

Chloroeremomycin is produced in *A. orientalis* (Figure 1.18), which had formerly been classified as a Nocardiae (see above), the largest and medically most important genus in the Actinomycetales (Lechevalier, 1986). The complete genome sequence is available for this species and the open reading frames (orf), thought to encode the proteins involved in the chloroeremomycin biosynthesis have been identified. As the variation and the specificity of the antibiotics are determined by the sugar moieties attached to the aglycone, we are interested in the enzymes catalysing the biosynthesis of the carbohydrate part. The starting material in the synthesis of L-epivancosamine is dTDP-glucose (Figure 1.19), which is converted to dTDP-6-deoxy-D-xylo-4-hexulose by RmlB, the second enzyme in the L-rhammnose biosynthesis pathway (Allard *et al.*, 2002). This material, in turn, is the starting sugar for the vancosamine biosynthetic pathway. The enzymes catalysing the conversion of dTDP-6-

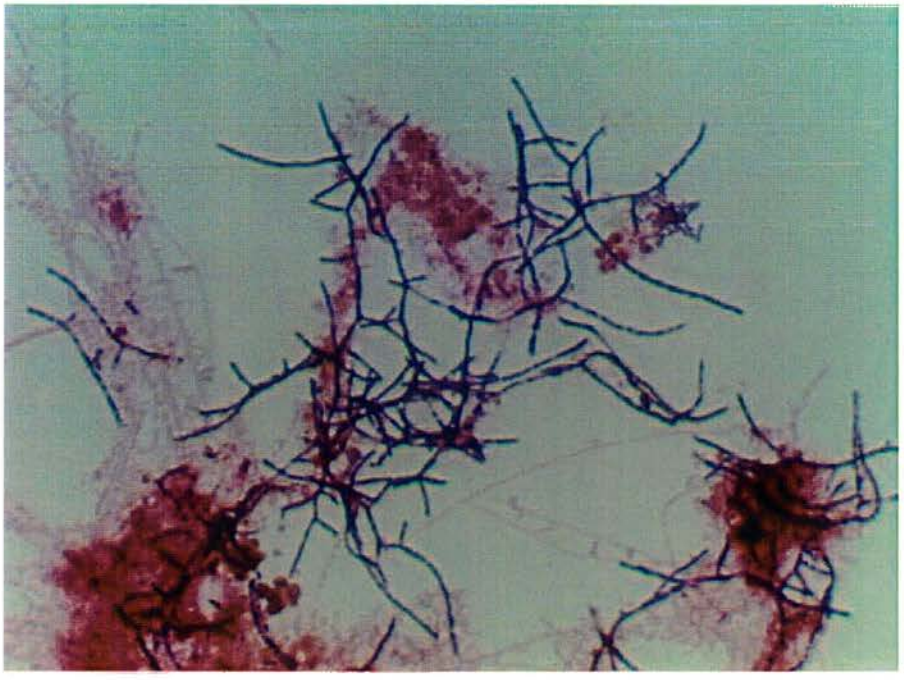


Figure 1.18 - ex-Nocardiae: Now classed as *Amycolatopsis orientalis*.

deoxy-D-xylo-4-hexulose to epivancosamine have been identified and are encoded by orfs 14, 23, 24, 25 and 26. The enzymes are known by a number of names, depending on the reference involved. From here onwards, they will be referred to as EvaA, EvaB, EvaC, EvaD and EvaE respectively. Table 1.3 summarises the enzyme names, their proposed functions (according to Chen *et al.*, 2000) and the corresponding orf in *A. orientalis*.

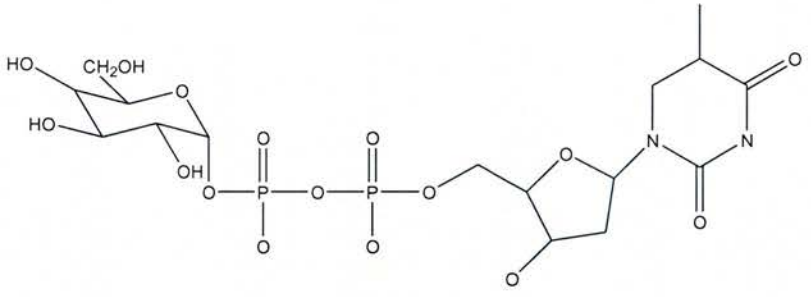


Figure 1.19: deoxy-Thymidine-5′-diphospho-D-glucose
(reproduced from Stryer, 1995)

Table 1.3 - Enzymes Involved in the Epivancosamine Pathway and Their Corresponding Orfs and Functions:

Enzyme name	Corresponding ORF	Proposed function
EvaA	23	2,3-dehydratase
EvaB	25	Aminotransferase
EvaC	14	SAM-methyltransferase
EvaD	24	Epimerase
EvaE	26	Reductase

The exact reaction pathway is still controversial, as two opposing views have been proposed in the literature (Chen *et al.*, 2000; Kirkpatrick *et al.*, 2000), the main difference being the order of the enzymes involved and therefore affecting each step of the pathway (Figure 1.20). The pathway proposed by Chen *et al.* has been verified by biochemical experiments and the identification of breakdown products of reaction intermediates. Kirkpatrick *et al.* based their proposition on similarities to known proteins. The differing order of the enzymes involved in each pathway results in changes of some details of the individual reaction steps. First of all, the substrate for

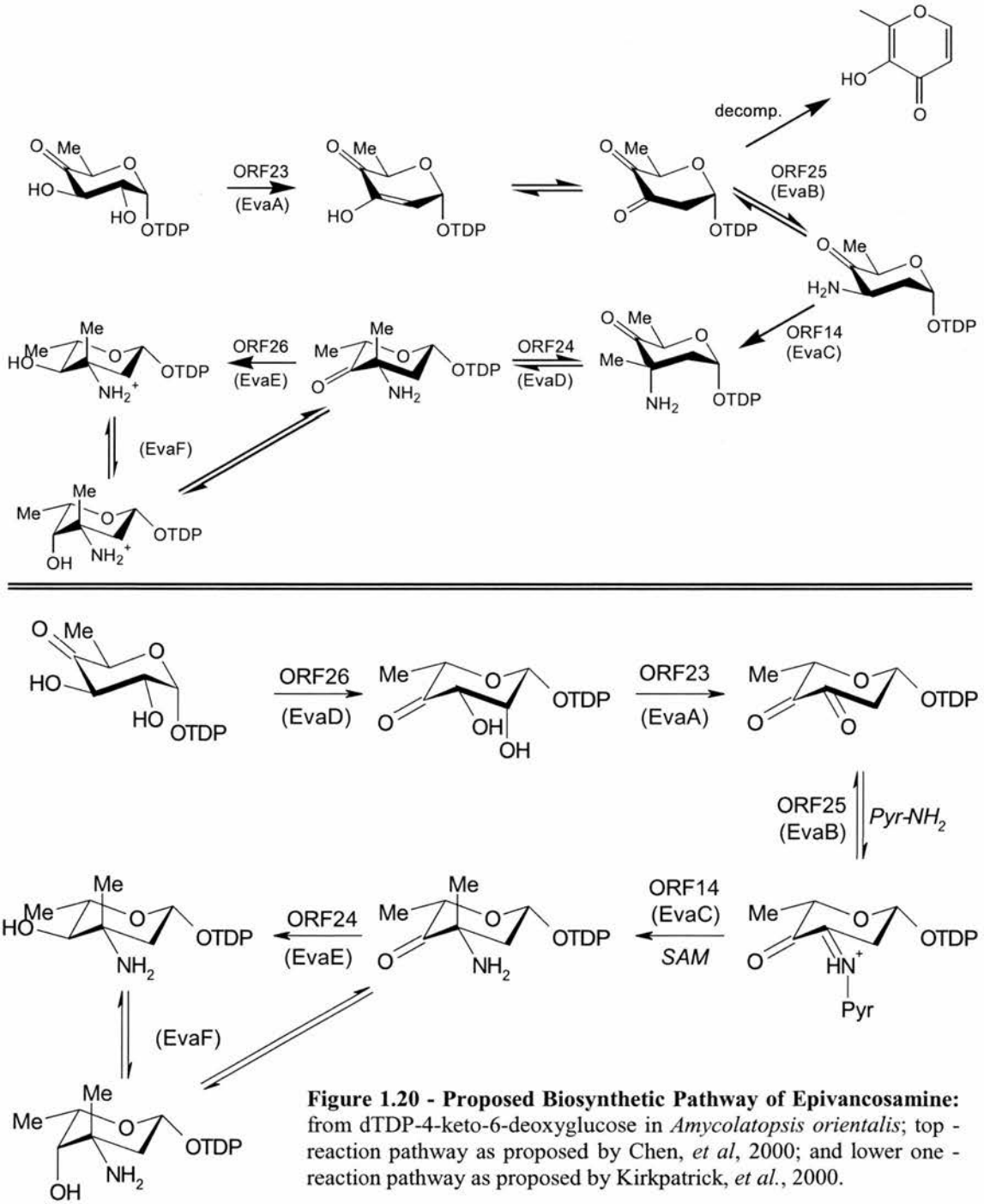


Figure 1.20 - Proposed Biosynthetic Pathway of Epivancosamine: from dTDP-4-keto-6-deoxyglucose in *Amycolatopsis orientalis*; top - reaction pathway as proposed by Chen, *et al*, 2000; and lower one - reaction pathway as proposed by Kirkpatrick, *et al.*, 2000.

each enzyme is slightly altered, as it is in a different processing state. For example, EvaD either catalyses the epimerisation of dTDP-3-amino-2,3,6-trideoxy-3*C*-methyl-D-*erythro*-hexopyranosyl-4-ulose, according to the pathway proposed by Chen *et al.*, or, according to Kirkpatrick *et al.*, it acts on dTDP-6-deoxy-D-xylo-4 hexulose. Not only is the substrate for epimerisation different but also the reaction itself is: the epimerisation occurs only on the C5′ according to the former group and on both the C3′ and C5′ according to the latter group's proposals. Both are possibilities and our studies may confirm one or the other pathway.

Another interesting difference between the two possible reaction schemes is the methylation step catalysed by EvaC. This enzyme is an S-adenosyl methionine (SAM) dependant methyltransferase, which catalyses the addition of a methyl group onto C3′ of the substrate. However, the major difference is the stereopositioning of the added methyl group: in the pathway elucidated by Chen *et al.*, the addition of methyl occurs in an equatorial fashion, whereas Kirkpatrick's scheme allows for an axial addition of the methyl group onto C3′. There is no precedent in literature for an equatorial methyl addition, as this should be prevented by the misalignment of orbitals.

It is highly likely that L-vancosamine, L-epivancosamine and 4-oxo-L-vancosamine (balhimycin) are synthesised in exactly the same manner, differing only in their final step of the enzyme pathway. Vancoasmine differs from L-epivancosamine only in the chirality at the C4′ position (equatorial in the case of chloroeremomycin and axial in the case of vancomycin). An enzyme, shown as EvaF in Figure 1.20, may convert the L-epivancosamine into L-vancosamine by converting the equatorial hydroxyl group on the C4′ into an axial one. The same idea may hold true for the production of 4-oxo-L-vancosamine in the biosynthesis of balhimycin. The interconversion of L-vancosamine to produce the different substituents for differing antibiotics is depicted in Figure 1.21.

The action of glycopeptides has been studied extensively and, like β-lactam antibiotics, they interfere with the cell wall synthesis in Gram-positive bacteria. However, unlike the latter antibiotic class, glycopeptides interfere with the second stage of the cell wall synthesis: the polymerisation of the phosphodisaccharide-pentapeptide lipid complex.

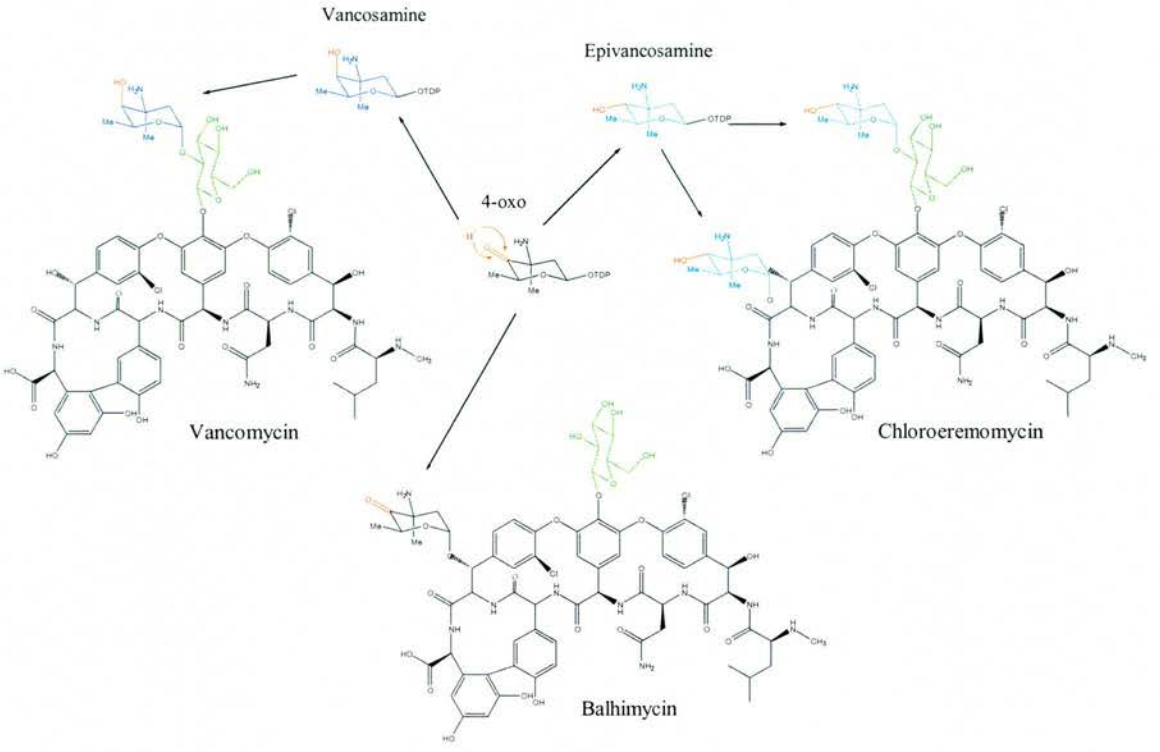


Figure 1.21 - The Interconversion of L-vancosamine: to L-epivancosamine and 4-oxo vancosamine to create the sugar substituents of the glycopeptide antibiotics vancomycin, chloroeremomycin and balhimycin respectively (figure adapted from Hubbard and Walsh, 2003).

In growing cells, transpeptidases catalyse the attack of an ε-amino group of a Lys at position 3 of a peptide chain onto the carbonyl group of D-Ala_4 in an adjacent chain. The new Lys-D-Ala isopeptide cross bridge is formed as the D-Ala_4-D-Ala_5 bond cleaves, which results in a net transpeptidation (Hubbard and Walsh, 2003).

Vancomycin, teicoplanin and other members of this family, with their rigidified cross-linked heptapeptide scaffolds, recognise the *N*-acyl-Ala_4-D-Ala_5 terminal D,D-dipeptide of un-crosslinked strands to form high-affinity complexes through 5 hydrogen bonds from the undersurface of the glycopeptide to the amide and carbonyl groups of the peptides (Figure 1.22). A space-filling model shows the complimentary surfaces leading to a tight fit of the antibiotic and the *N*-2,6-bis(acyl)-L-Lys-L-Ala-D-Ala ligand, and the methyl side chain of Ala_5 being packed tightly against the antibiotic cavity (Figure 1.23a). The D-Ala_4-D-Ala_5 amide linkage is sequestered in the complex with the glycopeptide and occluded from recognition by the cross linking transpeptidases (Sharman and Williams, 1997). Thus, the synthesis of peptidoglycan is blocked and membrane-bound lipid intermediates accumulate in the presence of vancomycin, making bacteria prone to osmotic pressure differences and lysis as the cell walls are destabilised during growth (Watanakunakorn, 1984). The sugar moieties on

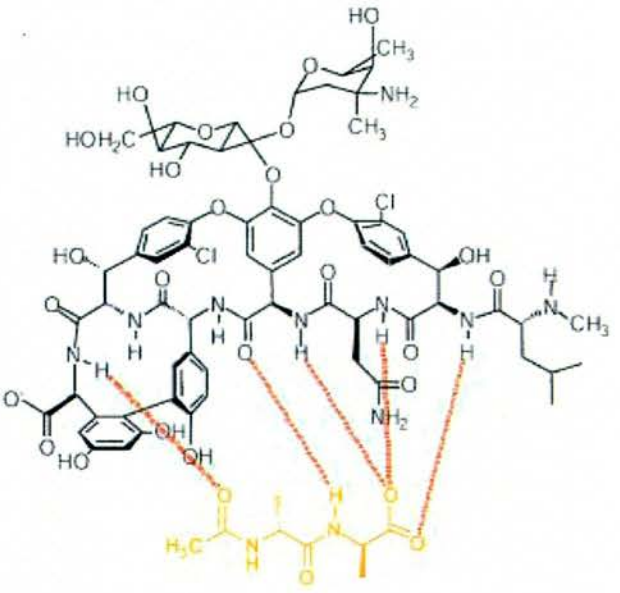


Figure 1.22 – Hydrogen Bonding for the Activity of Vancomycin: Hydrogen-bonds are formed between the glycopeptide antibiotic vancomycin and the free *N*-Acyl-D-Ala- D-Ala cell wall peptides (figure taken from Hubbard and Walsh, 2003).

vancomycin confer the antibiotic's specificity and selectivity and are necessary for the antibiotic dimerisation (Figure 1.23b) (Barna and Williams, 1984; Mackay *et al.*, 1994a; Mackay *et al.*, 1994b). Recent studies have also shown that the disaccharide may be bactericidal in its own right (Ge *et al.*, 1999; Hubbard and Walsh, 2003), and the aglycone by itself has a negligible antibiotic potential (Hubbard and Walsh, 2003). In addition to its primary mode of action, vancomycin-type antibiotics also alter the permeability of the cell membrane (Hancock and Fitz-James, 1964) and selectively inhibit RNA recognition (Jordan and Inniss, 1959).

Correlating with the increased use of glycopeptide antibiotics over the past decade, the resistance against these has increased dramatically and vancomycin-resistant Enterococci (VRE) are commonplace now. The latest CDC report from 2003 (CDC, 2003) estimated that over the past decade the hospital infections, caused by VRE, increased from 0.4 to 21.2% and 0.4 to 22.6% in non-intensive and intensive care units respectively. Enterococci were the first bacteria species found to have developed resistance to glycopeptide antibiotics; but since then, and with worse potential consequences, eight isolates of vancomycin-insusceptible *S. aureus* (VISA) and two isolates of vancomycin-resistant *S. aureus* (VRSA) have been found (Sieradzki *et al.*, 1999). The definitions of VISAs and VRSAs are based on laboratory tests assessing the minimum inhibitory concentrations (MICs) of the cultures for the

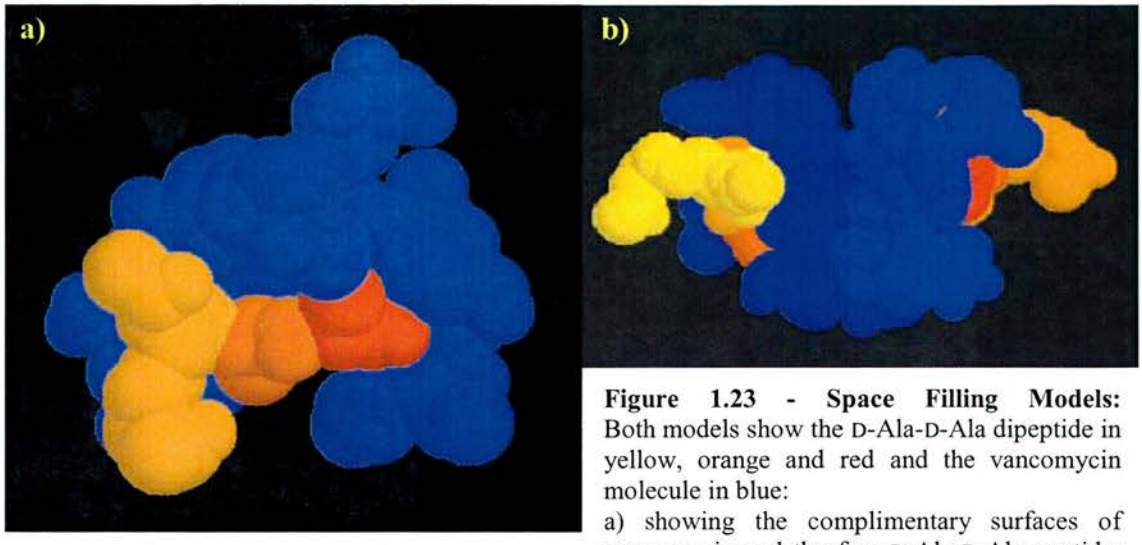


Figure 1.23 - Space Filling Models: Both models show the D-Ala-D-Ala dipeptide in yellow, orange and red and the vancomycin molecule in blue:
a) showing the complimentary surfaces of vancomycin and the free D-Ala-D-Ala peptide;
b) space filling model of antibiotic dimerisation (Hubbard and Walsh, 2003).

antibiotic under investigation (VISA 8 to 16μg ml^{-1} and VRSA ≥32μg ml^{-1}) (Appelbaum and Bozdogan, 2004). The first case was reported in 1997 in Japan and there have been more since (Sieradzki *et al.*, 1999). Luckily, so far all isolates have been susceptible to other antibiotics and the infections could be cleared.

Resistance to glycopeptide antibiotics occurs when the original target cell culture produces an altered peptide that cannot act as substrate for vancomycin-type antibiotics. The terminal D-Ala is replaced by D-Lactate, which is still a potent substrate for the transpeptidase in the organism but the potential for hydrogen-bonding with vancomycin is reduced (Figure 1.24). Several genes cause this build-up of D-Ala-D-Lactate peptides in the cell wall; generally a 5-enzyme cassette is associated with the resistance found in organisms. The resistance gene *van*X encodes a dipeptidase that hydrolyses D-Ala-D-Ala peptides, therefore removing the target of vancomycin. The gene product from *van*A, DdlM is a ligase that couples D-lactate to D-alanine to produce the depsipeptide D-Ala-D-Lactate. VanH is the enzyme that reduces pyruvate to lactate, producing more substrate DdlM. It is thought that these resistance genes were transferred to *S. aureus* from Enterococci through lateral gene transfer (Koonin *et al.*, 2001). The possibility of this occurring had been predicted as early as 1992, when Noble and co-workers demonstrated that *vanA*, as well as the other resistance genes, could be transferred to *S. aureus* from Enterococci species *in vitro* (Eom *et al.*, 2004; Noble, 1997; Noble *et al.*, 1992).

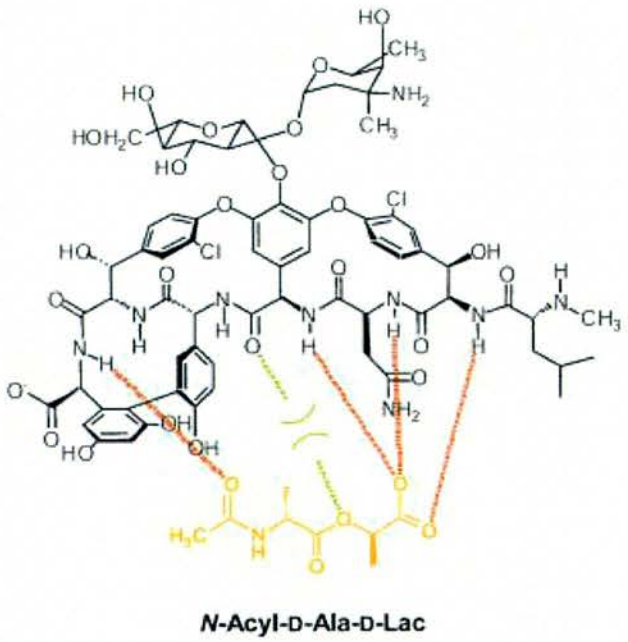


Figure 1.24 – Hydrogen Bonding Changes in Resistance: the hydrogen-bonding between vancomycin and the *N*-Acyl-D-Ala-D-Lac is reduced as compared to vancomycin and D-Ala-D-Ala.

The above were the first cases of resistance found to that antibiotic family in formerly susceptible bacteria; however, since then other bacteria have acquired the same resistance through lateral gene transfer and swapping of genetic material (Burke, 1998; Cassell and Mekalanos, 2001).

In addition to the emerging resistance to vancomycin and its derivatives, there are a number of drawbacks in the clinical use of certain members of the glycopeptide antibiotic family (Cha and Rybak, 2003). Only two of them - vancomycin and teicoplanin - are licensed for human use, and the former is a particularly bad drug as far as pharmacokinetics is concerned. Vancomycin is poorly absorbed when given orally and therefore has to be administered intravenously. Using this antibiotic orally has limited applications in treating certain infections of the gastrointestinal tract. The serum half-life of the active drug is about four hours, making frequent re-administrations necessary. Side effects are often severe, even though the manufacturing process has improved over the past few years to produce a purer product with a reduced toxicity profile (Tharp, 2004).

As mentioned above, genetic engineering may allow us to manufacture higher affinity analogues of vancomycin to overcome or manage bacterial resistance, as well as improve upon some of its inherent pharmacokinetic drawbacks. Such an approach, however, requires the detailed knowledge of both the structural basis of substrate recognition and the biosynthetic reaction at each step of the biosynthetic pathway (Low and Scheld, 1998). There are three main steps to the pathway: the activation step that yields the sugar nucleotide derivative, the transformation of the sugar moiety whilst attached to the nucleotide and, lastly, transfer of the monosaccharide from the sugar nucleotide to the acceptor (Liu and Thorson, 1994).

In naturally-occurring nucleotides the linkage between the activated sugar and the phosphate group involves the anomeric carbon. Various bases such as adenine, cytosine, guanine, uracyl and - as in the case under consideration - thimidine can participate in the sugar activation process as the corresponding nucleoside triphosphates. Thimidine is only found as the deoxyribonucleotide, whilst the others are ribonucleotides. The general reaction, as catalysed by pyrophosphorylases, occurs between the sugar-1-phosphate precursor and the nucleotidetriphosphate to yield the sugar-diphosphonucleoside (Escalante *et al.*, 1998).

The last step in the pathway is the transfer of the transformed sugar-nucleoside moiety onto the aglycone to yield the finished product. This step is catalysed by a variety of transferase enzymes, often positioned on the cell wall of the antibiotic-producing organism.

1.7 Aims

This thesis is concerned with the biosynthesis pathway which produces the unusual sugar nucleotide L-epivancosamine in A. orientalis. Two of the 5 enzymes in the epi-vancosamine pathway are being studied with the aim to elucidate their three-dimensional structures. The X-ray crystallographic structures will allow an insight into these enzymes' reaction mechanism. EvaD, according to literature a 3,5-epimerase is expected to be similar to the well established epimerase RmlC which is part of the L-rhamnose biosynthesis pathway. There is controversy of the exact vancosamine

biosynthesis pathway and knowing if EvaD is a competent 3,5 double epimerase will support one but not the other pathway. However, if it is not, the structure will give insights into its unique reaction mechanism and how it differs from RmlC. In order to gain an understanding of the reaction mechanism kinetic assays and deuterium incorporation studies are to be carried out.

The reaction mechanism for a 2,3-dehydratase (EvaA) is completely unknown and having an insight into its structure (especially its active site) may allow us to establish this unique reaction mechanism. This enzyme is expected to form a novel class of dehydratases which has many homologues in other Actinomycetales.

The final part of the thesis deals with an unconnected project: further structural characterisation of the KDPG aldolase from E. coli. The structure of this enzyme is well known but so far the reaction mechanism is not established beyond all doubts. To help gain a better understanding of this enzyme, X-ray crystallographic studies aiming at the structure determination with specific inhibitors and substrates are to be carried out. This will hopefully allow us to identify all amino acid residues that are involved in the reaction mechanism and establish a possible mechanism for the aldolase catalysed carbon-carbon bond formation.

Chapter 2

EvaD – 3,5 epimerase?

2.1 Summary

The first or fourth enzyme in the L-epivancosamine biosynthesis pathway in *A. orientalis* is an epimerase. This chapter describes the structure determination of this enzyme and biochemical assays carried out to further characterise the wild-type enzyme and some of its mutants.

As expected from its sequence similarity to RmlC, a 3,5 epimerase in the biosynthesis pathway of L-rhamnose, the three-dimensional structure of EvaD is essentially identical to RmlC. However, there was some controversy about the position of EvaD in the L-epivancosamine biosynthesis pathway and concurrently whether EvaD was carrying out the mono-epimerisation at C5′ only or the double epimerisation at C3′ and C5′ of its substrate.

The data presented in this chapter show that EvaD is capable of carrying out the double epimerisation as established in the archetypical RmlC enzyme activity assay, however with a 200 fold reduction in activity. In addition to those results, deuterium incorporation assays determined that EvaD preferentially incorporates deuterium at C5′ which we see as an indication for this enzyme to act as a mono-epimerase in the L-epivancosamine pathway. Structure determination of the wild-type enzyme and a co-complex with dTMP allowed modelling of the substrate analogues dTDP-xylose and dTDP-glucose into the active site, using RmlC from *S. suis* and *M. tuberculosis* as models. Comparing EvaD and its substrate analogue models to RmlC structures indicated one major change in the active site which we thought to be responsible for the difference between mono and double epimerisation as well as substrate recognition. The orientation of the active site Tyr, which is completely conserved in sequence and structure in RmlCs is changed in EvaD from the g+ rotamer found in RmlC homologues to the g- rotamer in EvaD. Subtle changes around the active site residues, particularly Met131 seemed to cause the orientation change in EvaD's catalytic Tyr. In an attempt to re-orientate this residue, mutants were made, one of which showed a 40% increase in double epimerisation activity in the kinetic assay despite displaying a 1.6 fold decrease in substrate affinity. In addition to its epimerisation activity towards dTDP-4-keto-6-deoxy-D-glucose, it exhibited the same amount of deuterium incorporation at C3′ and at C5′ as RmlC did at reaction equilibrium, indicating that this particular mutant may have re-estabilished double epimerisation.

2.2 Introduction

As discussed in the previous chapter, carbohydrates are ideally suited to molecular recognition events. By varying the stereochemistry of the hydroxyl substituents, the simple six-carbon, six-oxygen pyranose ring can exist as ten molecules. The biosynthetic pathways that lead to the production of these compounds are complex and involve multiple reaction steps such as oxidations, epimerisations, acetylations, dehydrations and reductions.

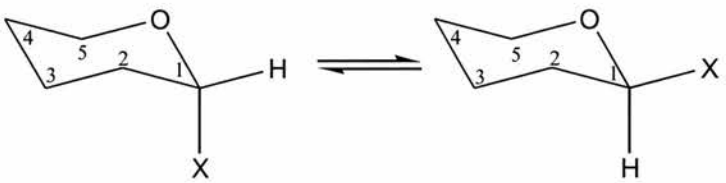


Figure 2.1 - Epimerisation Reaction: the inversion of configuration of an asymmetrically substituted carbon in a sugar (figure taken from Stryer, 1995).

Epimerisation is the inversion of configuration of an asymmetrically substituted carbon in linear or cyclic sugars (Figure 2.1); it can be thought of as a process of hydrogen removal from one face and its return to the opposite face of a central carbon. The reaction is very common throughout all branches of life: several different chemical strategies to invert chirality have evolved. Table 2.1 gives an overview of some enzymes and the reaction types they catalyse.

Table 2.1 - Epimerases: table depicting the type of epimerisation, enzyme examples and the site at which the respective epimerisation takes place.

Epimerisation Type	Enzyme	Epimerisation Site
By transient keto-intermediate	UDP-galactose 4-epimerase	C4
	ADP-L-glycero-D-mannoheptose 6-epimerase	C6
	CDP-tyvelose 2-epimerase	C2
Proton abstraction/re-addition	dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase (RmlC)	C3, C5
	GDP-4-keto-6-deoxy-D-mannose epimerase/reductase	C3, C5
	D-ribulose 5-phosphate 3-epimerase	C3
Nucleotide elimination and re-addition	UDP-*N*-acetylglucosamine 2-epimerase	C2
	N-acetyl-D-glucosamine 2-epimerase	C2
Carbon-carbon bond cleavage	L-ribulose-5-phosphate 4-epimerase	C4
Mutorotation (ring opening)	Galactose mutorotase	C1

The table is a summary of all references detailed on page 62 within the text

Hydride abstraction (oxidation) from one face of the stereocentre and hydride addition (reduction) to the opposite face of the same stereocentre involves the use of a redox active co-factor such as NAD or NADP. During turnover, a transient keto intermediate is generated and stabilised in the active site of the protein. This formation of keto-sugars helps to acidify the hydrogen α to the keto group, therefore lowering the pKa for the reaction (Figure 2.2a) (Thoden *et al.*, 1996a; Thoden *et al.*, 1996b; Thoden *et al.*, 1996c).

Epimerisation by proton abstraction and re-addition to the opposite face of the stereocentre involves the formation of a keto-intermediate which can be isolated from the reaction. The epimerisation occurs at the stereocentres α to the keto-group and not at the keto-group directly. In the case of RmlC (dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase), an epimerase involved in the rhamnose pathway, two epimerisations are catalysed, one on C3′ the other on C5′. The reaction requires that the hydrogens attached to C3′ and C5′ are removed from one face and then replaced on the opposite face. These reaction steps cannot function in a concerted fashion as the pKa of the remaining proton is significantly increased after one proton has been removed. Thus, the two carbons must be epimerised sequentially (Figure 2.2b) (Giraud *et al.*, 2000; Graninger *et al.*, 1999).

The interconversion of UDP-*N*-acetyl-glucosamine and UDP-*N*-acetylmannosamine involves an initial anti-elimination of UDP, generating the 2-acetamidoglucal intermediate, and a subsequent syn-addition of UDP to yield the epimerisation product (Figure 2.2c) (Morgan *et al.*, 1997; Sala *et al.*, 1996).

The epimerisation reaction that is carried out by the L-ribulose-5-phosphate 4-epimerase, proceeds via a carbon-carbon bond cleavage. The reaction requires a divalent cation as a co-factor, and it proceeds without the incorporation of solvent derived oxygen or hydrogen. The enzyme is structurally similar to the class II aldolase L-fuculose-1-phosphate aldolase, suggesting that epimerisation proceeds through an aldol reaction, i.e. a carbon-carbon bond cleavage (Figure 2.2d) (Deupree and Wood, 1975; Deupree and Wood, 1972; Lee *et al.*, 2000; Salo *et al.*, 1972).

The galactose mutorotase enzyme catalyses the ring opening of cyclic sugars and so allows them to interconvert between the α and β anomer in their linear form. Ring closure prevents the interconversion between α and β forms and thus creates a statistical mixture of both anomers (Figure 2.2e) (Beebe and Frey, 1998; Hucho and Wallenfels, 1971).

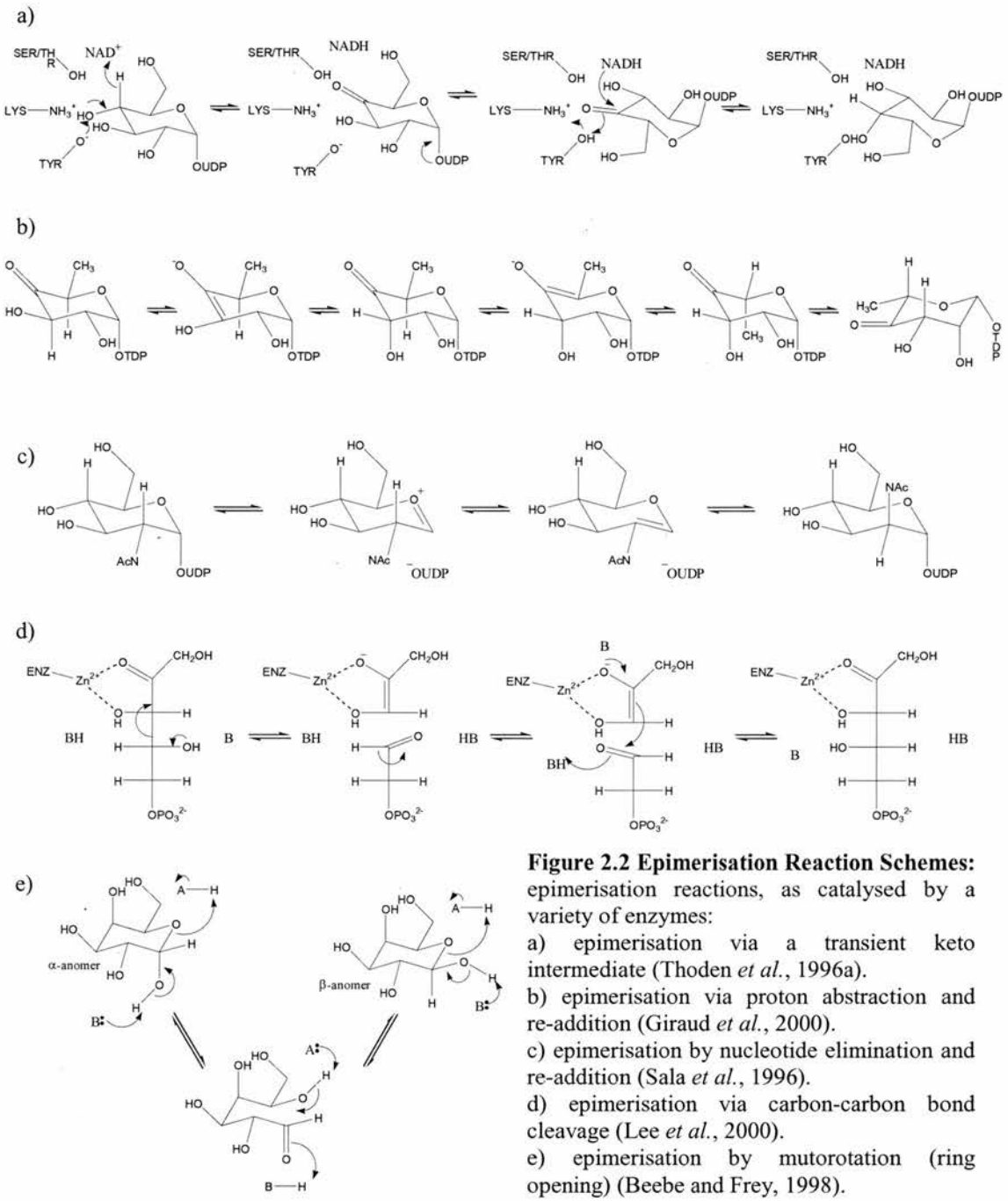


Figure 2.2 Epimerisation Reaction Schemes: epimerisation reactions, as catalysed by a variety of enzymes:
a) epimerisation via a transient keto intermediate (Thoden *et al.*, 1996a).
b) epimerisation via proton abstraction and re-addition (Giraud *et al.*, 2000).
c) epimerisation by nucleotide elimination and re-addition (Sala *et al.*, 1996).
d) epimerisation via carbon-carbon bond cleavage (Lee *et al.*, 2000).
e) epimerisation by mutorotation (ring opening) (Beebe and Frey, 1998).

The biosynthesis of epivancosamine from the RmlB product dTDP-6-deoxy-D-xylo-4-hexulose involves an epimerisation reaction either as its first or its fourth reaction step. The enzyme that has been identified to catalyse such a reaction in the chloroeremomycin biosynthetic pathway in *A. orientalis* is the dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase, EvaD. It is encoded by orf24 as identified by DNA sequencing of the organism's complete genome (Chen *et al.*, 2000). Its sequence is highly similar to that of RmlC which carries out the double epimerisation at C3′ and C5′ via a proton abstraction/re-addition mechanism. RmlC is the third enzyme in the biosynthetic pathway of L-rhamnose in bacteria and plants (Giraud *et al.*, 2000; Giraud

and Naismith, 2000) and the structures of several homologues from *Mycobacterium tuberculosis* (*M. tuberculosis*), *Streptococcus suis* (*S. suis*), *Pseudomonas putida* (*P. putida*), *E. coli* and others have been determined (Dong *et al.*, 2003; Giraud *et al.*, 2000). Figure 2.3 shows a phylogram relating the homologues from different species to one another. The tree was constructed by the web server CLUSTALW hosted at the EBI in Cambridge (Thompson *et al.*, 1994), using the amino acid sequences from Swiss-Prot/TrEMBL (Apweiler and Bairoch, 2004; Bairoch *et al.*, 2004; O'Donovan *et al.*, 2002) as input. The tree shows that the RmlCs from Streptococci species are least related to the other RmlCs. Nevertheless, RmlC from *S. suis,* as an example of these, is a completely competent and efficient RmlC indicating that EvaD is probably catalytically competent too. EvaD is most similar to RmlC from *M. tuberculosis*, displaying a sequence identity of 41% and a sequence similarity (allowing for conservatively substituted amino acids) of 57%. The residues that are known to be crucial to the catalytic activity of RmlC, His63, Lys73 and Tyr133 (EvaD numbering), are fully conserved in sequence throughout all these homologues.

Despite EvaD being similar to RmlC, a number of questions remain, which cannot be predicted from sequence similarities. The placement of this enzyme in the L-epivancosamine biosynthesis pathway is controversial as two opposing views have been published (Chen *et al.*, 2000; Kirkpatrick *et al.*, 2000). Concurrent with this, is the question if EvaD is a mono or di-epimerase, it may carry out only the C5′ mono epimerisation as predicted by Chen *et al.* or it may functionally belong to the RmlC enzyme class and carry out the archetypical double epimerisation at C3′ and C5′ of the carbohydrate substrate. The aim of this study was to clarify the exact function of EvaD through structural and biochemical analyses and identify the basis for substrate differentiation between EvaD and RmlC.

Phylogram

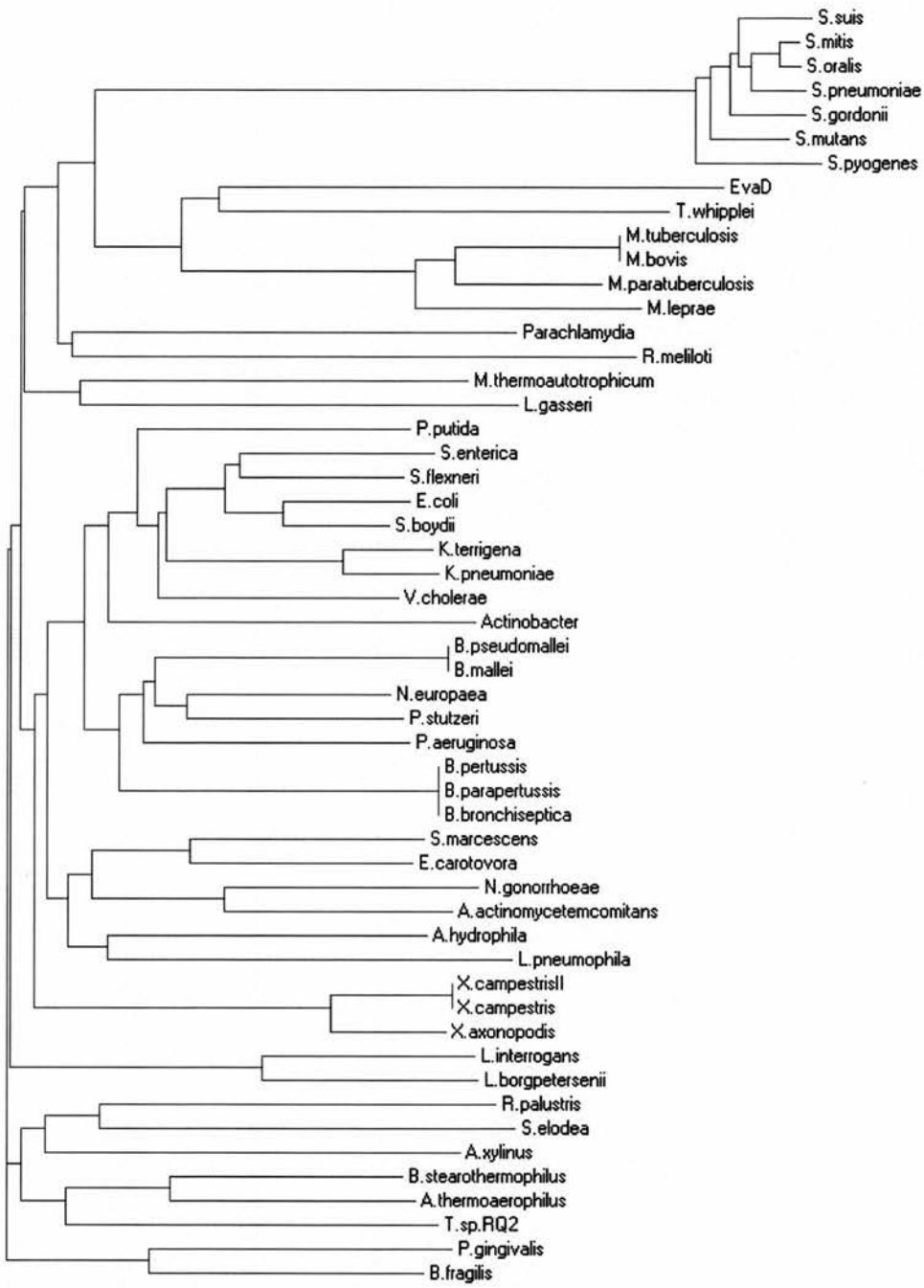


Figure 2.3 – EvaD Phylogram: showing the relationship between EvaD and the RmlC homologues from a number of different bacteria species (created by ClustalW – EBI Cambridge).

2.3 Materials and Methods

2.3.1 Site-directed Mutagenesis

All molecular biology protocols were carried out according to standard procedure as described in Sambrook and Russell, 2001. Orf24 (the gene encoding EvaD) cloned into the multiple coning site (MCS) of pET16b+, using the restriction enzymes *Nde* I and *Xho* I, was obtained from our collaborators Chris Walsh and Michael Burkart (Harvard Medical School).

The following mutants of EvaD were produced by site-directed mutagenesis; the primers were custom-made by Oswel (now Eurogentec Ltd.). Appendix A details the sequences of all primers designed.

- I50N
- H63A
- M131F
- Y133F
- L135A
- M131F/L135A
- I50N/M131F/L135A

In vitro site-directed mutagenesis, using the native EvaD double stranded miniprep plasmid as template, was carried out according to the Stratagene QuickChange® site-directed mutagenesis protocol. Figure 2.4 gives a graphical overview of the steps that were carried out according to that protocol. A standard polymerase chain reaction (PCR) in 100µl reaction volume containing forward and reverse primers (1.5µM each), 10ng vector template, 0.5mM nucleotide triphosphate (dNTP), 2U DNA polymerase from *Pyrococcus furiosus* (*Pfu* DNA polymerase, Promega), 10µl 10x PCR buffer and dH_2O was carried out under the following conditions:

Program:

95°C → 30sec

95°C → 30sec
55°C → 1min
68°C → 6.5min
} 25 cycles

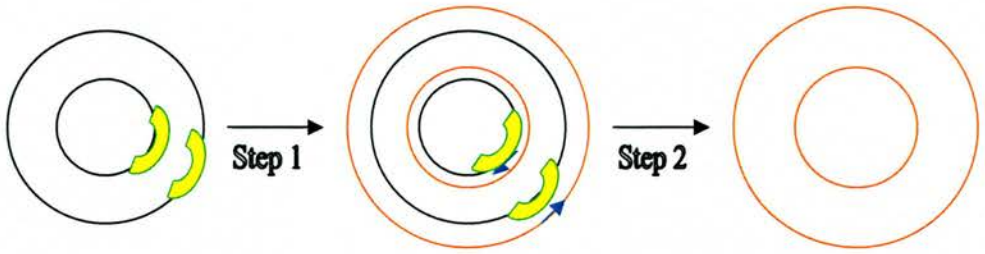


Figure 2.4 – Site-directed Mutagenesis Reaction Scheme: site-directed mutagenesis using primers (yellow) to the parent plasmid (black) that was to be mutated. Step 1: PCR - the primers were elongated into the synthetic, non-methylated DNA containing the desired mutation (red). Step 2: *Dpn* 1 was used to cut the methylated and hemimethylated parent DNA so only synthetic DNA was left to transform competent cells with.

Following amplification, the mixture of parental/newly formed DNA was cooled down to 37°C and treated with 2U *Dpn* 1 (Promega) for 1 hour at this temperature. The endonuclease specifically cleaves methylated and hemi-methylated DNA, i.e. the parental plasmid. Subsequently the mix was used to transform single shot TAM-1® *E. coli* cells (ActiveMotif, Belgium) using the standard heat shock protocol (see section 2.3.2). Successfully transformed cells were selected on ampicillin containing Luria Bertani (LB) agar, amplified in 10ml LB broth and the plasmid isolated using the Qiagen miniprep kit® (spin column method using a microcentrifuge). The DNA was eluted off the spin columns using 30µl dH_2O. The triple mutant was obtained by carrying out the I50N single site-directed mutation in the same way as described above, but using the M131F/L135A double mutant as template.

The Y133F mutant was created in a two-step route since the standard PCR described above, and sequential PCRs (starting with only one primer present in the initial reaction mix, and adding the second primer 10 cycles into the reaction) failed to yield any amplified product. Figure 2.5 presents an overview of the reaction steps involved. A first round of PCRs was carried out, using the same reactants and reaction conditions as the above PCR, but different primers: one reaction made use of the Y133F forward and T7 reverse primers, the other, the Y133F reverse and T7 forward primers. In this way, the region across the gene, including the mutation, was amplified all the way to the vector encoded T7 forward and reverse primer sites. The products were gel extracted using the Qiagen gel extraction kit® (centrifuge method using spin columns) and eluting the DNA off the columns into 30µl dH_2O. A second round of PCR was carried out in a 100µl reaction volume using 10ng of each first-round PCR product as template in addition to 1.5µM (each) T7 forward and reverse primers, 0.5mM dNTPs,

2U *Pfu* DNA polymerase, 10µl 10x PCR buffer and dH_2O. Subsequently, the product was gel extracted into 30µl dH_2O and then restriction digested. The double digest was carried out in 20µl reaction volume (containing 10µl of the gel extract) at 37°C for 1 hour using 2U of each *Nde* I and *Xho* I (both Promega) under standard buffer conditions. After the enzymes had been heat-denatured at 65°C for 20min, the insert was ready for ligation. The parent vector was double digested as above, but was gel extracted to remove the original insert DNA. Standard overnight ligations at 16°C were set up in 50µl reaction volumes containing vector and insert in 1:2 and 1:5 ratios, as well as, 2U T4 DNA ligase (Promega), 5µl 10x T4 DNA ligase buffer and dH_2O. The product was used to transform TAM-1® cells, which were selected on ampicillin containing LB agar plates, grown up in LB broth and the DNA extracted into 30µl dH_2O using the Qiagen gel extraction kit®. All vectors were sent for DNA sequencing to verify the mutations introduced.

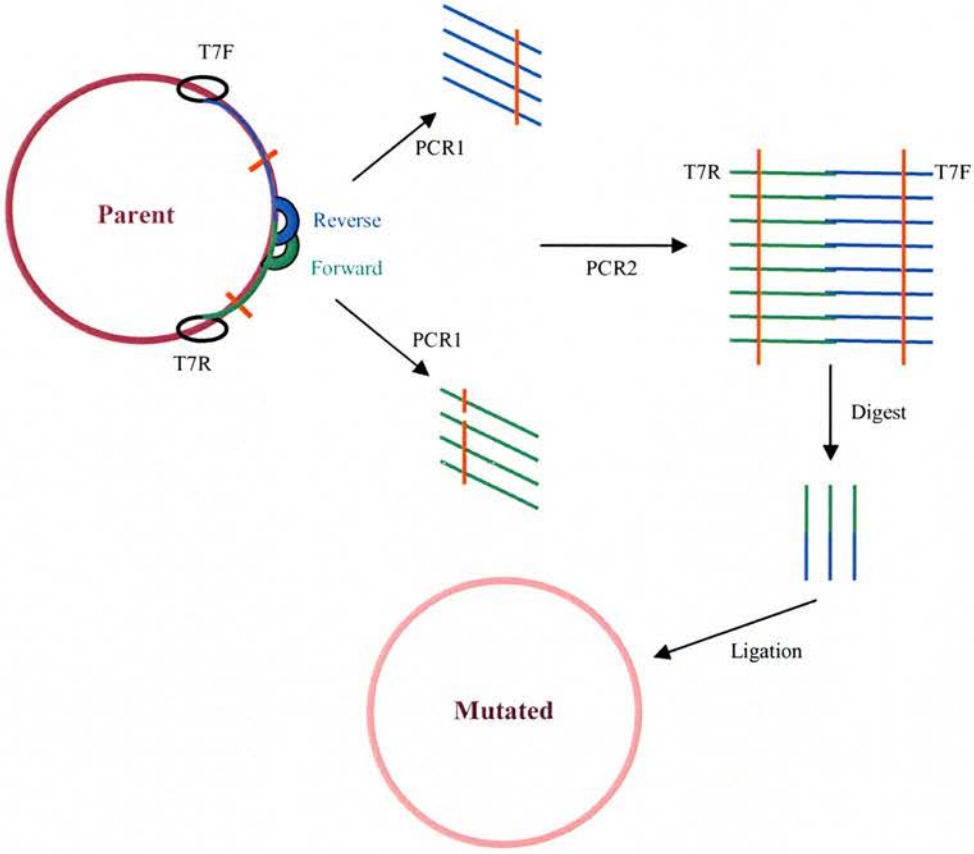


Figure 2.5 – Scheme Detailing the Two-step Site-directed Mutagenesis Reaction.

2.3.2 Transformations and Minipreps

Chemically competent cells were bought from ActiveMotif, Novagen or Promega. To transform these, 50-100µl cells were incubated with 200ng vector DNA, or 5 to 15µl ligation mix, for 1 hour on ice. The cells were heat-shocked at 42°C for 30sec and incubated on ice for another 5min. Subsequently, 250µl LB broth were added, the mix incubated at 37°C for 1 hour at 200rpm shaking, and then plated out on antibiotic containing LB agar plates to select successful transformants. The plates were incubated for 12 hours at 37°C.

Single colonies were picked and grown up in 10ml LB broth containing 50µg ml^{-1} antibiotic for subsequent DNA extraction. The cells were harvested by centrifugation (48,000*g*, 15min) and the DNA extracted as described in the protocol for the Qiagen miniprep kit®. The spin column and microcentrifuge method was used and the DNA was eluted off the column into 30µl dH_2O. To check the DNA quality, an analytical 1%w/v agarose gel containing 0.005%v/v ethidium bromide was run and the DNA was made visible under UV light. Purity of the miniprep DNA was checked by calculating its absorbance ratio at $\lambda = 280$ to $\lambda = 260$nm (laboratory standard: ratio 1.8 for pure DNA). The DNA concentration was calculated from its absorbance at $\lambda = 260$nm (Absorbance of 1.0 at $\lambda = 260$nm is equivalent to 50µg ml^{-1})

2.3.3 DNA Sequencing

All DNA sequencing was performed by The Sequencing Service (School of Life Sciences, University of Dundee, Scotland, www.dnaseq.ac.uk) using Applied Biosystems Big-Dye Ver. 3.1 chemistry on an Applied Biosystems model 3730 automated capillary DNA sequencer. For sequencing the EvaD mutants, T7 forward and reverse primers, as supplied by The Sequencing Service, were used.

2.3.4 Expression and Purification

Wild-type EvaD and all mutants were expressed and purified using the same methodology. EvaD in pET16b+ was used to transform chemically competent BL21(*DE3*) cells, and small scale (10ml) expression trials were performed. After determination of the optimum conditions, expression was scaled up to 500ml cultures. LB broth, containing 50μg ml^{-1} carbenicillin, was inoculated with 10ml overnight starter culture of BL21(*DE3*) cells containing EvaD in pET16b+. The culture was grown at 37°C, 200rpm shaking until the optical density measured at λ = 600 (OD_{600}) reached 0.8. Protein expression was induced using 1mM Isopropyl-β-D-thiogalactopyranoside (IPTG), temperature and agitation were kept constant. After 4 to 5 hours the cells were harvested by centrifugation at 13,000*g* for 20min (Beckman centrifuge Avanti J20-XP, JL8.100 rotor).

The cell pellet was re-suspended in buffer (100mM NaCl, 20mM Tris, pH7.5) and frozen overnight at -20°C. After thawing the sample, it was incubated on ice for one hour with 5mM Phenylmethylsulfonyl fluoride (PMSF), 2mM Dithiothreitol (DTT), 100μg ml^{-1} lysozyme and 20μg ml^{-1} deoxyribonuclease I (DNase). In addition to freeze thawing, the cells were lysed by sonication: five 30sec bursts of sonication at 18microns, with 30sec intervals of cooling on ice, were judged sufficient for complete cell lysis. Soluble and insoluble material were separated by centrifugation at 48,000g for 30min (Beckman centrifuge Avanti J20-XP, JA 25.5 rotor). As the pET16b+ vector encodes an N-terminal His·Tag, EvaD could be purified in two chromatographic steps – anion exchange and metal chelating chromatography. The dialysed sample (dialysis against an excess of 20mM Tris-HCl, pH7.5) was applied to a pre-equilibrated weak anion exchange column (DEAE Fast Flow Sepharose™ (Amersham Biosciences)). The column was washed with 5 column volumes (CV) of the equilibration buffer and the protein eluted using a linear NaCl gradient (0 to 2.5M). The EvaD containing fractions, as judged by sodium dodecylsulfate - polyacrylamide gel electrophoresis (SDS-PAGE), were pooled, concentrated, dialysed against buffer (20mM Tris-HCl, pH7.5, 100mM NaCl) and then applied to a metal chelating column. After charging the column (Metal chelating Fast Flow Sepharose™ or Ni-NTA™ metal chelating resin (Amersham Biosciences or Invitrogen) with $NiSO_4$, it was equilibrated with the same buffer. A linear imidazole gradient (0 to 1.0M) was used to elute EvaD.

The purity of all proteins was judged by Coomassie™ Blue (0.1% Coomassie blue R-250, 40% ethanol, 10% acetic acid) stained SDS gels, the integrity verified by matrix-assisted laser desorption ionisation time of flight (MALDI-TOF) mass spectrometry and their concentrations determined by Bicinchoninic Acid (BCA) (Smith and Krohn, 1985; Wiechelman, 1988) as well as Bradford (Bradford, 1976) assays.

The pure protein was dialysed against 20mM Tris-HCl, pH7.5, concentrated to 10mg ml^{-1} and then used for sparse matrix crystallisation trials (Hampton Research – Crystal Screens 1 and 2, Index 1 and 2, SaltRX, PEG/Ion; Emerald BioStructrures - Wizard 1 and 2) in sitting drop plates. The initial 'hits' were optimised by altering protein, salt and Polyethylene glycol (PEG) concentrations, PEG molecular weight, pH and temperature, as well as the vapour diffusion method (sitting drop, hanging drop and under oil crystallisation).

2.3.5 Circular Dichroism Spectroscopy

To verify protein folding, wild-type EvaD and all mutants were analysed by circular dichroism (CD) spectroscopy and the spectra compared to one another. The CD scans were carried out in the Protein Characterisation Facility, Institute of Biomedical & Life Sciences, University of Glasgow. Spectra were recorded on protein samples contained in 20mM Tris-HCl, pH7.5 and at protein concentrations of 1mg ml^{-1}.

2.3.6 Data Collection and Processing

A 0.5 x 0.3 x 0.1mm crystal was equilibrated in a cryo-protectant solution of 20%v/v glycerol in mother liquor, mounted in a cryo-loop and flash-frozen in a nitrogen gas stream. A 1.5Å resolution native X-ray dataset was recorded on BM14, at the European Synchrotron Radiation Facility (ESRF), at a wavelength of $\lambda = 0.933$Å using the ADSC Q4 CCD area detector (Area Detector Systems Corporation, Qantum 4, Charge Coupling Device). The beamsize was 50µm and the crystal to detector distance was 130mm. Data were recorded as 110 non-overlapping 15sec 1° oscillations.

The reflections were indexed in an orthorhombic Bravais lattice and integrated and merged in the Laue group P222 using DENZO SCALEPACK (Otwinowski and Minor, 1997). The data were scaled and re-indexed using SCALA, version 3.2.1 (Evans, 1993) as implemented in the Collaborative Computing Project Number 4 (CCP4) Program Suite (Bailey, 1994).

A similar crystal to the native one was equilibrated for 15min in cryo-protectant solution (20%v/v glycerol in mother liquor), containing 10mM dTDP-glucose. The crystal was then mounted in a cryo-loop and flash frozen in a nitrogen gas stream. A data set to 1.4Å resolution was recorded as 200 non-overlapping 15sec 1° oscillations at the ESRF, ID 14-2 beamline at a wavelength of λ = 0.932Å. The beam size was 50µm and the crystal to detector distance 130mm.

The reflections were indexed in an orthorhombic Bravais lattice and integrated and merged in the Laue group P222 using MOSFLM, version 6.2.4 (Leslie, 1992). The output reflection file was re-indexed and scaled in the Scala, version 3.2.1 (Evans, 1993).

A crystal of the M131F/L135A double mutant of the dimensions 0.2 x 0.2 x 0.2mm was equilibrated in cryo-protectant solution (20%v/v glycerol in mother liquor), mounted in a cryo-loop and flash frozen in a lnitrogen gas stream. A data set was collected in-house as 180 non-overlapping 10min 1° oscillations. The X-ray beam λ = 1.542Å was produced by a rotating copper anode; the detector was a Rigaku RaxislV++ image plate and the crystal to detector distance was 160mm. The data were collected to a resolution of 1.74Å, but had to be cut to 2.08 Å as the completeness dropped sharply in the higher resolution shells. Including all data to 1.74 Å decreased the overall completeness to 74%, and the three highest resolution shells were only 6.7, 20.7 and 34.5% complete respectively.

The reflections were indexed in space group c2 and integrated, merged and scaled using d*TREK, version 8.0 (CrystalClear 1.3.5©, Rigaku). The merged data was chosen to be outputted in the SHELX (hkl) format (Sheldrick, 1992), which could then be converted into the mtz file format using the f2mtz conversion program by M. Kjeldgaard in the CCP4 Program Suite (Bailey, 1994).

2.3.7 Structure Solution

Starting phases for the native data were calculated by molecular replacement using RmlC from *Salmonella enterica* serovar Typhimurium (*S. typhimurium*) as a search model (Protein Data Bank (PDB) entry: 1DZR); it displayed a 33% sequence identity (51% sequence similarity) with EvaD. The program MolRep (Vagin and Teplyakov, 1997), as implemented in the CCP4 Program Suite (Bailey, 1994), was used for molecular replacement using all the data in the resolution range 15 to 4Å. The initial phasing model was used in its monomeric form and as a partial polyalanine model (all non-identical residues had been substituted by alanine side chains). This had to be done as several molecular replacement attempts with the dimer, as well as with models containing all side chains, did not yield a clear solution. Two molecules per asymmetric unit (au) were searched for, as the active enzyme was thought to be a homodimer and the solvent content estimation (Matthews, 1968) agreed with that prediction. The cross-rotation search found a number of peaks, which were then subjected to a translational peak search, yielding potential solutions. The best one, as calculated by MolRep, was used to calculate the new coordinates of the model. This one, and its symmetry related objects according to its space group, were displayed in the graphics program O, version 7.0 (Jones *et al.*, 1991a) to check the crystal packing.

The dTDP-glucose soaked crystal had the same unit cell as the native one; therefore, molecular replacement was unnecessary. The model of the fully refined apo-structure was simply refined against the complex derivative data set.

The data for the M131F/L135A crystal displayed a different unit cell compared to that of the wild-type protein. Thus, molecular replacement had to be carried out again, using MolRep with one monomer of EvaD as search model, since the unit cell volume was only half that of the native protein crystal. The apo-model should fit the new data well; therefore, the rotation and translation searches were carried out in one step. All data between 15 and 4.0Å resolution were used and MolRep was asked to expect a 0.99 fraction completeness of the model with 0.99 fraction similarity to the input structure. The best solution was used to calculate the new model coordinates which were subsequently displayed in O, version 7.0, together with the model's symmetry related objects to check for crystal packing and generation of the homodimer.

2.3.8 Structure Refinement

Prior to all refinement steps 5% of the data were excluded for cross-validation (Brünger, 1992; Brünger, 1997; Kleywegt and Brünger, 1996) using the program uniquify as implemented in the CCP4 Program Suite (Bailey, 1994). This is done to check the validity of each step and the overall refinement; the excluded data is kept in a separate data pool and is not used for any refinement. Thus, if the R-value of this pool of data, denoted by R_{free}, decreases, it must be because the model has become better, while for the other 95% of data there is the danger that the R-value decreased because of overfitting of the free parameters of the model, leading to a false minimum.

The initial phases from molecular replacement were used for autotracing by the program ARP/wARP (Morris *et al.*, 2002) using 100 cycles of building with 10 refinement cycles in between. The model together with single and double difference electron density maps (F_o-F_c and $2F_o$-F_c) which had been calculated using FFT (Immirzi, 1966; Read and Schierbeek, 1988; TenEyck, 1973) were displayed in O, version 7.0 (Jones *et al.*, 1991b). The maps were averaged across the dimer using the programs MAMA (Kleywegt and Jones, 1993) and RAVE (Kleywegt and Jones, 1994). The former was employed to identify the monomers and create masks for each one; the latter was used to average the electron density across these masks. The improved maps were displayed in the graphics program O and 11 additional residues were fitted into the density manually. The five C-terminal residues (VPVST) in each monomer could not be located at this stage. Initial structure refinement was carried out in Refmac, version 5.0 (Murshudov *et al.*, 1997) using the rigid body refinement option, and the R_{factor} and R_{free} dropped to 29.5% and 33.4% respectively. After this refinement step the next residue in the sequence, a Val, could be located in each monomer. Restrained refinement using the weighting term, which defines the weight between X-ray and geometric part of the refinement residual, at 0.2 further improved the model and allowed the placement of another amino acid (Pro202) at the end of the polypeptide chain of each monomer. Subsequently, TLS refinement (Howlin *et al.*, 1989; Winn *et al.*, 2001) was employed to further improve the model. TLS refinement allows the reduction in displacement parameters, which would increase significantly by moving from individual isotropic to individual anisotropic refinement; it would yield a six-fold increase in parameters which might be problematic if a realistic data-to-parameter ratio could not be retained. TLS

parameterisation describes the translation, libration and screw-rotation displacements of a pseudo-rigid body, and therefore reduces the number of parameters by using collective variables rather than independent atomic variables. Even with this further refinement, and therefore better phases in the C-terminal region, no more residues could be positioned into the model as the density for the final three residues appears too disordered. Lastly, water and glycerol molecules were added into the structure and the anisotropic B-factor was refined.

The dTDP-glucose soaked model was refined in Refmac, 5.0 in the same way as described above for the apo-model. First of all, the same set of indices for cross validation was imported into the soaked dataset from the apo-dataset. Then rigid body and restrained refinement cycles were carried out and the model was continually checked manually against the calculated electron density maps. After the TLS refinement step, the model was displayed in O, version 7.0, and glycerol molecules were placed into the F_o-F_c density. The pyranose ring could not be located in the active site, but dTMP was found to be present, and it could be placed into the derivative model. Then water molecules were added and their positions confirmed in O. The completed model was fully refined using the anisotropic B-factor refinement option in Refmac 5.0.

The data for the double mutant was refined similarly to that of the wild-type EvaD after 5% of the data had been excluded for cross validation. Initial restrained refinement was carried out on the apo-model; subsequently, the model was displayed in O, version 7.0 and the final three residues (Val, Ser, Thr) missing from both the wild-type and derivative models could be placed into electron density, after 10 cycles of rigid body refinement in Refmac, 5.0 had been carried out. Further restrained refinement was carried out which allowed the placement of glycerol into the model. The TLS parameters were refined as chain A only, and refinement using the TLS option was carried out before water molecules were added to the model. A final round of refinement was carried out, but anisotropic B–factors could not be refined as the resolution of the dataset was too low.

The fully refined models, apo, complex and mutant EvaD, were validated using Procheck (Laskowski *et al.*, 1993) and WhatIf, version 5.0 (Hooft *et al.*, 1996) to check stereochemical quality of the model and SFcheck (Vagin *et al.*, 1999), which checks the model quality against the structure factor data.

2.3.9 Kinetic Assays

For the kinetic assays, all enzyme concentrations were determined by Bicinchoninic Acid (BCA) (Smith and Krohn, 1985; Wiechelman, 1988) and Bradford assays (Bradford, 1976) and their purity judged by Coomassie™ Blue stained SDS gels. Reagents were bought from Sigma-Aldrich, stock solutions made up and immediately frozen at –20°C, to keep all variables at a minimum.

The kinetic analysis itself was based on the RmlC assay, first published by Graninger *et al.*, 1999 and refined in our laboratory as published by Blankenfeldt *et al.*, 2002 and Iain Kerr (University of St. Andrews, PhD Thesis 2003). The three-enzyme coupled assay was modified and RmlC was replaced with wild-type or mutant EvaD; the other enzymes involved (RmlB and RmlD), were kept in excess to ensure that EvaD - or RmlC in the control assays - was rate-limiting. Figure 2.6 gives an overview of the reaction scheme for the kinetic analysis using the three enzyme coupled assay.

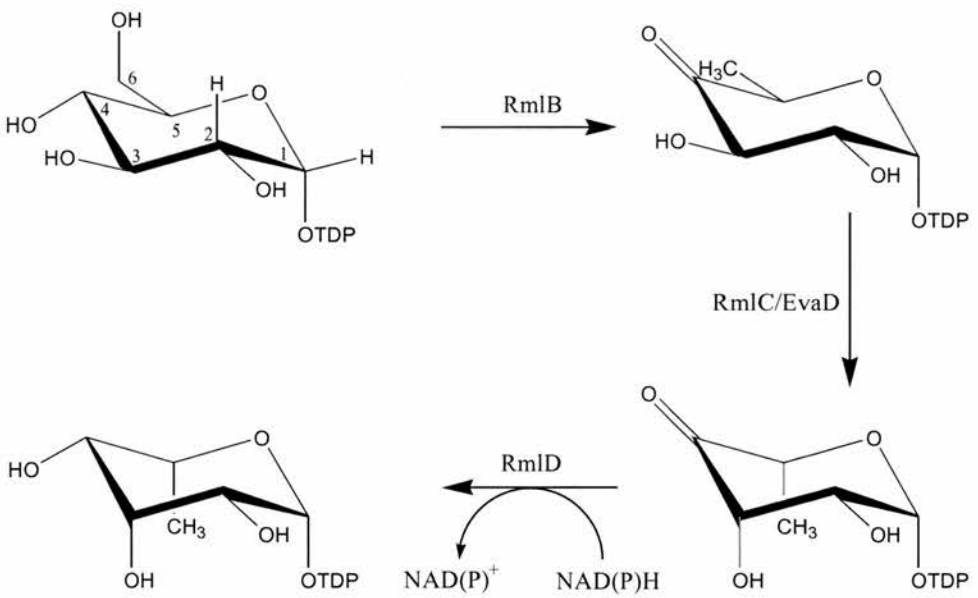


Figure 2.6 – Reaction Scheme for EvaD Kinetics: reaction scheme for the chemical pathway, that was carried out to determine the kinetics of EvaD. To determine their kinetic parameters, EvaD or RmlC were kept at reaction limiting concentrations (scheme adapted from Kerr, 2003).

Enzyme activity was determined by measuring the oxidation of NADPH at A_{340} by RmlD: this enzyme recognises only the doubly epimerised (at C3′ and C5′) product; therefore, a measurable reaction rate shows that EvaD is capable of catalysing the double epimerisation.

The decrease in the absorbance at λ = 340nm was monitored continuously over a 5min period in a Unicam spectrophotometer (Helios α), using the Unicam software Vision32.lnk and the rate method. The background rate was determined in the following two ways, which were averaged: the NADPH oxidation rate of the assay mixture in the presence of enzyme without substrate, and the oxidation rate of NADPH by the assay mixture in the presence of substrate but without enzyme, were ascertained. Each reaction was carried out in triplicate and with a range of substrate concentrations using a 50µl quartz cuvette.

After determining an appropriate EvaD concentration for the assay, the reaction rate dependency on EvaD (or RmlC) was confirmed. This was done by keeping EvaD constant and in turn altering the RmlB and RmlD enzyme concentrations between one tenth of, and ten times, their respective concentrations. Table 2.2 shows the amounts of each component in the reaction mix for the native enzyme assay.

The reaction rates were calculated for wild-type EvaD and RmlC. Rates were also determined for the EvaD mutants: I50N, H63A, M131F, Y133F, L135A, M131F/L135A, and I50N/M131F/L135A.

A mutant was confirmed as being inactive if 500 times the normal substrate concentration did not result in a measurable reaction rate.

Table 2.2 - Concentrations of Reaction Components

Reaction component	Concentration of stock solution	Amount used in final assay
RmlB	4.7mg ml^{-1}	1.3µM
EvaD	3.4mg ml^{-1}	12.5µM
(RmlC)	3mg ml^{-1}	0.03µM
RmlD	4.6mg ml^{-1}	1.0µM
NADPH	100mM	0.3mM
$MgCl_2$	100mM	9mM
Tris-HCl, pH7.5	1M	20mM
dTDP-glucose	100mM	Variable
dH_2O	-	Made up to 70µl

2.3.10 Deuterium Incorporation

The incorporation of deuterium into dTDP-6-deoxy-D-xylo-4-hexulose as catalysed by RmlC had previously been carried out by James Errey (University of East Anglia, PhD Thesis 2003). The same assay was carried out substituting EvaD for RmlC; RmlC catalysed deuterium incorporation was used as the positive control for the assay.

The substrate dTDP-6-deoxy-D-xylo-4-hexulose was prepared by an adaptation of the procedure described by Marumo, 1992. A 500µl reaction mixture was prepared containing 3mg RmlB from *S. typhimurium* or *S. suis*, 1mM NAD^+ and 15mmol dTDP-glucose in 50mM Tris-HCl, pH7.5. This was incubated for 8 hours at 37°C and the reaction quenched by filtering the reaction mixture through an Amicon YM3,000 spin column, removing the enzyme and retaining the flow-through containing the substrate. To confirm the production of dTDP-6-deoxy-D-xylo-4-hexulose, 20µl of the flow-through were removed for analysis by electrospray ionisation (ESI) mass spectrometry. The sample was then lyophilized and reconstituted with D_2O.

The assays were prepared by an adaptation of the procedure described by Daffe *et al.*, 1990. For each sample, 30µl of the substrate solution, reconstituted in D_2O, was used; 15µg of sample enzyme were added, resulting in a final concentration of ~91% deuterium enriched water. The reaction mixture was incubated for 2 hours at 37°C, and the reaction was quenched by the addition of 50µl ethanol. The sample was spun at 31,200*g* for 5min and the supernatant collected.

To analyse the samples by gas chromatography coupled to mass spectrometry (GC-MS), the ketosugar nucleotides were reduced with sodium borohydride and cleaved with trifluoroacetic acid (TFA). The resulting hexoses were further reduced to the corresponding alditols and then per-*O*-acetylated.

GC-MS sample preparation

The supernatant was added to a sample vial containing 2mg sodium borohydride and stirred at room temperature for 2 hours. The reaction was quenched by the addition of glacial acetic acid, followed by 20µl 10% acetic acid in methanol. The samples were subsequently freeze dried.

The sample was hydrolysed with 250µl 2M TFA solution containing 25µg myo-inositol (used as the internal standard). The samples were then dried on the speed vac (Savant Speed Vac System SC210), and the final amounts of TFA were removed by co-evaporation with isopropanol. This reaction step required the addition of copious amounts of isopropanol and heating of the speed vac to 40°C. The sample was then reduced to the corresponding alditol by the addition of 250µl 1M ammonium hydroxide containing 10mg ml^{-1} sodium borohydride. The solution was allowed to stand at 21°C for 1 hour, after which the excess borohydride was converted to borate by the addition of 2 to 3 drops of glacial acetic acid.

250µl 10% acetic acid in methanol were added to the sample and evaporated using the speed vac. Three further evaporations using the same solution were carried out, followed by four evaporations using 250µl methanol. *O*-acetylation of the resulting alditols was achieved by treating them with 250µl acetic anhydride and 250µl dry pyridine and heating them to 121°C for approximately 20min. To remove the final traces of pyridine, two co-evaporation steps, each using 200µl toluene, were carried out. The per-*O*-acetylated alditols were partitioned between dichloromethane (DCM) and dH_2O (500µl of each). The DCM layer was transferred to a clean tube using a pasture pipette, and the samples were freeze-dried once more before they could be analysed by GC-MS.

Following the reduction and per-acetylation, four different products (Figure 2.7) may be observed by GC-MS analysis, depending on the position(s) of the deuterium exchange.

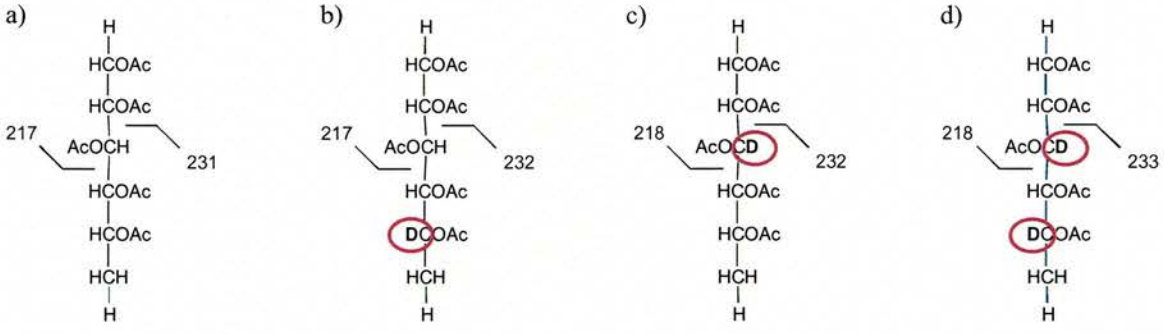


Figure 2.7 – GC-MS Fragments: GC-MS fragments of alditol per-acetates associated with deuterium incorporation. a) Starting material: 217/231; b) selective D incorporation at C5′: 217/232; c) selective D incorporation at C3′: 218/232; d) double D incorporation at C3′ and C5′: 218/233.

The samples were re-dissolved in DCM and analysed by GC-MS against the internal myo-inositol standard as published by Daffe *et al.*, 1990. Samples were injected in splitless mode from 200µl GC-MS sample vials, using an autosampler into a gas chromatograph connected to a mass selective detector (Agilent). The injection port was 290°C and the transfer line 280°C. The column was a 30m Hewlett Packard HP-5 MS fused capillary column (5% Phenylmethyl Siloxane; 30m x 0.25mm; 0.25µm film thickness), and the oven was programmed to hold at 50°C for 1min followed by a 30°C per minute increase to 165°C followed by a 10°C per minute rise to 280°C. The mass spectrometer was set to scan from 50 to 400 atomic mass units (AMU) in the ion scanning mode, scanning for the following mass peaks: 217, 218, 231, 232 and 233.

Since each reaction was carried out in triplicate, the average of each mass peak was calculated. These data were then used to calculate the relative deuterium incorporation at each position; this in turn was compared to the background deuterium incorporation and the one determined for RmlC under similar conditions. RmlC from *S. suis* served as the positive control for double deuterium incorporation at C3′ and C5′.

To directly compare deuterium incorporation to epimerisation to following assumptions had to be made when converting the deuterium incorporation results:

- All deuterium incorporation could be detected (counts equivalent to incorporation)
- EvaD monomer = 22,500Da
- RmlC monomer = 20,663Da
- 12.5µg enzyme used in both assays
- The deuterium incorporation is linear throughout the experiment
- Temperature (reaction rate doubles for every 10˚C increase in temperature
- No isotope effects, no other interference in the spectrum
- The yield is 100%

2.4 Results and Discussion

2.4.1 Wild-type and Mutant EvaD: Purification and Folding

All desired mutants (I50N, H63A, M131F, Y133F, L135A, M131F/L135A and I50N/M131F/L135A) were successfully produced, as verified by DNA sequencing; no additional mutations had been inserted accidentally into the gene during PCR.

EvaD was purified to single-band purity in two steps as seen on a Coomassie™ Blue stained SDS gel. The first step used an anion exchange column (DEAE Fast Flow Sepharose™ Amersham Bioscience) and the second chromatographic step was a metal chelating one using $NiSO_4$ charged Poros resin (BioRad), metal chelating Fast Flow Sepharose ™ (Amersham Biosciences) or Ni-NTA resin (Invitrogen). Figure 2.8 shows a typical chromatography trace, as obtained during the metal chelating purification step.

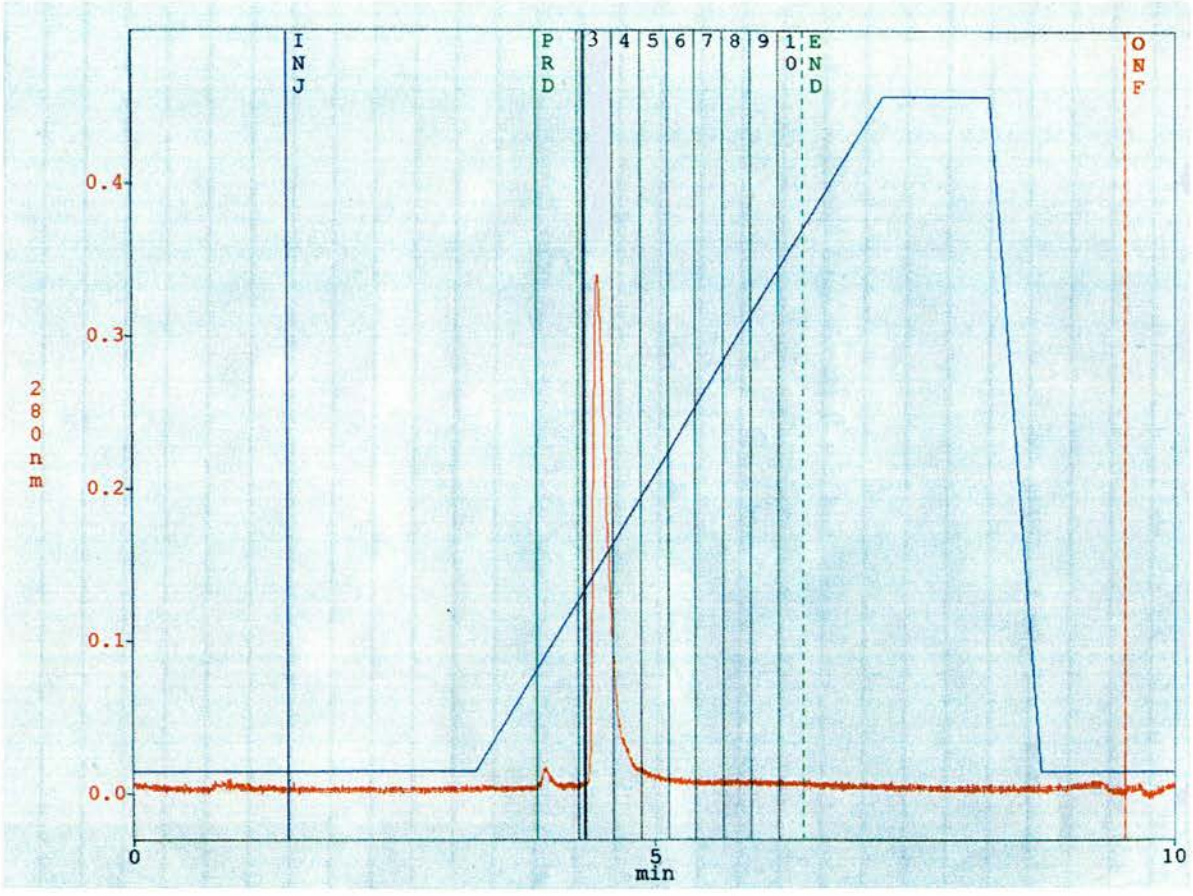


Figure 2.8 – HPLC Metal Chelating Column Trace: metal chelating column purification step (HPLC). The blue line is the imidazole gradient (up to 1.0M), the red line represents the absorbance of the eluent at λ = 280nm. The peak (fraction 3) is EvaD.

Figure 2.9a shows a Coomassie™ Blue stained SDS gel for the EvaD purification, Figure 2.9b shows pure EvaD as obtained from the metal chelating column.

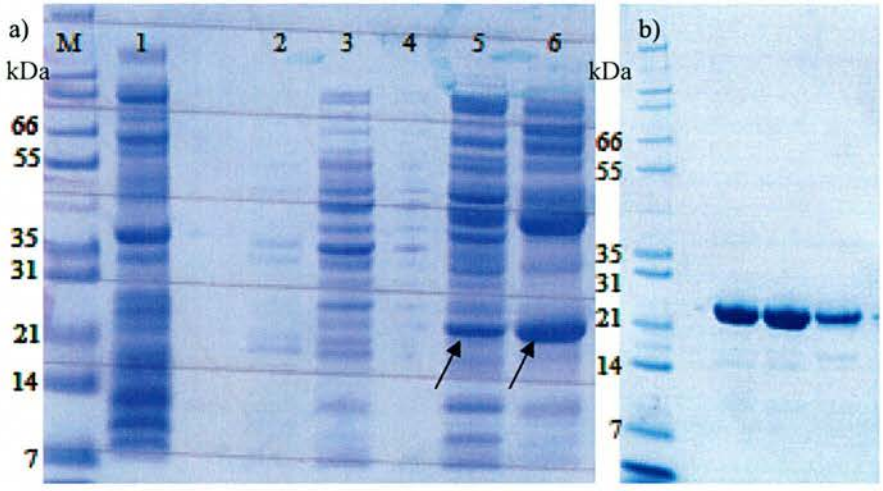


Figure 2.9 - Gels for the Purification of EvaD:
a) DEAE Sepharose anion exchange column:
M – molecular weight marker (BioRad), 1 DEAE column load; 2 – unbound; 3 and 4 peaks off the column, not containing EvaD; 5 & 6 elution peak containing EvaD (arrows)
b) pure EvaD as obtained from the Ni column (peak shown in figure 2.8)

All mutants were purified in an identical fashion, and Figure 2.10 shows an SDS gel of a selection of the purified mutants as compared to wild-type EvaD. ESI, as well as, trypsin-digest MALDI-TOF mass spectrometry on the pure protein in solution, were carried out to confirm that the right protein with the desired mutation had been expressed and purified.

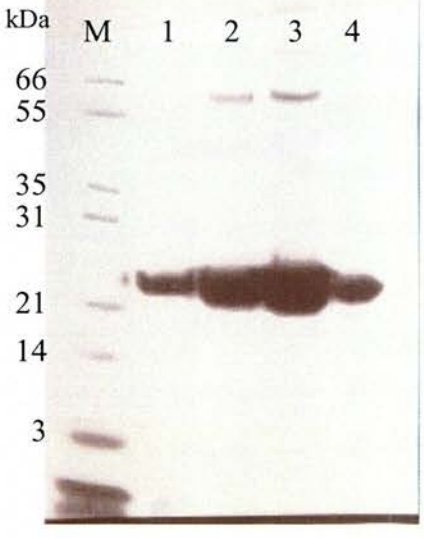


Figure 2.10 - SDS PAGE Gel of EvaD and a Selection of Purified Mutants: Coomassie Blue™ Stained SDS gel
M – Molecular weight marker (BioRad)
1 – Y133F
2 – native EvaD
3 – M131F/L135A
4 – H63A

Before crystallisation trials and biochemical assays were carried out, the secondary and tertiary structures of wild-type and mutant EvaD were analysed by CD

spectroscopy. All samples, with the exception of the triple mutant, displayed CD signals in the far and near UV regions, which were consistent with a protein exhibiting both secondary and tertiary structure. All spectra were similar to one another and to that of the wild-type protein, and therefore, could be assumed to be displaying correct folding. Figure 2.11 presents some of the CD spectra obtained for the mutant enzymes in comparison to the wild-type EvaD, whose spectrum is shown in red.

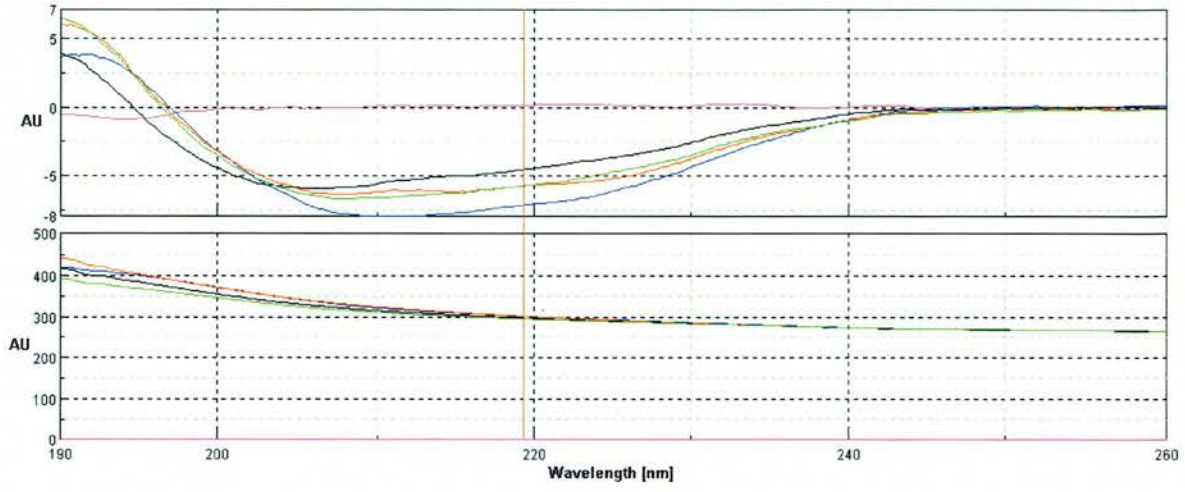


Figure 2.11 - CD Spectra for Wild-type EvaD and its Mutants: the spectra for wild-type EvaD are red, those for L135A are blue, M131F are green, I50N are black and for the triple mutant they are pink. The lower spectrum is the base line to give a comparison

2.4.2 Crystallisation

Following optimisation of the initial 'hits', the best crystals for wild-type EvaD were obtained from the following precipitant conditions: 100mM Tris, pH8.0, 100mM $MgCl_2$, 25%w/v PEG4000 using the sitting drop vapour diffusion method, a protein concentration of 10mg ml^{-1} and incubating the crystallisation plate (Hampton Research - CrystalClear Strips) at 4°C. In this case the incubation temperature was crucial to crystal quality: crystals grown under identical precipitant conditions but at 20°C gave small needle clusters, which diffracted very poorly, were highly mosaic and proved difficult to optimise; whereas at 4°C large, high-quality crystals of the dimensions 0.5 x 0.3 x 0.1mm appeared within 8 to 10 days (Figure 2.12).

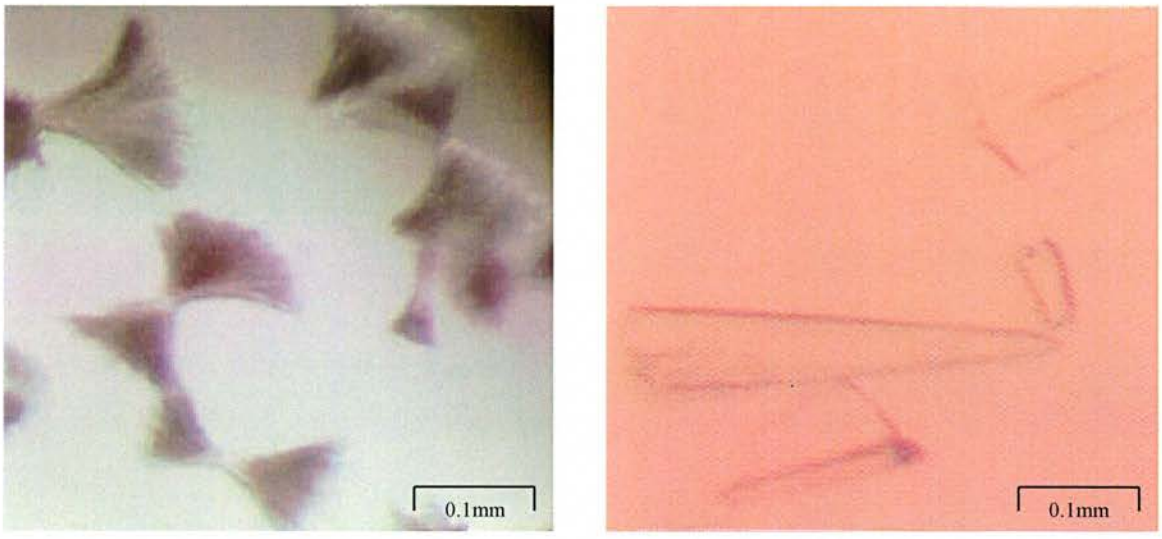


Figure 2.12 - EvaD Crystals: differences observed with altering incubation temperature; left-hand side: low-quality needle clusters obtained at 20°C; right-hand side: well diffracting single crystals obtained under the same precipitant conditions at 4°C.

Despite the correct folding of all but one of the mutant enzymes, only crystals for the wild-type protein and the M131F/L135A double mutant could be obtained. There is no real explanation for this, since crystallisation conditions for all mutants were re-screened completely in sparse-matrix crystallisation screens and at incubation temperatures of 20°C and 4°C.

The M131F/L135A double mutant was found to crystallise under the following precipitant conditions using protein at a concentration of 10mg ml^{-1}, the hanging drop vapour diffusion method and a drop size of 2+2µl: 100mM Tris, pH8.0, 150mM NaCl, 25%w/v PEG8000. The crystallisation trays were incubated at 20°C and crystals of the dimensions 0.2 x 0.2 x 0.2mm were found to form within approximately 10 days.

2.4.3 Data Collection and Processing

A total of 292,414 measurements, 65,259 of which were unique reflections, were recorded of the native crystal to 1.5Å. The reflections were indexed in an orthorhombic space group of the following dimensions: a = 98.6Å, b = 72.0Å, c = 57.1Å, $\alpha = \beta = \gamma = 90^{o}$. The space group was identified as $P22_12_1$ by the observation of n = 2 systematic absences at the k and l axes but not at the h-axis. The data were re-indexed using Scala, version 3.1.2 (Evans, 1993) to allow the expression of the space group in the more conventional format, $P2_12_12$ (space group no. 18). To estimate the solvent content of the crystal, the Matthews volume was calculated (Matthews, 1968). V_m was

found to be 2.3Å^3 per Dalton (Da) giving a solvent content of 45%; this assumed that 2 molecules, i.e. one dimer, were present in the au and a molecular weight of 22,000Da per monomer.

Of the substrate-soaked crystal, a data set to 1.4Å resolution was recorded, consisting of a total of 325,948 reflections, of which 81,444 were unique. They were indexed, integrated, merged, scaled and re-indexed as described above for the apo-data, using MOSFLM, version 6.2.4 (Leslie, 1992) and Scala, version 3.2.1.

For the M131F/L135A double mutant, a total of 27,815 reflections were measured, of which 9,756 were unique. The reflections were cut to 2.08Å, indexed, integrated and merged using d*TREK, version 8.0 (CrystalClear 1.3.5©, Rigaku) in space group c2; the unit cell dimensions were as follows: a = 51.44Å, b = 79.09Å, c = 48.68Å, α = 90.00°, β = 98.72°, γ = 90.00°. Calculation of the Matthews volume (Matthews, 1968) gave V_m = 2.2Å^3 Da^{-1} and a solvent content of 43.6% assuming 1 molecule per au and a molecular mass of 22,000Da.

Table 2.3 details the data for the individual resolution shells for all data sets collected; and Table 2.4 summarises the overall data collection statistics for the native dataset, as well as, the ones collected on the dTDP-glucose soaked crystal and the M131F/L135A double mutant.

Table 2.3 - Data Collection Details for Individual Resolution Shells:

Resolution Shell (Å)	Measured Reflections	Completeness (%)	Multiplicity	R_{merge}[a]	I/sigma	Mn(I)/sd
EvaD – Apo Data Set						
57.74 – 6.71	3,097	99.3	3.7	0.047	11.5	30.9
6.71 – 4.74	5,990	100.0	4.2	0.046	11.8	32.2
4.74 – 3.87	7,829	100.0	4.4	0.043	12.6	33.2
3.87 – 3.35	9,382	100.0	4.4	0.045	12.4	33.6
3.35 – 3.00	10,550	100.0	4.5	0.044	12.7	32.7
3.00 – 2.74	11,815	100.0	4.5	0.044	12.7	30.4
2.74 – 2.54	12,672	100.0	4.5	0.048	11.5	28.3
2.54 – 2.37	13,640	100.0	4.5	0.052	10.7	26.3
2.37 – 2.24	14,503	100.0	4.5	0.055	11.1	25.3
2.24 – 2.12	15,256	100.0	4.5	0.058	10.4	23.8
2.12 – 2.02	16,120	100.0	4.5	0.062	10.4	22.2
2.02 – 1.94	16,616	100.0	4.5	0.070	9.2	19.2
1.94 – 1.86	17,521	100.0	4.5	0.085	7.7	16.1
1.86 – 1.79	18,070	100.0	4.5	0.099	6.5	13.7
1.79 – 1.73	18,632	100.0	4.5	0.118	5.9	11.4

Table 2.3 continued - Data Collection Details for Individual Resolution Shells:

Resolution Shell (Å)	Measured Reflections	Completeness (%)	Multiplicity	R_{merge} [a]	I/sigma	Mn(I)/sd
EvaD – Apo Data Set						
1.73 – 1.68	19,397	100.0	4.5	0.145	4.8	9.3
1.68 – 1.63	19,886	100.0	4.5	0.172	4.0	7.8
1.63 – 1.58	20,584	100.0	4.5	0.202	3.5	6.4
1.58 – 1.54	21,088	100.0	4.5	0.239	3.0	5.4
1.54 – 1.50	19,464	100.0	4.1	0.271	2.5	4.3
57.74 – 1.50	**292,661**	**100.0**	**4.4**	**0.058**	**7.1**	**17.5**
EvaD – dTDP-glucose Soak						
34.10 – 5.01	6,696	98.9	3.5	0.052	10.0	27.4
5.01 – 3.54	12,992	99.6	3.9	0.048	11.7	29.4
3.54 – 2.89	17,072	99.9	4.0	0.051	11.1	27.3
2.89 – 2.51	20,261	99.9	4.0	0.054	10.8	24.3
2.51 – 2.24	23,044	99.9	4.1	0.058	10.7	21.6
2.24 – 2.05	25,152	99.8	4.1	0.065	9.9	19.6
2.05 – 1.89	27,535	99.7	4.1	0.081	8.2	16.2
1.89 – 1.77	29,333	99.8	4.1	0.102	6.7	13.2
1.77 – 1.67	31,181	99.7	4.1	0.135	5.2	9.9
1.67 – 1.58	32,661	99.8	4.0	0.186	3.8	7.4
1.58 – 1.51	34,391	99.8	4.0	0.252	2.8	5.8
1.51 – 1.45	35,789	99.9	4.0	0.330	2.1	4.3
1.45 – 1.40	29,841	99.9	4.0	0.451	1.5	3.2
34.10 – 1.45	**325,948**	**99.8**	**4.0**	**0.074**	**6.3**	**13.4**
EvaD – M131F/L135A Double Mutant						
48.22 – 7.79	526	95.3	2.4	0.074	10.6	30.0
7.79 – 5.51	1,001	99.4	2.5	0.069	10.4	25.6
5.51 – 4.50	1,319	99.5	2.5	0.065	11.1	26.9
4.50 – 3.90	1,546	99.8	2.5	0.070	8.8	25.9
3.90 – 3.48	1,722	99.8	2.5	0.064	11.2	24.0
3.48 – 3.18	1,914	99.8	2.4	0.063	11.8	21.4
3.18 – 2.95	2,043	99.5	2.4	0.065	12.4	17.5
2.95 – 2.75	2,244	99.3	2.5	0.081	10.6	14.5
2.75 – 2.60	2,321	99.2	2.4	0.095	8.5	12.5
2.60 – 2.46	2,467	99.1	2.4	0.108	7.9	9.9
2.46 – 2.35	2,575	98.6	2.4	0.123	7.2	9.5
2.35 – 2.25	2,638	98.1	2.4	0.141	6.3	7.9
2.25 – 2.16	2,831	97.9	2.4	0.179	5.1	6.7
2.16 – 2.08	2,668	95.3	2.3	0.180	5.1	5.9
48.22 – 2.08	**27,815**	**98.6**	**2.4**	**0.071**	**7.2**	**17.0**

[a] $R_{merge} = \Sigma\Sigma I(h)j - (I(h)) / \Sigma\Sigma I(h)j$, where I(h) is the measured diffraction intensity, and the summation includes all observations.

Table 2.4 - Data Collection Statistics

	Native Data Set (ESRF, BM14)	***dTDP-glucose* (ESRF, ID14-2)**	**M131F/L135A (in-house)**
Wavelength (Å)	0.933	0.932	1.542 (Cu anode)
Resolution (Å)	57.74 - 1.50	34.10-1.40	48.22-2.08
Space group	$P2_12_12$ (No. 18)		c2 (No. 5)
Unit-cell Parameters (Å, °)	a = 98.6, b = 72.0, c = 57.1 α = β = γ = 90°		a=51.44, b=79.09, c=48.68 α=γ 90.00, β=98.72
V_m (mol/au ($Å^3$ Da^{-1}))[a]	2.3	2.3	2.2
Percentage solvent[a]	44.9	44.9	43.6
Total measurements	292,414	325,948	27,815
Unique reflections	65,259	81,444	11,455
I/σ	7.1 (2.5)	6.3 (1.5)	7.2 (5.1)
Average redundancy	4.4 (4.1)	4.0 (4.0)	2.4 (2.3)
Data completeness (%)	100 (100)	99.8 (99.9)	98.8 (95)
R_{merge}[b] (%)	5.8 (27.1)	7.4 (45.1)	7.1 (18.0)
Wilson B factor ($Å^2$)	13.78	11.61	26.1

[a] V_m and solvent content were calculated assuming 2 monomers per au for the native and the dTDP-glucose soaked data sets and 1 monomer per au for the M131F/L135A double mutant.
[b] $R_{merge} = \Sigma\Sigma I(h)j - (I(h)) / \Sigma\Sigma I(h)j$, where I(h) is the measured diffraction intensity, and the summation includes all observations.
Parentheses denote data in the highest resolution shell.

2.4.4 Structure Solution

Phases for the apo-structure of EvaD were calculated, as detailed in section 2.3.7 by molecular replacement using MolRep (Vagin and Teplyakov, 1997). The ten best cross-rotation peaks are detailed in Table 2.5. These were then subjected to a translational peak search yielding potential solutions, which were given in fractional coordinates and Eularian angles (Table 2.6). The best one, which had an R_{factor} of 0.471 and a correlation coefficient of 0.512, was used to calculate the new, re-orientated model coordinates.

This model and its symmetry related objects, according to the point group $P2_12_12$, were displayed in the graphics program O, version 7.0 (Jones *et al.*, 1991) to check the crystal packing. As no steric clashes could be observed, the model was assumed to be the correct one.

Table 2.5 – Cross-rotation Solutions for the Native Data Set: the ten best solutions for the cross rotation search, as computed by MolRep using all data in the resolution range 15 to 4Å.

Rotation sol. No	theta	phi	chi	alpha	beta	gamma	R_{fac}	R_{fac} / sigma
1	139.09	179.75	127.29	33.00	71.86	213.49	2.141	3.65
2	146.08	-59.83	166.64	128.21	67.32	67.86	2.069	3.53
3	165.60	-164.32	144.85	33.78	27.42	182.43	1.768	3.02
4	154.29	-70.07	134.44	134.91	47.16	95.06	1.525	2.60
5	149.51	-117.72	158.60	74.65	59.81	130.09	1.497	2.55
6	38.43	131.95	106.39	88.26	59.69	4.37	1.481	2.53
7	135.00	-43.93	179.86	136.17	90.00	44.02	1.443	2.46
8	136.60	-69.29	64.05	176.27	42.74	134.85	1.432	2.44
9	125.73	-163.73	36.73	95.30	29.64	242.76	1.343	2.29
10	58.88	179.39	76.29	111.48	63.85	292.70	1.290	2.20

Table 2.6 – Translation Solutions for the Native Data Set: a selection of the translation solutions given by MolRep using the ten best cross-rotation peaks (show in Table 2.5).

Solution	alpha	beta	gamma	x_{frac}	y_{frac}	z_{frac}	TF / sig	R_{fac}	Corr. C.
1 – 1	33.00	71.86	213.49	0.011	0.239	0.340	4.05	0.669	0.168
1 – 2	33.00	71.86	213.49	0.120	0.244	0.000	3.27	0.625	0.181
1 – 3	33.00	71.86	213.49	0.258	0.081	0.462	3.08	0.483	0.356
1 – 4	33.00	71.86	213.49	0.125	0.234	0.308	3.03	0.580	0.273
1 – 5	33.00	71.86	213.49	0.263	0.085	0.218	2.85	0.577	0.229
1 – 6	33.00	71.86	213.49	0.008	0.247	0.500	2.59	0.771	0.041
1 – 7	33.00	71.86	213.49	0.400	0.249	0.221	2.46	0.586	0.242
1 – 8	33.00	71.86	213.49	0.725	0.232	0.488	2.15	0.632	0.173
1 – 9	33.00	71.86	213.49	0.310	0.463	0.151	1.96	0.616	0.073
1 – 10	33.00	71.86	213.49	0.157	0.386	0.465	1.58	0.624	0.169
2 – 1	128.21	67.32	67.86	0.490	0.252	0.182	3.91	0.639	0.230
2 – 2	128.21	67.32	67.86	0.237	0.083	0.240	3.60	0.540	0.336
2 – 3	128.21	67.32	67.86	0.250	0.094	0.029	3.32	0.661	0.111
2 – 4	128.21	67.32	67.86	0.164	0.294	0.252	3.29	0.612	0.259
2 – 5	128.21	67.32	67.86	0.260	0.293	0.000	2.84	0.647	0.145
2 – 6	128.21	67.32	67.86	0.092	0.236	0.468	2.72	0.557	0.250
2 – 7	128.21	67.32	67.86	0.491	0.239	0.000	2.40	0.633	0.153
2 – 8	128.21	67.32	67.86	0.261	0.481	0.120	2.12	0.624	0.057
2 – 9	128.21	67.32	67.86	0.363	0.108	0.263	0.69	0.471	0.512

Phases for the M131F/L135A double mutant were calculated in MolRep as detailed in section 2.3.7. Table 2.7 and 2.8 detail the top solutions for cross-rotation and translational peak searches. The best solution displayed an R_{factor} of 0.448 and correlation coefficient of 0.597. The new model coordinates were calculated automatically and displayed in O, version 7.0 together with the symmetry related objects, according to the space group c2. No steric clashes were observed and crystal packing was sensible, since the homodimer was formed by a symmetry-related object in the adjacent unit cell.

Table 2.7 – Cross-rotation Solutions for the Double Mutant: the ten best solutions for the cross rotation search as computed by MolRep using all data in the resolution range 15 to 4Å.

Rotation sol. No	theta	phi	chi	alpha	Beta	gamma	R_{fac}	R_{fac} / sigma
1	46.78	-71.32	110.43	243.26	73.53	205.91	0.4727	15.80
2	43.63	-79.74	103.44	232.79	65.59	212.27	0.1705	5.70
3	40.46	149.67	127.70	116.83	71.26	357.50	0.1415	4.73
4	142.75	99.52	168.48	286.74	74.06	267.70	0.1307	4.37
5	132.40	37.58	97.34	270.10	67.36	14.94	0.1291	4.32
6	116.48	-119.91	90.20	125.98	78.70	185.81	0.1270	4.25
7	117.93	39.84	102.19	279.70	86.87	20.03	0.1265	4.23
8	136.00	34.64	98.56	164.76	63.54	15.48	0.1247	4.17
9	147.88	82.53	159.46	274.60	63.08	289.54	0.1228	4.10
10	44.36	155.70	121.33	117.53	75.11	346.13	0.1220	4.08

Table 2.8 – Translation Solutions for the Double Mutant: a selection of the translation solutions given by MolRep using the ten best cross-rotation peaks (shown in Table 2.7).

Solution	alpha	beta	gamma	x_{frac}	y_{frac}	z_{frac}	TF / sig	R_{fac}	Corr. C.
1 – 1	243.26	73.53	205.91	0.232	0.000	0.347	12.16	0.448	0.597
1 – 2	243.26	73.53	205.91	0.262	0.000	0.394	5.42	0.555	0.391
1 – 3	243.26	73.53	205.91	0.312	0.000	0.309	5.17	0.567	0.374
1 – 4	243.26	73.53	205.91	0.288	0.000	0.355	5.06	0.548	0.404
1 – 5	243.26	73.53	205.91	0.208	0.000	0.301	4.83	0.578	0.345
1 – 6	243.26	73.53	205.91	0.259	0.000	0.305	4.37	0.561	0.370
1 – 7	243.26	73.53	205.91	0.236	0.000	0.435	4.20	0.572	0.360
1 – 8	243.26	73.53	205.91	0.208	0.000	0.390	3.65	0.556	0.387
1 – 9	243.26	73.53	205.91	0.182	0.000	0.262	3.28	0.565	0.369
1 – 10	243.26	73.53	205.91	0.283	0.000	0.439	3.03	0.570	0.361
2 – 1	232.79	65.59	212.27	0.250	0.000	0.322	3.59	0.643	0.167
2 – 2	232.79	65.59	212.27	0.270	0.000	0.381	2.85	0.645	0.163
2 – 3	232.79	65.59	212.27	0.220	0.000	0.371	2.63	0.641	0.165
2 – 4	232.79	65.59	212.27	0.218	0.000	0.286	2.52	0.658	0.141
2 – 5	232.79	65.59	212.27	0.241	0.000	0.430	2.05	0.659	0.125
2 – 6	232.79	65.59	212.27	0.291	0.000	0.427	1.93	0.638	0.180
2 – 7	232.79	65.59	212.27	0.194	0.000	0.318	1.76	0.654	0.140
2 – 8	232.79	65.59	212.27	0.267	0.000	0.469	1.38	0.667	0.124
2 – 9	232.79	65.59	212.27	0.325	0.000	0.282	1.28	0.658	0.134
2 – 10	232.79	65.59	212.27	0.198	0.000	0.409	1.01	0.654	0.152

2.4.5 Structure Refinement and Validation

Wild-type EvaD was auto-traced by the program ARP/wARP (Morris *et al.*, 2002); the first building cycle identified 325 of the 410 expected residues with a connectivity index of 0.92. The procedure converged after 41 cycles, 389 residues in eight separate chains were identified and the connectivity index was 0.96. Electron density map averaging across the dimer, improved the electron density and allowed placement of additional residues, as well as amino acid side chains, into the

model manually, using O, version 7.0. Rigid body refinement in Refmac, version 5.0 (Murshudov *et al.*, 1997) decreased the R_{factor} and R_{free} to 29.5% and 33.4% respectively. Restrained refinement using the weighting term at 0.2, and subsequent TLS refinement, further improved the model. Subsequently, 3 glycerol molecules were added manually into the electron density, before 532 water molecules were added to the model using ARP/wARP as part of Refmac, 5.0. Lastly anisotropic B-factors were refined; the final R_{factor} and R_{free} were 14.4% and 17.8% respectively.

The model for the dTDP-glucose soaked data was refined similarly to the apo-model. After TLS refinement, glycerol and dTMP were placed into the model manually prior to the addition of 676 water molecules. The pyranose ring could not be located in the active site. Anisotropic B-factor refinement decreased the R_{factor} and R_{free} to its final values of 13.1% and 16.6% respectively.

The double mutant was refined, as described above for the wild-type model; initial rigid body refinement and the placement of 3 further residues to the end of the polypeptide chain, decreased the R_{factor} to 28%. TLS refinement decreased the R_{factor} to 22% and the R_{free} to 28%. At this stage, one glycerol molecule was placed into its position as seen in the single difference electron density map. The addition of 103 water molecules and final rounds of restrained refinement resulted in a final R_{factor} and R_{free} of 17.6% and 24.7% respectively.

The final refinement weights during apo, complexed and double mutant model refinement, had been adjusted to achieve an ideal balance between weighting terms for model geometry and crystallographic data. The final rmsds of bond lengths were 0.02Å for all models and the rmsd bond angles were 1.71° for the apo, 1.85° for the complex and 1.99° for the double mutant model.

Validation of the final models together with their original structure factors showed that the final bond angles and lengths met the set targets. More than 90% of the residues of each model fall into the core region of the ϕ/ψ Ramachandran plot (Ramachandran and Sasisekharan, 1968), all other residues are within the allowed or generously allowed regions (Figure 2.13). The other statistics for the models, such as main chain bond lengths, main chain bond angles, residue planarity, side chain *chi1* and *chi2* angles, etc. were within, or better than, the normal limits observed for protein

structures. The model coordinates, as well as the structure factors, have been deposited in the PDB. The accession codes are 1OFN (native), 1OI6 (dTMP) and 1WA4 (double mutant). Table 2.9 details the refinement and validation statistics for the three models.

Table 2.9 - Refinement and Validation Statistics

	Native Data Set	**dTDP – glucose Data Set**	**M131F/L135A Data Set**
R_{factor}	14.4	13.1	17.62%
R_{free}	17.8	16.6	24.79%
Rmsd bonds (Å) / angles (°)	0.02/1.71	0.02/1.85	0.02 / 1.997
B factor deviation bonds/angles ($Å^2$)			
Main chain	1.5/2.2	2.0/2.8	1.02/1.58
Side chain	3.2/4.9	3.8/5.6	2.88/4.24
% Residues in Ramachandran core[a]	93.1	94.2	91.4
% Residues in additionally / generously allowed region	6.9/0.0	5.8/0.0	8.0/0.6
No. of protein atoms	3123	3123	1581
No. of water molecules	532	676	103
No. of other molecules	3 glycerol	1 dTMP 3 glycerol	1 glycerol
Average B factor ($Å^2$)	12.21	10.84	13.27
PDB accession code	1OFN	1OI6	1WA4

[a] Ramachandran core refers to the most favoured region in the ϕ/ψ Ramachandran plot as defined by Laskowski *et al.*, 1993

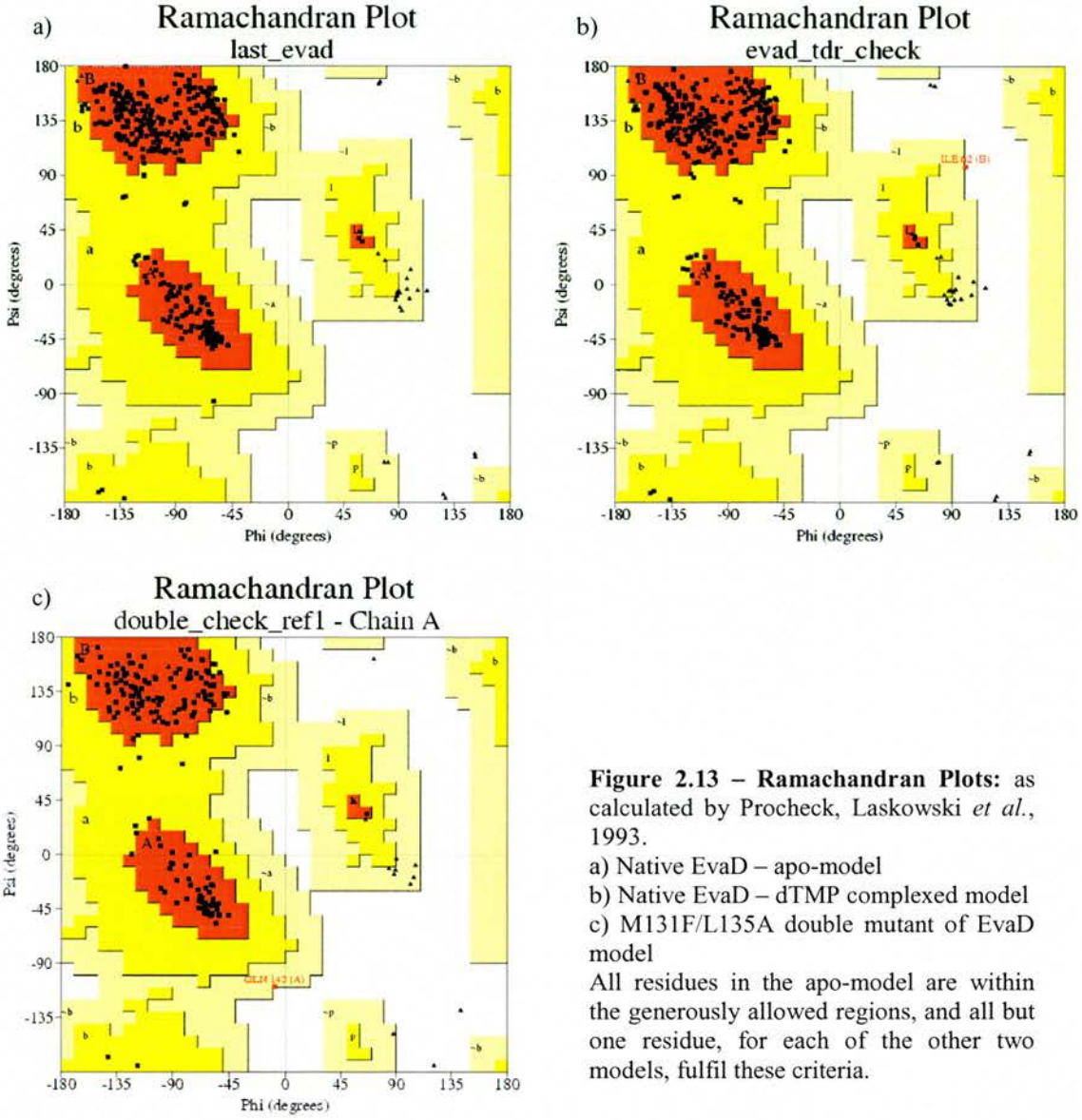


Figure 2.13 – Ramachandran Plots: as calculated by Procheck, Laskowski *et al.*, 1993.
a) Native EvaD – apo-model
b) Native EvaD – dTMP complexed model
c) M131F/L135A double mutant of EvaD model
All residues in the apo-model are within the generously allowed regions, and all but one residue, for each of the other two models, fulfil these criteria.

2.4.6 Gross Structure of EvaD

As the phylogram shows (section 2.2), EvaD is very similar to the RmlC homologues and may be classified as an RmlC; the sequence alignment shown in Figure 2.14 confirms this. All residues known to be crucial to the catalytic activity of RmlC enzymes are fully conserved and highlighted in the sequence alignment: His63, Tyr133 and Lys 73 (EvaD numbering). They align completely with the respective catalytic base, acid and the residue for the stabilisation of the reaction intermediate as published for RmlC (Giraud *et al.*, 2000). The secondary structure given above the sequence alignment in Figure 2.14 was calculated from the fully refined EvaD structure using the ESPript2.2 server at Toulouse, France (Gouet *et al.*, 2003) (http://prodes.toulouse.inra.fr/ESPript/cgi-bin/ESPript.cgi). α denotes the presence of alpha helices, β that of beta sheets and T that of those residues involved in turns, this assignment correlated well with the secondary structures of the RmlC homologues.

Our three-dimensional model we obtained by X-ray crystallographic methods agrees with what we had predicted from the sequence alignment: EvaD is very similar to the RmlC class of epimerases. Figure 2.15 shows a structural alignment of the EvaD carbon alpha backbone (in yellow) with that of several RmlC homologues. This alignment was created in LSQMAN, Uppsala Software Factory (Kleywegt and Jones, 1997a; Kleywegt and Jones, 1997b) using the brute force alignment option; it shows that the gross structure and folding are largely conserved and EvaD is no more different from the RmlC enzymes than these enzymes are from one another. The rms deviation between the main chain of EvaD and those of published RmlCs falls within the range observed for authentic RmlC structures. For example, the main chain rms deviation between EvaD and RmlC from *M. tuberculosis* is 1.41Å which is closer than the main chain rms deviation between the authentic RmlCs from *S. typhimurium* and *S. suis* (rmsd distance = 1.56). Table 2.10 shows how the main chains of different homologues compare to one another and to EvaD (as calculated by LSQMAN, Uppsala Structure Factory).

The overall topology of the monomers is essentially identical to that of authentic RmlC structures. The protein monomer is composed mainly of β-sheets and has a jelly-roll-like topology with overall dimensions: 47Å x 34Å x 51Å.

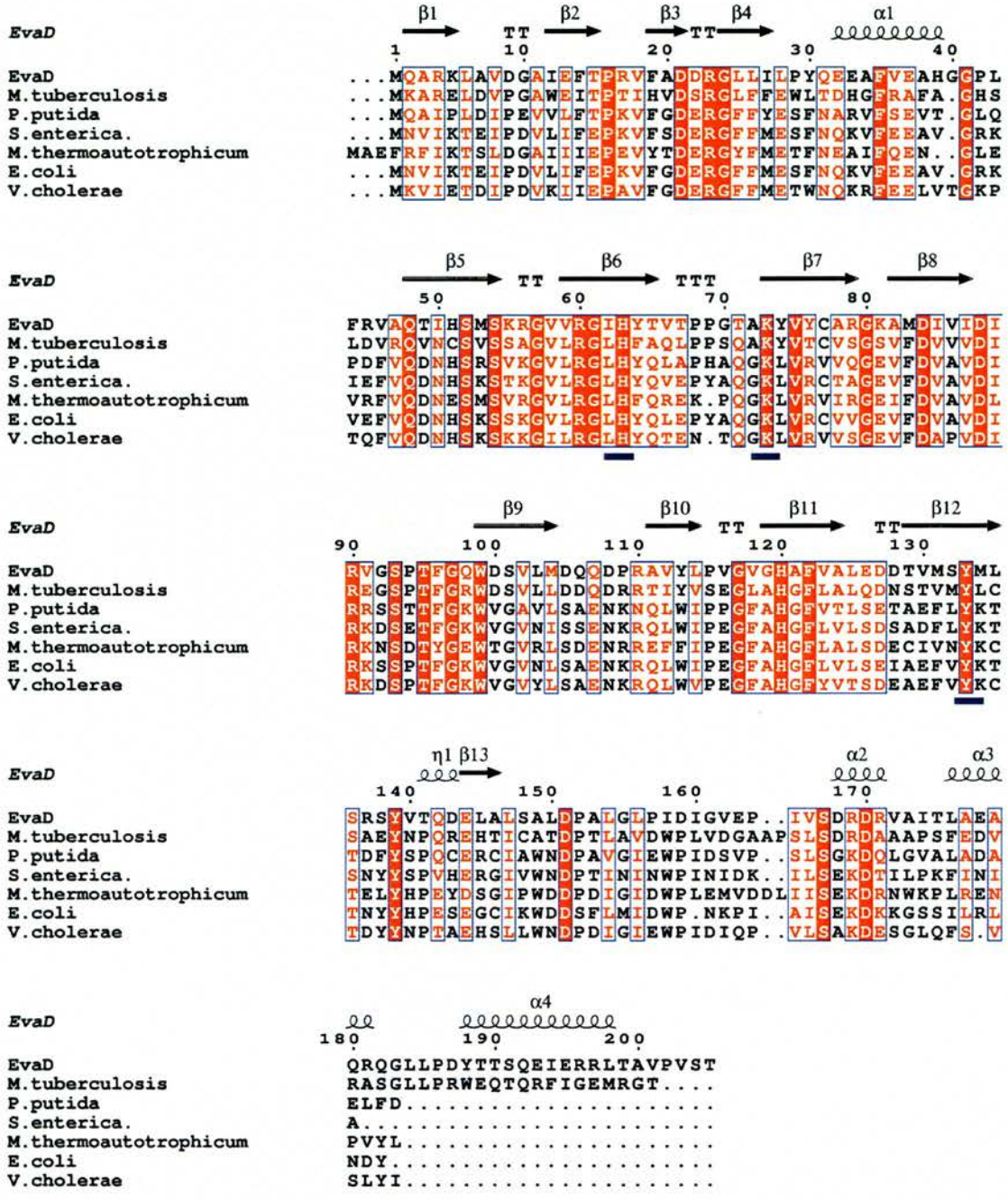


Figure 2.14 - Sequence and Structural Alignment of EvaD and RmlC Homologues: the boxes in the sequence alignment highlight fully conserved amino acids in white letters on red shading and conservatively substituted amino acids in red. Above the alignment the secondary structure of EvaD is given: α denotes α-helices, β - β-sheets and T- turns. The underlined residues are known to be crucial to RmlC activity and are fully conserved throughout this enzyme family as well as in EvaD.

Table 2.10 – Rms Deviation Comparisons: comparison of rmsd distances (in Å) between the carbon alpha backbones of different RmlC homologues and EvaD. Sequence identities between the homologues are given in brackets.

	EvaD	*M. tuberculosis*	*P. aeruginosa*	*S. suis*	*S. typhimurium*
EvaD	-	1.41 (41%)	1.50 (32%)	1.47 (24%)	1.47 (33%)
M. tuberculosis	1.41 (41%)	-	0.97 (36%)	1.55 (30%)	1.00 (34%)
P. aeruginosa	1.50 (32%)	0.97 (36%)	-	1.42 (31%)	0.79 (59%)
S. suis	1.47 (24%)	1.55 (30%)	1.42 (31%)	-	1.56 (31%)
S. typhimurium	1.47 (33%)	1.00 (34%)	0.79 (59%)	1.56 (31%)	-

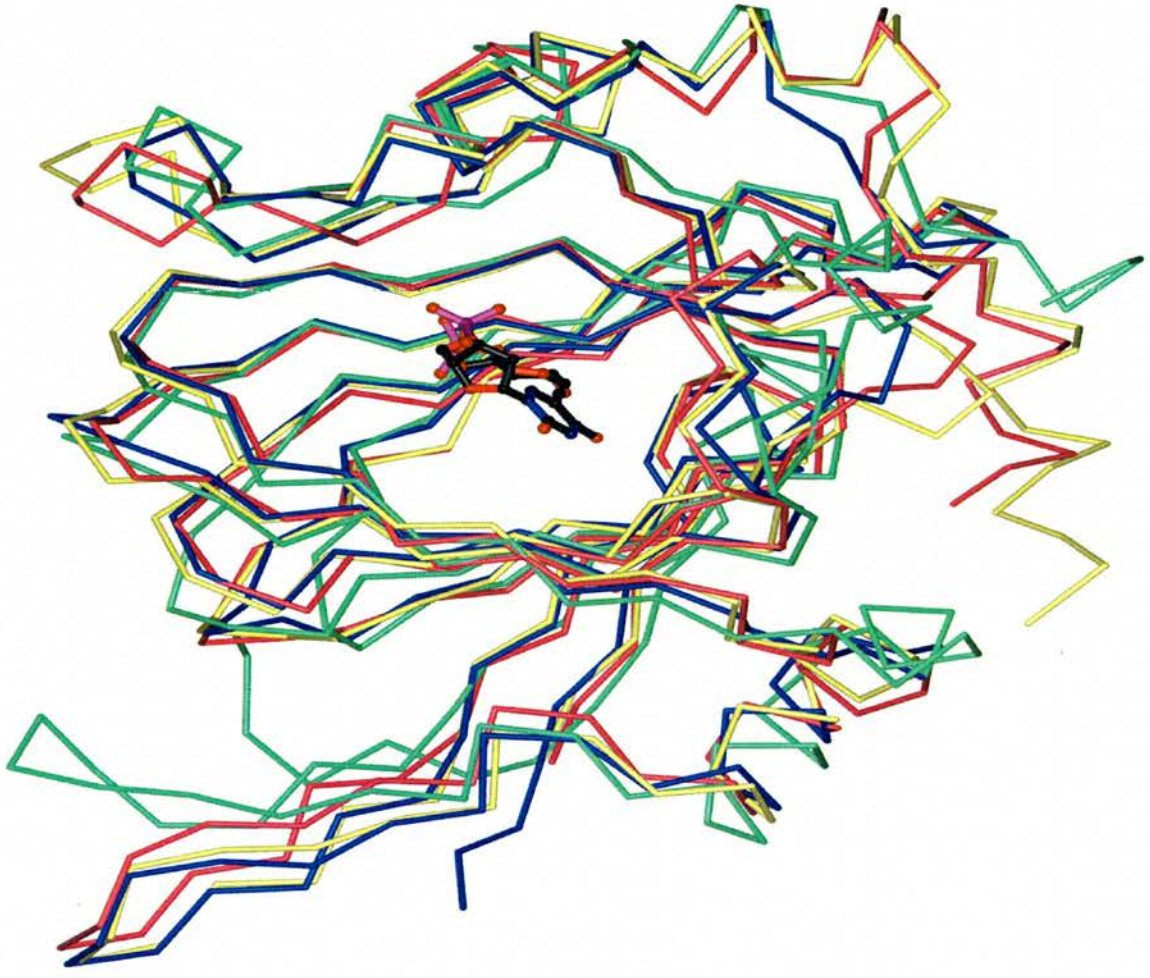


Figure 2.15 – Cα backbones of EvaD and RmlCs: backbone superposition of one monomer of EvaD (yellow) and the equivalent monomers of a selection of RmlCs from different species; *S. suis* – green, *P. aeruginosa* – blue and *M. tuberculosis* – red. dTDP-xylose – an RmlC substrate analogue is bound to the active site of RmlC from *S. suis* for reference.

Each monomer (Figure 2.16) of the homodimer can be divided into three separate regions: the N-terminal, core and C-terminal regions. The N-terminus (residues 1 to 47) consists of an antiparallel β-sheet made up of strands β1, β2, β3 and β4 and a two-turn α-helix α1. The core of the monomer consists of two twisted antiparallel β-sheets made up of strands β5 to β13; together these form a flattened barrel. β-sheet 1 consists of strands β5, β13, β8 and β11 and β–sheet 2 consists of strands β10, β9, β12, β6 and β7. One end of the barrel, formed by strands β5 and β13, is open and the inside is lined with polar residues. The other side is more obscured: strands β1 and β2 from the N-terminal region fold over the entrance and β10 from β-sheet 2 sits at the bottom of the barrel. These strands are made up of bulky, hydrophobic residues that seal the entrance to the barrel; there is also a closely packed cluster of hydrophobic residues inside the barrel.

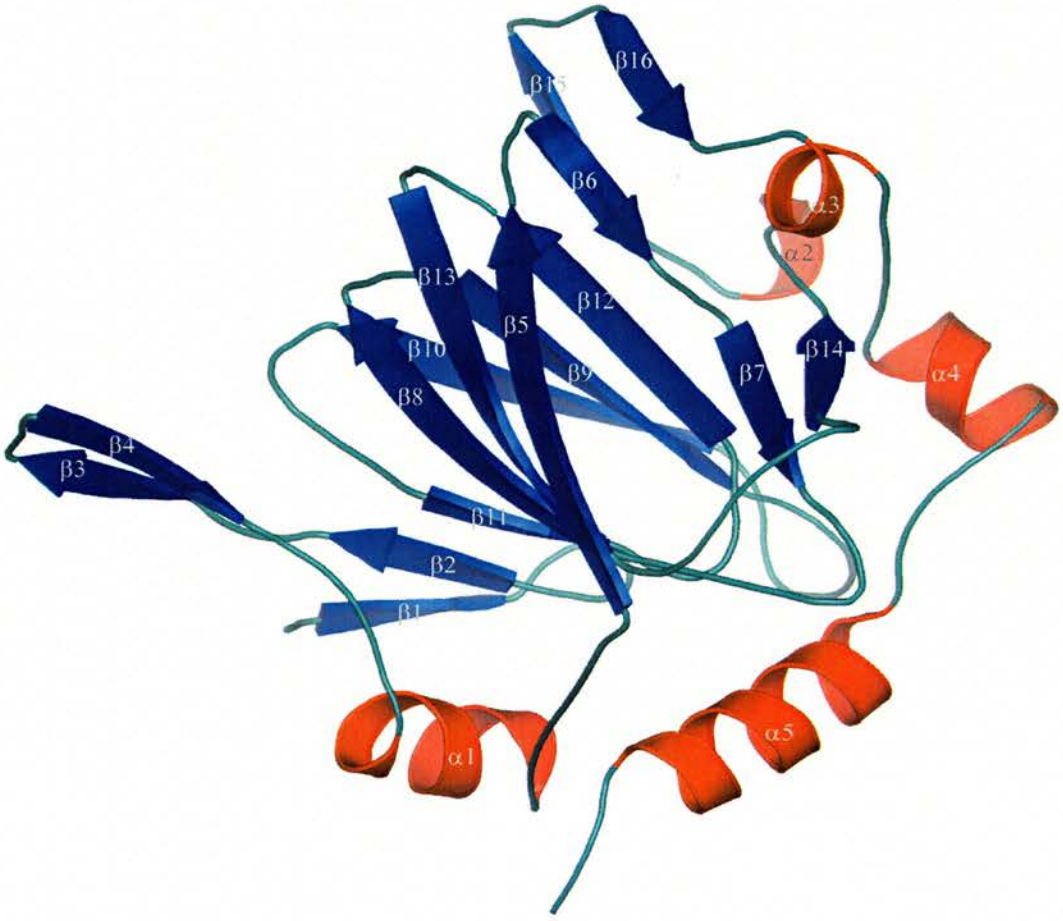


Figure 2.16 - EvaD Monomer: depicting all secondary elements; α-helices are coloured in red, β-sheets in blue and turns display a turquoise line colour.

The C-terminal region (residues 168 to 202) is located on the outside of β-sheet 2; it is flexible and made up of two helical turns, α2 and α3, two short β-strands, β14 and β15, and a short α-helix, α4. Another α-helix, α5, links the C-terminus to the core region of the monomer; it connects the C-terminus to β5 of β-sheet 1 by wrapping around the core region. This helix is a three-turn α-helix consisting of residues 188-198; it is not conserved throughout the RmlC enzymes, some of them contain this final helix, such as the homologue from *M. tuberculosis* which has a two-turn α-helix, whilst others, such as RmlC from *S. suis* and *P. aeruginosa*, do not display this motif at all.

Like RmlC, EvaD is a homodimer with an interface formed by the antiparallel interaction of two β-strands from differing monomers, as shown in Figure 2.17. The dimer interface buries 16.6% (1682Å²) of the monomer's accessible surface area as calculated using the Protein-Protein Interactions Server at UCL by Laskowski, *et al.* (http://www.biochem.ucl.ac.uk/bsm/PP/server) (Laskowski, 1995; Lee and Richards, 1971; McDonald and Thornton, 1994). The monomers are held together by two salt bridges and a total of 24 hydrogen bonds (14 between the main chains of β-strands

5^A and 7^B, and 5^B and 7^A); additionally, a number of bridging water molecules are contained in the interface. Overall, a six-stranded anti-parallel sheet is formed across the two monomers, made up of strands $\beta2^B$, $\beta3^B$, $\beta5^A$, $\beta13^A$, $\beta8^A$ and $\beta11^A$ (Figure 2.18).

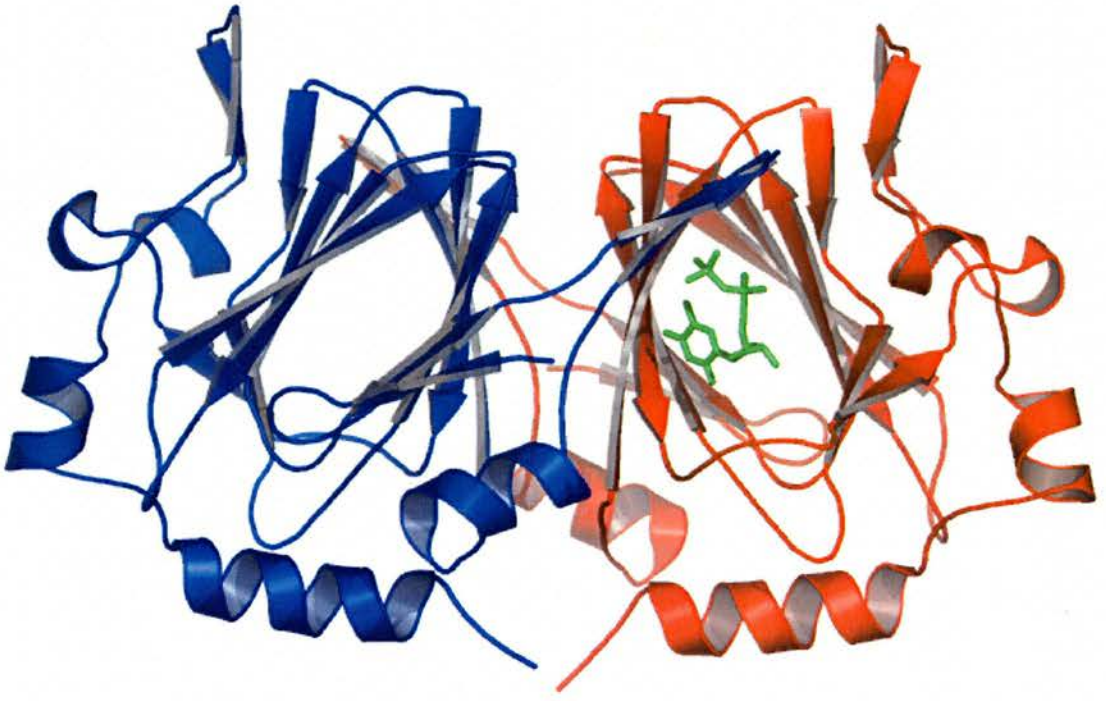


Figure 2.17 - EvaD Dimer: the interaction between the monomers is clearly seen as two β-strands from one monomer (blue) fold across to interact with the other monomer (red). dTMP bound in the β-barrel of one of the monomers is shown for reference.

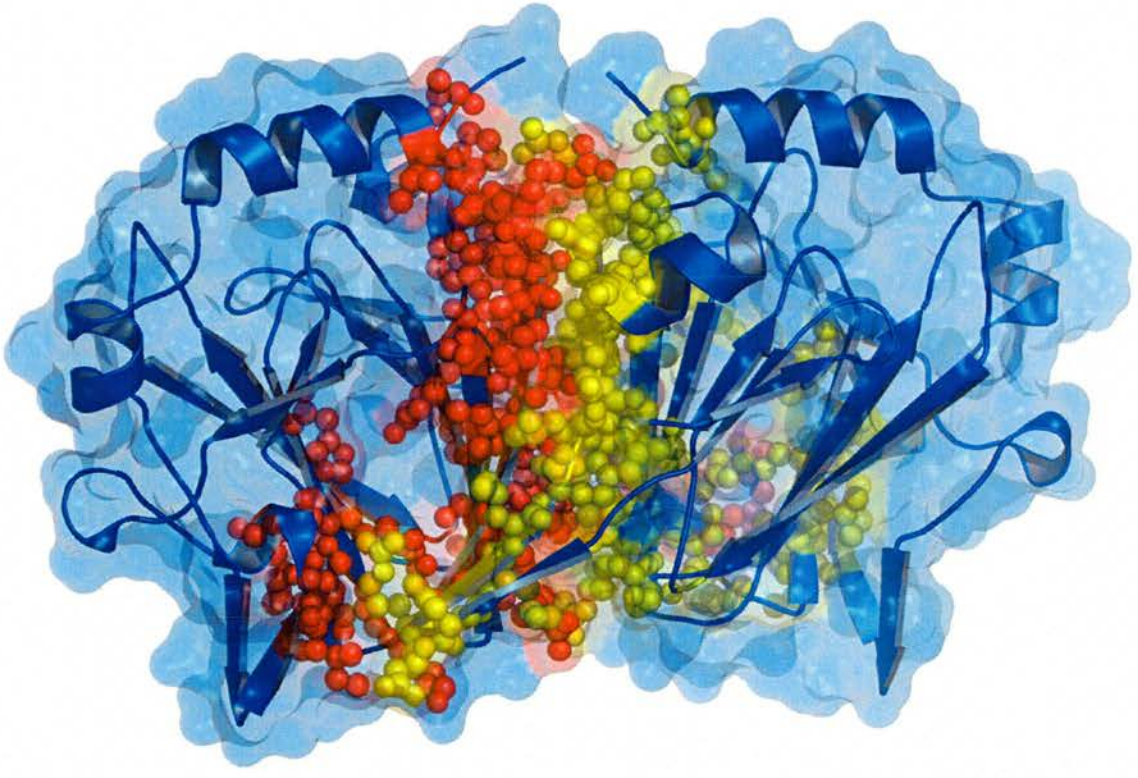


Figure 2.18 – EvaD Dimer Interface: EvaD-dimer, all residues involved in the dimer interface are highlighted. Yellow spheres represent the residues of monomer A and red spheres those of monomer B which are involved in the dimer interface.

2.4.7 dTMP Binding

As previously mentioned, EvaD crystals were soaked with dTDP-$_D$-glucose and the structure was determined to 1.4Å; however, only electron density for dTMP was observed in the active site, as shown in Figure 2.19. It is unclear whether disorder, contamination or degradation is responsible for the lack of the pyranose. Nevertheless, the complex structure, with dTMP bound to the protein, shows that the location of the nucleotide-binding region is identical to that previously published for RmlC enzyme complexes (Christendat *et al.*, 2000; Dong *et al.*, 2003; Giraud *et al.*, 2000). In all RmlCs, the dTDP part of the substrate is always found in an identical position in the active site; this binding is conserved also in EvaD.

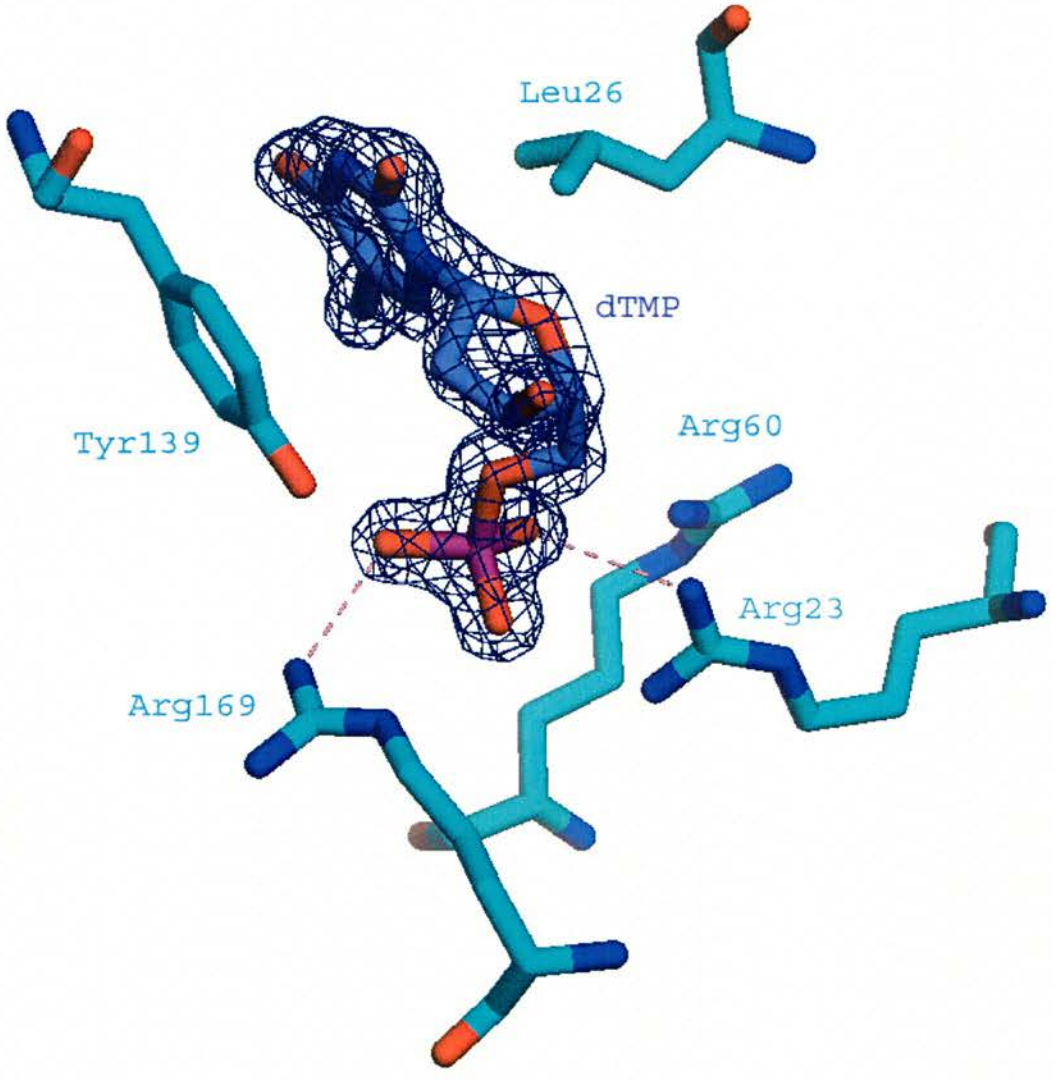


Figure 2.19 – dTMP Binding: part of the EvaD active site showing the TMP bound to the nucleotide binding region. The 2Fo-1Fc electron density is shown at 1σ. Tyr 139, Arg 169 and Arg60 are part of one monomer whereas Arg 23 and Leu26 are part of the other. Hydrogen bonds between the phosphate oxygens Arg23 and Arg169 are indicated; Arg60 is ideally positioned to hydrogen bond with the second phosphate (not present in this model).

No gross structural changes in EvaD between the apo- and the dTMP-complex structures could be observed. The rms deviation between the main chains was calculated as 0.17Å and the rms delta B as 2.20Å^2. The thymidine ring of the substrate is sandwiched between Tyr139 and Leu 26*, where * denotes that this particular residue is part of the other monomer. This is different from the active site of all RmlC enzymes in which two aromatic residues sandwich the thymidine ring by π-stacking on both sides. Loss of the second aromatic side chain in EvaD increases the size of the active site, allowing for the increased substrate size of EvaD compared to that of RmlC. Consequently, the plane of the thymidine ring is offset by approximately 50° relative to the plane of thymidine ring in the RmlC complexes (Dong *et al.*, 2003). The altered nucleotide orientation perturbs the positions of the ribose ring and phosphate groups. Despite these changes, it is clear that the dTMP portion of the ligand is bound in a very similar manner to that seen in RmlC enzymes (Figure 2.20); the ribose ring is not specifically recognized. The phosphate group of dTMP is hydrogen bonded to Arg169 and Arg23*; the absolutely conserved Arg60 is ideally positioned to form a hydrogen bond with the second phosphate group which would be present in the real substrate.

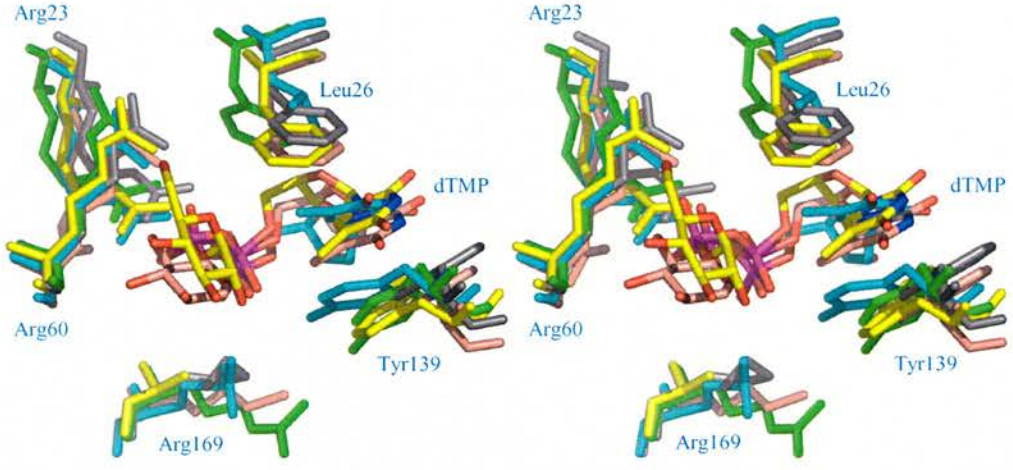
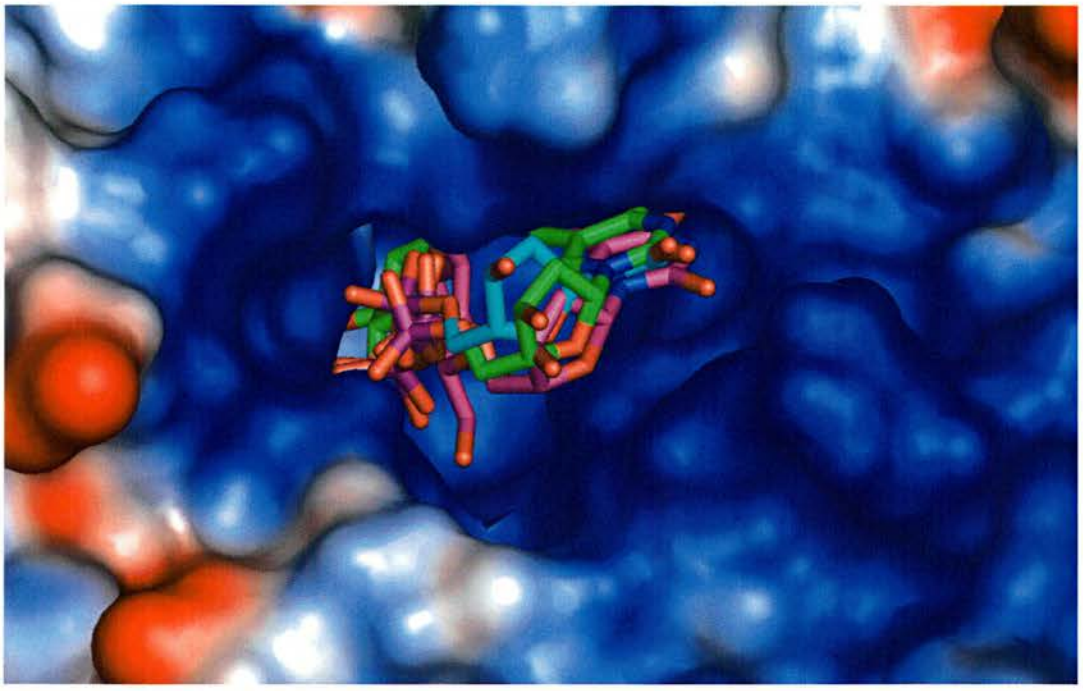


Figure 2.20 – Stereoimage of RmlC and EvaD Residues Involved in Binding of the dTMP Portion of the Substrate: EvaD in complex with dTMP – cyan; RmlC from *M. tuberculosis* – green; RmlC from *P. aeruginosa* – grey; RmlC from *S. suis* in complex with dTDP-xylose – pink; RmlC from *S. suis* in complex with dTDP-glucose – yellow. Labels refer to EvaD residues, Arg23 and Leu26 are part of one monomer, Arg60, Tyr139 and Arg169 part of the other.

The pocket into which the substrate binds is relatively negatively charged. Figure 2.21 shows how dTMP and other substrates (modelled dTDP-xylose and dTDP-glucose) fit into the active site pocket.



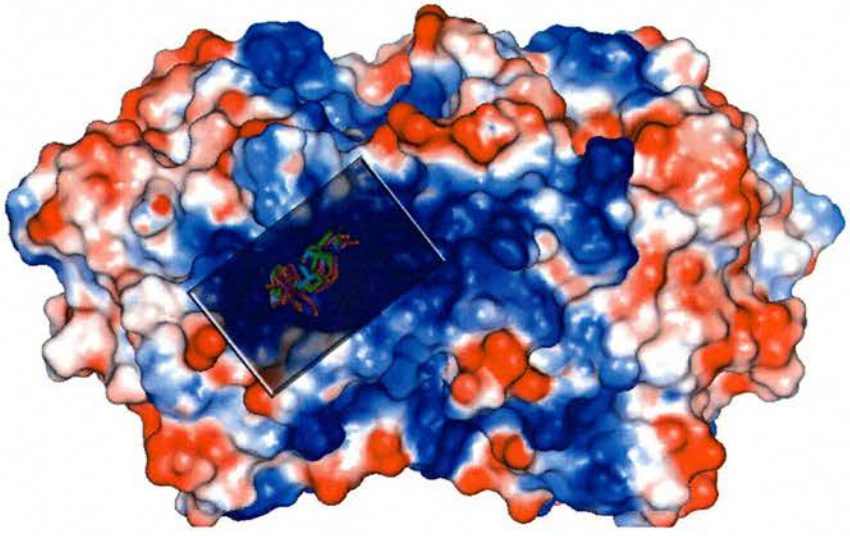


Figure 2.21 - Surface Charge of EvaD: The relative charge of the active site of EvaD is shown - with dTMP (cyan) and dTDP-glucose (green) and dTDP-xylose (magenta) bound. The lower picture serves for orientation: the darker, rectangular region is magnified above.

As I have been unable to obtain a complex with a pyranose ring detectable at the active site of EvaD, the dTMP complex of EvaD and the apo-structure, together with the RmlC-dTDP-xylose, dTDP-glucose and dTDP-rhamnose complexes, were used to generate models of pyranose molecules in the EvaD catalytic site. These were achieved by aligning the carbon alpha backbone of the active site residues of each enzyme with one another. The structures of RmlC from *S. suis* with the substrate analogues (dTDP-glucose and dTDP-xylose) are broadly similar and the sugar nucleotide is bound in a U-shaped conformation (Dong *et al.*, 2003). Subtle changes are observed between the axial dTDP-xylose and dTDP-glucose complexes as well as the equatorial nucleotide linkage sugar ring (dTDP-rhamnose): the pyranose ring is rotated by about 20° for the axial ones, and by 36° for the equatorial nucleotide linkage sugar ring around the glycoside

linkage between the thymidine and pyranose (Chang-Jiang Dong, University of St. Andrews, PhD Thesis 2002). Modelling them all into the active site of EvaD gives a good overview of pyranose binding to the active site of EvaD, and allows us to make comparisons with RmlC.

2.4.8 Catalytic Residues

In RmlCs, the catalytic base (His) is thought to act in a diad with the fully conserved Asp. His/Asp diads have been extensively studied as part of catalytic triads (Ser-His-Asp/Glu) which are found in serine proteases such as trypsin, elastase and α-lytic protease (Frey *et al.*, 1994; Steitz and Shulman, 1982). In triads, the role of the imidazole ring of the His is to act as a general base catalyst to remove the proton from the serine hydroxyl group. The importance of the carboxylic acid in the catalytic triad is less clear. In simulation of the serine esterase cutinase, the active site residues exhibit a great deal of mobility when the Asp-His hydrogen bond is broken (Lau and Bruice, 1999). The catalytic acid in RmlC is a fully conserved Tyr residue (Tyr140 in *S. suis*; Tyr131 in *P. aeruginosa*).

dTDP-glucose is located at the entrance to the β sandwich and the carbohydrate ring is completely enveloped with over 90% of the surface area of the sugar being withdrawn from solvent upon complexation. In the RmlC complex, the O2 atom is bound to two water molecules that are part of an extensive hydrogen bonding network.

The O3 atom is hydrogen bonded to the NZ atom of an absolutely conserved Lys (Lys 71 in *P. aeruginosa*; Lys82 in *S. suis*) (Figure 2.22). This interaction is commonly observed in protein carbohydrate complexes and correlates with affinity (Vyas, 1991).

Two aromatic residues Tyr138 and Phe129 make van der Waals contacts with C6′ and O6 of the substrate. Although not absolutely conserved in RmlC proteins, these residues are very commonly found as Phe (>95% for Phe129 and >50% Tyr138); the substitutions found are conservative for bulky, hydrophobic or aromatic residues.

The O4 atom (where the enolate forms) is close to a fully conserved Lys residue (Lys71 in *P. aeruginosa*; Lys82 in *S. suis* and Lys 72 in *M. tuberculosis*) which may act to stabilise the reaction intermediate (Figure 2.22). However, another residue could act to stabilise the enolate - His119 (in *M. tuberculosis*) is close to the O4 atom too. This

His is almost fully conserved in RmlC, an exception being the homologue from *S. suis*, which comprises an Asn residue in this position (Asn127): however, the ND2 atom of the Asn may take on the role of the ND1/NE2 of His, aiding the stabilisation of the enolate intermediate.

Comparing this to the EvaD crystal structures, as well as models of EvaD with dTDP-xylose and dTDP-glucose positioned in the active site, shows that the residues, crucial for catalysis are completely conserved. In the catalytic site the modelled pyranose ring sits above the ND1 atom of His63, thus the axial groups at the C3′ and C5′ positions point towards this residue, allowing for its activity as the catalytic base. His also acts in a diad with Asp170 whose function is probably to enhance the basic nature of His. The catalytic acid, Tyr133 is conserved in its position compared to RmlC; however, a major change in its orientation is observed, which will be discussed in detail in the next section.

The model shows that the O3 atom may also form a hydrogen-bond with NZ of Lys73; its distance is approximately 2.7Å. The enolate intermediate, which forms on the O4 of the pyranose, may be stabilised by either Lys73 or His120. The proximity of Lys73 and His120 to the O4 atom of the modelled pyranose substrate is shown in Figure 2.22; their distances from the enolate forming O4 are 2.9Å and 2.64Å respectively.

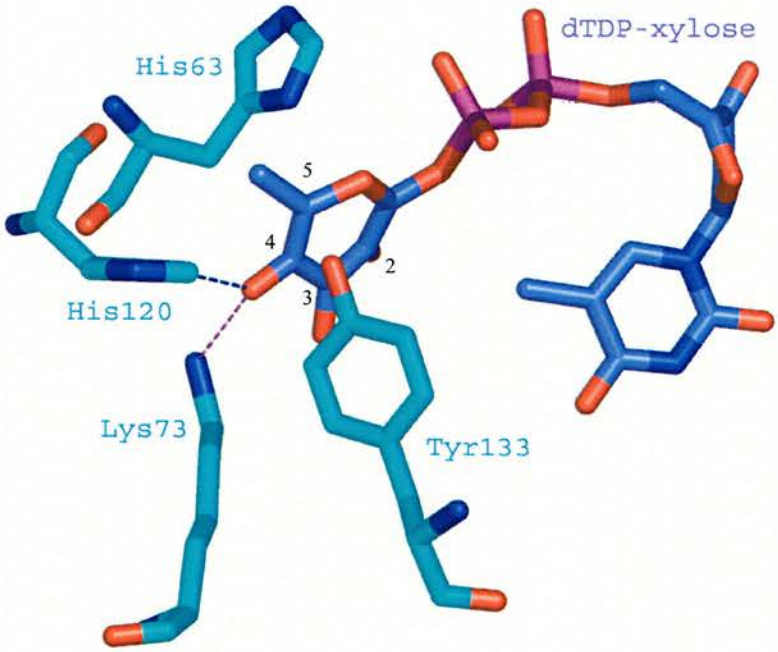


Figure 2.22 – Enolate Stabilisation: the close proximity between the O4 atom of the substrate's pyranose ring and the Lys73 and His120 is visible; either of these may aid in the stabilisation of the enolate intermediate; the distances are 2.9Å and 2.65Å for Lys and His respectively.

The methyl group at C6′ points towards a hydrophobic region made up of Phe122, Met131 and the aliphatic portion of the side chain of Arg60 (Figure 2.23).

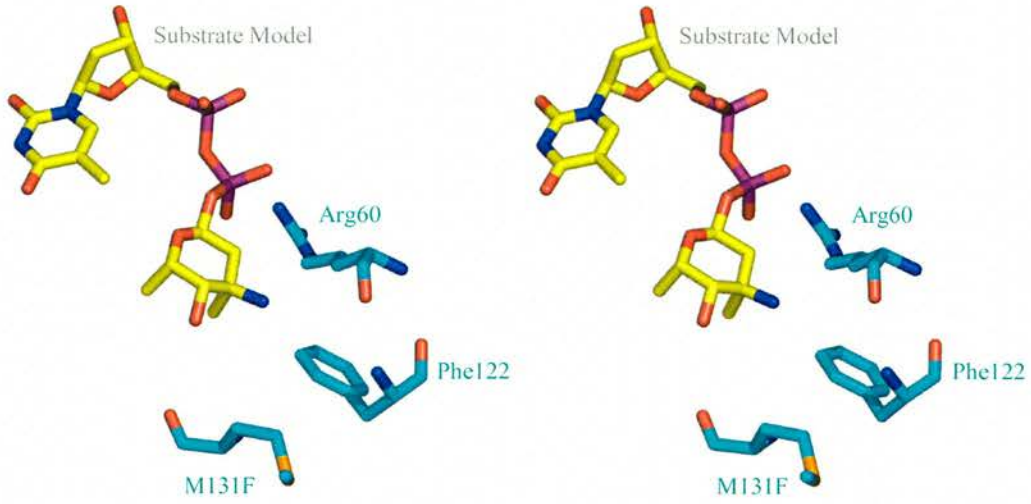


Figure 2.23 – Stereoimage of the Hydrophobic Pocket at the Active Site: the methyl group of the substrate pyranose ring (C6′) points towards a hydrophobic pocket created by Phe122, Met131 and the aliphatic portion of Arg60. All these residues are part of the same monomer. The EvaD substrate model was generated in PyMol using the dTDP-xylose model.

A difference between EvaD and RmlC is the substitution of the conserved aromatic residue Tyr138 (*S. suis* numbering) by a Met in EvaD (Met131) (Figure 2.24); the significance of this will become clear in the next section.

2.4.9 Positioning of the Catalytic Tyrosine

Comparing the active site of the EvaD model to those of RmlC enzymes, there is one striking difference: the catalytic Tyr133 is positioned in a different manner to that found in any of the RmlC enzymes (Figure 2.24). In the RmlC orientation, the phenol group of EvaD's Tyr would clash with the C5′ atom and any equatorial substituent (< 2.0 Å) of the sugar substrate. However, the Tyr in EvaD is rotated in *chi1* relative to the RmlC active site Tyr to give the g- conformer.

The Protein Sidechain Conformation Analysis site (Dunbrack, 1998) (http://dunbrack.fccc.edu/bbdep/confanalysis.php) presents in detail a description of local steric interactions that influence sidechain rotamer and *chi* angle 'choice' of amino acids (Dunbrack and Cohen, 1997). The side chain conformations of amino acids are

limited by the close steric interactions between the delta carbons and the backbone N and C of the same residue. Aromatic sidechain *chi2*'s are perturbed by interaction between XD1 and XD2 and backbone N and C; without perturbation, *chi2* would be near +90 or –90. When the *chi1* rotamer (r1) is trans, interaction between backbone N and XD2 and XD1 at *chi2* = 120 or –60 pushes the average for *chi2* to values below 90, or below –90 respectively. Similarly, for r1 of g-, the averages are pushed to values above *chi2* = 90 and *chi2* = -90 by interactions of N and XD1 and XD2 when *chi2* is 60 or –120. For g+ rotamers, interaction at *chi2* = 60 and 120 exert steric conflict about equally, so that *chi2* averages 90 degrees. The r1 for all sidechains except Pro are defined as follows:

- $0 \leq chi1 < 120$ g+
- $120 \leq chi1 < 240$ trans
- $-120 \leq chi1 < 0$ g-

The probability that an amino acid should adopt a certain conformer has been studied extensively (Dunbrack and Cohen, 1997). For Tyr residues, the g- rotamer is the energetically most favourable and the probability of an amino acid adopting this conformation is almost 0.6. In contrast, the g+ conformer is the energetically least favourable, with a probability of the amino acid adopting that conformation of only 0.15.

Therefore, it is interesting that the catalytic Tyr of EvaD adopts the g- conformer (*chi1* = -53°, *chi2* = -117°), whereas that of all RmlC homologues adopts the far less favoured g+ conformer. In RmlC enzymes, the Tyr cannot adopt the g- conformation because an aromatic residue (Tyr138 in *S. suis*) occupies the space. In an examination of more than 80 RmlC sequences, an aromatic residue is found in 71%, a Val in 26% and an Ile in 3% in this particular position (i.e. 2 residues prior to the catalytic acid). All residues (Phe, Thr, Trp, Val, Ile) occupying the space have a tertiary Cβ atom and would clash with a g- conformation of the Tyr residue.

In EvaD, the blocking aromatic residue is changed to Met131. This amino acid has a secondary Cβ atom and is positioned in such a way as to create a pocket for the g- conformer. Interestingly, in this conformation, Tyr133 is close to the side chain of His120, as the orientation of the imidazole ring is fixed by Asp84. Although it does not make a hydrogen bond, Tyr133 is actively prevented from adopting the g+ conformer as it would clash with the side chain of Leu135 in the orientation (<1.0Å). In over 80

RmlC sequences the residue at this position is confined to one of the following: Thr, Val or Cys, all of which are small or flexible enough to accommodate the g+ conformation of the active site Tyr. The hydroxyl group of the Tyr in RmlC is also hydrogen bonded (3.0Å) to a conserved Asn (Asn52 in RmlC from *M. tuberculosis*), which itself is part of a hydrogen bonding network (Glu50, Trp140 (main chain N)). This presumably helps hold the Tyr in the g+ conformation. This hydrogen bonding network is absent in EvaD, as Ile50 is equivalent in position to Asn52. Thus, the catalytic Tyr appears to be locked into different conformations in RmlC and EvaD (Figure 2.24). This is expected to have an impact upon catalytic activity of EvaD compared to the RmlC homologues.

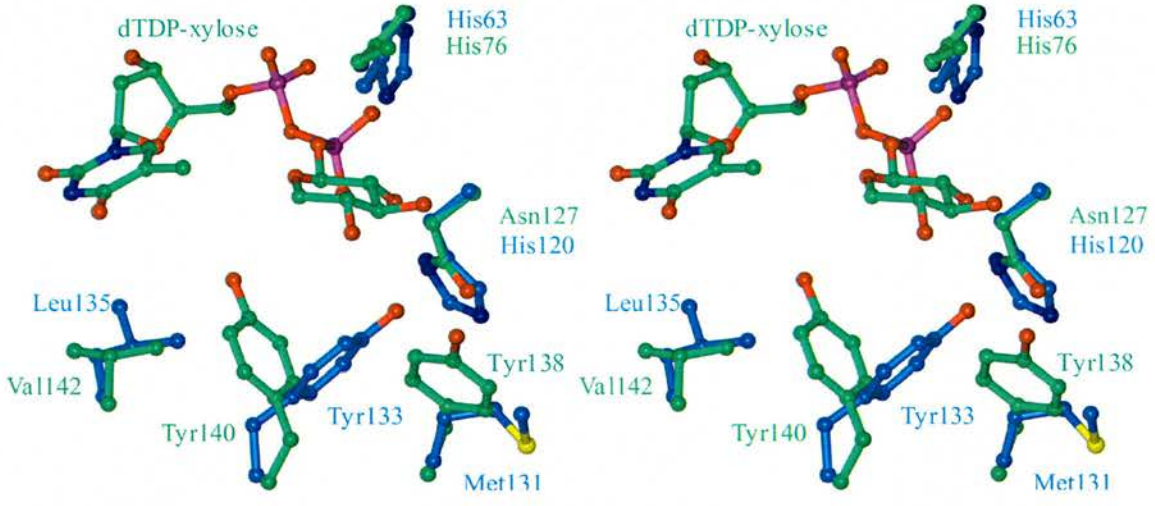


Figure 2.24 – Stereoview of the Active Site of EvaD and RmlC from *S. suis*: dTDP-xylose is shown for reference. EvaD residues are blue, and those of RmlC are green. The difference in *chi1* between Tyr140 (RmlC from *S. suis*) and Tyr133 (EvaD) is clearly seen.

2.4.10 Structure of the M131F/L135A Double Mutant

The overall structure of the double mutant is essentially identical to that of the wild-type enzyme; the rms deviation of the main chains was calculated as 0.475Å and the rms delta B as 5.615Å^2. The electron density for the side chains of residues Phe131 and Ala135 confirms that this is indeed the double mutant (Figure 2.25). However, contrary to our predictions, the catalytic Tyr does not seem to have changed position at all, compared to wild-type EvaD, as Figure 2.26 shows.

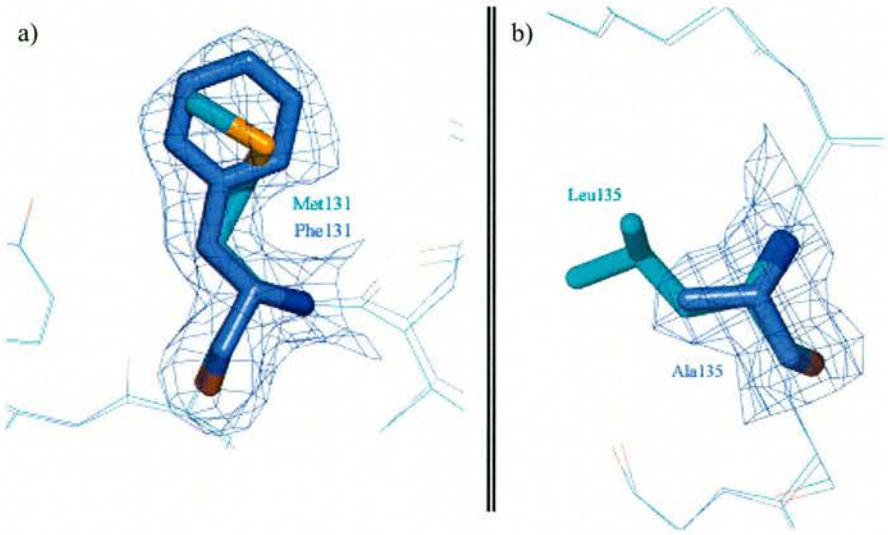


Figure 2.25 – Side Chains of the EvaD Double Mutant: the changes in the side chains a) M131F and b) L135A can be clearly seen. Wild-type EvaD is in cyan, the double mutant in purple. The 2Fo-Fc density is contoured at 1σ.

The loop made up of residues 140 to 143 may be slightly distorted, compared to wild-type EvaD, but the density is ambiguous. This loop is flexible in RmlC, depending on substrate binding. It may be associated with the substrate entering and leaving the active site (Chang-Jiang Dong, University of St. Andrews, PhD Thesis 2002). Additionally, Lys73, which is thought to be important to enolate stabilisation, is rotated by approximately 15° around *chi1*, as compared to wild-type EvaD (Figure 2.26).

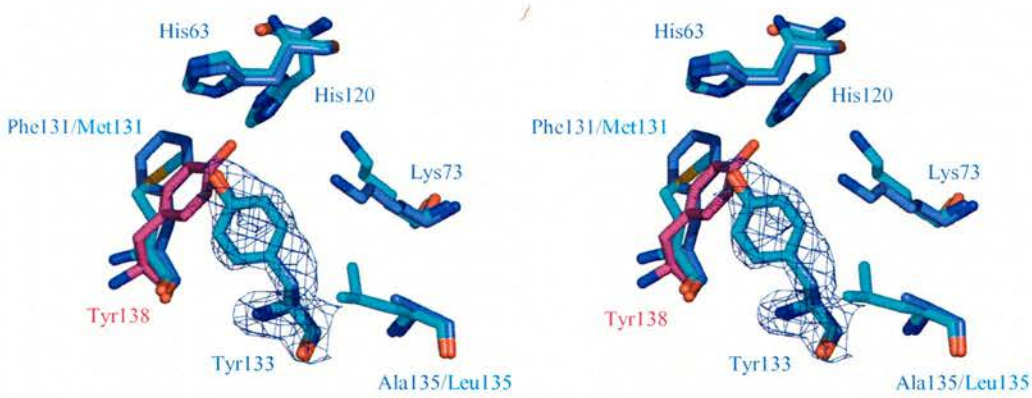


Figure 2.26 – Stereoimage of the Active Site of the EvaD Double Mutant: wild-type EvaD is coloured in cyan, the double mutant in purple; all active site residues are depicted. The 2Fo-Fc density for Tyr133 of the double mutant is contoured at 1σ. The mutation did not assume the position of Tyr138 (red) of *S. suis* to block the pocket.

We had predicted, that Tyr133 would assume the g+ position if we mutated the residues responsible for holding it in the g- position; a major role that we had proposed for Met131 (as discussed in section 2.4.9). Mutating it to Phe which has a tertiary Cβ was thought to hinder the g- positioning of Tyr133. We had predicted that the mutated

residue should assume a position equivalent to that observed in RmlC. However, surprisingy, our crystal structure demonstrated that the presence of a tertiary Cβ in the pocket is compatible with the g- rotamer of the catalytic Tyr. The Phe is rotated by approximately 75° around *chi1* compared to the equivalent Tyr138 in RmlC from *S. suis*; this means that contrary to our predictions, the tertiary Cβ is not blocking the Tyr pocket by adopting a position similar to the original Met which does not have a tertiary Cβ.

This orientation of Phe131 could not have been predicted from our native crystal structure but the physical basis for it became apparent from the details of the double mutant crystal structure. A steric clash was observed between the CE1 atom of His120 and the predicted position of Phe131; the distance between these two atoms is only 1.5Å, preventing Phe131 from adopting the RmlC conformation. In the RmlC homologue from *S. suis*, the His is changed to Asn127, allowing hydrogen bonding to hold the Tyr in place. All other RmlCs display a His residue in this position but a bit further removed from the active site than EvaD. The shift in the main chain is only minimal and reasons for it cannot easily be observed by superimposing EvaD and RmlC enzymes.

2.4.11 Biochemical Characterisation

Wild-type EvaD, as well as its mutants, were used in biochemical assays to compare their respective activities to those of RmlC enzymes. The active site mutants H63A and Y133F were made to confirm the amino acids' roles as catalytic base and acid respectively. To restore the hydrogen-bonding network with the hydroxyl group of Tyr133, Ile50 was mutated to Arg (I50N). The residue that would clash with the g+ rotamer, Leu135, was removed (L135A) to allow the repositioning of Tyr133 into the g+ conformation. To force Tyr133 out of its g- conformation, the Met that creates a pocket for it to go into, by having a secondary Cβ, was mutated to introduce a tertiary Cβ, which should clash with the g- conformation of the catalytic residue (M131F). A double mutant (M131F/L135A) and a triple mutant (I50N/M131F/L135A) were made to incorporate all the above changes. Table 2.11 summarises the relative turnover rates of the mutants, as compared to wild-type EvaD. (The raw data is presented in Appendix D). Table 2.12 summarises the enzyme assay results for EvaD and its M131F

mutant. Both enzymes display typical Michaelis Menton kinetics (Cornish-Bowden, 1999) and the apparent K_M and k_{cat} are given. The M131F mutant displayed a 40% increase in reaction rate even though the binding affinity for the substrate (apparent K_M) was decreased by about 1.6-fold compared to wild-type EvaD.

Table 2.11 - Relative Reaction Rates of the Different EvaD Mutants Compared to Wild-type EvaD

EvaD mutant enzyme	Relative reaction rate compared to wild-type
H63A	Zero
Y133F	Zero
I50N/M131F/L135A	Zero
I50N	Decreased
L135A	Decreased
M131F/L135A	Decreased
M131F	Increased

As expected, the H63A and Y133F mutants are completely inactive, confirming the roles of His and Tyr as the catalytic base and acid respectively. In addition, this served as another negative control for the assay, i.e. the epimerisation seen for the wild-type mutant enzymes is due to EvaD and not to any contaminants that may still be present in the sample even though not seen on an SDS gel. The triple mutant was inactive too; this had been expected as the CD spectrum detected the absence of any secondary and tertiary structural characteristics, indicating unfolded protein.

Table 2.12 - Enzyme Assay Results

Enzyme	Apparent K_M (mM)	k_{cat} (1/s)	k_{cat}/K_M	Protein (μM)
EvaD	0.274 ± 0.033	0.08 ± 0.003	0.31	12.5
EvaD M131F	0.455 ± 0.022	0.12 ± 0.002	0.27	12.5
S. suis RmlC	0.029 ± 0.003	10.4 ± 0.3	357	0.027
S. enterica RmlC	0.081 ± 0.008	19.2 ± 0.5	236	0.029

The biochemical assay shows that wild-type EvaD, as well as, the M131F mutant are capable of epimerising C3′ and C5′; however, they are not very efficient. The activity of wild-type EvaD is more than 200 fold reduced, as compared to RmlC. Even though RmlC from *S. suis* is structurally and in sequence the least conserved enzyme in the enzyme family, it is efficient in the archetypical RmlC double epimerisation reaction (double epimerisation of dTDP-6-deoxy-4-keto-D-glucose at C3′ and C5′). This indicates that despite EvaD being similar to RmlCs in structure and

sequence, it cannot be a true member of this enzyme family. The M131F mutant's turnover rate (apparent k_{cat}) in the assay is increased by 1.5 fold, despite its 1.6 fold decrease in apparent substrate affinity. All the other mutants displayed a decreased activity compared to wild-type EvaD.

As EvaD probably only catalyses the C5′ epimerisation in its native pathway, deuterium incorporation studies were carried out to analyse C3′ and C5′ epimerisation, in terms of deuterium incorporation, separately. The incorporation of deuterium at these positions should be equivalent to the epimerisation; thus, single epimerisation at either position can be detected. Since we have no way of producing the real substrate for EvaD (dTDP-3-amino-2,3,6-trideoxy-3C-methyl-D-erythro-hexopyranosyl-4-ulose), the substrate for RmlC, dTDP-4-keto-6-deoxy-D-glucose, which can be conveniently produced from dTDP-glucose using RmlB, was used in the assay. Figure 2.27 gives a graphical representation of the deuterium incorporation results, which are shown in Table 2.13.

Table 2.13 – Percentage Deuterium Incorporation

	None	C3′	C5′	C3′ and C5′
Negative control (first)	79.3 ± 3.5	8.2 ± 2.1	5.4 ± 2.2	7.1 ± 3.7
RmlC (positive control)	11.2 ± 3.0	5.9 ± 2.2	15.3 ± 3.5	67.6 ± 3.4
EvaD 30min	21.4 ± 1.5	7.9 ± 1.3	62.2 ± 2.0	8.6 ± 0.8
EvaD 120min	17.8 ± 2.6	10.4 ± 2.2	64.0 ± 0.7	7.8 ± 0.6
M131F 30min	22.7 ± 5.2	14.9 ± 9.0	46.8 ± 1.5	15.7 ± 4.6
M131F 120min	14.8 ± 2.9	6.1 ± 1.5	16.1 ± 1.6	63.1 ± 2.8
H63A 120min	76.5 ± 3.6	7.8 ± 1.9	7.6 ± 2.2	8.1 ± 3.5
I50N 120min	50.1 ± 4.6	6.9 ± 3.2	33.9 ± 1.1	9.1 ± 2.5
L135A 120min	53.4 ± 1.7	5.9 ± 2.9	32.5 ± 4.9	8.2 ± 3.2
M131F/L135A 120min	58.9 ± 3.3	10.5 ± 3.6	24.0 ± 2.5	6.6 ± 2.3
Negative control (second)	73.9 ± 3.3	0.0 ± 1.1	13.1 ± 2.6	13.1 ± 1.7
Y133F 30min	74.8 ± 2.1	0.3 ± 0.9	12.3 ± 2.0	12.6 ± 1.1
Y133F 120min	74.8 ± 0.7	0.0 ± 0.1	13.0 ± 0.3	12.5 ± 0.5

The second graph shows that neither the H63A nor the Y133F mutants show any deuterium incorporation. These results confirm the above kinetic results and show that these residues are crucial to catalysis; this concurs with the reasoning that they are catalytic base and acid respectively. As with the kinetic assays, the 'dead' mutants also confirm the validity of the assay.

The native enzyme preferentially incorporates deuterium at C5′, whereas RmlC catalyses the characteristic double incorporation at C3′ and C5′. RmlC shows some incorporation at C5′ only, but mainly displays the archetypical double deuterium incorporation reaction. Wild-type EvaD incorporates no deuterium at C3′, since neither the C3′ incorporation, nor the double incorporation at either time point, is significantly different from the background rate. However, there is a significant amount of deuterium incorporation at C5′, indicating that single epimerisation will be carried out at this carbon atom.

Initially, as the 30 minute time point shows, the EvaD M131F mutant shows less deuterium incorporation at C3′ and it is much more similar to that of native EvaD than to RmlC. However, as equilibrium approaches (after 120min in this particular assay) the deuterium incorporation at C5′ as well as C3′ is very similar to the one seen for the RmlC controls. This means that this mutant re-establishes the catalytic double epimerisation; however, is not as efficient as RmlC. This again is in accordance with the kinetic results presented above: the substrate affinity for EvaD, as well as its single amino acid mutant M131F, is very much reduced compared to RmlC; both enzymes are capable of carrying out the double epimerisation, but M131F is more active than the wild-type enzyme.

The I50N, L135A and M131F/L135A mutants display reduced deuterium incorporation. No significant fragments, showing the 3′ or double incorporation of deuterium into the substrate, could be detected. There is some deuterium incorporation at C5′, but it is significantly less than that observed for the wild-type enzyme. All these results are in line with the kinetic data obtained but the deuterium incorporation identified the epimerisation position(s) close to equilibrium of the reaction.

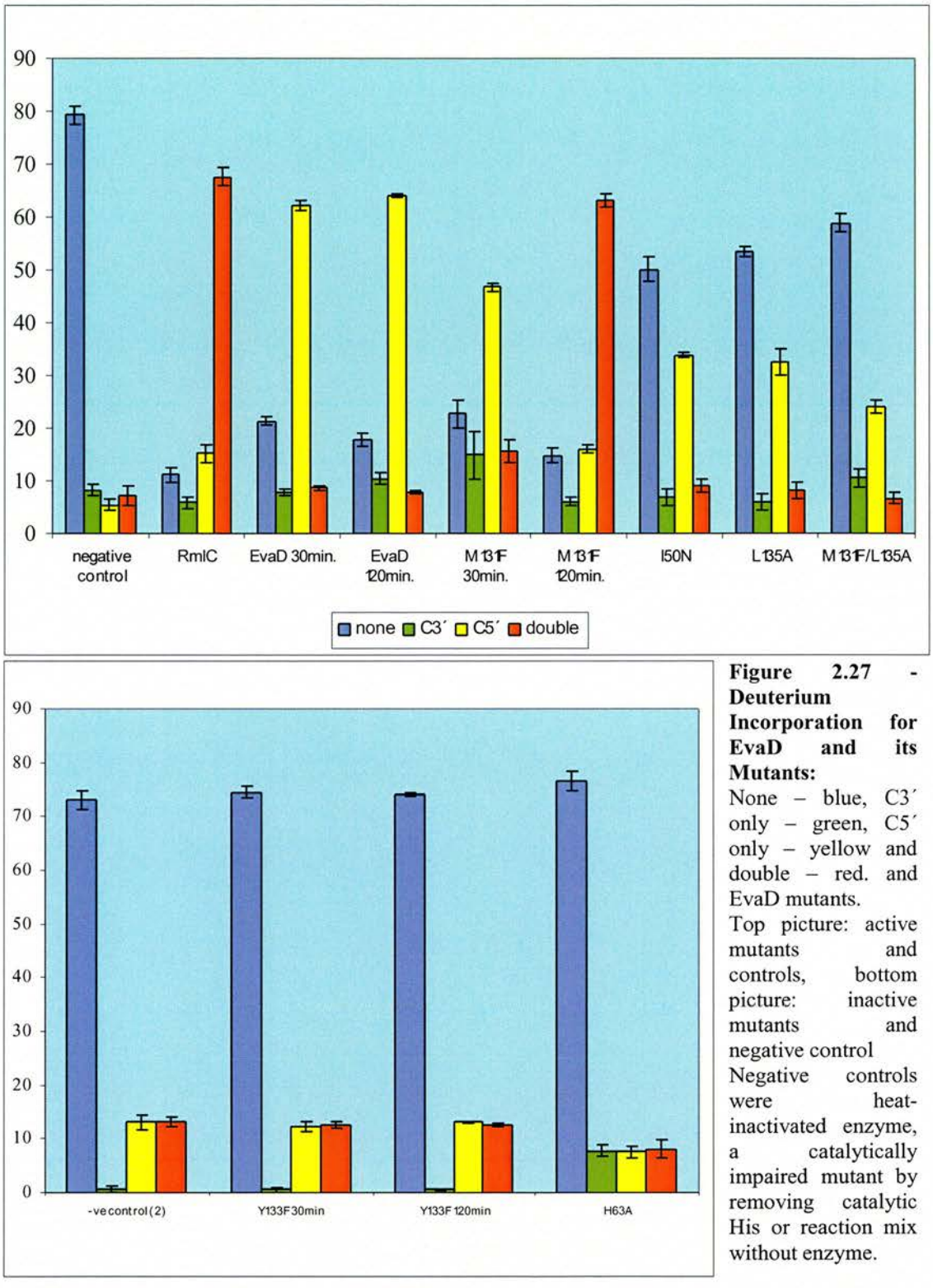


Figure 2.27 - Deuterium Incorporation for EvaD and its Mutants:
None – blue, C3´ only – green, C5´ only – yellow and double – red. and EvaD mutants.
Top picture: active mutants and controls, bottom picture: inactive mutants and negative control
Negative controls were heat-inactivated enzyme, a catalytically impaired mutant by removing catalytic His or reaction mix without enzyme.

A comparison of a deuterium incorporation time-course experiment with kinetic data for EvaD, indicated that deuterium incorporation proceeds ten-fold faster than the epimerisation reaction. However, these values should be treated with caution as the two experimental systems utilise different conditions.

2.5 Conclusions and Discussion

Prior to this study, two suggestions had been made for the placement of EvaD in the dTDP-epivancosamine biosynthetic pathway. One proposal stemmed from the 40% homology of EvaD to RmlC in the dTDP-L-rhamnose pathway and the previously demonstrated activity of RmlC as a 3′, 5′ epimerase acting on the dTDP-4-keto-6-deoxyhexose intermediate in that pathway (Graninger *et al.*, 1999). Kirkpatrick *et al.*, 2000, reported that EvaD could catalyse the exchange of protons at C3′and C5′ and proposed that EvaD was functionally an RmlC. However, another investigation with purified EvaA to E, may have established the full reconstitution of the pathway from dTDP-4-keto-6-deoxy-D-glucose to dTDP-L-epivancosamine. It suggested that EvaD would take dTDP-3-amino-3-methyl-4-keto-6-deoxy-D-glucose and epimerise the C5′ centre to generate the L-sugar, which can then be reduced at C4′ by EvaE to complete the L-epivancosamine pathway (Chen *et al.*, 2000).

Thus, while EvaD may possess the capacity to generate C3′ and C5′ carbanion intermediates (i.e. C3-C4 / C5-C4 enolates), it normally works on a 3-methyl-3-amino substrate, and hence the 5′ epimerase activity would be the relevant one.

Determining the three-dimensional structures of wild-type EvaD, EvaD with dTMP bound, and the double M131F/L135A mutant, has given a structural insight into the differences in catalytic activity. Also, biochemical studies with wild-type and mutant EvaD, have confirmed the enzyme's role as the 5′ epimerase. Residues which are thought to be of catalytic importance were identified: His63 is thought to act as the base, abstracting the proton from the 5′ position of the substrate and Tyr133 as the catalytic acid for re-protonation of the substrate on the other face of the sugar. The residues that may stabilise the enolate reaction intermediate Lys73 and His120 are positioned in close proximity to O4 allowing either of them to stabilise the enolate reaction intermediate. Enzyme assays and deuterium incorporation studies have shown that EvaD is efficient only at carrying out the C5′ epimerisation, confirming the reaction pathway for the synthesis of L-epivancosamine as proposed by Chen, *et al.*, 2000 (Figure 2.28). This puts EvaD into the fourth position in the pathway, carrying out only the 5′ epimerisation of dTDP-3-amino-2,3,6-trideoxy-3*C*-methyl-D-*erythro*-hexopyranosyl-4-ulose.

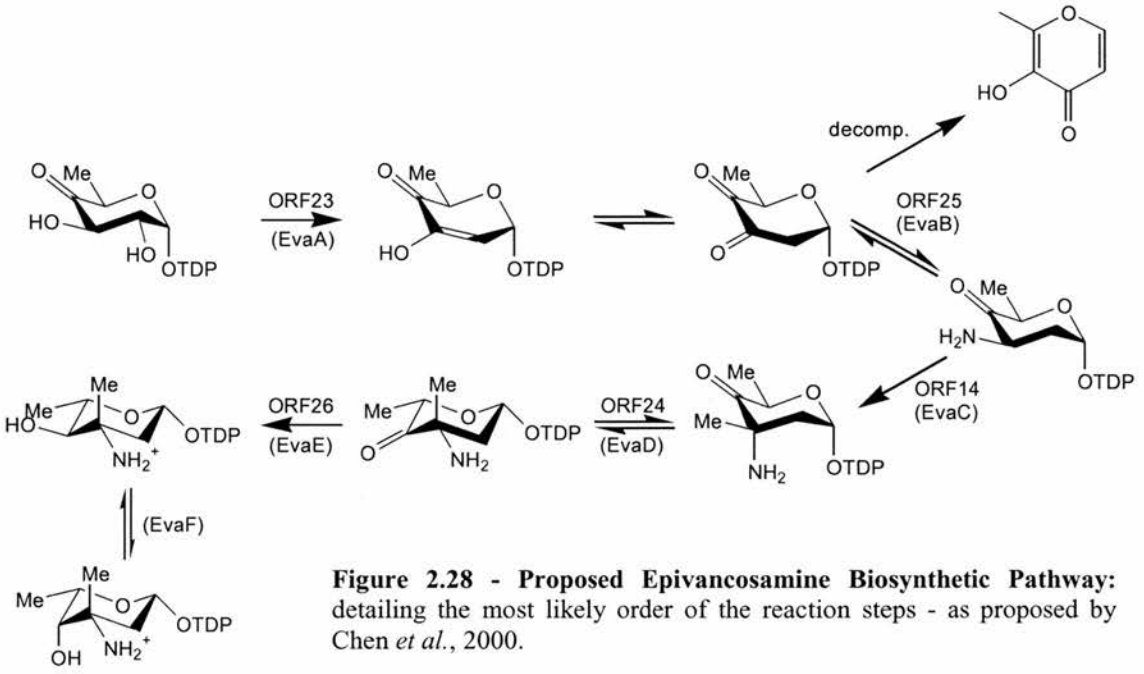


Figure 2.28 - Proposed Epivancosamine Biosynthetic Pathway: detailing the most likely order of the reaction steps - as proposed by Chen *et al.*, 2000.

The binding of thymidine phosphate in the active site is identical to that found in the RmlC class of enzymes, which carry out the double epimerisation at C3′ and C5′. The residues involved in the phosphate and thymidine binding have been identified and are conserved throughout all RmlC homologues, as well as in EvaD, with the exception of one aromatic residue that creates a π-stacking sandwich in RmlCs. In EvaD, this residue is replaced by a smaller, hydrophobic one (Leu26) to open up the binding site, allowing the accommodation of EvaD's larger natural substrate. At the same time, this results in a rotation of the pyranose ring around the glycoside linkage between the thymidine and pyranose. This rotation does happen to some degree in RmlC, which seems to bind equatorial sugars differently to axial ones, but it seems to be increased in EvaD.

The main difference in the active site between EvaD and RmlC has been identified as the positioning of the catalytic Tyr residue. It is in its g- conformation in EvaD, but in the less commonly observed g+ conformation in all RmlC enzymes. Mutations of residues surrounding the Tyr seem to be responsible for that difference. A Met (Met131) opens up the pocket for the Tyr to go into in the case of EvaD. This pocket is blocked by residues that have a tertiary β-carbon. In more than 80 RmlC homologues, 71% contain an aromatic, 26% a Val and 3% an Ile in this position, all of these residues sterically interfere with the formation of the g- conformer of the catalytic Tyr. At the same time EvaD contains a Leu on the other side of the Tyr, thus preventing

it from flipping into the g+ conformation. In addition to this, the Tyr in RmlCs is held in place by a hydrogen bonding network, made up of three amino acids. In EvaD the first one of these residues is mutated to an Ile which breaks the hydrogen bonding network as this amino acid has no potential for hydrogen bonding with the Tyr133. The catalytic Tyr may be held in place by a His120, but this cannot be confirmed, as our crystallographic data does not allow us to make a distinction between the two possible His conformations, i.e. we cannot determine which way round the His faces (i.e. N or C towards Tyr133); both conformers are equally plausible when taking the surrounding environment into account.

In an attempt to reposition the catalytic Tyr and, therefore, potentially recreate the capacity of EvaD to carry out the double epimerisation efficiently, several mutations at and around the active site were made and the mutants assayed in the archetypical RmlC assay. Their capacity for incorporating deuterium at C3′ and at C5′ was also analysed.

The repositioning of the Tyr is somewhat debatable as I have not been able to crystallise the M131F mutant. However, this one is catalytically more active than the wild-type enzyme, even though its affinity for the RmlC substrate is greatly reduced. It incorporates deuterium at C3′ and at C5′ compared to the 5′ only incorporation that is seen for wild-type EvaD. However, this incorporation is achieved more slowly than that seen for its RmlC counterpart. The M131F/L135A double mutant, as well as the other single mutants made, is less efficient than EvaD in both the kinetic and the deuterium incorporation assay.

The C3′ and C5′ incorporation of deuterium by M131F supported our theory of repositioning of the catalytic Tyr residue by mutating the surrounding amino acids to block the pocket. However, since the assays were done, crystallographic data was obtained for the double mutant: it shows that mutation of the Met131 and Leu135 do not reposition the Tyr133. Its active site is not changed visibly as compared to the native enzyme, only the Lys73, which is thought to stabilise the reaction intermediate, is rotated by about 15° around *chi1*.

The Phe of the M131F mutation did not adopt the predicted position that was found for all other RmlC enzymes; rather it adopted a position more similar to the original EvaD Met so the tertiary Cβ introduced in this position did not actually

sterically hinder the g- rotamer conformation of the catalytic Tyr133. The reason for Phe being rotated by approximately 75° around *chi1* compared to the equivalent residue in RmlCs is due to the steric hindrance imposed by His120: the distance between the CE1 and the predicted Phe was only about 1.5Å. In RmlCs the main chain around the equivalent His is moved further away from the active site, allowing the rotation of Phe/Tyr to the position observed.

Additionally, the activity of the double mutant in the enzyme assay compared to that of wild-type EvaD and single M131F mutant is reduced, as is its ability to incorporate deuterium at C3′ and C5′. All in all, we cannot explain these findings and it demonstrated that the results of single amino acid changes on the overall structure of an enzyme are harder to predict than originally thought.

Before obtaining the structural data for the double mutant, we regarded the biochemical and deuterium incorporation results as evidence that our structural interpretation is correct and that fine tuning of the Tyr133 position allows EvaD to distinguish between the wrong substrate and dTDP-3-amino-2,3,6-trideoxy-3*C*-methyl-D-erythro-hexopyranosyl-4-ulose. Now, the interpretation of all the data obtained is less clear.

However, the discrimination imposed by the catalytic Tyr rotamer may be required because it is not in the organism's interest to epimerise the true RmlC substrate (dTDP-6-deoxy-4-keto-glucose) to any significant degree. This substrate is part of the dTDP-L-epivancosamine pathway, most likely being the substrate for EvaA, the hexose 2,3-dehydratase; its double epimerisation would reduce the yield of L-epivancosamine by creating an alternative synthetic route. The subtle change in the position of an active residue is an elegant method of achieving such discrimination.

Interestingly, WbcA (SwissProt, TrEMBL entry: Q56862) from *Yersinia enterolitica* is predicted to be a mono-epimerase at C3′ only (Reeves *et al.*, 1996; Zhang *et al.*, 1997). This epimerase is involved in the biosynthesis of dCDP-6-deoxy-D-gulose, the precursor of D-gulose, a deoxysugar present in the *O*-antigen part of *Y. enterolitica*. The enzyme's substrate is dCDP-6-deoxy-D-xylo-4-hexulose. WbcA is homologous to EvaD and RmlC but lacks the conserved catalytic Tyr which is replaced by a Cys (Cys133) as shown by sequence alignment. Of course, this Cys cannot act as the

catalytic acid for the epimerisation. A notion for RmlC activity has been that the C5′ is re-protonated by the active site Tyr and the C3′ by a conserved water molecule in the active site and not the Tyr residue. This idea is in accordance with all data we have obtained so far: the conserved water does not exist in EvaD which carries out the C5′ epimerisation only; WbcaA may well contain this water molecule but does not have a Tyr and therefore only carries out the C3′ epimerisation. This would be a very elegant way of nature to determining the C3` versus the C5′ and the double epimerisation.

However, this is only a prediction since we do not have enough structural data to confirm all this. There is some evidence against this notion as well: the putative 3′ mono-epimerase NDP-hexose 3-epimerase TylJ (SwissProt, TrEMBL entry: Q9S4D4) from *Streptomyces fradiae* (Bate and Cundliffe, 1999; Fouces *et al.*, 1999), which is homologous to EvaD (42% sequence identity and 61% sequence similarity) and RmlCs, contains a Tyr at the right position in an alignment with EvaD and RmlC enzymes (Figure 2.29). Nevertheless, its structure has not been determined and to my knowledge this enzyme is a putative 3′ epimerase, and therefore may turn out to epimerise other position(s).

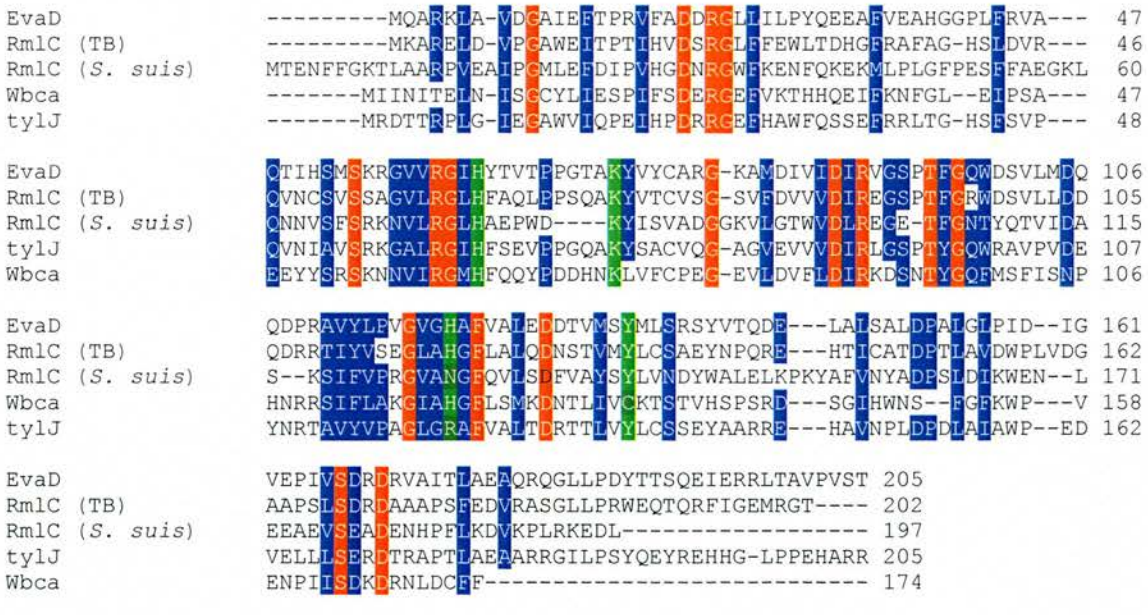


Figure 2.29 – Sequence Alignment of EvaD, RmlC, WbcA and TylJ: complete conservations throughout all epimerases are highlighted on red and conservative substitutions on blue background. The catalytic residues in RmlC and EvaD, as well as those, that align with them are highlighted on green, where the non-conserved ones are shown on dark green.

2.5 Further Work

The conformation of the catalytic Tyr133 in EvaD in the M131F single mutant needs to be identified from structural studies. This will explain why this mutant is more active than wild-type enzyme and hopefully confirm our theory of repositioning the catalytic Tyr. Secondly, it would be very interesting to see the difference in the active site between WbcA, TylJ and RmlC and EvaD, as it may allow us to verify the difference in the active sites of 3′, 5′ mono- and double epimerases. This may give more insight into the differences between epimerisation positions as well as the substrate selectivity of the enzyme in question, which is extremely important since EvaD has to differentiate between its 'right' and 'wrong' substrate, which are both present in the same pathway.

All enzymes in the L-epivancosamine pathway need to be purified, their activities confirmed and the crystal structures determined. This will allow the establishment of the full pathway and we will thus gain further insight into substrate selectivity and the reaction mechanisms as catalysed by the different enzymes at every reaction step.

Chapter 3

EvaA – 2,3-dehydratase

3.1 Summary

EvaA, encoded by orf23, is thought to be the first enzyme in the L-epivancosamine biosynthesis pathway in *A. orientalis*. It belongs to the family of NDP-hexose 2,3-dehydratases and catalyses the 2,3-dehydration of dTDP-4-keto-6-deoxyglucose, using the co-factor flavin mononucleotide. Enzymes in this family are made up of two similar domains which may have arisen from domain duplication.

The 2,3-dehydration reaction is without precedence in the literature and structure determination will allow us to determine how the reaction is catalysed. The aim was the structure determination of EvaA and some of its homologues to identify the basis for substrate selectivity.

Attempts at solubly expressing EvaA on its own in *E. coli* failed; however, the protein could be expressed as a fusion protein with the maltose binding protein from *E. coli*. After the fusion protein had been purified chromatographically using amylose resin, it was cleaved by TEV protease. Subsequently, MBP, TEV protease and EvaA were separated using anion exchange chromatography. For this step, only ANX Fast Flow Sepharose™ could be used since it allowed weak binding of the proteins via a particularly long diethylaminopropyl linker region.

Since no suitable model for molecular replacement is known, experimental phasing for EvaA has to be carried out. For this reason, SeMet preparations of EvaA were made using the metabolic inhibition pathway. SeMet protein was purified in the same way as wild-type protein adding additional DTT to the buffers to keep the protein reduced. More than 90% SeMet incorporation has been achieved which will be suitable for phasing if crystals of the derivative protein can be grown.

Crystal trials for EvaA in its MBP bound and free form have been carried out and protein crystals obtained. To date, these crystals have not yielded any X-ray diffraction, except some powder diffraction typical of spherulites.

Several EvaA homologues from Actinomycetales were chosen and the bacteria revived from freeze dried samples. The genomes of five of these were isolated and the genes encoding the EvaA homologues amplified. Several genes were successfully cloned into plasmids for recombinant protein expression in *E. coli*. The homologues proteins were purified similarly to EvaA and crystallisation trials were set up.

3.2 Introduction

In the course of a dehydration reaction a water molecule is eliminated from the substrate by a dehydratase enzyme. The first step (according to Chen *et al.*, 2000) or possibly the second step (according to Kirkpatrick *et al.*, 2000) in the biosynthesis of L-epivancosamine involves such a reaction. It is catalysed by EvaA, dTDP-4-keto-6-deoxyglucose 2,3-dehydratase, encoded by orf23 in *A. orientalis*. It belongs to a family of NDP-hexose 2,3-dehydratases, which include a range of proteins from antibiotic production pathways, such as the ones encoded by *dnm*T, *ery*BVI and *sno*H. These are part of glycopeptide producing gene clusters for the antibiotics daunorubicin (Adriamycin), erythromycin and nogalamycin in the species *Streptomyces peucetius*, *Saccharypolyspora erythraea* and *Streptomyces nogalater* respectively. The catalytic activity of enzymes in this family is dependant on the co-factor flavin mononucleotide (FMN) (Figure 3.1) (Hubbard and Walsh, 2003). The proteins are composed of two copies of a 250 amino acid long unit as can be seen by aligning the two domains against one another (Figure 3.2). This motif has probably arisen by domain duplication as it is often found throughout all kingdoms of life (Apic *et al.*, 2003; Long, 2001). In EvaA, the two domains are 27% identical and 62% similar, when taking conservatively substituted residues into account.

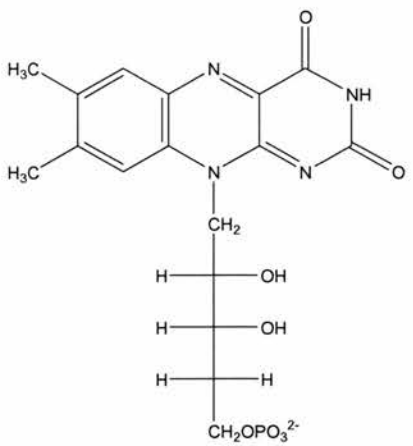
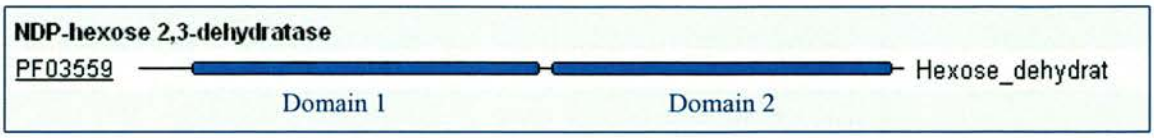


Figure 3.1 - Flavin Mononucleotide (FMN)

For cleavage of bonds between a carbon atom and a poor leaving group such as a hydroxyl group, activation of the substrate by removal of a hydrogen atom from the neighbouring carbon atom is required. The C-H bond can be activated by an acid-base mechanism or by abstraction of a hydrogen atom using the high reactivity of radicals or metal containing co-factors.




```
Domain 1   MSSFVVPSLTAVRPRDHHDYADRIALSAATTDGVQMRTEDVRAWIAERRDANVFHVERIP  60
Domain 2   ------------------------------------QLLSWFTNERSRHDVRVRRIP 281

Domain 1   FADLDQWWFEGVTGNIVFRSGRFFTIEGLHVIEHDGPHGDGPYREWQQPVIRQPEVGILG 120
Domain 2   LADVCGWKQG--AEEIEHEDGRYFKVLAVAVKGSNREK-----ISWTQPLVESVDLGVVA 334

Domain 1   ILAKEFDGVLHFLMQAKMEPGNPNLVQLSPTVQATRSNYTKAHGGTNVKLIEYFAPPDPE 180
Domain 2   FLVRKIDGVPHVLVQARVDGGFLDTVELAPTVQCTPLNYAHLPAEERPPFLDLVQNAPRS 394

Domain 1   RVIVDVLQAEQGSWFFRKSNRNMIVETVDDVPLWDDFCWLTLGQIAELMHEDETINMNSR 240
Domain 2   RIRYEAIHSEEGGRFLGVRARYLVIDADEAIDPPPGYAWVTPAQLTALTRHGHYVNVEAR 454

Domain 1   SVLSCLPYQDITPRALFSDV 260
Domain 2   TLLACINAAAAQPRGGA--- 471
```

Figure 3.2 – Two Domains of EvaA: the top figure shows the two domains of EvaA as found by InterProScan (EBI, Cambridge) (Zdobnov and Apweiler, 2001); the lower figure is an alignment of the two domains of EvaA against one another; the residues on red are conserved and the ones on blue background are conservatively substituted.

The structures of a number of dehydratases have been solved and published and can be divided into several classes according to their catalytic mechanisms.

One class of dehydratases is exemplified by 5-aminolaevulinic acid dehydratases, which have been isolated from *E. coli* and *P. aeruginosa* and properties extensively characterised. They are metalloenzymes, requiring a bivalent metal ion for activity and use a covalent step in their catalysis, forming a Schiff base intermediate with the substrate. The metal ion (most commonly Zn^{2+}) is thought to act either as a Lewis acid to polarise the carbonyl group of the substrate, facilitating deprotonation, or alternatively, could function in the form of a zinc hydroxide and participate as base in the deprotonation. As the metal ion is bound to three cysteine residues, it would be very unusual for it to act as a Lewis acid; for this reason, the metal hydroxide mechanism is more likely. This idea is substantiated by the finding that the zinc metal can be replaced by cadmium in the active site whilst still retaining the enzyme's activity (Schoolingin-Jordan *et al.*, 2002).

A second class of dehydratase enzymes is B_{12} dependant, with a reaction mechanism involving free radicals. Diol and glycerol dehydratases belong to this enzyme family and catalyse the dehydration of 1,2-diols and glycerol respectively. Adenosylcobalamin (co-enzyme B_{12}) serves as the cofactor for the enzymatic radical reaction: it is triggered when an adenosyl radical is formed in the active site by

homolytic cleavage of the coenzyme Co-C bond. This radical helps to activate the C-H bond for proton abstraction which subsequently allows hydroxyl elimination from the adjacent carbon of the substrate (Toraya, 2000).

A further group of dehydratases is the RmlB-like enzyme class, whose members belong to the extended family of short chain dehydrogenase/reductase enzymes (SDR). RmlB is the second enzyme in the L-rhamnose biosynthetic pathway and catalyses the dehydration of dTDP-D-glucose to dTDP-4-keto-6-deoxy-D-glucose and requires the co-factor nicotinamide adenine dinucleotide (NAD) for activity. It has long been established that the redox potential, driving the reaction, is due to the formation or destruction of the aromaticity of nicotinamide by the gain or loss of hydride from the C4′ of the pyridine ring. NAD^+ is thought to abstract a hydride from C4′ of the substrate, and then a tyrosine acts as the catalytic base, deprotonating the C4′ leading to the formation of a keto-sugar intermediate. The nicotinamide ring lowers the pKa of the active site Tyr; a Thr may act as a proton shuttle as the Tyr is too far removed from the C4′ to directly act as the catalytic base. Next, a Glu removes a proton from C5′ of the sugar ring which has been acidified by the adjacent keto group. This leads to the formation of an enolate intermediate; a hydride is transferred onto C6′ resulting in the inversion of the configuration of the glucose and protonation of C5′ by the active site Glu (Allard *et al.*, 2002; Cleland and Kreevoy, 1994; Hegeman *et al.*, 2001; Simon Allard, University of St. Andrews, PhD Thesis 2001)

EvaA and its homologues from other Streptomyces species have no sequence similarity to any of the dehydratases described above. They do not contain the motif required for SDRs (YXXXK); a necessity for divalent metal ions is unknown but unlikely as they do not share any sequence similarity with metalloenzymes and an involvement of a free radical mechanism is unlikely as they have no known requirement for vitamin B_{12}. For these reasons, EvaA and its homologues probably form a novel class of dehydratase enzymes. In addition to this, the 2,3-dehydration reaction is not precedented in the literature. Thus, EvaA is anticipated to exhibit a novel catalytic mechanism which is probably used throughout Streptomyces species to synthesise deoxysugars that are part of glycopeptide antibiotics.

An RmlB-type dehydratase is necessary to produce the substrate for EvaA from dTDP-glucose in the L-epivancosamine biosynthesis pathway. The structure of such an SDR dehydratase - the dTDP-glucose 4,6-dehydratase from *Streptomyces venezuelae* -

has been published (Allard *et al.*, 2004). This enzyme is involved in the formation of desosamine (3-dimethylamino-3,4,6-trideoxyhexose) which is part of various macrolide antibiotics. However, the structure of the next enzyme in the pathway (*Ery*BVI), an EvaA-like dehydratase, is unknown; it displays a sequence identity of 44% and a sequence similarity of 58% to EvaA.

The structure of EvaA would certainly be that of a novel dehydratase enzyme to which a large number of enzymes in the Actinomycetales' antibiotics biosynthesis pathways will belong (Aguirrezabalaga *et al.*, 2000; Baltz and Seno, 1988). No fold motif could be identified for the amino acid sequence of EvaA but it is likely to contain a Rossman fold which is necessary for recognition and binding of nucleotides such as FMN (Ahn et al., 2004; Brakoulias and Jackson, 2004).

The determination of the EvaA structure will allow an insight into the 2,3-dehydration mechanism. Figure 3.3 shows an overview of the reaction scheme that has been proposed by Chen et al., 2000 for the 2,3-dehydration reaction.

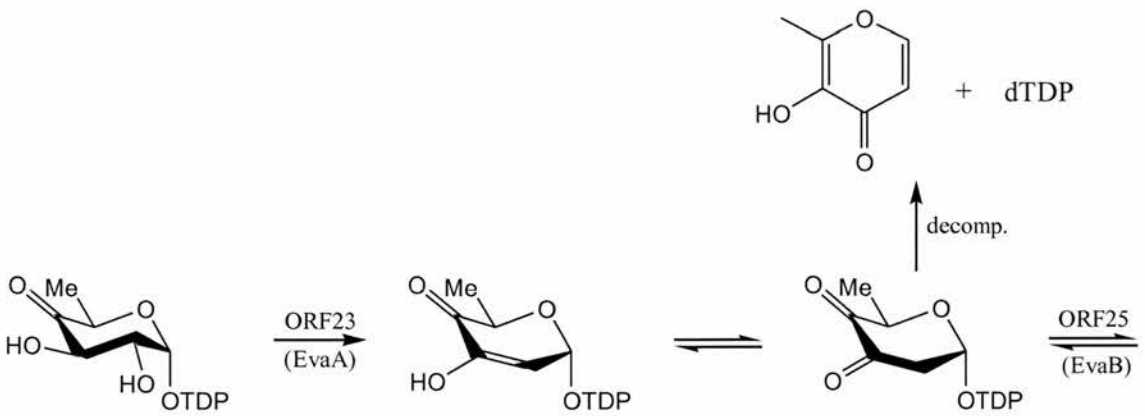


Figure 3.3 – EvaA Reaction Scheme: reaction scheme for the dehydration reaction catalysed by EvaA, the 2,3-dehydratase, as proposed by Chen *et al.*, 2000.

3.3 Materials and Methods

3.3.1 Homologue Choice

Protein sequence homologues were identified using the web server WU-Blast located at the EMBL, Heidelberg (Yuan *et al.*, 1998). Six homologues from different species were chosen to work with: these range from close homologues, e.g. orf8 from *Amycolatopsis mediterranei*, to more distant ones, e.g. *Ery*BVI from *Saccharypolyspora erythraea*. Table 3.1 states the species chosen, the corresponding accession numbers (Swiss-Prot Protein knowledgebase) (Bairoch *et al.*, 2004), the gene's respective position in the species' genome, the restriction sites that were introduced for subsequent cloning and their specific sequence homology to EvaA. Appendix A gives an overview of the primers designed for each homologue.

Table 3.1 - Overview of the EvaA Homologues Chosen

Species Name	Accession no.	Gene	Restriction sites	Similarity[a] (%)	NCIMB species code
Amycolatopsis mediterranei	Q939X7	Orf 8 49,364-50,776	*Nco* I/*Eco*R I	89/92	9613
Streptomyces antibioticus	Q9L6C6	orf 10 1,173 – 2,597	*Bsp*H I/*Hin*d III	47/63	11506
Streptomyces avermitilis	Q950P1	*aveBVI* 5,988 – 7,397	*Bsp*H I/*Bam*H I	62/72	12804
Saccharypolyspora erythrea	O50540	***Ery*BVI** 3,308-4,870	*Bsp*H I/*Eco*R I	44/58	8594
Streptomyces fradiae	Q96454	***Urd*S** 3,196 – 4,505	*Nco* I/*Bam*H I	68/76	8233
Streptomyces nogalater	O54255	*snog*H 152 – 1,618	*Bsp*H I/*Eco*R I	50/62	9489
Streptomyces violaceoruber	Q9ZA32	orf27 (*gra*) 23,869 – 25,200	*Bsp*H I/*Eco*R I	48/63	9622

[a] Similarity is given as percentage sequence identity and similarity to EvaA
All details have been reproduced from http://us.expasy.org/ and links therein.

Figure 3.4 gives a phylogram showing the correlations of these genes in a more graphical way as calculated by ClustalW (EBI, Cambridge) (Thompson *et al.*, 1994).

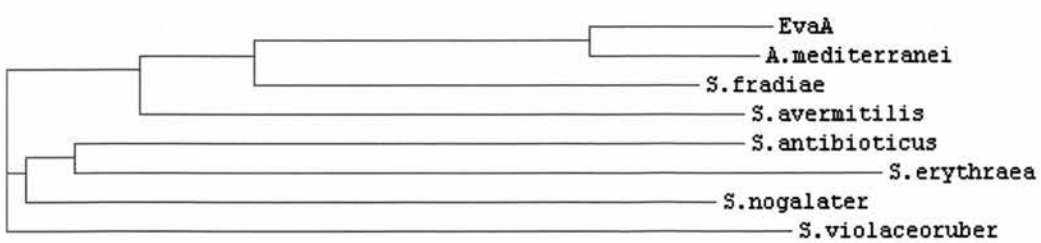


Figure 3.4 - Phylogram Tree: showing the relationship between EvaA and its homologues

3.3.2 Growing Streptomyces

No genomic or cosmid DNA for any of the chosen species was commercially available. For this reason, all species were obtained from the National Collection of Industrial Food and Marine Bacteria (NCIMB) in the form of freeze dried bacteria from which the DNA had to be isolated. Table 3.1 details the NCIMB Culture Collection species code for the different homologues. Insights into the optimum growth conditions were gained from a number of sources including web pages, books (Chater and Merrick, 1979; Kieser *et al.*, 2000), research articles and suppliers, such as the ATCC (American Type Culture Collection, http://www.atcc.org or http://www.lgcprochem.com/atcc), NCIMB (http://www.ncimb.co.uk), DSMZ (Deutsche Sammlung von Mikroorgansismen und Zellkulturen, http://www.dmsz.de/dsmzhome.htm). An amalgamation of all growth conditions found was used: the species required a temperature of 20 to 25°C and a nutrient-rich growth medium. Special media, both solid and liquid (Table 3.2) were made up and sterilised by autoclaving at 121°C (15psi) for 20min. All work with the Actinomycetales was carried out under sterile conditions to prevent cross-contamination with other bacteria and yeast grown in the laboratory as no antibiotic selectivity can be used when growing the Actinomycetales species.

Table 3.2 - Growth Medium Composition Used for the Different Actinomycetales Species

Species	Composition of solid medium (per litre)	Composition of liquid medium (per litre)
A. mediterranei, *S. erythrea*, *S. antibioticus*, *S. fradiae*, *S. nogalater*, *S. violaceoruber*	4.0g glucose 4.0g yeast extract 10.0g malt extract 2.0g $CaCO_3$ 12.0g agar dH_2O pH to 7.0	4.0g glucose 4.0g yeast extract 10.0g malt extract dH_2O pH to 7.0
S. avermitilis	10.0g ground oat flakes 10.0g fine oat meal 15.0g agar dH_2O pH to 7.0	10.0g ground oat flakes 10.0g fine oat meal dH_2O pH to 7.0

The freeze dried samples were re-suspended in a small volume of liquid media (typically 500µl) and plated out on solid agar and incubated at 25°C for several days until growth was visible. Single colonies were picked and used for successive rounds of

plating out to gain a homogenous culture. To reduce the accumulation of revertants, it is undesirable to propagate Streptomyces by successive rounds of mass culturing, as variant forms may grow or sporulate better than the original phenotype. Instead they should be plated or streaked out and a single colony taken to start the next culture (Kieser *et al.*, 2000). To provide solid support for these soil organisms, liquid cultures were maintained in tissue culture bottles, promoting growth of the bacteria at the bottom in low volumes of liquid medium (approximately 10ml). Incubation was carried out at 25°C with maximum aeration (200rpm) for several days. The identity of the species was judged by the distinct 'forest-like' smell (caused by geosmins - secondary metabolites); additionally *S. violaceoruber* displayed a distinctive red colour that allowed identification. As DNA isolation proved impossible for bacteria cultures grown in the oatmeal medium, (due to the clogging of the Genomic tips – see later), *S. avermitilis* was picked from the solid oatmeal agar and grown in the liquid medium specified for all other Actiomycetales.

3.3.3 Isolation of Genomic DNA

Once cell density (as judged by eye) was sufficient for DNA isolation, the cells were harvested by centrifugation (7,200*g*, 20min) and cells that had grown on the toothpick from inoculation were scraped off and added to the cell pellet for lysis. Isolation of the genomic DNA was carried out via Genomic tips 100/G® (Qiagen) broadly following the instructions delivered with this product. Cells were lysed using lysis buffer (Buffer P1 out of the Qiagen MaxiPrep® kit), 100µg ml^{-1} lysozyme and 50µg ml^{-1} proteinase K; RNase was added to prevent RNA from binding to the Genomic-tips 100/G®. The DNA was washed and eluted off the Genomic tip 100/G® using the buffers supplied with the Qiagen kit. It was then precipitated in isopropanol and pelleted by centrifugation (35,000*g*, 20min). The DNA pellet was washed with cold 70% ethanol and air-dried before being re-dissolved in 500µl dH_2O overnight at room temperature with continuous agitation. The DNA concentrations were calculated by determining the samples' absorbances at a wavelength λ = 260nm. Quality of the DNA was checked on a 0.5%w/v agarose gel containing 0.005%v/v ethidium bromide to allow visualisation of DNA under UV light; purity was determined by calculating the ratio of absorbance at λ = 260nm and λ = 280nm.

3.3.4 Gene Amplification

The homologue genes were amplified from their respective genomes using PCR. Appendix A details the primers used for each homologue. To facilitate primer binding, the genomic DNA was partially unwound before adding it to the PCR mix by heating it to 70°C for 5min. For initial tests, 20µl PCR mixtures were prepared containing 10 to 20ng genomic DNA, 1-3µM forward and reverse primers, 0.5mM dNTPs, 0.5U DNA polymerase, 2µl 10x PCR buffer and dH_2O. In addition to varying DNA and primer concentrations, *Pfu*, *Vent* (both Promega) and BioXact (Bioline) DNA polymerases were used. Also, a number of additives, such as DMSO (1 to 10%w/v), $MgCl_2$ (2 to 10mM), 1x HiSpec Additive and 1x PolyMate Additive (Bioline), were tested to increase yield and purity of the PCR products. The reaction was carried out in the thermocycler (Applied Biosystems GeneAmp PCR System2400) under varying conditions but always using 25 cycles of melting II, annealing and extension I steps:

Program: Melting I: 94 – 95°C → 5 – 10min
Melting II: 94 – 95°C → 30sec – 1.5min
Annealing: 5 – 15°C below primer melting temperature → 30sec – 1min
Extension I: 66 – 74°C → 1.5 – 3min
Extension II: same temperature as extension I → 7min
Hold: 4°C

The success of the PCR was judged under UV light after separation of the sample on a 1% agarose gel containing 0.005%v/v ethidium bromide, the size of the products were judged against the standard 1kb ladder (Promega) run on the same gel. The reaction mixtures that gave positive PCR results were identified and scaled up to a 100µl volume keeping all ratios constant.

3.3.5 Subcloning of EvaA from pET22b+

Custom made primers (Appendix A) were obtained from Oswel (now Eurogentec) to subclone EvaA from the pET22b+ vector (obtained from our collaborators Chris Walsh, Harvard Medical School) whilst introducing appropriate restriction sites. *Nco* I and *Eco*R I sites were introduced 5′and 3′ to the gene

respectively for cloning into the pEHISTEV vector (see Appendix B). *Nco* I and *Bam*H I sites were introduced 5′and 3′to the gene for cloning into the pMAL-c2X, pMAL-p2X and pLou-3 vectors. Both amplifications were carried out in the same way.

Standard PCRs were performed in 100µl reaction volumes containing forward and reverse primers (1.5µM), 8ng vector template, 0.5mM dNTPs, 2U Vent polymerase, 10µl 10x PCR buffer and dH_2O. Reactions were carried under the following conditions:

Program:

95°C – 7min	
95°C – 1min	25 cycles
46°C – 1min	
72°C – 2min	
72°C – 7min	
4°C - hold	

The success of the PCR was judged as before. Amplified bands of the appropriate size (1,500bp), as judged against the standard 1kb ladder (Promega), were cut out and the DNA gel-extracted using the Qiagen gel extraction kit® (spin column method using a microcentrifuge). DNA was eluted from the spin column in a final volume of 30 to 50µl dH_2O.

3.3.6 Restriction Digests and Ligation Reactions

Following PCR, isolated DNA was digested with the appropriate restriction enzymes, obtained from New England Biolabs (NEB) and/or Promega. Double digests were carried out by incubating 300ng DNA with 2U of each restriction enzyme, appropriate buffer and dH_2O for 1 to 2 hours at 37°C.

Sequential digests were carried out by incubating 300ng DNA with 2U of one restriction enzyme, 3µl 10x buffer and dH_2O in a total reaction volume of 30µl at 37°C for 1 hour. The primary restriction enzyme was then inactivated by heating the reaction mix to 65°C for 20min. Subsequently, 2U of the second enzyme were added together with the appropriate buffer and dH_2O, increasing the total reaction volume to 50µl. This mixture was incubated for 1 hour at 37°C.

Single digests were carried out by incubating 300ng DNA with 2U restriction enzyme, 5µl 10x buffer and dH_2O in a total reaction volume of 50µl for 1 hour at 37°C. The DNA was run on a 1% agarose gel, containing 0.005%v/v ethidium bromide for visualisation under UV light. The DNA band was extracted from the agarose gel using the Qiagen gel extraction kit® (spin column method using a microcentrifuge). The DNA was eluted off the spin column into 50µl dH_2O, all of which was then used for the second restriction digest. This was carried out in a total volume of 100µl using the gel extracted DNA, 2U restriction enzyme, 10µl 10x buffer and dH_2O. Reactions were incubated for 1 hour at 37°C and subsequently gel purified.

In later digests, DNA was purified between reactions using only the spin columns of the Qiagen gel extraction kit®. This was particularly important for vector digests which were not gel purified.

To prevent star activity, the digest reaction volume was increased to reduce the amount of glycerol present in the sample. When *Bam*H I was used, 0.25 to 0.5mM Bovine Serum Albumin (BSA) was added to enhance the activity of the restriction enzyme.

Ligations were carried out by adding restriction digested vector and DNA insert in ratios 1:1, 1:3, 1:5 and 1:10 to 20µl reaction mixtures containing 2U T4 DNA ligase (Promega), 2µl 10x buffer and dH_2O. The reaction mix was incubated for 8 to 12 hours at 16°C or room temperature.

3.3.7 Transformations

Chemically competent *E. coli* cells were bought from ActiveMotif, Novagen or Promega (depending on *E. coli* strain used). For each transformation, 50-100µl cells were incubated with 200ng vector DNA or 5 to 15µl ligation mix for 1 hour on ice. Cells were heat-shocked at 42°C for 30sec and incubated on ice for another 5min. Subsequently, 250µl LB broth was added, transformation reactions incubated at 37°C for 1 hour with 200rpm shaking and then plated out on LB agar plates containing an appropriate antibiotic. The plates were incubated for 12 hours at 37°C.

For DNA extraction, single colonies were picked and grown overnight a 37°C, with 200rpm shaking, in 10ml LB broth containing 50μg ml^{-1} antibiotic. The cells were harvested by centrifugation (20,000*g* for 15min) and DNA extracted using the Qiagen miniprep kit®.

3.3.8 DNA Sequencing

All DNA sequencing was performed by The Sequencing Service (School of Life Sciences, University of Dundee, Scotland, www.dnaseq.ac.uk) using Applied Biosystems Big-Dye Ver. 3.1 chemistry on an Applied Biosystems model 3730 automated capillary DNA sequencer. We provided the M13 and MalE primers for sequencing of the pMAL-c2X, pMAL-p2X and pLou-3 vectors together with our samples. The other vectors were sequenced with the T7 forward and reverse primers which were provided by The Sequencing Service.

3.3.9 Cloning and Expression Vectors

The following vectors were used for cloning and expression of orf23 (encoding EvaA) in *E. coli* cells:

- **pET22b+** (Novagen): it carries an N-terminal *pelB* signal sequence for potential periplasmic translocation plus an optional C-terminal His·Tag® sequence. It also encodes the *bla* coding sequence providing the host cells with ampicillin resistance. As with all pET expression systems, target gene expression is induced by IPTG via the T7 promoter. This promoter requires the presence of T7 RNA polymerase (bacteriophage origin) whose expression is induced by IPTG (Figure 3.5).
- **pET32a+** (Novagen): the vector is designed for high-level expression of peptide sequences fused with the 109 amino acid Trx·Tag™ thioredoxin protein (LaVallie *et al.*, 1993) which can be cleaved off using thrombin. The vector also carries the cleavable His·Tag® and S·Tag® sequences for protein detection and purification. The T7 promoter is present for gene expression and the *bla* coding sequence for ampicillin resistance.

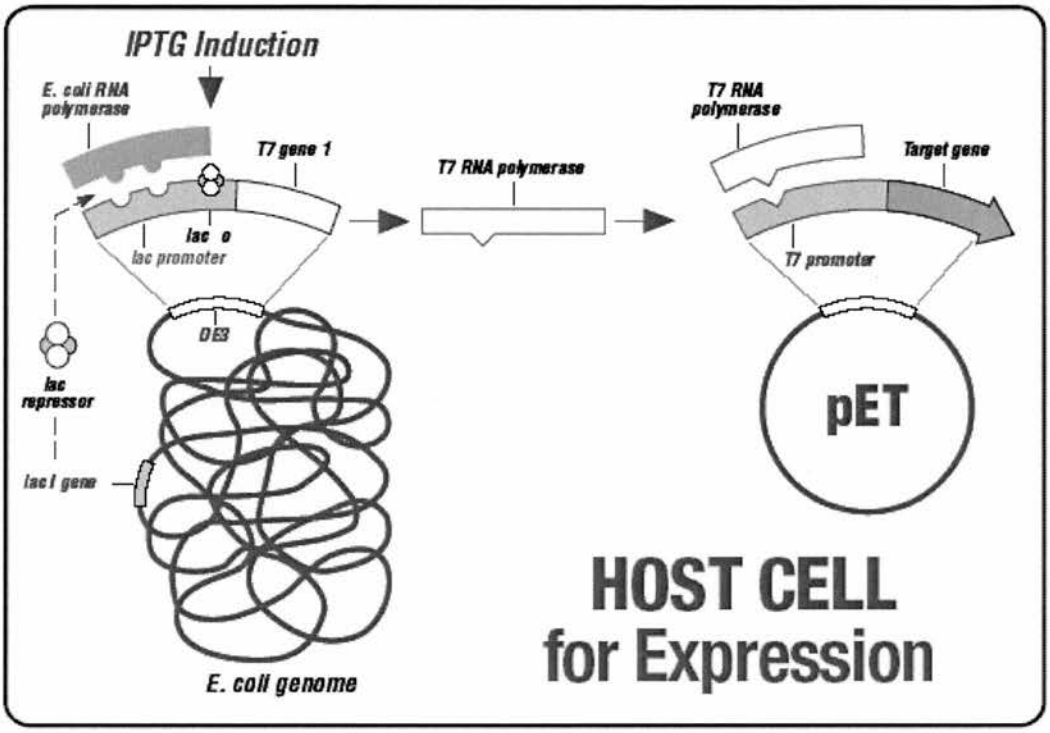


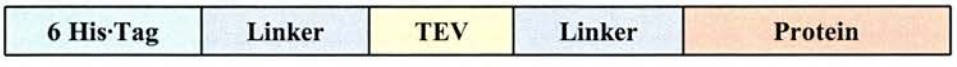
Figure 3.5 – Protein Expression from pET Vectors: protein expression from a pET vector requires the use of expression optimised *E.coli* host cells containing the DE3 construct of the T7 RNA polymerase gene, under control of the *lac* promoter. The T7 RNA polymerase is required for expression from the T7 promoter. Overview taken from Novagen Catalogue (2002-2003), p.88.

- **pETBlue-1** (Novagen): this vector allows the convenient subcloning of target genes as well as the T7*lac* promoter based expression of unfused genes containing the ATG start codon. The cloning region is within the alpha-peptide coding region of beta-galactosidase. Insertional inactivation of the alpha-peptide allows recombinant clones to be directly identified by blue/white screening (Maas, 1999). The vector also carries the *bla* coding sequence for ampicillin resistance.
- **pGEM®-T Easy** (Promega): this vector system contains single 3′-T (thymidine) overhangs on both ends of the linearised plasmid. This facilitates the ligation of PCR products created by polymerases which add a single deoxyadensosine to the product. The high copy number vector contains T7 and SP6 RNA polymerase promoters flanking the MCS. This in turn, is within the alpha-peptide coding region of the enzyme beta-galactosidase allowing blue/white screening of clones (Maas, 1999). The vector carried the *bla* coding sequence for ampicillin resistance.
- **pMAL-c2X** and **pMAL-p2X** (NEB): these vectors are designed to produce maltose binding protein (MBP) fusion proteins, where the protein of interest can be cleaved from MBP with the specific protease Factor Xa. The gene of interest was cloned into the vector C-terminal to and in the same translational reading frame as the *malE*

gene (encoding MBP). The polylinker region between the *malE* gene and the vector MCS encodes the four amino acid recognition site for Factor Xa (IDGR). The vector also encodes the *bla* sequence for ampicillin resistance. Insertion of the target gene into the MCS disrupts the pre-existing *malE-lacZ* fusion facilitating screening of recombinant clones. The only difference between pMAL-c2X and pMAL-p2X vectors is the deletion of the *malE* signal sequence in the former. As a result, the fusion protein remains in the cytoplasm rather than being exported as is the case with the latter vector.

- **pEHISTEV** (H. Liu, St. Andrews): this vector is based on the pET30 vector (Novagen) whose ribosome binding site, MCS, His·Tag®, linker region, TEV cleavage site and MCS have been modified. It encodes a His·Tag® with a long polylinker region which can be cleaved by the TEV protease (Nuclear Inclusion, Tobacco etch virus). The His·Tag® has been shown to improve solubility of some proteins and the cleavage with TEV is efficient and cheap; therefore, convenient for large scale protein production. The vector encodes the *kan* gene sequence for kanamycin resistance. Appendix B details the vector map and Figure 3.6a shows a simplified overview of the protein expressed from pEHISTEV.
- **pLou-3** (L. Major, St. Andrews): this vector is based on the pMAL-c2X vector (NEB) but encodes a similar polylinker region to the one in the pEHISTEV vector allowing the fusion protein to be cleaved with TEV. Additionally, the vector encodes a His·Tag® C-terminal to MBP facilitating purification of the fusion protein. Appendix B details the vector map and Figure 3.6b shows a simplified overview of the protein expressed from pLou-3.

a) pEHISTEV



b) pLou-3

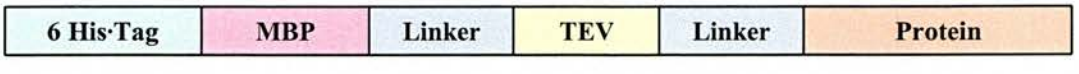


Figure 3.6 – Overview of Expressed Protein:
a) expressed from pEHISTEV vector, b) expressed from pLou-3 vector.

3.3.10 Protein Expression Trials

For the initial protein over-expression trials, *E. coli* host cells were transformed with pET22b+ or pEHISTEV and selected on antibiotic containing LB agar plates. Single colonies were picked and grown up in 10ml LB broth at 37°C and 200rpm shaking. After 12 hours, 100µl aliquots were used to inoculate fresh growth medium (10ml samples) for the over-expression trials. Expression was induced using IPTG once the cell density had reached an OD_{600} of 0.5 to 0.8. The following variables were altered to optimise expression of soluble EvaA from the two vectors.

- **Temperatures:**

  18°C; 25°C; 37°C

- **Growth media:**

  Luria-Bertani Medium, Terrific Broth, 2x Tryptone Yeast Extract (2x TYE), Hyper Broth™, Power Broth™, Superior Broth™, Turbo Broth™, M9 Minimal Medium, Glucose M9CA, Miller Medium, Tryptone Phosphate Medium

- **Growth medium additives:**

  Phosphate (50-250mM), Glucose (1-5%w/v), Sucrose (1-5%w/v), Ethanol (1%v/v), Sarcosyl (0.5-1.5%w/v)

- ***E. coli* Strains:**

  BL21(*DE3*), BL21trx(*DE3*), BL21-SI; BL21 Star™, BL21(*DE3*)pLysS, BL21(*DE3*)pLysE, BL21(*DE3*)-Gold, BL21(*DE3*)-Gold-pLysS, BL21(*DE3*)RP-codon+, Rosetta(*DE3*)

- **IPTG concentrations:**

  None (leaky expression), 0.1mM, 0.5mM, 1mM, 5mM, 10mM

- **Expression times:**

  20min, 1hr, 2hrs, 3hrs, 4hrs, 5hrs, 18hrs

- **Heat shock and cold shock protocols:**

  No heat or cold shock protocols used,

  42°C for 20min, then immediate induction at expression temperature,

  42° for 20min, then cool down to 25°C before induction,

  42°C for 20min, then 4° for 15min, then induction,

  4°C for 20min, then induction

- **Lysis buffer additives:**

  Polyoxyethylene sorbitainmonolaureate (TWEEN 20) (0.1-1%w/v), t-Octylphenoxypolyethoxyethanol (Triton X-100) (0.1-1.5%w/v), n-Octylpolyoxyethylene (OPOE) (0.05-0.5%w/v), n-Octyl-β-D-glucopyranoside (BOG) (0.05-0.5%w/v)

Expression trials with each of the two vectors, (pET22b+ and pEHISTEV) were carried out altering the variables given in the list above against one another.

In addition to the standard overexpression trials the OvernightExpress™ Autoinduction System (Novagen) was used to express EvaA. The system is optimised for very tight expression control and induction at high cell density.

From all expression trials, 1ml samples were taken at regular intervals for analysis. The cells were harvested by centrifugation (31,200*g*, 10min) and the cell pellet was re-suspended in buffer (100mM NaCl, 20mM Tris-HCl, pH7.5). The cells were lysed by sonication at 18 microns for 20sec and the insoluble cell debris separated from the soluble fraction by centrifugation (31,200*g* for 10min). The insoluble matter was solubilised in 1ml 1%SDS and together with the soluble fractions and non-induced sample analysed by SDS-PAGE. Standard molecular weight markers (BioRad) were run on the same gel for weight comparisons; all gels were stained with Coomassie™ Blue stain and then destained with dH_2O containing 40% ethanol and 10% acetic acid. The expected molecular weight of EvaA was ~55kDa and bands of the appropriate size were cut out. The identity of the excised protein was judged by trypsin digest MALDI-TOF mass spectrometry against a calculated trypsin digest signature.

Similar expression trials to the ones described above were carried out for the fusion proteins: EvaA in pET32a+, fused to thioredoxin, and EvaA fused to MBP expressed from the pMAL-c2X, pMAL-p2X and pLou-3 vectors. Again, starter cultures were grown in LB broth and then used to inoculate fresh growth medium containing 50μg ml^{-1} ampicillin or carbenicillin. After induction of protein expression with IPTG, samples were taken every hour to judge optimum protein expression. Analysis was carried out by SDS-PAGE and trypsin digest MALDI-TOF mass spectrometry of the resulting bands on the protein gels.

3.3.11 Protein Purification

All proteins expressed with a His·Tag® were purified chromatographically using a metal chelating column. After expression, cells were harvested by centrifugation (16,800*g*, 20min) and the cell pellet suspended in lysis buffer solution containing 100mM NaCl, 50mM Tris-HCl, pH7.5, 100μg ml^{-1} lysozyme, 20μg ml^{-1} DNase, 5mM PMSF, 3mM DTT and one Complete EDTA-free protease inhibitor cocktail tablet (Roche). The suspension was incubated on ice for 1 hour and the cells lysed by sonication using five 30sec cycles at 18microns and 30sec cooling on ice between cycles. The soluble and insoluble cell matter were separated by centrifugation (48,000*g*, 20min) and the cell debris discarded. The soluble matter was dialysed against two changes of buffer (50mM Tris-HCl, pH7.5) and then applied to a pre-equilibrated (50mM Tris-HCl, pH7.5) anion exchange column (diethylaminoethyl (DEAE) Sepharose or Poros HQ20 (both Amersham Biosciences)). The column was run on a High Performance Liquid Chromatography (HPLC) BioCAD 700E workstation (Global Medical Instrumentation – GMI). Proteins were eluted from the column using a linear 0 to 2.5M NaCl gradient and fractions collected. Throughout the run, the absorbance at λ = 280nm was monitored to detect protein eluting from the column; protein containing fractions were analysed by SDS-PAGE. After this initial clean-up step, the protein was applied to a metal chelating column (Sepharose metal chelating resin (Amersham Biosciences) or Ni-NTA resin (Invitrogen)) that had been charged with $NiSO_4$ and equilibrated with 100mM NaCl and 50mM Tris-HCl, pH7.5. Protein was eluted off the column using a linear 0 to 1.0M imidazole gradient. Load, unbound, wash and fractions were monitored at λ = 280nm and protein containing fractions were analysed by SDS-PAGE.

Protein fused to MBP was purified to single band purity using amylose resin (NEB). Cells expressing the MBP-EvaA construct were harvested by centrifugation, lysed by sonication at 18microns in amylose column buffer (ACB) (200mM NaCl, 1mM EDTA, 50mM Tris-HCl, pH7.5), containing 100μg ml^{-1} lysozyme, 20μg ml^{-1} DNase, 3M DTT and one Complete EDTA-free protease inhibitor cocktail tablet (Roche). The soluble and insoluble cell debris were separated by centrifugation (48,000*g*, 20min) and the supernatant applied to a column containing pre-equilibrated

amylose resin. The column was washed with 15CV ACB, and then the protein was eluted using 1CV to 2CV ACB containing 10mM maltose (Figure 3.7).

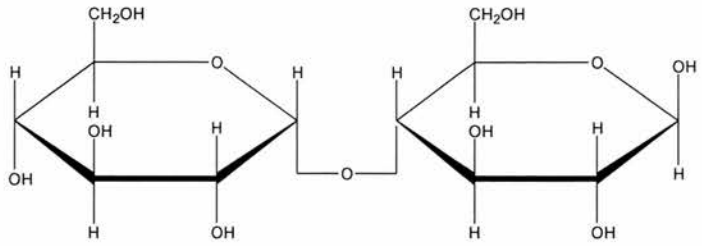


Figure 3.7 - Maltose: [*O*-α-$_D$-glucopyranosyl-(1→4)-$_D$-glucopyranose] (reproduced from Stryer, 1995, p.471)

The MBP-EvaA fusing construct was cleaved using either Factor Xa or TEV protease depending on the vector from which the protein had been expressed. The cleavage with Factor Xa was carried out in accordance with the protocol supplied by NEB, i.e. at room temperature in ACB using 5µg Factor Xa per milligram fusion protein.

Initially, cleavage of the MBP-EvaA construct with TEV protease was done by adding 2µg protease per milligram fusion protein and dialysing the sample against PBS buffer containing 300mM NaCl, 0.5mM EDTA, 1mM DTT and 50mM Tris-HCl, pH7.5 at 4°C. Reiterative optimisation of the TEV protease cleavage showed, that the enzyme could be added directly to the fusion protein in ACB containing 1mM EDTA and 5mM DTT. The reaction mix was then stirred at 4°C for 12 to 16 hours. Initially, 0.05%w/v OPOE was added to the protein mixture to keep EvaA stable in solution after cleavage. Following several attempts, the addition of detergent was deemed unnecessary and omitted from later cleavage reactions.

Progress of the reaction was analysed by SDS-PAGE; after cleavage of MBP from EvaA by TEV protease, it was necessary to devise a protocol enabling separation of EvaA from MBP and TEV.

The protocol for purifying MBP fusion constructs (NEB, 2004) suggested extensive protein dialysis and re-applying the cleaved protein mixture to a second pre-equilibrated amylose column to which MBP but not the fusion partner should stick. As EvaA could not be purified in this way (see section 3.4.6), other chromatography columns were tested on a BioCad system in order to separate EvaA from MBP and TEV protease. A large number of chromatography columns, buffer conditions and pHs were screened in an attempt to purify EvaA. Protein was eluted from all columns, with the

exception of the gel filtration columns, using linear salt gradients. Table 3.3 lists the columns used and their respective pH and buffer conditions tested for purification.

Table 3.3 - Purification Attempts for EvaA: The following columns were used under the following conditions in order to separate EvaA from MBP

Column	Buffers used	pH[§]	Elution method
Ni-NTA	100mM Tris-HCl 100mM $NaPO_4$ (+0.2-1.5M NaCl)	6.5 – 8.5	Imidazole gradient (0.1-2M) + 4M NaCl
Ni-Fast Flow Sepharose™	100mM Tris-HCl 100mM $NaPO_4$ (+0.2-1.5M NaCl)	6.5 – 8.5	Imidazole gradient (0.1-2M) + 4M NaCl
Anion exchange (hiprep)	100mM Tris-HCl 100mM $NaPO_4$	6.5 – 9.5	NaCl gradient (0-5M)
Anion exchange (Poros HQ20)	100mM Tris-HCl 100mM $NaPO_4$	7.0 – 9.5	NaCl gradient (0-5M)
Anion exchange (DEAE Sepharose)	100mM Tris-HCl 100mM $NaPO_4$	7.0 – 9.5	NaCl gradient (0-5M)
Anion exchange (ANX ff Sepharose)	100mM Tris-HCl	7.5	NaCl gradient (0.2-5M)
Cation exchange	100mM Tris-HCl 100mM $NaPO_4$	4.5 – 7.5	NaCl gradient (0-5M)
Hydrophobic	100mM Tris-HCl 100mM $NaPO_4$	7.0 - 8.0	Reverse $AmSO_4$ gradient – 2M-0M
Gel filtration (S100) sephadex	100mM Tris-HCl 100mM $NaPO_4$ (+0.2-1.5M NaCl)	7.5 – 8.5	Size exclusion
Gel filtration (S200) sephacryl	100mM Tris-HCl 100mM $NaPO_4$ (+0.2-1.5M NaCl)	7.5 – 8.5	Size exclusion

[§] The pH range of the column buffer was tested in 0.5 pH value increments

3.3.12 Selenomethionine Incorporation

Selenomethionine (SeMet) was incorporated into the MBP-EvaA fusion protein by metabolic inhibition. Cells were grown in minimal medium (M9) according to the protocol by van Duyne *at al.*, 1993. A 5x stock of M9 medium was made up and autoclaved at 121°C (15psi) for 20min (one litre of 5x M9 stock contained 30g Na_2HPO_4, 15g KH_2PO_4, 5g NH_4Cl, 2.5g NaCl). 1 litre of culture medium contained 200ml 5x M9 stock, 800ml dH_2O to which 50µg ml^{-1} ampicillin, 1ml 1M $MgSO_4$ (autoclaved separately), 10ml 40% glucose, 100µl 0.5%w/v thiamine vitamin and 1ml 0.4g ml^{-1} $FeIISO_4$ solution (all solutions were sterile filtered) were added.

The overnight starter culture of BL21(DE3) cells containing pLou3-EvaA in LB medium was harvested by centrifugation (7,200*g*, 10min), washed twice with M9 medium and used to inoculate the M9 expression medium. The cells were grown at 37°C with good aeration (200rpm) until an OD_{600} of 0.6 was reached. Amino acids and SeMet were then added (Thr, Lys, Phe each 100mg l^{-1}; Leu, Ile, Val and SeMet each 50mg l^{-1}) and the culture incubated for a further 20min at 37°C at 200rpm. Subsequently, protein expression was induced using 1mM IPTG and the temperature reduced to 25°C; expression was terminated after 12 hours and the cells harvested by centrifugation (16,800*g*, 20min).

The purification of SeMet protein was identical to that of wild-type EvaA. However, 5mM DTT was added to all buffers since the SeMet had to remain reduced. SeMet incorporation was checked by MALDI-TOF and ESI-TOF mass spectrometry on the fusion protein as well as purified EvaA.

3.3.13 Circular Dichroism Spectroscopy

The secondary and tertiary structure elements of purified EvaA were analysed by CD spectroscopy. The scans were carried out in the Protein Characterisation Facility, Institute of Biomedical & Life Sciences, University of Glasgow. The protein concentrations for this procedure were 1mg ml^{-1} and the protein was kept in 20mM Tris-HCl, pH7.5. For the fusion protein, the known MBP spectrum was subtracted from the overall CD spectrum to obtain the spectrum for EvaA.

3.3.14 Crystallisation

As the purification of the cleavage product proved initially difficult, the crystallisation of the fusion protein was attempted. As there is no structure known which is similar enough in sequence to suggest any relation to EvaA, phases would either have to be calculated from MBP in the case of the fusion protein crystals, or by experimental phasing if EvaA on its own was crystallised. The former of these two methods would be the 'easier' one as it allows the phasing from normal protein, not

relying on the expression, purification and crystallisation of protein into which sufficient SeMet had been incorporated.

All initial crystallisation screens were performed using the sitting drop vapour diffusion method, 2+2µl crystallisation drops, a protein concentration appropriate for crystallisation as determined by the Pre-Crystallisation Test (Hampton Research) (approximately 5mg ml^{-1}) and incubation temperatures of 4°C and 20°C. The following sparse matrix crystallisation screens were carried out:

- Hampton Research – Crystal Screens 1 and 2
- Hampton Research – Index Screens 1 and 2
- Hampton Research – PEG/Ion Screen
- Hampton Research – Salt/RX Screen
- Emerald BioSciences – Wizard 1 and 2
- Sigma-Aldrich – Crystallisation Screens 1 and 2

Initial crystallisation trials for the EvaA-MBP fusion construct were carried out at protein concentrations of 5 mg ml^{-1} and 10mg ml^{-1}.

Concentration of EvaA separated from MBP proved difficult and sparse matrix crystal trials were set up using EvaA at a concentration of 3mg ml^{-1}. However, adding between 0.5 and 1mM FMN to EvaA allowed the concentration of the protein to 10mg ml^{-1}. Crystallisation trials with EvaA at 5mg ml^{-1} and 10mg ml^{-1}, each containing 0.5mM FMN were performed.

3.3.15 Data Collection on EvaA-MBP

Data were collected at the ESRF, ID14.1. A single crystal of 0.5 x 0.5 x 0.5mm dimensions was equilibrated in cryo-protectant solution of 10%v/v glycerol in mother liquor, mounted in a loop and flash-frozen in a liquid nitrogen gas stream. The crystal was exposed to X-rays at 0.934Å for 8sec intervals (2 oscillations per image); 360 0.5° oscillations were collected. The data were indexed in a monoclinic space group and integrated and merged using MOSFLM, version 6.2.4 (Leslie, 1992) and then Scala, version 3.2.1 (Evans, 1993) as implemented in the CCP4 Program Suite (Bailey, 1994). The resolution was cut to 2.4Å and unit cell dimensions were used to calculate the solvent content of the crystal (Matthews, 1968).

3.3.16 EvaA Activity Assay

A two-enzyme assay was established to determine EvaA activity. The assay contained 5µg EvaA, 5µg RmlB and 10mM dTDP-D-glucose, which is converted to the EvaA substrate, dTDP-4-keto-6-deoxy-D-glucose. The assay was prepared in 50µl total volume and the mixture incubated at 37°C for 3 hours. Table 3.4 details the reaction mixture components. Following incubation, the enzymes were removed from the assay mixture by concentration through an Amicon YM10,000 spin column. The column flow-through containing the products was retained for analysis by ESI-TOF mass spectrometry. The dehydration reaction catalysed by EvaA, as well as the RmlB product should be visible since hydroxyl groups are abstracted from the products (Figure 3.8).

Table 3.4 – Concentrations of Reaction Components

	Concentration in Assay Mixture
Tris-HCl, pH7.5	20mM
dTDP-glucose	10mM
RmlB	5µg
NAD^+	2mM
dH_2O	Made up to a total volume of 50µl
EvaA, EvaA-MBP, FMN	5µg

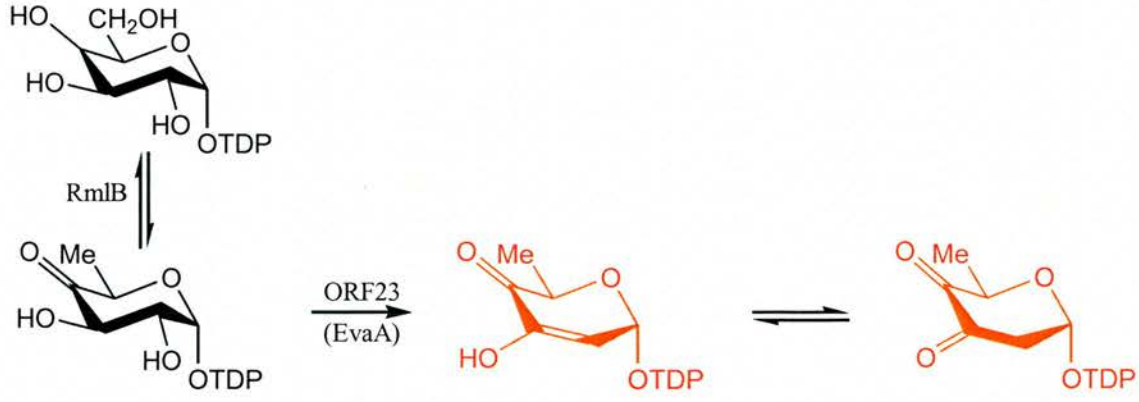


Figure 3.8 – Reaction Pathway for the EvaA Activity Assay: the possible reaction products are shown in red, they should be equivalent when analysed by mass spectrometry.

EvaA in the presence and absence of FMN was analysed as well as the EvaA-MBP fusion construct. Reaction mixtures without enzymes and with denatured EvaA served as negative controls. To analyse the substrate tolerance of EvaA dTDP-glucose was incubated with EvaA in the absence of RmlB.

A 5µl portion of the above column flowthrough, containing the reaction products but no enzymes, was analysed by ESI-MS in negative mode. Injections were made using a Waters 2795 autosampler and the analysis was carried out on a Micromass LCT instrument. Signals for M-H and M-H+Na-H were obtained for all samples and compared.

3.4 Results

3.4.1 EvaA Sequence Analysis

EvaA belongs to a family of NDP-hexose 2,3-dehydratases which are structurally completely uncharacterised. Knowledge of the structures of different EvaA homologues would allow us to draw up comparisons between active site topology and substrate recognition. Analysing the sequence similarity between EvaA and the six homologues shows that these enzymes are more similar to one another than would be expected for enzymes from different Actinobacteridae (Figure 3.9). It is particularly interesting that the homologue from *S. fradiae* is very similar to the Amycolatopsis proteins since it belongs to the Streptomycineae rather than the Pseudonocardineae. At the same time, *S. erythraea*, which belongs to the Pseudonocardineae produces a protein much more similar to the one from *S. antibioticus* which is a Streptomycineae. Also, one would have expected all the Streptomycetaceae proteins to be most similar to one another, to indicate that they have co-evolved. This is in line with the findings by Egan *et al.*, 2001, and Wellington *et al.*, 1992 who concluded from their studies on Streptomyces that antibiotic gene clusters are very prone to lateral gene transfer between different species and are therefore more similar to one another than would be expected from the similarity of their 16S rRNA.

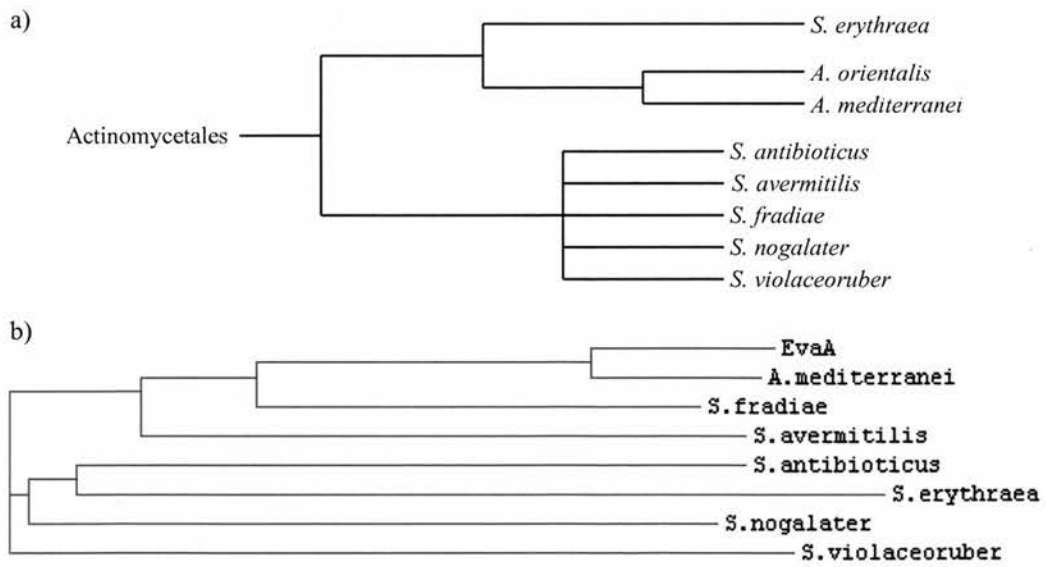


Figure 3.9 - Phylogram Trees: a) expected phylogram, according to 16s rRNA classification of Actinomycetales (NEWT (EBI, Cambridge) and NCBI taxonomy ID2037; Buchanan, 1917; Felske *et al.*, 1999; Skerman *et al.*, 1980; Stackebrandt *et al.*, 1997).
b) according to sequence similarity between the EvaA homologues from the same species.

Proteins in this family consist of two copies of a 250 amino acid long units which is most likely to have arisen by domain duplication (Apic *et al.*, 2003; Long, 2001). In EvaA, these two domains are 27% identical and 62% similar (see section 3.1). The first of the two domains is more conserved amongst the different homologues than the second one displaying between 93% and 53% sequence identity to EvaA, compared to between 86% and 44% sequence identity to EvaA for the second domain of the homologues. However, one significant motif (SEEGGRF) is completely conserved in the second domain but only exists partially in the first one (SGRFF). Figure 3.10 shows a ClustalW alignment (server at EBI, Cambridge) (Thompson *et al.*, 1999) of EvaA and its homologues; the domain break between domains 1 and 2 is highlighted in yellow.

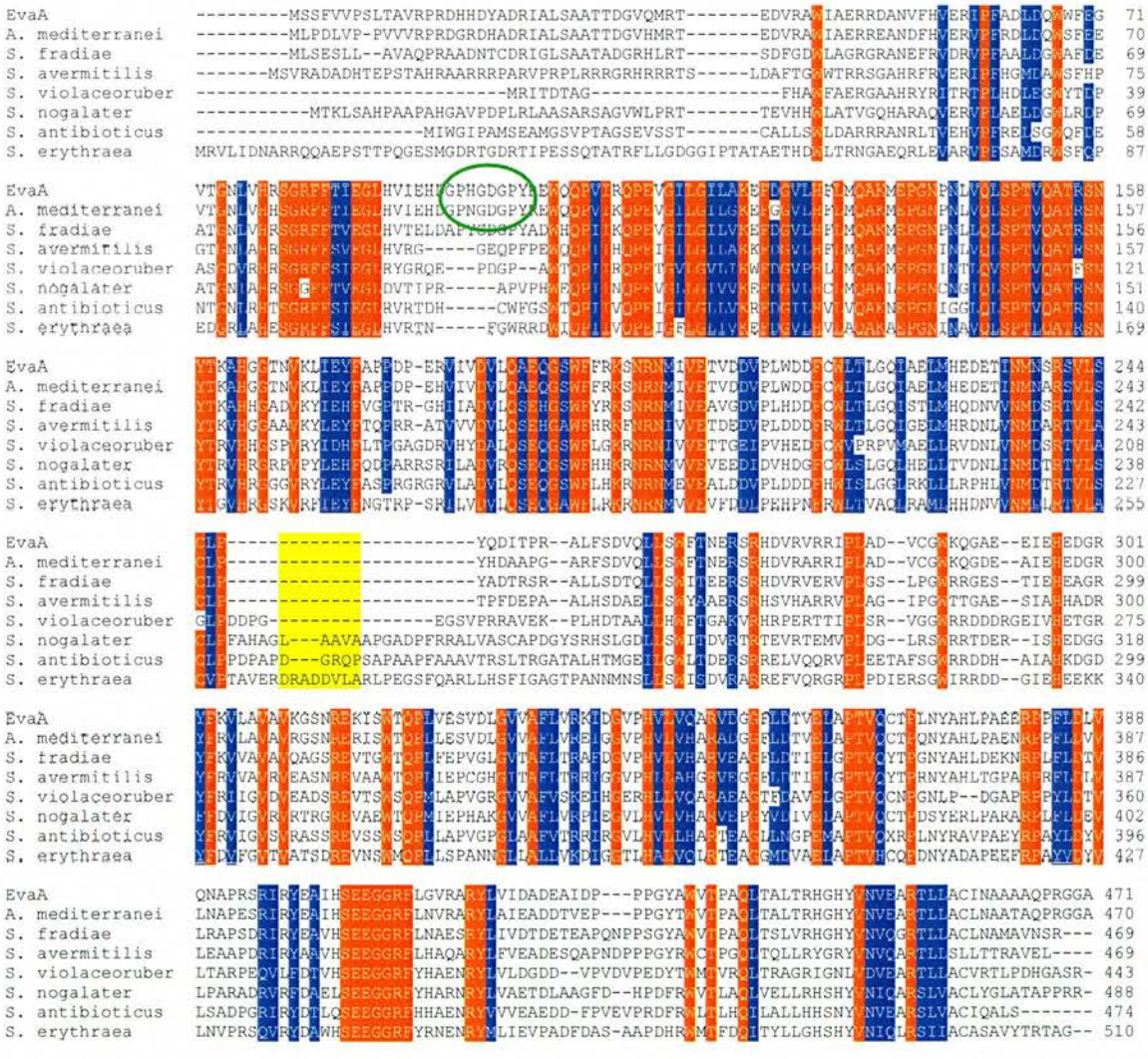


Figure 3.10 – EvaA Sequence Alignment: Sequence alignment of EvaA and a number of homologues. Completely conserved residues are denoted in red, conservative substitutions in blue; yellow denotes the domain break; the green ellipsoid denotes the GXXGXG motif in the two Amycolatopsis species.

Enzymes of this family are proposed to rely on FMN as the co-factor for their activity. However, no FMN binding motif as such is known, but nucleotide binding to a protein is normally via a Rossmann fold (Rossmann *et al.*, 1974). Several attempts at searching for this motif by a variety of database search programs (MotifScan, InterProScan (Zdobnov and Apweiler, 2001) and Hits (Pagni *et al.*, 2001)) have failed. This is not unexpected as the Rossmann fold motif displays a low sequence conservation and is simply defined as an α-helix, β-sheet, α-helix motif, which may be difficult to predict from the sequence alone.

Several domain searches, especially with MotifScan (Falquet *et al.*, 2002) (http://www.myhits.isb-sib.ch/cgi-bin/motif_scan) as part of the ExPASy Server (Gasteiger *et al.*, 2003), for the whole protein sequence and single domains were carried out, but nothing with the exception of the matching 2,3-dehydratase profile could be found. Subsequently, the more powerful three-dimensional domain threader at Imperial College, (3D-PSSM Server, http://www.sbg.bio.ic.ac.uk/servers/3dpssm, Kelley *et al.*, 2000) was employed. It was used for each complete homologue as well as individual domains to analyse possible sequence and fold pattern similarities to known proteins. This pinpointed to the possibility of FMN binding to the first but not the second domain, since several similarities to flavodoxin and ferrodoxin reductases (PDB entries d1fdr and d1que) have been found in this domain. Similarly, the second domain may be responsible for sugar binding, as it aligns with several carbohydrate processing enzymes, such as the Threonine dehydrogenase, an isomerase from *E. coli* and the *E. coli* adp-ribose2 pyrophosphatase (95% confidence in an alignment with the second domain from *S. nogalater*). A metal binding motif may be present as several predictions have included Zn^{2+}- dependant enzymes, such as Zn^{2+}-dependant dehydratases and proteases.

The secondary structure was predicted using Jpred and Jnet (Web server at the University of Dundee, http://www.compbio.dundee.ac.uk/~jpred, Cuff *et al.*, 1998; Cuff *et al.*, 2000) and Hierachical Neuronal Network Protein Sequence Analysis (NPS@), at Lyons http://npsa-pbil.ibcp.fr/cgi-bin/secpred_hnn.pl (Combet *et al.*, 2000). All programs came up with largely the same overall structure (Figure 3.11), the most defined of which seemed to be three to four α-helices: one long one at each end of the complete protein and one running across the two domains, or more likely, one α-helix at the end of domain one, followed by a short coiled region and another α-helix at the start

of the second domain. The remainder of both domains is more likely to adopt β-sheet folds. This is in line with the thought of domain duplication, as the fold motifs repeat across the two domains. The Hierachical Neuronal Network Analysis identified another alpha helix about 100 residues into domain 2 (approximately between residues 350 and 365) (the other programs annotated it as disordered coil or beta strand). These fold prediction have to be treated with some caution as they do not take all the surrounding residues into account, for example, the occurrence of a beta-strand two amino acids in length within an extended region predicted to be an alpha helix is highly unlikely.

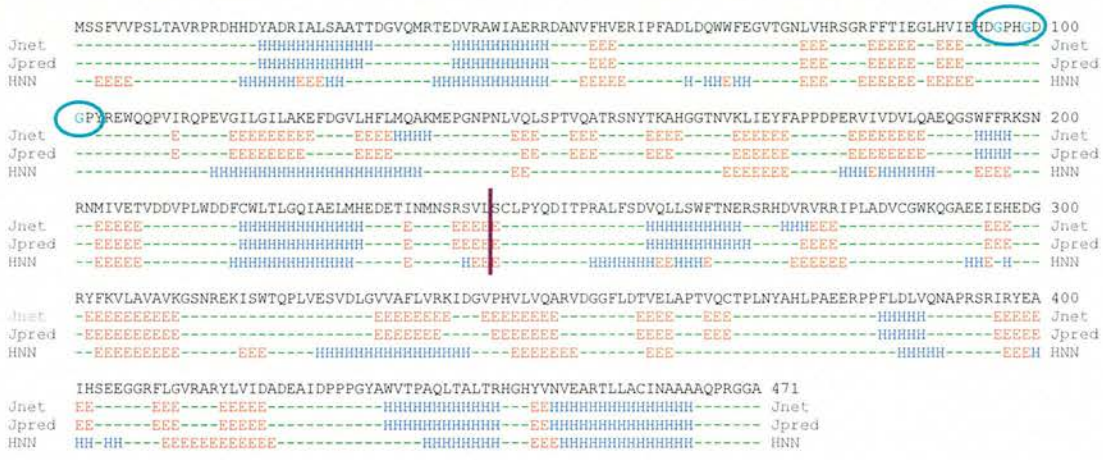


Figure 3.11 – Secondary Structure Prediction for EvaA: Jpred, Jnet and the Hierachical Neuronal Network (HNN) were used to predict secondary structure elements. α-helices (H) are denoted in blue, β-strands (E) in red and random coil regions (-) in green below the amino acid sequence. The domain break is denoted by a purple vertical line and the GXXGXG motif is highlighted by turqoiuse ellipsoids.

As indicated in the introduction, no Rossmann fold for nucleotide binding could be identified. A motif (GXXGXG), similar to the Wierenga motif (GXXGXXG) (Wierenga *et al.*, 1985; Wierenga and Hol, 1983) was found in the domain that was predicted to bind the FMN co-factor and may therefore adopt a Rossmann fold (Rossmann *et al.*, 1974). This motif is important in the correct positioning of the pyrophosphate backbone of nucleotides when found within a Rossmann fold (Iain Kerr, University of St. Andrews, PhD Thesis 2003). However, the significance of this is not yet understood as the motif is only conserved in the Amycolatopsis species but not the other Actinomycetes.

Doing this sort of structural analysis from the sequence alone, it is very important not to place too much emphasis on the prediction results, since protein folding is not completely understood (Protein Folding Problem) (Branden and Tooze, 1998) and therefore drawing conclusions from the sequence alone may be misleading (Novotny *et al.*, 2004).

3.4.2 Genome Isolation

Genomic DNA from five of the seven chosen species was successfully isolated (Figure 3.12). Inoculating medium with *S. erythraea* did not result in the growth of any culture. The *S. fradiae* cell culture was contaminated with yeast after the second round of culturing, as judged by scent and turbidity. The following list details the genomes obtained and their respective yields:

- *A. mediterranei* 62.5µg (125ng μl^{-1})
- *S. antibioticus* 206.5µg (413ng μl^{-1})
- *S. avermitilis* 128.0µg (256ng μl^{-1})
- *S. nogalater* 73.5µg (147ng μl^{-1})
- *S. violaceoruber* 49.0µg (98ng μl^{-1})
- *S. erythraea* did not grow
- *S. fradiae* yeast contamination

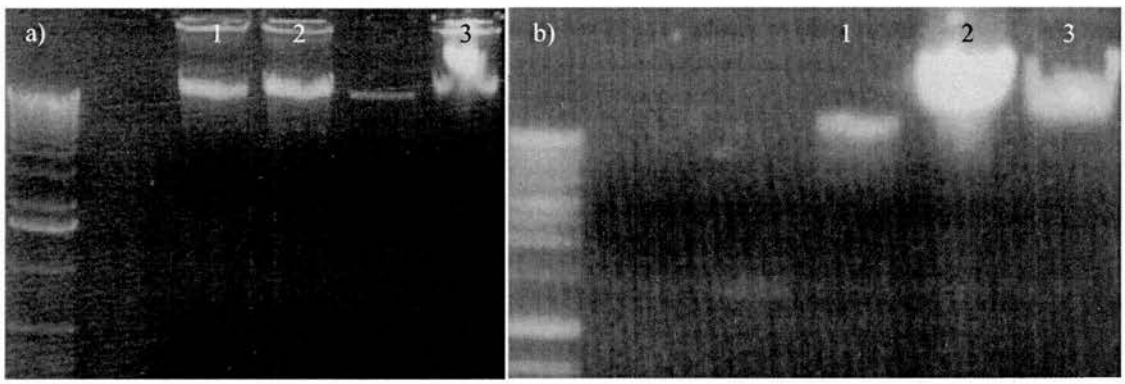


Figure 3.12 - Genomic DNA Extracted from Actinomycetes:

a)
1 – *A. mediterranei*
2 – *A. mediterranei*
3 – *S. nogalater*

b)
1 – *S. violaceoruber*
2 – *S. antibioticus*
3 – *S. avermitilis*

3.4.3 Amplification of Homologues

Carrying out PCR for G+C rich genomes is intrinsically difficult. It was found that the only DAN polymerase capable of amplifying genes from the genomic templates was the BioXact DNA polymerase (Bioline). PCR was only successful if the genome had been heated to 70°C for several minutes to uncoil it before adding it to the reaction

mix. Depending on the template, optimum $MgCl_2$ concentrations were determined for amplification. All genes also required the addition of the HiSpec additive (Bioline), which was used according to the manufacturer's instructions. For the amplification of *S. violaceoruber* DNA, the addition of DMSO was necessary. The PolyMate additive (Bioline) did not improve the amplification when used in concert with the BioXact DNA polymerase. Furthermore, other DNA polymerases (*Pfu* and *Vent*) did not yield any useful PCR products. The number of cycles (25), melting (94°C to 95°C) and extension temperatures (68°C) and times (30sec to 1min melting, 1min annealing, 2.5min extension) did not appear to make much difference to the outcome of the PCR. Throughout the trials, forward and reverse primer concentrations were maintained in excess.

The EvaA homologues were between 1400bp and 1600bp in size; Figure 3.13 shows some of the successfully amplified genes. Positive controls were carried out for all PCRs by amplifying the EvaA gene from its pET22b+ vector at the same time under similar conditions. Table 3.5 details the optimum reaction conditions for the amplified genes (in 100µl reaction volumes):

Table 3.5 - Optimum PCR Conditions for the Different Genome Templates

	A.mediterranei	*S. antibioticus*	*S. avermitilis*	*S. nogalater*	*S.violaceoruber*
Genomic template	550ng	900ng	750ng	450ng	490ng
DNA Polymerase	BioXact	BioXact	BioXact	BioXact	BioXact
Forward and Reverse Primer conc.	5µM	5µM	5µM	5µM	5µM
HiSpec additive	1x	1x	1x	1x	1x
$MgCl_2$ conc.	3mM	3mM	4mM	3mM	8mM
DMSO conc.	none	none	none	none	7.5-10%
Melting temperature	94°C	95°C	95°C	94°C	95°C
Annealing temperature	58°C	53°C	54°C	50°C	61°C (1^{st} cycle) 63°C
Extension temperature	68°C	68°C	68°C	68°C	68°C

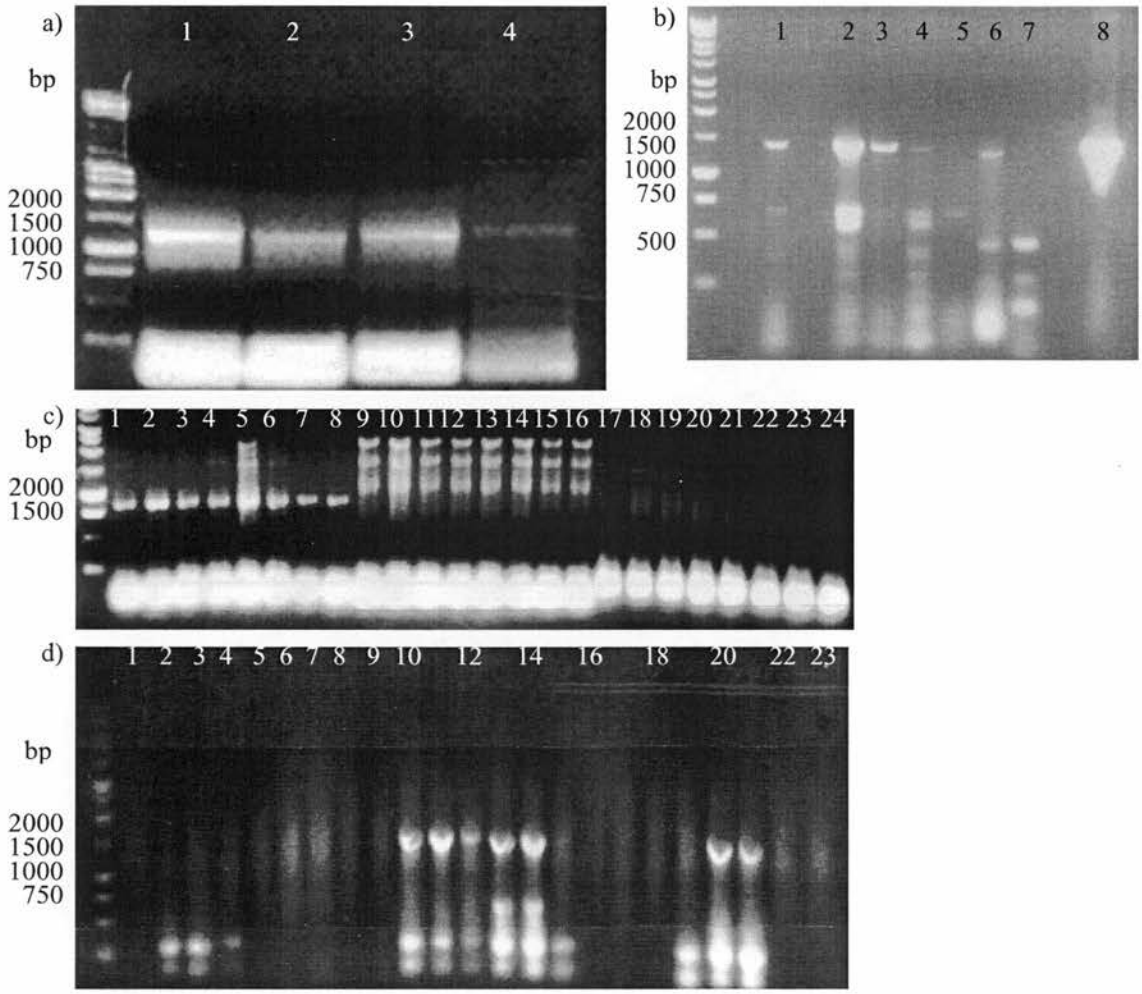


Figure 3.13 - PCR Results for the Different Homologues:

a) *S. avermitilis*:
varying [$MgCl_2$]
1: 4mM $MgCl_2$, 2: 6mM $MgCl_2$, 3: 8mM $MgCl_2$, 4: 10mM $MgCl_2$

b) *S. nogalater*:
varying [$MgCl_2$] and annealing temperatures
1: 50˚C and 2mM $MgCl_2$, 2: 48˚C, 3mM $MgCl_2$, 3: 50˚C and 5mM $MgCl_2$, 4: 52˚C and 5mM $MgCl_2$, 5: 54˚C and 5mM $MgCl_2$, 6: 52˚C and 8mM $MgCl_2$, 7: 52˚C and 10mM $MgCl_2$, 8: positive control - pET vector with EvaA insert

c) 1-8 *A. mediterranei*; 9-16 *S. antibioticus*; 17-24 *S. violaceoruber* varying [$MgCl_2$], no DMSO used
1: 1mM $MgCl_2$, 2: 3mM $MgCl_2$, 3: 4mM $MgCl_2$, 4: 5mM $MgCl_2$, 5: 7mM $MgCl_2$, 6: 8mM $MgCl_2$, 7: 10mM $MgCl_2$, 8: 15mM $MgCl_2$,
9: 1mM $MgCl_2$, 10: 2mM $MgCl_2$, 11: 4mM $MgCl_2$, 12: 5mM $MgCl_2$, 13: 7mM $MgCl_2$, 14: 8mM $MgCl_2$, 15: 10mM $MgCl_2$, 16: 15mM $MgCl_2$,
17: 1mM $MgCl_2$, 18: 2mM $MgCl_2$, 19: 3mM $MgCl_2$, 20: 5mM $MgCl_2$, 21: 8mM $MgCl_2$, 22: 10mM $MgCl_2$, 23: 15mM $MgCl_2$, 24: 20mM $MgCl_2$

d) *S. violaceoruber* varying [$MgCl_2$] and [DMSO]
1: 2mM $MgCl_2$ and no DMSO, 2: 4mM $MgCl_2$ and no DMSO, 3: 8mM $MgCl_2$ and no DMSO, 4: 4mM $MgCl_2$ and 2% DMSO, 5: 6mM $MgCl_2$ and 2% DMSO, 6: 8mM $MgCl_2$ and 2% DMSO, 7: 4mM $MgCl_2$ and 5% DMSO, 8: 6mM $MgCl_2$ and 5% DMSO, 9: 8mM $MgCl_2$ and 5% DMSO, 10: 6mM $MgCl_2$ and 7.5% DMSO, 11: 8mM $MgCl_2$ and 7.5% DMSO, 12: 4mM $MgCl_2$ and 10% DMSO, 13: 6mM $MgCl_2$ and 10% DMSO, 14: 8mM $MgCl_2$ and 10% DMSO, 15: 15mM $MgCl_2$ and 10% DMSO, 16: 4mM $MgCl_2$ and 15% DMSO, 17: 6mM $MgCl_2$ and 15% DMSO, 18: 8mM $MgCl_2$ and 15% DMSO, 19: 15mM $MgCl_2$ and 15% DMSO, 20: 8mM $MgCl_2$ and 6% DMSO, 21: 15mM $MgCl_2$ and 6% DMSO, 22: 15mM $MgCl_2$ and 6%DMSO

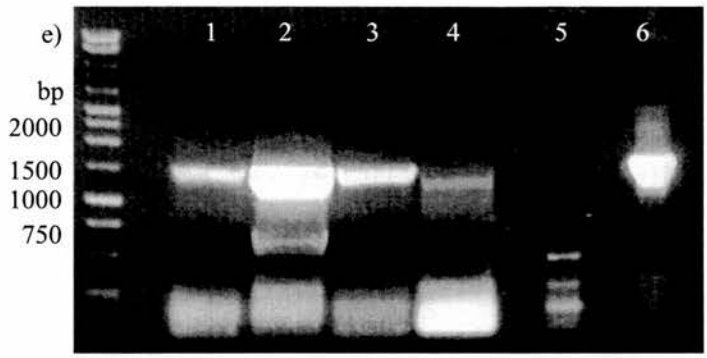


Figure 3.13 continued - PCR Results for the Different Homologues:
e) *S. antibioticus* at varying [$MgCl_2$]
1: 2mM $MgCl_2$, 2: 3mM $MgCl_2$, 3: 4mM $MgCl_2$, 4: 8mM $MgCl_2$, 5: 12mM $MgCl_2$, 6: positive control, EvaA in pET vector

3.4.4 Cloning of EvaA and its Homologues

All amplified genes were gel extracted using the Qiagen gel extraction kit®, eluted off the spin column into dH_2O and prepared for ligation by restriction digestions with the appropriate restriction enzymes.

EvaA, amplified from pET22b+, was double digested in the appropriate buffers with *Nco* I and *Bam*H I (for the pEHISTEV vector) and *Nco* I and *Eco*R I (for the pMAL-c2X, pMAL-p2X and pLou-3 vectors) then ligated into the vector as detailed in section 3.3.6. The reaction mix was subsequently used to transform highly competent Tam-1 *E. coli* cells (ActiveMotif). To determine the presence of inserts in the vectors, colonies were used either for colony PCR using the cloning primers; or the plasmid DNA was isolated by miniprep and insert presence and size determined by double restriction digest. Figure 3.14 shows two analytical DNA gels to verify the presence of EvaA in the pEHISTEV and pLou-3 vectors.

During cloning experiments it became apparent that the method by which the pEHISTEV vector was purified, following restriction digest, affected its ability to be ligated to an appropriate insert. If extracted from agarose gels, ligation into the pEHISTEV vector became impossible. This finding was borne in mind for cloning of all other genes into this vector.

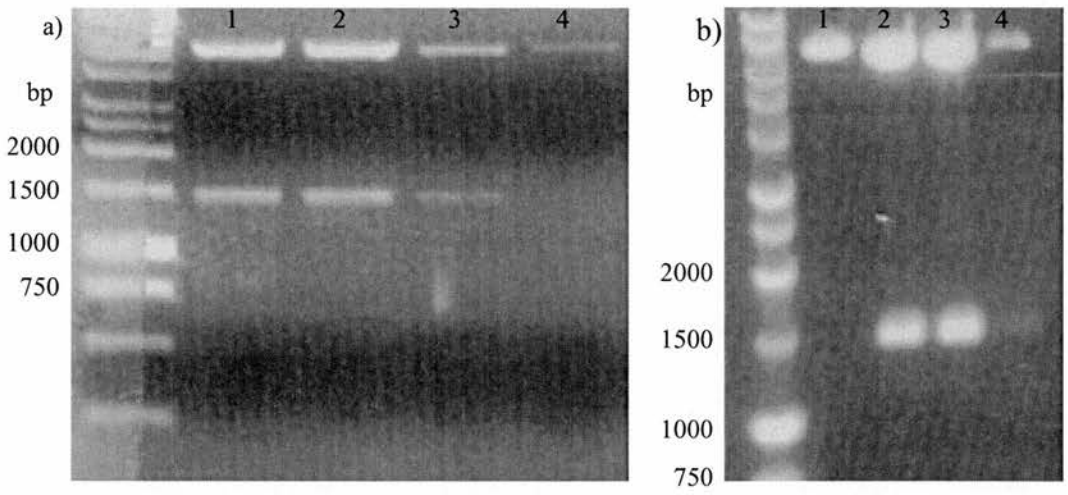


Figure 3.14 - Double Digests to Determine the Presence of orf23 in a) pEHISTEV and b) pLou-3
a) lanes 1, 2 and 3: insert present, lane 4: no insert present
b) lanes 2 and 3: insert present, lanes 1 and 4: no insert present

The *A. mediterranei* gene was cloned into the pHISTEV and pLou-3 vectors using a sequential restriction digest of the PCR product and then a standard ligation reaction. As restriction digests proved difficult for the other PCR products, all other homologues were cloned in a two-step strategy. First, the amplified gene product, containing adenosine overhangs, was ligated into a T-tailed vector (pETBlue-1 or pGem-T-easy), which was then amplified and digested. During primer design, consideration had been given to the overhang length required for restriction enzyme action. However, the digests (double, single or sequential) were far more efficient using a vector template. Digesting the gene out of the parent vector also ensured that the gene had been completely digested and was ready for ligation into the host vector. After agar gel electrophoresis, the gene was gel-extracted prior to ligation. In this way the EvaA homologues from *S. antibioticus* and *S. nogalater* were cloned into the pEHISTEV and pLou-3 vectors. The presence of insert was checked as before (Figure 3.15) and all vectors containing inserts were sent for DNA sequencing. Several subcloning attempts for the EvaA homologue from *S. violaceoruber* failed as verified by DNA sequencing; a number of small genes of *E. coli* origin were identified.

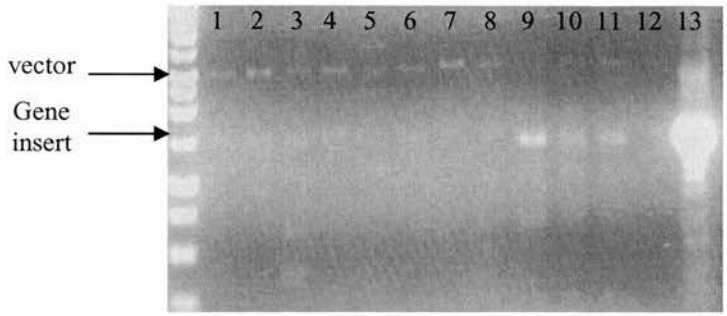


Figure 3.15 - Colony PCR: to confirm the presence of gene insert in the pLou-3 vector.

1 – *S. antibioticus*
2 – *S. antibioticus*
3 – *S. antibioticus*
4 – *S. nogalater*
5 – *S. nogalater*
6 – *S. nogalater*
7 – *S. violaceoruber*
8 – *S. violaceoruber*
9 – *A. mediterranei*
10 – *A. mediterranei*
11 – *A. mediterranei*
12 – *A. mediterranei*
13 – positive control

3.4.5 Expression of Wild-type EvaA

Expression trials were carried out for wild-type EvaA cloned in pET22b+ and a number of *E. coli* strains under a variety of expression conditions. None of the 565 conditions tested yielded any soluble protein. The majority of conditions generated expression of large amounts of EvaA as judged by SDS-gel electrophoresis (mass 55,000Da) and trypsin-digest MALDI-TOF mass spectrometry. However, all of the protein was found in inclusion bodies, i.e. the insoluble cell fraction. Even the addition of mild detergents, such as OPOE and BOG, to the cell lysate did not yield a reasonable amount of soluble protein (Figure 3.16) (Blackwell and Horgan, 1991; Frankel *et al.*, 1991; Georgiou and Valax, 1996; Schaffner *et al.*, 2001; Schein, 1990).

Some conditions, such as using the *E. coli* strain BL21*trx*B(*DE3*), did not result in any protein being over-expressed at all whilst others only gave insoluble EvaA. This particular strain possesses the same thioredoxin reductase mutation as the AD494 strains in the protease deficient BL21 background. This mutation facilitates the disulfide bond formation and their use may increase the fraction of properly folded protein. However, no general all encompassing idea could be established connecting these two opposing results to expression variables (other than BL21*trx*B(*DE3*) cells). The expression trials that relied on leaky protein expression rather than IPTG induction to gain soluble proteins also failed. The opposite, i.e. usage of *E. coli* strains that prevent

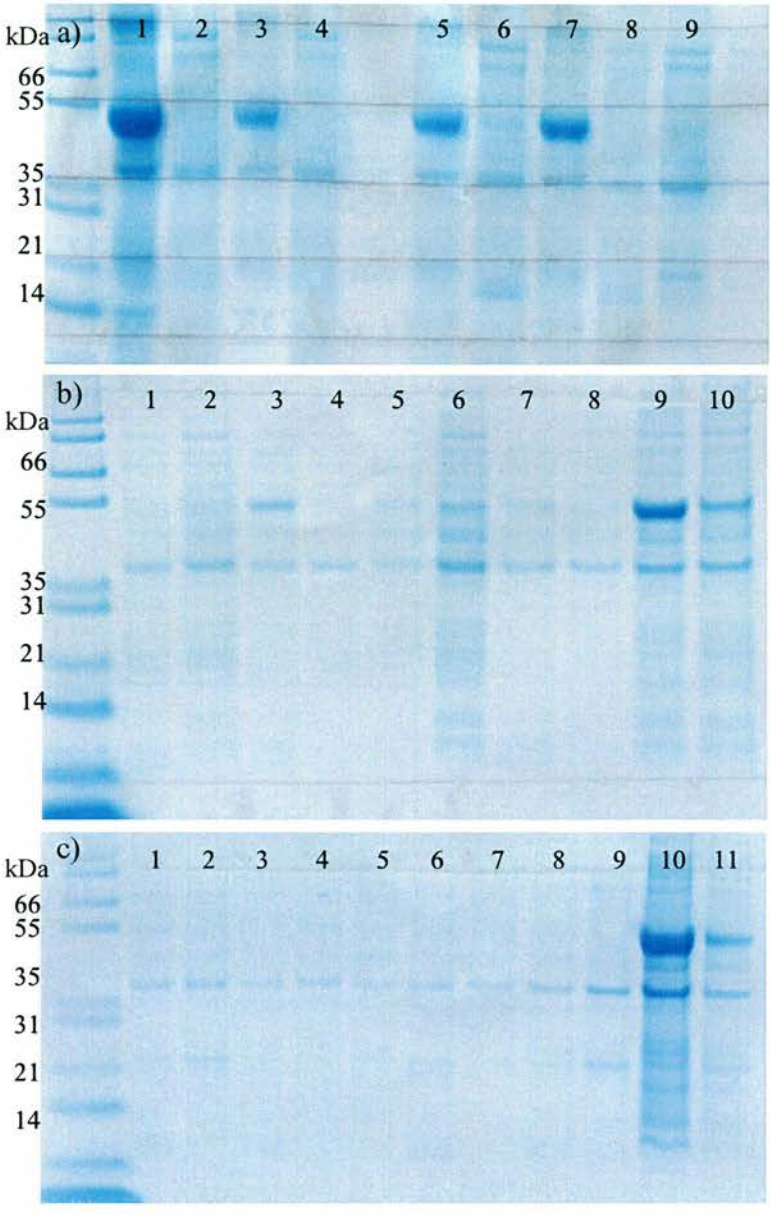


Figure 3.16 - Expression Trials for EvaA from pET 22b+:
a) BL21(*DE3*)
1 – LB 5hrs, insoluble
2 – LB, 5hrs, soluble
3 – LB, 3hrs, insoluble
4 – LB, 3rs, soluble
5 – LB, 2hrs, insoluble
6 – LB, 2hrs, soluble
7 – LB, 1hr, insoluble
8 – LB, 1hr, soluble
9 – LB o/n sarter culture
b) BL21 Star™
1 – LB o/n starter culture
2 – TB, 1hr, soluble
3 – TB, 1hr, insoluble
4 – TB, 2hrs, soluble
5 – TB, 3hrs, soluble
6 – TB + sucrose, 1hr, insoluble
7 – TB + sucrose, 1hr, soluble
8 – TB + sucrose, 3hr, soluble
9 – TB + sucrose, 3hr, insoluble
10 – TB o/n starter culture
c) BL21-SI
1 – LB + sucrose, 25°C, sol.
2 – LB + sucrose, 37°C, sol.
3 – LB + glucose, 25°C, sol.
4 – LB + glucose, 37°C, sol.
5 – LB + gluc. + sucr., 25°C, sol.
6 – LB + gluc. + sucr., 37°C, sol.
7 – minimal med., 25°C, sol.
8 – minimal med., 37°C, sol.
9 – minimal med. + gluc., 25°C, soluble
10 – LB + glucose + sucrose, 25°C, insoluble
11 – minimal med. + glucose, 25°C, insoluble
a) and b) expression at 25°C

leaky protein expression (BL21(*DE3*)pLysS and BL21(*DE3*)pLysE), also did not result in soluble protein. The OvernightExpress™ Autoinduction System which is optimised for tight expression control and induction at high cell density, in case the expressed protein is toxic to the cells, did not yield any soluble protein.

The over-expressed protein is unlikely to be toxic to the bacteria as expression was not preventing cell growth, as judged by the increase in OD_{600} during protein expression. It is more likely that the recombinant protein was not folded correctly, possibly aggregated and therefore contained in inclusion bodies in the host cells.

The DNA sequence of orf23 was analysed for the presence of rare codons (Appendix C) using the rare codon calculator (http://66.75.109/cgi-bin/cmura/racc.html)

(Schenk *et al.*, 1995). It contained 15 rare codons in total, five of which were rare Arg codons, one rare Ile codon and the remainder rare Pro codons; no rare Arg codons were direct repeats of one another. This result lead to the use of codon-optimised *E. coli* strains, such as BL21(*DE3*)RP-codon+ and Rosetta(*DE3*). Once again, this did not make any difference to the solubility of the expressed protein.

The pEHISTEV vector had been successfully proven to allow the soluble expression of 'difficult' proteins, i.e. ones that do not express solubly under standard conditions (H. Liu, University of St. Andrews, personal communication). For this reason, EvaA was cloned into the pEHISTEV vector and expression trials were undertaken. Despite the presence of the His·Tag, the TEV cleavage site and the long linker region, encoded by the vector, the recombinant protein expressed from this vector was insoluble (Figure 3.17). The slight expression bands at approximately 55kDa in the soluble fractions were analysed by trypsin digest MALDI-TOF mass spectrometry and their trypsin digest signatures identified them as proteins of *E. coli* origin.

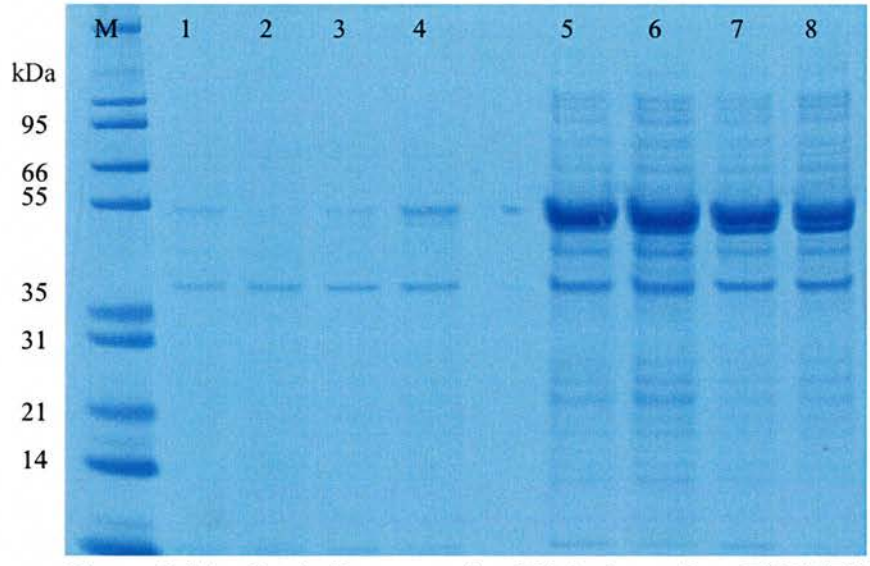


Figure 3.17 - EvaA Overexpression Trials from the pEHISTEV Vector: using BL21(*DE3*) *E. coli* cells:

M – Marker (BioRad)
1 – LB
2 – LB + 1%w/v glucose
3 – TB
4 – TB + 1%w/v glucose
5 – LB
6 – LB + 1%w/v glucose
7 – TB
8 – TB + 1%w/v glucose

Soluble fractions

Insoluble fractions

The Trx·Tag™ had been reported to improve the solubility of proteins expressed as the thioredoxin fusion construct (Davis *et al.*, 1999). However, it was not possible to get any expression of EvaA from this vector at all. As a control experiment, the empty vector was used for expression trials and thioredoxin protein could be expressed. Changes of expression time, temperature, medium, additives, cell lines and IPTG concentration did not make any difference to this result.

No recombinant protein expression was achieved using the pMAL-p2X vector either. However, according to the vector documentation, it is not uncommon to get little or no protein exported from the cytoplasm when trying to express recombinant protein from this vector.

Expressing soluble EvaA fused to MBP was possible using the pMAL-c2X vector (Figure 3.18). Optimisation of the conditions showed that expression of the fusion construct was uncomplicated. For example, large amounts of soluble protein could be obtained from any cell line (Tam-1, NovaBlue, DH5α). This was possible as the vector contained an intrinsic *E. coli* promoter, making the need for expression in optimised cell lines obsolete. EvaA fused to MBP also expressed well in the BL21(*DE3*) cell line, allowing the production of greater quantities of material from the

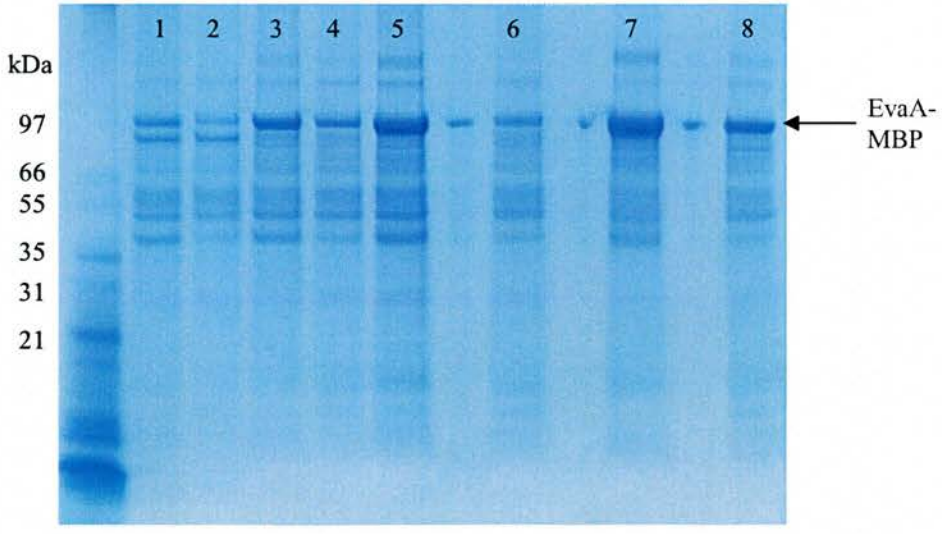


Figure 3.18 - Expression of EvaA Fused to MBP: expression rom the pMal-c2X vector - the arrow shows the expression of EvaA fused to MBP (all exxpression carried out in LB at 25°C using 1mM IPTG for induction)

1 – non-induced, total	5 – 2 hours, total
2 – non-induced, soluble	6 – 2 hours, soluble
3 – 1 hour, total	7 – 5 hours, total
4 – 1 hour, soluble	8 – 5 hours, soluble

same volume of cell culture. However, the EvaA-MBP fusion protein expressed in BL21(*DE3*) cells was very resistant to TEV protease cleavage and therefore, EvaA-MBP was expressed using slow-growing cloning cells, such as Tam-1. The expression medium was not a critical component of the expression, as LB, TB and 2x TYE all worked equally well. After induction the temperature was decreased to between 20 and 25°C to give the expressed protein the best possible chance of folding correctly. Expression time was not deemed critical either and therefore protein was expressed overnight.

Expression was analysed by SDS-PAGE and the 97kDa bands were cut out of the gel and their contents determined by trypsin digest MALDI-TOF mass spectrometry. As trypsin digest fingerprints were present for MBP as well as EvaA, it was concluded that the correct protein had been expressed solubly.

3.4.6 Protein Purification

Even though little expression could be seen on protein gels, His·tagged proteins are simple to purify, even when present in small quantities. For this reason, purification trials were set up for EvaA expressed from the pEHISTEV vector. After cell lysis, separation of soluble material from insoluble cell debris and dialysis, the lysate was applied to a pre-equilibrated anion exchange column (DEAE Fast Flow Sepharose™, Amersham Biosciences). Protein was eluted (Figure 3.19) using a linear NaCl gradient from 0 to 2.5M. Fractions were collected and protein containing ones analysed by SDS-PAGE. Mass spectrometry was employed to determine the presence of EvaA but only *E. coli* proteins were found in any of the samples.

The purification of the EvaA-MBP fusion construct relied on the MBP's affinity for amylose and maltose. The fusion protein was purified to single band purity in one step (Figure 3.20a). Cleavage of the construct using TEV was also straight forward and successful (Figure 3.20b), whilst the separation of EvaA from MBP was less straight forward. EvaA itself was found to bind to the amylose column preventing it to be separated from MBP on that particular column. Even increasing the maltose concentration on the column in 0.02mM steps did not allow separation.

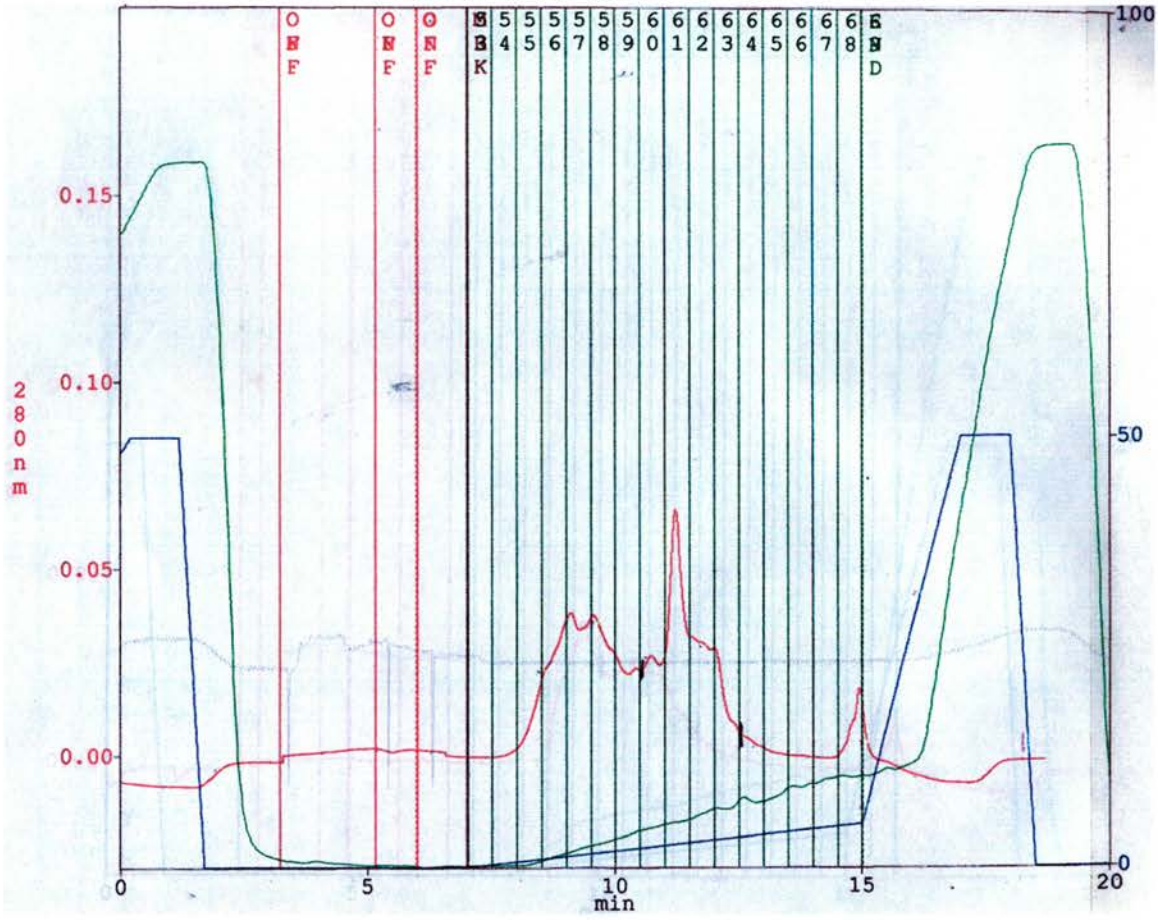


Figure 3.19 - Anion Exchange Trace for EvaA: red line: absorbance at λ = 280nm, green – salt concentration, blue: percentage of 4M NaCl used (elution gradient), green: conductivity – equivalent to up to 2M NaCl. Fractions 55 to 64 and 68 were analysed by SDS-PAGE.

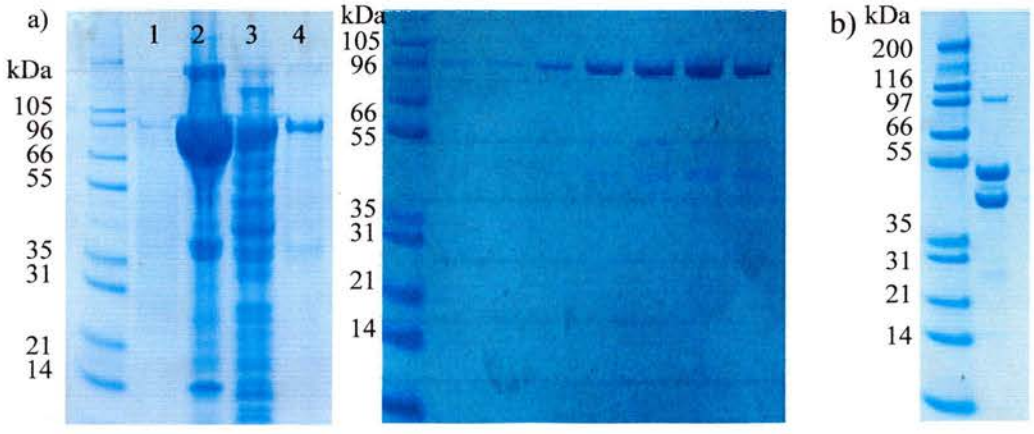


Figure 3.20 – Purification and Cleavage of EvaA-MBP:
a) Purification of the EvaA-MBP fusion protein:
1 – non-induced
2 – whole cell fraction (5hrs)
3 – soluble matter (amylose column load)
4 – fusion protein bound to amylose
Right hand gel: EvaA-MBP eluted off the amylose column.
b) Successful cleavage of the EvaA-MBP fusion construct using the TEV protease.

The purification attempts using any of the columns described in section 3.2.11 also failed. EvaA bound to all columns tested, in an irreversible manner:

- Metal chelating: Ni-NTA, Ni fast flow Sepharose
- Anion exchange: Poros HQ20, DEAE Sepharose Fast Flow
- Cation exchange
- Hydrophobic
- Gel filtration: S100 Sephadex, S200 Sepharose

The protein could only be eluted from these columns using denaturing conditions; 50% ethanol, 0.5%w/v SDS, 6M Urea or 8M Guanidine-HCl. The solution to this problem was the use of the ANX Fast Flow Sepharose™ (Amersham Biosciences) column for the second purification step. This column was equilibrated with ACB and EvaA was eluted using 1.2M NaCl in ACB; MBP did not bind at all to the column under the equilibration conditions and came off in the flow-through (Figure 3.21).

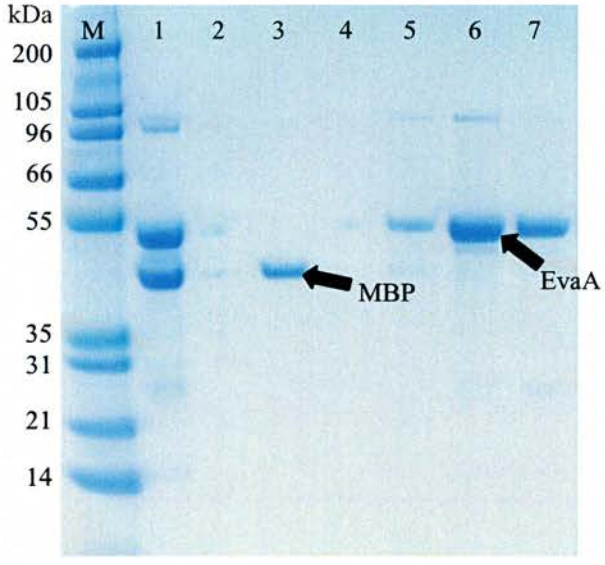


Figure 3.21 - Separation of EvaA from MBP: using ANX FF anion exchange chromatography
1: column load, 2: column flow-through, 3: wash, 4: wash, 5: elute (1.2M NaCl), 6: elute (1.2M NaCl), 7: elute (1.2M NaCl)

The ANX Fast Flow Sepharose™ is based on 4% highly cross-linked beaded agarose. The active end of the charged group is the same for the DEAE Fast Flow Sepharose™ and the ANX Fast Flow Sepharose™, the difference is the length of the carbon chain of the charged group. DEAE Sepharose™ has a diethylaminoethyl group

bound to the agarose whilst ANX Sepharose™ has a diethylaminopropyl group attached. This single carbon atom in the linker region prevents EvaA from binding irreversibly to the ANX Sepharose™ resin and thus allowed purification.

After purification, EvaA could only be successfully concentrated to levels suitable to protein crystallisation trials in the presence of 0.5mM FMN. However, the presence of FMN did not aid purification of EvaA as it did not co-purify with the protein. Adding FMN before the final chromatography step did not allow purification of EvaA via any of the other columns described above.

3.4.7 Circular Dichroism Spectroscopy

EvaA separated from MBP and the fusion construct were both analysed by CD spectroscopy (section 2.3.5). The near and far UV spectra of both EvaA preparations showed the peaks characteristic for secondary and tertiary structures, indicating correct folding (Figure 3.22). There is no difference between the spectrum for EvaA-MBP fusion and free EvaA, indicating that EvaA remains correctly folded after cleavage.

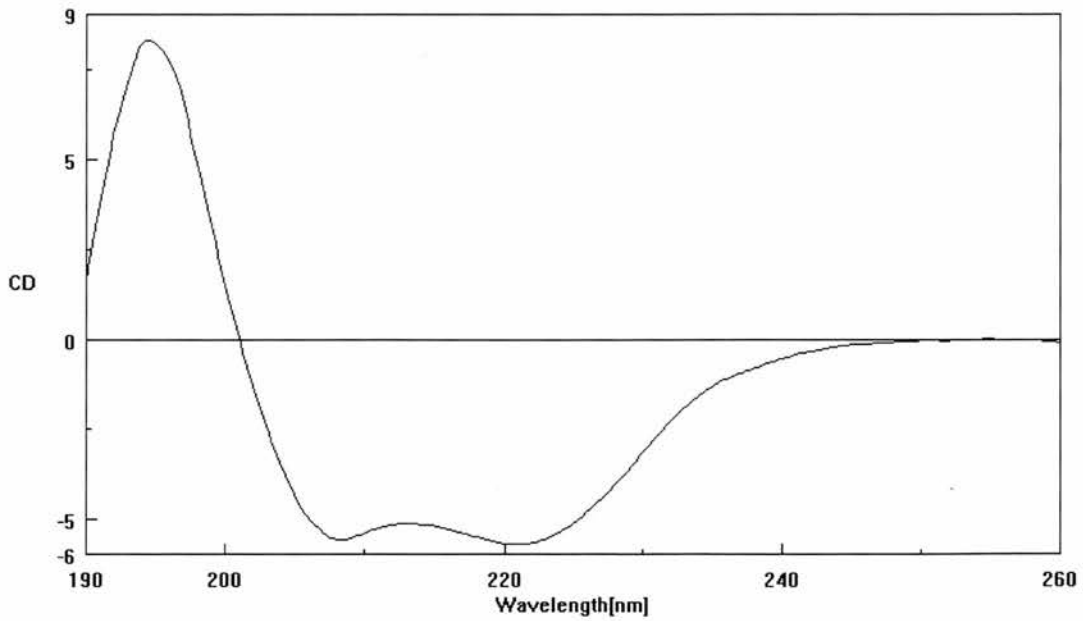


Figure 3.22 - CD Spectrum for free EvaA: showing the presence of secondary and tertiary structure elements in the protein, indicating folding.

3.4.8 Structural Data on EvaA-MBP

Crystals of the EvaA-MBP fusion construct were obtained under the following conditions: 1.5M Ammonium sulfate, 0.1M Tris, pH8.5, 12% Glycerol (anhydrous) and a protein concentration of 5mg ml^{-1}.

Data were collected at the ESRF, ID14-2 (λ = 0.932Å) and indexed in a monoclinic space group (a = 57.30Å, b = 45.69Å, c = 57.93Å, $\alpha = \gamma = 90°$, $\beta = 94.06°$); the reflections were integrated, merged and scaled using MOSFLM, version 6.2.4 (Leslie, 1992) and then Scala, version 3.2.1 (Evans, 1993). The cell volume was found to be 151,306$Å^3$; calculating the Matthews number for proteins of 97kDa, 55kDa and 42kDa respectively, showed that only the MBP (42kDa) could fit into the unit cell. Its Matthews number was 1.8 and its solvent content was calculated as 31.2% for a single monomer of MBP per unit cell. Using 55kDa (EvaA) for calculating the solvent content resulted in a percentage unacceptably low for a protein crystal. The fusion protein (97kDa) did not fit into the cell volume at all. The fact that only MBP had crystallised was confirmed by molecular replacement using MolRep (Vagin and Teplyakov, 1997). MBP (PDB entry 1MH4) was used in its entirety as a search model and the rotation and translation searches were carried out as a one-step process. The solution statistics were good enough to assume the crystal structure as solved and correct crystal packing was verified for the solution obtained by visualising it, together with its symmetry related objects in the graphics program O, version 7.0 (Jones *et al.*, 1991). Table 3.6 gives an overview of the data collection and structure solution.

Table 3.6 - Data Collection and Structure Solution Statistics for 'EvaA' MBP

Wavelength (Å)	0.932
Resolution (Å)	45.6-2.4
Space Group	Monoclinic P2
Unit-cell Parameters (Å, °)	a=57.30, b=45.69, c=57.93, $\alpha=\gamma=90°$, $\beta=94.06°$
V_m (1mol/au ($Å^3$ Da^{-1}))	1.8
Percentage solvent	31.2
Unique reflections	11,808
Data completeness (%)	99.8
Molecular Replacement	
R_{factor}	0.547
Correlation Coefficient	31.4%
Rotation peak (Rf/sig)	16.01
Translation function	0.236

The molecular replacement statistics were not completely convincing and the data could have come from other fragments of the EvaA-MBP fusion protein. However, carrying out 20 rounds of restrained refinement on the model against the crystallographic data using Refmac, 5.0 (Murshudov *et al.*, 1997) caused a drop in the R_{factor} to 32%. This would not have happened if some part of EvaA for example, had crystallised and we had tried to solve it using MBP. If the fold was similar enough (and it may well be as both bind carbohydrate substrates) we could have obtained a reasonable molecular replacement solution but under no circumstances would it have been possible to refine it. Thus the crystal obtained was definitely of MBP. I have got no reasonable explanation for this as the EvaA-MBP fusion complex is stable in solution at 4°C for months as determined by SDS-PAGE.

3.4.9 Homologue Expression and Purification

As expected, the EvaA homologues were not soluble, following expression from the pEHISTEV vector using cell lines and other expression variables described earlier. The EvaA homologues from *S. nogalater* and *S. antibioticus* were expressed in the pLou-3 vector in the same way as EvaA. Purification was also carried out in an identical manner to the one described for EvaA and purity was judged by SDS-PAGE. Protein integrity was analysed by trypsin digest MALDI-TOF mass spectrometry. Folding was verified by CD and a comparison to the standard EvaA far and near UV spectra displayed secondary and tertiary structure elements as would be expected for a correctly folded protein.

3.4.10 Crystallisation of EvaA

Sparse matrix crystallisation trials were set up for EvaA at 5mg ml^{-1} containing 0.5mM FMN in 20mM Tris-HCl, pH7.5. Small, yellow crystals (Figure 3.23) were obtained within 1 week of incubation at 20°C. The following conditions were used: sitting drop vapour diffusion method, a drop size of 2+2µl, 300mM Magnesium formate, 100mM Tris, pH8.5. They are not salt crystals as salt crystals of that size would have produced the characteristic salt X-ray diffraction pattern of salt: the small

unit cell typical of a salt crystal would have given intense diffraction spots, following exposure to X-rays, especially at high oscillation angles, i.e. 5°.

The crystals were distinguished as protein crystals as they stained blue upon application of Izit Crystal Dye™ (Hampton Research). Also they did behave as one would expect of protein crystals. Optimisation screens have, so far, yielded minor improvements on the crystals obtained. However, it is encouraging that they are reproducible from different batches of protein. These crystals do not diffract as, the best we have obtained to date, is the powder diffraction pattern, typical for spherulites (Hamodrakas *et al.*, 2004). Sparse matrix as well as optimisation crystal trials with higher protein concentrations (up to 10mg ml^{-1}) have been set up.

No crystals have been obtained for EvaA without FMN. This was possibly due to the protein concentration, as it was not possible to concentrate EvaA to more than 3mg ml^{-1} in the absence of FMN. Only 20% of the drops containing EvaA under these conditions showed any precipitate, indicating that the protein concentration was too low for crystallisation.

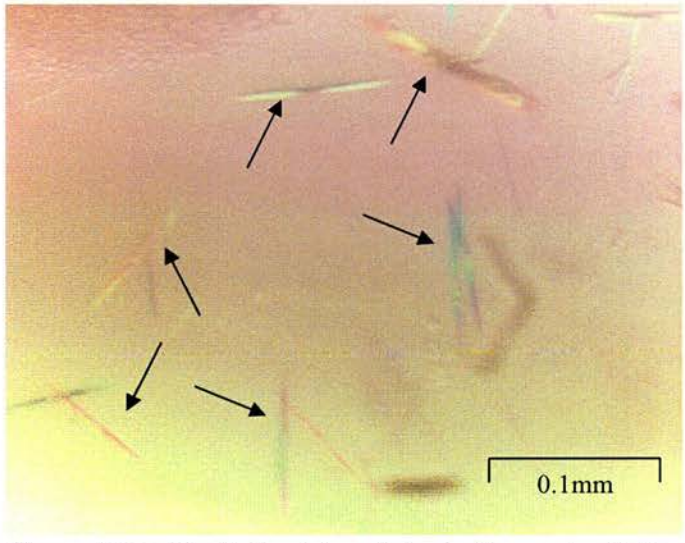


Figure 3.23 - EvaA Crystals: obtained after pre-incubation of the protein with 0.5mM FMN.

3.4.11 Selenomethionine Incorporation

SeMet EvaA was expressed and purified as described. The fusion protein, as well as purified EvaA was identified beyond doubt by trypsin digest MALDI-TOF mass spectrometry. The expected trypsin digest signature was detected. The fusion protein was analysed by ESI-TOF mass spectrometry and its mass compared to that of wild-

type EvaA. The mass difference between the wild-type fusion protein and the SeMet fusion protein was 660Da (95,910Da vs. 96,570Da) which accounted for the incorporation of 13.8 out of the expected 15 SeMet. This result was equivalent to 92% SeMet incorporation (Figure 3.24). This should be enough for structure solution by multiple anomalous difference (MAD) phasing if crystals of this derivative protein can be grown.

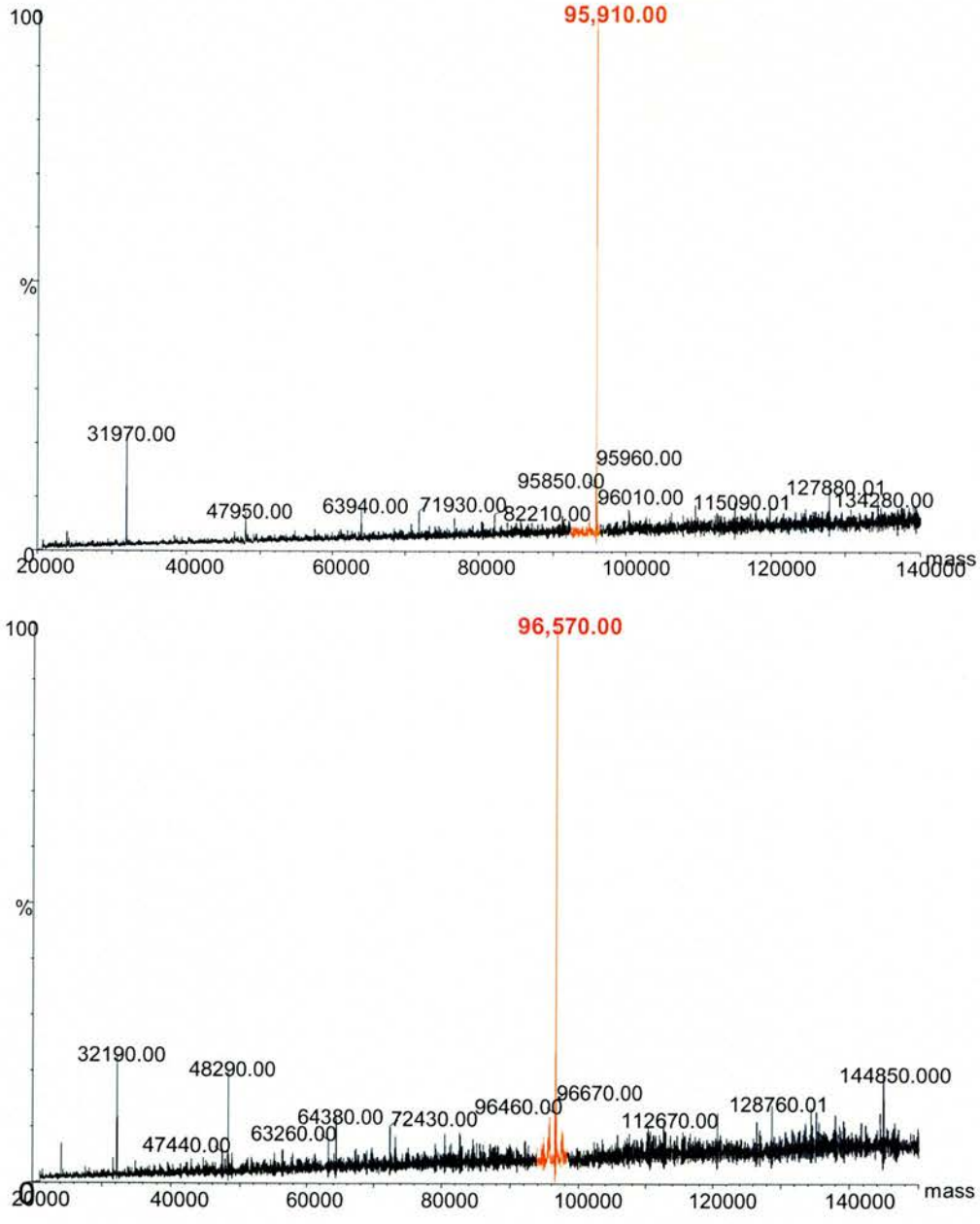


Figure 3.24 – SeMet Incorporation into EvaA-MBP: top spectrum: wild-type EvaA-MBP fusion for mass comparison; lower spectrum: EvaA-MBP fusion containing 92% SeMet incorporation

The SeMet derivative seemed to 'behave better' during purification compared to wild-type EvaA: concentration of the enzyme to 10mg ml^{-1} was more straight forward as the protein was less sticky. Sparse matrix crystallisation trials (Hampton Research: Crystal Screens 1 and 2, Index Screens 1 and 2, PEG/Ion Screen, Salt/RX Screen; EmeraldBioSciences: Wizard 1 and 2; Sigma: Crystallisation Screens 1 and 2) for the SeMet protein containing 0.5mM FMN were set up at protein concentrations of 5mg ml^{-1} and 10mg ml^{-1}. So far (4 weeks maximum incubation time) no crystals have been obtained however, the wells showed either precipitation or clear drops.

3.4.12 Enzyme Activity Assay

The original assay to determine the activity of EvaA relied on the production of maltol (highlighted in red in Figure 3.25). As described by Chen *et al.*, 2000, the reaction mixture was extracted twice with ethyl acetate, the organic solvent removed and the sample analysed by reverse phase HPLC using a Vydac C_{18} small pore analytic column. However, we thought that the formation of TDP adducts would occur preferentially and overshadow the presence of maltol, making its detection difficult if not impossible. For this reason, we decided to carry out a mass spectrometry assay using RmlB to produce the EvaA substrate dTDP-4-keto-6-deoxyglucose from dTDP-glucose. The assay was set up to monitor the removal of oxygen from the dTDP-glucose and dTDP-4-keto-6-deoxy-D-glucose starting materials by mass spectrometry.

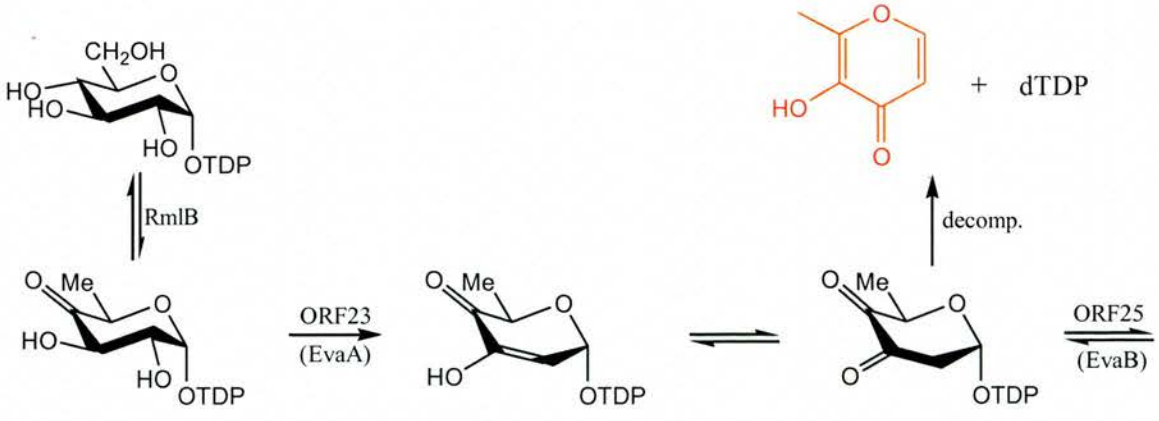


Figure 3.25 – EvaA Reaction Scheme: starting from dTDP-4-keto-6-deoxyglucose, we may have been able to determine the presence of maltol (highlighted in red) by reverse phase HPLC, but the presence of TDP-adducts would probably prevent the identification of the right procduct.

After incubation, all samples were analysed by ESI-TOF mass spectrometry (Figure 3.26). The original substrate, dTDP-glucose, was shown to display its major peak at 562.9m/z. Some of the sugar was turned over by RmlB as was seen from the 544.9m/z peak which accounts for the loss of H_2O (18 units), i.e. the dTDP-4-keto-6-deoxy-D-glucose. The other large peak (539.9m/z) was accounted for by the presence of NAD required for RmlB activity. This was confirmed by a control assay containing only NAD and dTDP-glucose.

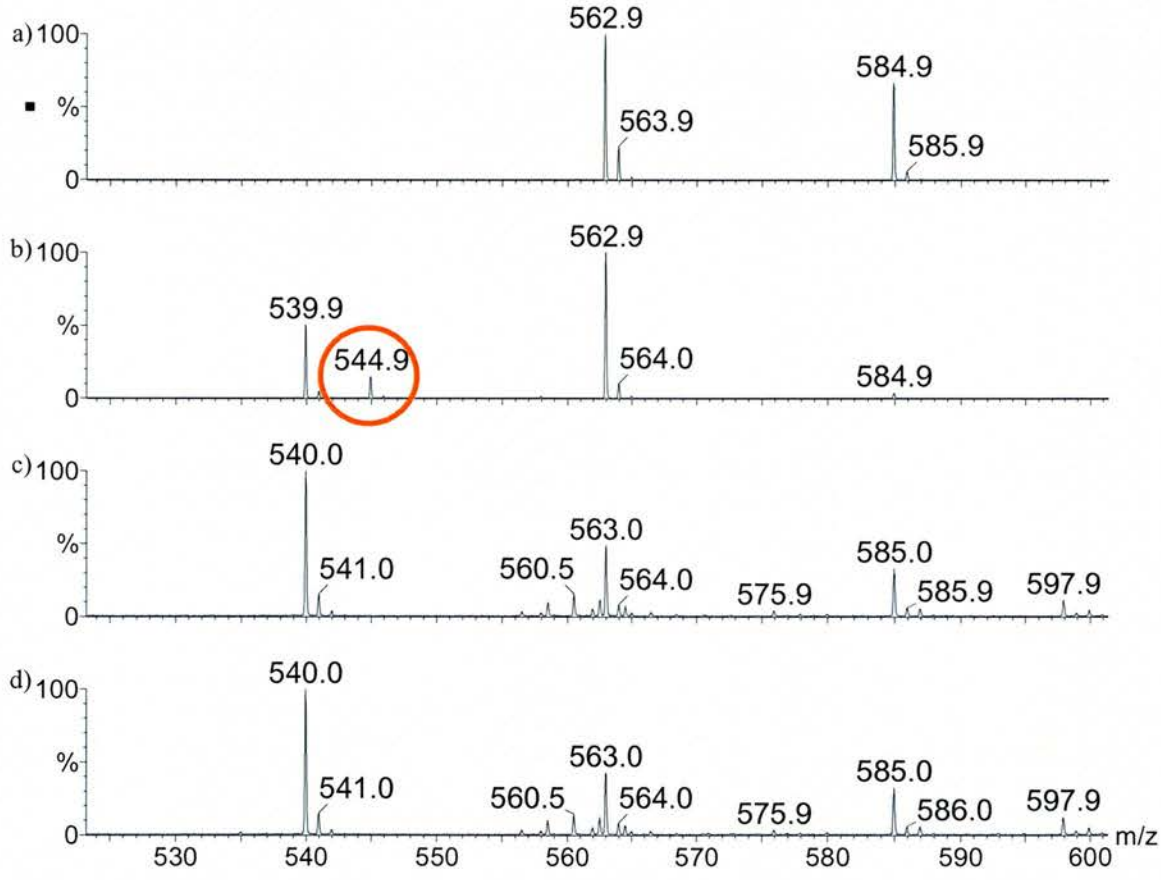


Figure 3.26 – ESI-TOF Mass Spectrometry Results for the EvaA Activity Assay: a) assay control – dTDP-D-glucose only, b) negative control – RmlB with NAD only, c) EvaA with FMN, d) EvaA without FMN.

The substrate peak (544.9m/z) disappeared in the presence or EvaA on its own or as its fusion protein form with MBP, confirming that the enzyme was active. It was not possible to account for the EvaA product as it was unstable and most likely to form dTDP-adducts (Prof. Rob Field, University of East Anglia, personal communication). Nevertheless, the enzyme did not seem to have an absolute requirement for FMN since samples with and without the co-factor displayed activity. This finding may be explained by one of several reasons: EvaA, displaying a Rossmann fold for nucleotide

binding, may be able to use NAD as a co-factor; it may have carried through enough FMN from the host cells during protein expression to be active, or its activity may depend on a different co-factor or substance, such as the presence of divalent metal ions. The presence of FMN in the EvaA sample is possible but unlikely since a spectrum of purified protein across the whole range (190-1000nm) was taken and did not result in a absorbance peaks typical for FMN (450nm (Barquera *et al.*, 2002)) or the reduced form of FMN (FMNH) (570nm (Riefler and Smets, 2002). Since the mass spectrometry assay was qualitative rather than quantitative turnover rates could not be determined. Therefore, it was not possible to judge changes in activity according to the co-factor present and EvaA may be able to use NAD as a co-factor to some extend.

Also, it was not possible to determine the substrate specificity of EvaA. The glucose peak was greatly reduced in the assays that did contain EvaA compared to the ones that contained only RmlB but again, the assay is qualitative not quantitative.

3.5 Conclusions and Discussion

There are various reasons why EvaA may not have been expressed solubly in *E. coli*. It is unlikely that the expressed protein was toxic to the host cells since they did not die or remain stagnant after protein expression had been induced. This was shown by the continued rise in absorbance at OD_{600}. To confirm that no EvaA at all had been expressed solubly, attempts to purify the His·tagged protein were made – these failed too. For this reason, we can assume that not even a small amount of soluble protein had been expressed from the vector not encoding protein fusions.

One reason could be differences in codon bias between the expression species *E. coli* and the species of origin *A. orientalis*. The EvaA gene contains 15 rare codons which may have prevented soluble protein expression. However, *E. coli* host strains, optimised for rare codon usage, such as BL21(*DE3*)codon+ and Rosetta(*DE3*), did not make any difference to the expression.

There may also have been problems with the DNA sequence around the start codon. To express genes of Actinomycetales or other non-*E. coli* species in *E. coli* host strains, some laboratories routinely mutate the first 20 bases to frequent codons (Berezovsky *et al.*, 1999). This is thought to aid ribosome binding: the ribosome contacts 30 nucleotides of mRNA (Steitz, 1969) and 34 nucleotides at and around the start codons (20 upstream and 11 downstream from the start codon) showing significant bias in nucleotide use and suggesting a role for ribosome binding (Schneider *et al.*, 1986). The level of translation is also affected by the sequence after the start codon, possibly due to efficiency of initiation (Martin-Farmer and Janssen, 1999; Strenström, *et al.*, 2001; Louise Major, University of Otago, PhD Thesis 2001). During the elongation of protein synthesis, errors may occur, one of which is the drop-off of peptidyl-tRNA from the ribosome (Heurgué-Hamard *et al.*, 1998; Karimi *et al.*, 1998). This free peptidyl-tRNA may be toxic to the cell for various reasons and is therefore often kept in inclusion bodies within the cell (Menez *et al.*, 2000).

These reasons may also explain why EvaA was expressed solubly as the fusion construct with MBP. The start codon and codon bias for *E. coli* MBP are optimised for expression. Drop-off of the peptidyl-tRNA from the ribosome is far less likely to occur once the ribosome has actually bound tightly; translation will then continue on from MBP through to the stop codon at the end of the EvaA gene.

A second problem with the project was the separation of EvaA from MBP once the fusion construct had been cleaved by TEV protease. It is not surprising that EvaA bound strongly to amylose and could not be separated from MBP in that way as EvaA functions *in vivo* as a sugar binding protein. The fact that the protein stuck to all other column resins is much harder to explain. The protein did not unfold when cleaved from the fusion construct as indicated by the continued presence of secondary and tertiary structure elements in the CD spectrum. However, it may have been possible that in the first round of purification, the steric bulk of MBP obscured EvaA binding to the amylose resin; following MBP removal, EvaA was able to bind. Interestingly, EvaA could be separated from MBP using the ANX Fast Flow Sepharose™, which is differentiated from the DEAE Fast Flow Sepharose™ by only one carbon unit in the linker region. This one carbon atom seems to have prevented the extremely tight binding of EvaA to the resin. The fact that the presence of FMN stabilises EvaA and allows concentration of the protein to 10mg ml^{-1} substantiates the above argument: FMN may block the binding site too. However, FMN did not purify together with EvaA when added to the protein preparation before the final purification step, indicating that it was not bound properly

The next problem with EvaA was the crystallisation of the fusion construct. There was no sensible way of telling if the whole protein had crystallised until data had been collected and the unit cell could be determined. There had been no reason to expect the two parts of the fusion protein to separate as it had been stable in solution for months as determined by SDS-PAGE. Acid hydrolysis was an unlikely cause as the crystallisation condition was mildly basic with a pH of 8.5. It was more likely that there had been some kind of protease present in the protein or crystallisation solution which may have caused the cleavage.

Yellow protein crystals for EvaA which had been pre-incubated with 0.5mM FMN have been obtained from several different protein preparations. Unfortunately, to date, optimisation of these has not resulted in diffracting crystals, the best result to date, has been the observation of powder diffraction patterns, typical for spherulites.

The gene from *S. violaceoruber* and *S. avermitilis* were successfully amplified from their respective genomes and ligated into T-tailed cloning vectors. All attempts of subcloning them into the pLou-3 vector failed as verified by DNA sequencing. The

other homologues from *A. mediterranei*, *S. antibioticus* and *S. nogalater* were expressed and purified in an identical fashion to EvaA. They behaved in exactly the same way throughout purification. Crystal screens have been set up; however, no protein crystals of these homologues have yet been obtained.

Bioinformatic studies using the amino acid sequence of EvaA and its homologues have pinpointed to the first domain being responsible for FMN binding and the second one for sugar binding and turnover. Secondary structures have been predicted but no Rossmann fold could be identified beyond doubts. To verify these results, determination of the three-dimensional crystal structure of EvaA is necessary.

Throughout the difficulties encountered with EvaA, there has always been the notion of refolding the protein. However, there is no guarantee that refolding would have worked; the main problem with it would have been to determine correct folding as no simple enzyme assay was available at that stage of the work. Without this verification, all crystal structures obtained would have been difficult to justify if there was uncertainty about the biological relevance of the protein in its particular state.

3.5 Further Work

Now, since we have EvaA to EvaD available to us, we can start determining an enzyme assays for the whole pathway, maybe using RmlD for detection of activity since soluble expression and purification of EvaE is still to be carried out to complete the L-epivancosamine biosynthesis pathway. Enzyme assays along the lines of EvaD and the rhamnose pathway enzymes need to be carried out for EvaA to test activity, since the result from mass spectrometry is qualitative rather than quantitative. However, it would mean carrying out a 6-enzyme assay; keeping all the right variable in excess would be a great challenge. A better approach would be developing a new assay which allows detection of EvaA activity in quantitative terms.

To understand the basis for enzyme activity, the crystal structure of EvaA has to be determined. Experimental phasing is required since no known folds are available for molecular replacement; therefore SeMet crystals as well as native crystals of reasonable quality have to be grown. Initial crystallisation conditions have been identified but these need optimisation possibly via another purification step to remove all traces of uncut MBP-EvaA fusion protein which may interfere with the order of the crystals. Maybe, the determination of novel crystallisation conditions to obtain diffracting crystals will prove to be the way forward.

For structure comparisons and to understand the selectivity of the different homologous enzymes, the homologues need to be crystallised and their structures determined.

Chapter 4

KDPG Aldolase from *E. coli*

Chapter 4: KDPG Aldolase

4.1 Summary

Synthetic organic chemists are increasingly turning to enzymes to make complex stereocentres, rather than using entirely traditional methods. However, although enzymes have exquisite regio- and stereo-control, they tend to display a limited substrate specificity, which has limited their usage by chemists. Thus, the ability to modify or design enzymes to perform custom synthesis has become a 'Holy Grail' of organic chemistry. For a rational engineering approach, understanding of the mechanism of action and the underlying causes of substrate selectivity need to be fully appreciated.

To fully understand the mechanism of, and the selectivity imposed on its substrate by, the KDPG aldolase from *E. coli*, crystallographic studies were carried out on wild-type and mutant protein. A way of understanding the mechanism is to trap substrate intermediates in the active site, as it allows the identification of residues involved in substrate binding and turn-over. However, as enzymes tend to turn their substrates over on a time-scale inappropriate for protein crystallography, studies using inhibitors and substrate analogues, as well as catalytically impaired enzymes, were carried out.

Co-crystallisation experiments failed to yield any high-quality crystals, thus soaking experiments on apo-crystals were carried out. X-ray diffraction data shows, however, that no substrate could be trapped in the active site of the enzyme in this way.

4.2 Introduction

Aldolases are subject of continuous interest owing to their central place in catalysing carbon-carbon bond formation in living organisms (Allen *et al.*, 1992; Breuer and Hauer, 2003).

Most aldolases are dimeric or tetrameric enzymes, but the KDPG aldolases from *E. coli* and from *P. putida* are homotrimers in solution. The structure of the KDPG aldolase from *P. putida* was determined to 2.8Å in 1982 (Mavridis *et al.*, 1982), but the active site residues in this model could not be identified. A new and revised model of this structure was published (Bell *et al.*, 2003), confirming the proposition by Wymer *et al.*, 2001 that the original structure had been traced incorrectly. This enzyme is closely related to its homologue from *E. coli*, they share 45% sequence identity. The crystal structure of the latter homologue in complex with pyruvate was determined by Allard *et al.*, 2001, and the active site residues were identified beyond doubt. The gross structure of the two homologues is very similar, and the rms deviation between the two Cα backbones is 1.5Å.

The KDPG aldolase is a key enzyme in the Entner Doudoroff pathway catalysing the reversible production of KDPG from G3P and pyruvate through a class I Schiff base mechanism (see Chapter 1.3, pages 31-37). The reaction proceeds by several ordered steps (Figure 4.1) (Allard *et al.*, 2001; Conway, 1992; Meloche *et al.*, 1975):

I. Schiff base (iminium ion) formation between the carbonyl pyruvate and the ε-amino group of an essential Lys residue (Lys133 in *E. coli*);
II. Enamine formation after proton abstraction at C3′in the iminium ion;
III. Enamine condensation with the carbonyl of G3P to form a second Schiff base;
IV. Hydrolysis of this second iminium ion to form KDPG (S-configuration at C4′) which is then released.

The enzyme displays a broad substrate specificity, making it a useful catalyst for stereocontrolled carbon-carbon bond formation between pyruvate and a range of unnatural aldehydes (Allen *et al.*, 1992; Machajewski and Wong, 2000; Shelton *et al.*, 1996). Despite this usefully broad substrate profile, the enzyme fails to convert electrophilic substrates that lack polar functionality at C2′, C3′ or C4′ and

nucleophilic substrates other than oxobutyrate or fluoropyruvate (Wymer *et al.*, 2001). Directed evolution resulted in the identification of a KDPG aldolase double mutant (K133Q/T161K) whose active site is relocated with a concurrent change in substrate specificity. The mutant was active against KDPG in the lactate dehydrogenase coupled retro-aldol assay of enzymatic activity. Although, in terms of enzymes with natural substrates, the values of k_{cat} are low ($1x10^{-8}$ sec^{-1}; K_M = 26.3mM), they are synthetically useful (Wymer *et al.*, 2001).

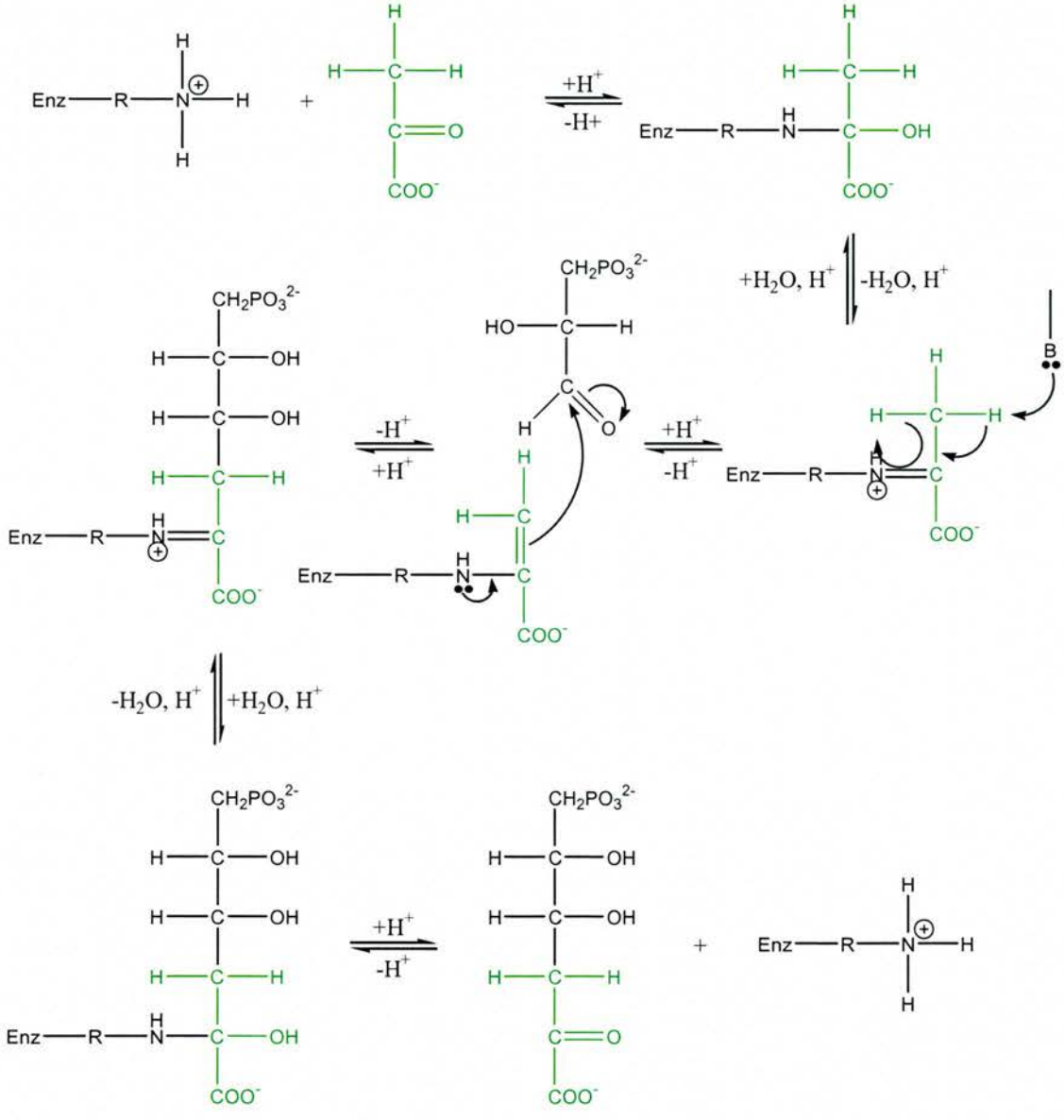


Figure 4.1 - Reaction Mechanism for the Aldol Condensation: between pyruvate and G3P. This is catalysed by the KDPG aldolase. The mechanism was proposed by Allard *et al.* in 2001 (figure taken from Allard *et al.*, 2001)

To date, substrate-bound co-complexes have been reported for aldolases from several sources such the transaldolase from *E. coli* and the human muscle aldolase (Dalby *et al.*, 1999; Dalby *et al.*, 2001; Jia *et al.*, 1997; Thorell *et al.*, 2002). Our mechanistic understanding was further advanced by the publications of a bound carbinolamine intermediate at the active site of the *E. coli* KDPG aldolase. The structure identifies Glu45 as the only plausible base, although the orientation of the carbinolamine precludes direct deprotonation by the residue. Because the structure was trapped by crystallisation at low pH, where deprotonation of the carbinolamine nitrogen inhibits imine formation, the relevance of this rotameric form to catalysis is unclear (Allard *et al.*, 2001). Another major advance towards an understanding of the aldolase mechanism came from a study of DERA. This work proposed a complex reaction mechanism involving the cyclic relay of protons between water and three residues during turnover (Heine *et al.*, 2001; Heine *et al.*, 2004). However, their proposed mechanism cannot be entirely general since two of the three residues (Asp102 and Lys201) are not conserved in other aldolases.

It has been shown that the accumulation of KDPG in bacteria is correlated with an immediate and significant decrease in growth (Fuhrman *et al.*, 1998), which suggests that this enzyme may be considered as a target for the development of new bacteriostatic and/or bactericidal drugs (Braga *et al.*, 2004). To date few inhibitors of KDPG aldolases have been developed with the exception of bromopyruvic acid (Meloche, 1967) and fluorodinitrobenzene (Barran and Wood, 1971). Other groups have considered the possibility of obtaining slow-binding inhibitors of this enzyme. These inhibitors were proposed to act through the stabilisation of reaction intermediates (iminium ion or enamine) within the enzyme by the use of pyruvate analogues bearing a β-dicarbonyl structure (Gefflaut *et al.*, 1996). The reaction intermediates may be stabilised through hydrogen bonding, as well as the equilibrium between several structures. A number of such analogues have been synthesised and their biochemical properties investigated (Braga *et al.*, 2004). The reaction scheme for the formation of the iminium ion and conjugated enamine, i.e. the mechanism, by which these compounds are thought to inhibit the KDPG aldolase from *E. coli*, are depicted in Figure 4.2.

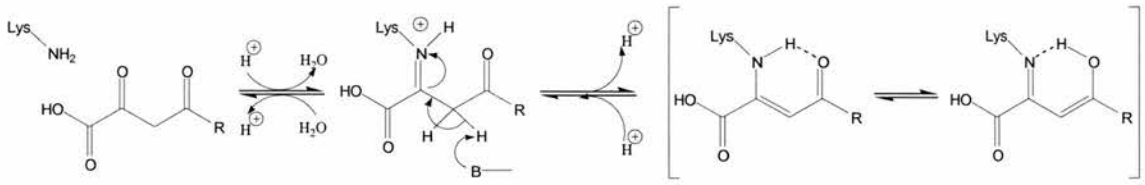


Figure 4.2 - Expected Mechanistic Pathway for the Inhibition: inhibition of the KDPG aldolase by synthetic compounds, such as 2,4-dioxo-4-[3-(trifluoromethyl)phenyl]butanoate (figure taken from Braga *et al.*, 2004.

Crystallisation of the KDPG aldolase with substrate and intermediate complex analogues, as well as inhibitors may further our understanding of the aldol and retro-aldol mechanisms by trapping substrate intermediates in the active site. For the same reason, the catalytically impaired mutant E45N, as well as two mutants with changed substrate and activity profiles, K133Q/T161K and KA3-L2 (containing the active mutations T84A/I92F/V118A/G141S/T105I) (Fong *et al.*, 2000), were attempted to be crystallised in the presence of substrate, substrate analogues and inhibitors.

Initial crystallisation conditions had been identified by Louise Buchanan (Buchanan, 2000; Buchanan *et al.*, 1999; Wymer *et al.*, 2001) but citrate contained in the precipitant solution bound tightly to the active site, making a displacement of this molecule by substrate analogues and inhibitors impossible. For this reason, new crystallisation conditions, which are more favourable to substrate binding, had to be identified.

4.3 Methods and Results

4.3.1 KDPG Aldolase Mutants

Three different mutants of the KDPG aldolase have been made (Figure 4.3); these display different catalytic properties and substrate profiles. Structural analysis of the three mutants may allow us to gain additional insight into the catalytic mechanism of the KDPG aldolase and to identify the basis of selectivity of the donor and acceptor substrates.

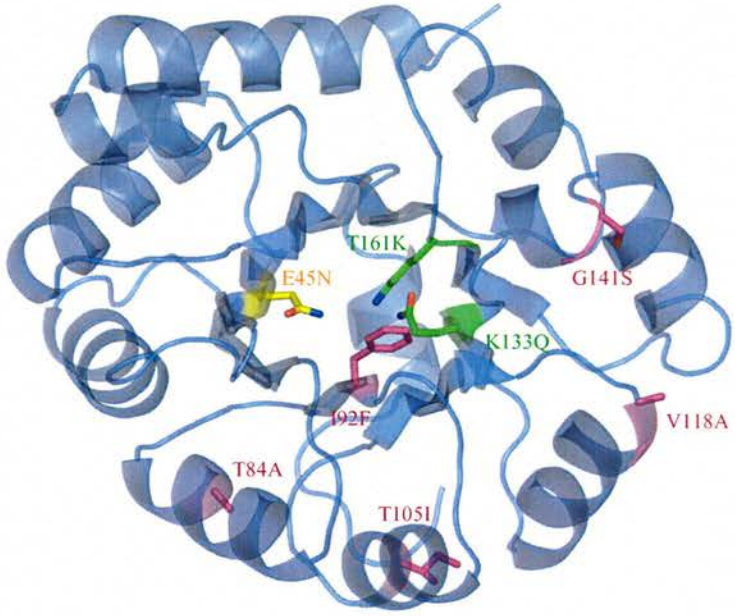


Figure 4.3 – KDPG Aldolase Mutants: the different mutations are shown as part of one KDPG aldolase monomer: yellow indicates the E45N, green the K133Q/T161K and pink the KA3-L2 (T84A/I92F/T105I/V118A/G141S) mutations.

E45N

Glu45 is thought to be the principal catalytic base, abstracting a proton from the imine form of the Schiff base, to result in the reactive enamine. Replacing the Glu by an Asn rendered it inactive; use of this inactivated KDPG aldolase may allow trapping of substrate intermediates in the active site.

K133Q/T161K

This mutant displayed a changed substrate profile: the wild-type enzyme failed to convert benzaldehyde to the corresponding aldol adduct at a rate greater than 0.01% that of G3P. In contrast, the mutant converted benzaldehyde with a k_{cat} of $1x10^{-8}sec^{-1}$ and a K_M of 26.3mM. Despite its k_{cat} being low compared to natural KDPG aldolase

activity, the mutant displays enhanced activity against pyridine carboxaldehydes, benzaldehyde and α-ketobutyrate. The single K133Q mutant was inactive, indicating that the Schiff base forming Lys was relocated to residue 161 of the double mutant. This resulted in a relocation of the active site Lys to strand β4 from β6, which was located on the other side of the principal base Glu45 changing the active site considerably, compared to the wild-type enzyme (Wymer *et al.*, 2001).

KA3-L2

This mutant was developed by directed evolution (Jaeger and Reetz, 2000; Nixon and Firestine, 2000) when screening for KDPG aldolase mutants with an increased activity against 2-keto-3-deoxy-galactonate (KDG) (Fong *et al.*, 2000). It incorporated the following amino acid changes: T84A, I92F, V118A, T105I and G141S. It displayed a 70-fold increased catalytic activity against KDG compared to wild-type (three-fold increase in k_{cat} and 23-fold reduction in K_M). Even though the mutant was 10-fold less active in consuming phosphorylated glyceraldehyde, it was 10-fold more active towards glyoxylate. This suggested to the authors (Fong *et al.*, 2000) that the change in activity may not have been the result of a change in the electrostatic environment in the binding site. Overall, the mutant was far less specific for the substrates it accepts; for example, it catalysed the addition of L-glyceraldehyde and methyl pyruvate to yield 3-deoxy-L-*threo*-2-hexulosonic acid in >90% diastereomeric excess (Figure 4.4). It was capable of accepting the D as well as the L stereoisomer of glyceraldehyde in its non-phosphorylated form.

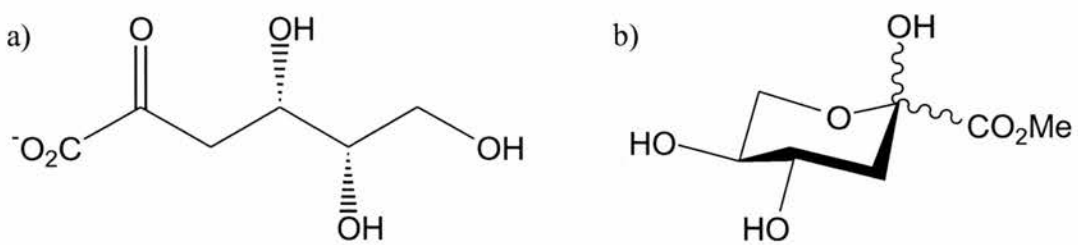


Figure 4.4 – Substrate and Product of the KA3-L2 Aldolase Mutant: a) KDG – readily accepted by KA3-L2; b) product of the addition of L-glyceraldehyde and methyl pyruvate: 3-deoxy-L-*threo*-2-hexulosonic acid.

4.3.2 Protein Expression and Purification

We were provided with the wild-type KDPG aldolase as well as its mutants E45N and K133Q/T161K cloned into the pET30b+ vector by our collaborators Eric Toone (Duke University, North Carolina, USA). The aldolase created by directed evolution (KA3-L2), cloned into pTrcHis2A was kindly provided by Chi-Huey Wong (The Scripps Research Institute, CA, USA). All vectors were used to transform chemically competent BL21(*DE3*) *E. coli* cells, which were then selected for, on antibiotic containing LB agar. Single colonies were picked, grown up in liquid medium and used for protein expression.

Expression was carried out in 500ml cultures of TB or LB broth containing 50μg ml^{-1} carbenicillin (for the proteins cloned into pET30b+) or 50μg ml^{-1} kanamycin (for the KDPG aldolase mutant PKA3-L2 expressed from pTrcHis2A). The 500ml cultures were inoculated with 500μl of overnight starter cultures grown in LB broth, containing the appropriate antibiotic. The cultures were incubated at 37°C, 200rpm shaking, until the OD_{600} of approximately 0.8 was reached. At this point, protein expression was induced using 1mM IPTG, and the incubation temperature was reduced to 25°C. After 8 hours the cells were harvested by centrifugation at 13,200*g* for 20min (Beckman centrifuge, Avanti J20-XP; JL8.100 rotor).

The cell pellet was re-suspended in buffer (100mM NaCl, 20mM Tris, pH7.5), and frozen overnight at -20°C. After thawing the sample, it was incubated on ice for one hour with 5mM PMSF, 2mM DTT, 100μg ml^{-1} lysozyme and 20μg ml^{-1} DNase. In addition to freeze thawing, the cells were lysed by sonication: five 30sec bursts of sonication at 18microns with 30sec intervals of cooling on ice were judged sufficient for complete cell lysis. Soluble and insoluble material were separated by centrifugation at 48,000*g* for 30min (Beckman centrifuge Avanti J20-XP, JA 25.5 rotor).

All KDPG aldolase enzymes (wild-type, E45N, K133Q/T161K, KA3-L2) were purified to single band purity in two liquid chromatography steps, making use of the vector-encoded N-terminal His·Tag for the first column. The soluble lysate was applied onto a metal chelating column (Ni-NTA™ metal chelating resin (Invitrogen), or Poros metal chelating resin (BioRad)), that had been charged with an excess of $NiSO_4$ and pre-equilibrated with buffer (200mM NaCl, 50mM Tris, pH7.5). After protein

application, the column was washed with 10CV equilibration buffer and the proteins eluted using a linear imidazole gradient (0 to 1.0M).

The eluent from the column was monitored at λ = 280nm and all protein-containing fractions were analysed by SDS-PAGE. The fractions containing the aldolase were concentrated and dialysed against 50mM Tris-HCl, pH7.5. The protein was then applied to a pre-equilibrated (50mM Tris-HCl, pH7.5) gel filtration column (Sephadex S200, Amersham Biosciences) which was run at a flow-rate of 1ml min^{-1} on a High Performance Liquid Chromatography (HPLC) BioCAD 700E workstation (GMI Global Medical Instrumentation) using the same buffer. Again, all eluent from the column was monitored at λ = 280nm and the peak collected, analysed by SDS-PAGE and then concentrated to between 10 and 30mg ml^{-1}, using a YM10,000 spin column (Amicon).

Use of the gel filtration column was necessary to allow the concentration of protein to above 3mg ml^{-1}. Even though the peak fraction did not seem any purer on a Coomassie™ Blue stained SDS gel. Without this column step the protein was found to precipitate at concentrations above approximately 3mg ml^{-1}. Crystallisation of protein at concentrations that low had proven to be problematic initially as all crystallisation drops were clear indicating too low a concentration of protein or precipitant. Figure 4.5 gives and overview of the purification stages.

The integrity and identity of the proteins was confirmed by trypsin digest MALDI-TOF mass spectrometry.

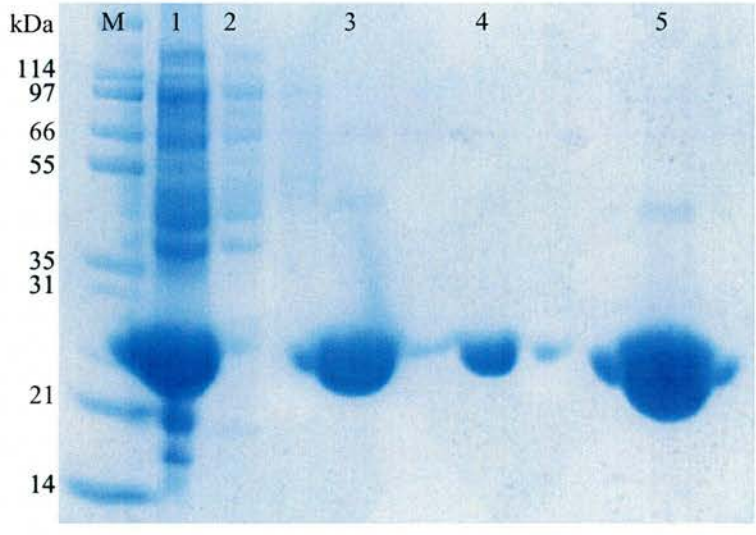


Figure 4.5 – Example of the Purification of the KDPG Aldolase: Purification steps for the wild-type aldolase:
M –Molecular Weight Marker (BioRad)
1 – whole cell lysate
2 – Ni column flow-through
3 – Ni-column peak
4 – Gel filtration peak
5 – Concentrated gel filtration peak ready for crystal trials.

4.3.3 Inhibitors, Substrates and Analogues

KDPG, G3P and pyruvate were commercially obtained from Sigma-Aldrich. All substrate analogues and inhibitors were kindly provided by our collaborators. The following compounds were used for crystallisation trials and are depicted in figure 4.6:

- KDPG
- G3P
- Pyruvate
- 2-oxo-4-pyridin-2-yl-butanoate (**1**)
- 4-hydroxy-2-oxo-octanoate (**2**)
- 3-hydroxy-2-oxo-4-pyridin-2-yl-butanoate (**3**)

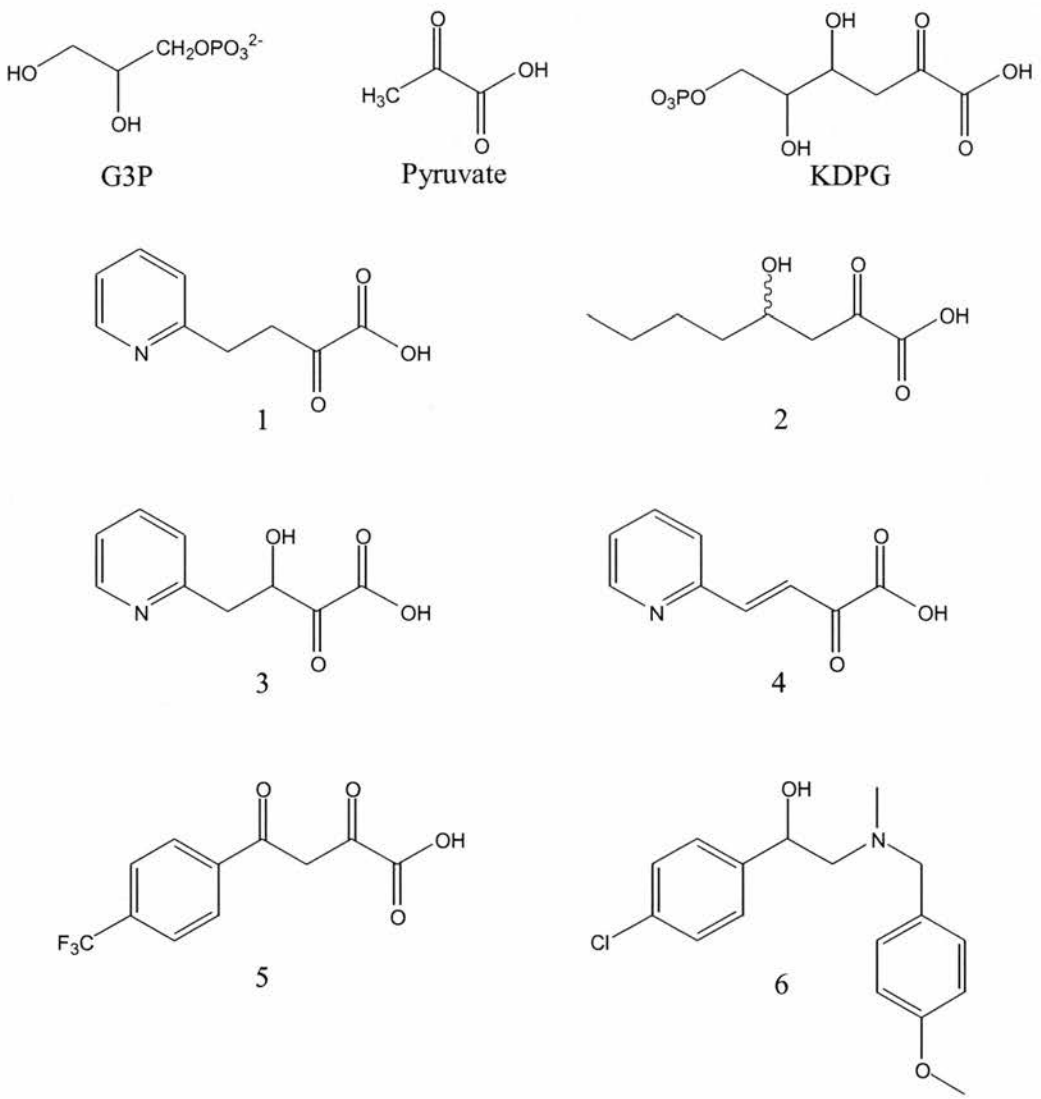


Figure 4.6 – Substrates and Inhibitors: compounds used for co-crystallisation trials and substrate soaks with the wild-type and mutant aldolase enzymes. G3P, pyruvate and KDPG are the natural substrates and products, the other compounds are inhibitors and substrate analogues.

- 2-oxo4-pyridin2-yl-but-3-enoic acid (**4**)
- 2,4-dioxo-4-[3-(trifluoromethyl)phenyl]butanoate (**5**)
- 1-(4-chloro-phenyl)-2-[(4-methoxy-benzyl)-methyl-amino]-ethanol (**6**)

Compound **5** is a so-called slow-binding inhibitor which is thought to inhibit the KDPG aldolase through the stabilisation of reaction intermediates (iminium or enamine) within the enzyme by the use of pyruvate analogues bearing a β-dicarbonyl structure (Gefflaut *et al.*, 1996). The stabilisation of the intermediate is possible through hydrogen bonding and equilibrium between several structures (Braga *et al.*, 2004) (see section 4.2 and Figure 4.2).

4.3.4 Crystallisation

Crystallisation conditions had been identified in our laboratory by Louise Buchanan: 20% PEG6,000, 0.075M citrate, pH4.0, protein concentration 3mg ml^{-1}, sitting drop vapour diffusion, 20°C. Structure determination (2.2Å resolution, R_{factor} 20%) using crystals grown under these conditions showed that citrate was bound to the active site of the enzyme (Buchanan *et al.*, 1999). Citrate preferentially bound to the active site which prevented soaking and co-crystallisation experiments with inhibitors and substrate analogues. For this reason, new crystallisation conditions, not containing any citrate, had to be identified.

A better starting condition was thought to be the one identified in the publication by Allard *et al.* 2001. This group managed to trap a covalently bound pyruvate intermediate in the active site of the aldolase in their crystals, when employing substrate soaking experiments using low pH (1.95Å resolution, R_{factor} 20%). These crystals had been obtained using the hanging drop vapour diffusion method, after equilibration of the protein against 10mM Tris-HCl, pH8.0, in 18.5% PEG3500, 0.2M Ammonium sulfate and 0.1M Sodium acetate. We reproduced crystals grown from these conditions at different protein concentrations; the conditions were optimised by alternatively changing PEG molecular weight, PEG and salt concentrations, pH and the buffer, as well as, using the three Hampton Additive Screens.

New sparse matrix crystallisation trials using the sitting drop vapour diffusion method were carried out at 4°C and 20°C using protein at concentrations of 1mg ml^{-1}, 2mg ml^{-1}, 5mg ml^{-1}, 7mg ml^{-1}, 10mg ml^{-1}, 15mg ml^{-1}, 20mg ml^{-1}, 25mg ml^{-1} and 30mg ml^{-1}. The following crystallisation screens were used: Crystal Screens 1 and 2, Index Screens 1 and 2, SaltRX Screen, PEG/Ion Screen (all Hampton Research); Wizard Screens 1 and 2 (Emerald BioStructures) and Crystallisation Screens 1 and 2 (Sigma-Aldrich). Initial hits were optimised by changing protein and precipitant concentrations and pH. Additive Screens 1, 2 and 3 (Hampton Research) were also employed to improve crystal quality.

Identical crystallisation trials were set up for the wild-type protein and the mutants described above. Crystals obtained from these conditions were used for soaking experiments. Table 4.1 details the conditions under which crystals were grown.

Table 4.1: KDPG Aldolase Crystallisation Conditions:

Protein Conc.	Crystallisation Conditions Salt	PEG	Buffer	Advantages	Drawbacks
1-10mg ml^{-1}	None	10-25% PEG 6000	75mM Citrate pH4.6	2.5Å resolution	Citrate present
1-20mg ml^{-1}	0.2M $AmSO_4$ 0.1M NaAc	15-20% PEG 3350	10mM Tris, pH8.0	2.5Å resolution	Sulfate ions
1-5mg ml^{-1}	0.1M $AmSO_4$	10-20% PEG 3350	20mM Tris, pH 7.5		Sulfate ions
1-20mg ml^{-1}	0.2M $AmSO_4$ 0.1M NaAc	15-20% PEG 4000	10mM Tris, pH8.0	2.5Å resolution	Sulfate ions
1-5mg ml^{-1}	0.1M $AmSO_4$	10-20% PEG 4000	20mM Tris, pH 7.5		Sulfate ions
5-10mg ml^{-1}	0.2M KH_2PO_4	10-20% PEG8000	20mM NaAc, pH5.6	3.0Å resolution	Phosphate ions
5-10mg ml^{-1}	0.2M K_2HPO_4	10-20% PEG8000	20mM NaAc, pH5.6		Phosphate ions
10-20mg ml^{-1}	0.1M KSCN	30% PEG MME2k	None	No phosphate or sulfate	10Å resolution
10-20mg ml^{-1}	0.15M KBr	30% PEG MME2k	None	No phosphate or sulfate	10Å resolution
10-20mg ml^{-1}	0.15M KBr	30% PEG MME2k	10mM Tris, pH7.5	No phosphate or sulfate	10Å resolution

Co-crystallisation experiments were carried out using mainly the crystal conditions identified for the apo-protein, but also a number of sparse matrix crystallisation screens. The aldolases were pre-incubated for 2 to 12 hours with inhibitors **1** to **6** (Figure 4.6) as well as KDPG, G3P and pyruvate at 2mM, 10mM and 50mM concentrations. However, no crystals could be grown in this way.

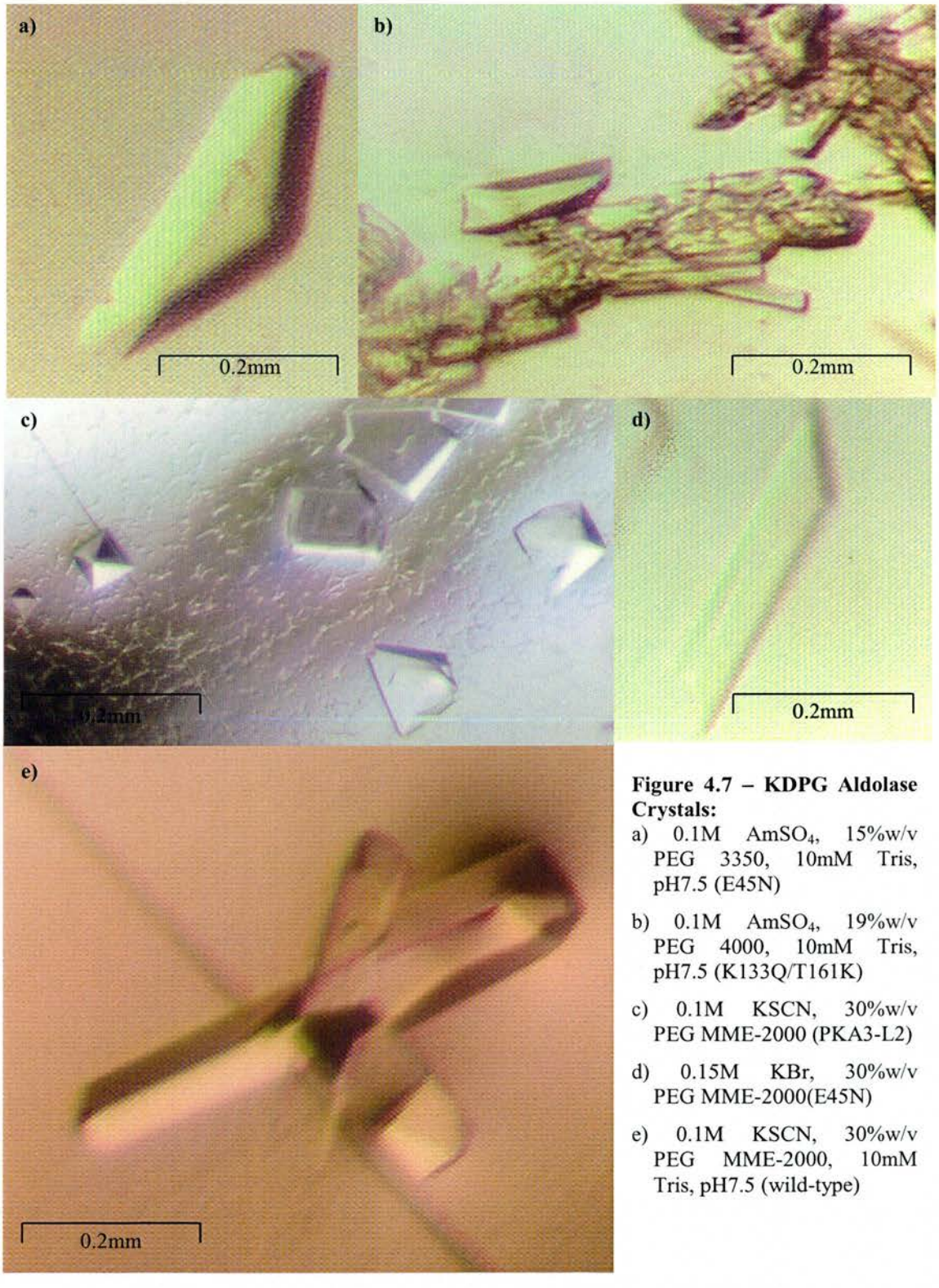


Figure 4.7 – KDPG Aldolase Crystals:

a) 0.1M $AmSO_4$, 15%w/v PEG 3350, 10mM Tris, pH7.5 (E45N)

b) 0.1M $AmSO_4$, 19%w/v PEG 4000, 10mM Tris, pH7.5 (K133Q/T161K)

c) 0.1M KSCN, 30%w/v PEG MME-2000 (PKA3-L2)

d) 0.15M KBr, 30%w/v PEG MME-2000(E45N)

e) 0.1M KSCN, 30%w/v PEG MME-2000, 10mM Tris, pH7.5 (wild-type)

In addition to the problems with identifying conditions that had no phosphate or sulfate ions present, there was also a significant variability of crystal quality. The same well gave a number of crystals, some of which diffracted; others did not go below 15Å. This problem had been described by Louise Buchanan and could not be solved in the crystallisation attempts detailed above. The presence of additive did not improve crystal quality and temperature change did not seem to make a difference either.

4.3.5 Soaking Experiments

Prior to data collection, apo-crystals of the aldolase were equilibrated in cryo-protectant solution made up of 70 to 80% mother liquor containing 30%v/v MPD or 20%v/v glycerol and 10 to 50mM inhibitor, substrate or substrate analogues. Disintegration of the crystal in the cryo-protectant solution was used as an indication for a change in the protein which may have been binding; this process was timed and prior to data collection, the crystals from the same well were soaked for a little less time (1 to 10min in total) in order to obtain high-quality crystallisation data with soaked compound present.

4.3.6 Data Collection and Processing of E45N with KDPG

After a 0.5 x 0.5 x 0.2mm apo-crystal of the E45N mutant was obtained from the following precipitant conditions: 0.1M Ammonium sulfate, 15%w/v PEG4000, 10mM Tris-HCl, pH8.0, (protein concentration 20mg ml^{-1}, sitting drop vapour diffusion method). It was soaked in cryo-protectant solution (20%v/v glycerol in mother liquor), containing 30mM KDPG (as described above) and then mounted in a cryo-loop and flash-frozen in a nitrogen gas stream. Data were collected in-house as 180 non-overlapping 10min 1° oscillations. The X-ray beam λ = 1.542Å was produced by a rotating copper anode; the detector was a Rigaku RaxisIV++ image plate and the crystal to detector distance was 240mm.

A data set to 2.8Å resolution, consisting of 168,016 measurements, of which 15,069 were unique reflections, was collected in 5 separate batches. These were indexed in a primitive hexagonal Bravais lattice and integrated in Laue group P321 (no. 150)

using MOSFLM, version 6.2.4 (Leslie, 1992). The batches were indexed, merged together and scaled using Scala, version 3.2.1 (Evans, 1993), as implemented in the CCP4 Program Suite (Bailey, 1994). Using the unit cell dimensions a = b = 105.2Å, c = 51.4Å, $\alpha = \beta = 90°$, $\gamma = 120°$ allowed the calculation of the Matthews volume (Matthews, 1968) of the crystal and therefore the solvent content. The V_m of the crystal was 3.6 and the solvent content 65.3%, assuming one monomer in the au and a molecular mass of 21,000Da. Table 4.2 details the data collection statistics for the individual resolution shells, and Table 4.3 gives an overview of the complete data collection statistics.

Table 4.2 - Data Collection Details for Individual Resolution Shells for the KDPG Aldolase E45N Mutant (KDPG soak)

Resolution Shell (Å)	Measured Reflections	Completeness (%)	Multiplicity	R_{merge} [a]	I/sigma	Mn(I)/sd
45.55 – 12.52	169	94.6	1.7	0.031	17.1	25.2
12.52 – 8.85	306	95.9	1.8	0.028	19.5	25.9
8.85 – 7.23	386	94.6	1.8	0.029	17.8	23.1
7.23 – 6.26	483	95.2	1.8	0.035	17.1	20.3
6.26 – 5.60	516	94.5	1.8	0.037	18.0	17.2
5.60 – 5.11	582	95.0	1.8	0.036	17.7	17.2
5.11 – 4.73	636	96.5	1.8	0.038	16.9	16.8
4.73 – 4.43	717	97.0	1.7	0.041	15.5	16.7
4.43 – 4.17	722	96.4	1.7	0.043	16.5	14.9
4.17 – 3.96	807	97.7	1.7	0.052	12.8	13.2
3.96 – 3.78	824	97.5	1.7	0.078	4.5	10.7
3.78 – 3.61	860	96.8	1.7	0.106	6.9	8.7
3.61 – 3.47	925	97.2	1.8	0.130	5.7	5.8
3.47 – 3.35	879	95.2	1.7	0.162	3.8	5.2
3.35 – 3.23	1,001	96.8	1.7	0.174	4.3	4.4
3.23 – 3.13	998	95.3	1.7	0.210	3.6	3.5
3.13 – 3.04	1,011	96.9	1.8	0.244	3.1	3.1
3.04 – 2.95	1,066	95.1	1.7	0.325	2.3	2.4
2.95 – 2.87	1,080	96.9	1.7	0.379	2.0	1.9
2.87 – 2.80	1,101	93.8	1.7	0.461	1.6	1.5
45.55 – 2.80	**15,069**	**96.1**	**1.7**	**0.062**	**8.5**	**9.1**

[a] $R_{merge} = \Sigma\Sigma I(h)j - (I(h)) / \Sigma\Sigma I(h)j$ where I(h) is the measured diffraction intensity and the summation includes all observations

Table 4.3 - Data Collection Summary for the KDPG Aldolase E45N Mutant (KDPG soak)	
Wavelength (Å)	1.542 (in-house Cu-anode)
Resolution (Å)	45.55 – 2.80
Space Group	P321 (no. 150)
Unit-cell Parameters (Å, °)	a = b = 105.2Å, c = 51.4Å α = β = 90°, γ = 120°
V_m ($Å^3$ Da^{-1}) (1mol per au)	3.6
Percentage solvent	65.3
Total measurements	168,016
Unique reflections	15,069
I/σ	8.5 (1.6)
Average redundancy	1.7 (1.7)
Data completeness (%)	96.1 (93.8)
R_{merge}[a] (%)	0.062 (0.461)
Wilson B factor ($Å^2$)	70.36

[a] R_{merge} = ΣΣ I(h)j – (I(h)) / ΣΣ I(h)j where I(h) is the measured diffraction intensity and the summation includes all observations
Parentheses denote data for the highest resolution shell

4.3.7 Structure Solution, Refinement and Validation

One monomer of the wild-type KDPG aldolase from *E. coli* (PDB entry: 1FQ0) was used to calculate starting phases by molecular replacement in MolRep (Vagin and Teplyakov, 1997) All data between 15 and 4.0Å were used, expecting a 0.99 fraction completeness of the model with a 0.99 fraction similarity to the input structure. The solution displayed a correlation coefficient of 0.637 and an R_{factor} of 0.368. The fractional coordinates and Eularian angles identified was used by the program to automatically calculate re-orientated model coordinates.

Prior to all refinement steps 5% of all data were excluded for cross-validation using the program uniquify (Brünger, 1992; Brünger, 1997) as implemented in the CCP4 Program Suite (Bailey, 1994). Initial structure refinement was carried out in Refmac, version 5.0 (Murshudov *et al.*, 1997); using the rigid body refinement option decreased the R_{factor} and R_{free} to 20.2% and 25.3% respectively. Single and double difference electron density maps (F_o-F_c and $2F_o$-F_c) were calculated using FFT (Immirzi, 1966; Read and Schierbeek, 1988; TenEyck, 1973) which were then displayed in O, version 7.0 together with the improved model. Cycles of checking the refined model in O and subsequently using the restrained refinement option in Refmac improved the model further, indicated by the concurrent decrease in R_{factor} and R_{free}.

Further rounds of TLS refinement (Howlin *et al.*, 1989; Winn *et al.*, 2001) in Refmac, 5.0, were employed to further improve the model. One sulfate molecule could be located in the model but no KDPG or turnover product could be seen in the electron density maps. 103 water molecules were added to the model using ARP/wARP (Morris *et al.*, 2002), as part of Refmac, version 5.0, which is in line with normal expectations for a model at this resolution. The R_{factor} and R_{free} of the fully refined model were 17.9% and 24.6% respectively; its PDB accession number is 1WAU.

The refinement weights during restrained model refinement were adjusted to 0.29 to achieve an ideal balance between weighting terms for model geometry and crystallographic data. The final bond length rmsd was 0.019Å and the rmsd bond angle was 1.832°. Validating the model in Procheck (Laskowski *et al.*, 1993) showed that 88.7% of all side chains were contained in the Ramachandran core region (Ramachandran and Sasisekharan, 1968), which increased to 99.4% and 100% when including the allowed and generously allowed regions (Figure 4.8).

All other statistics for the model, such as main chain bond lengths and bond angles, residue planarity, etc. were within or better than the normal limits observed for protein structures. Table 4.4 summarises the refinement and validation statistics for the E45N model.

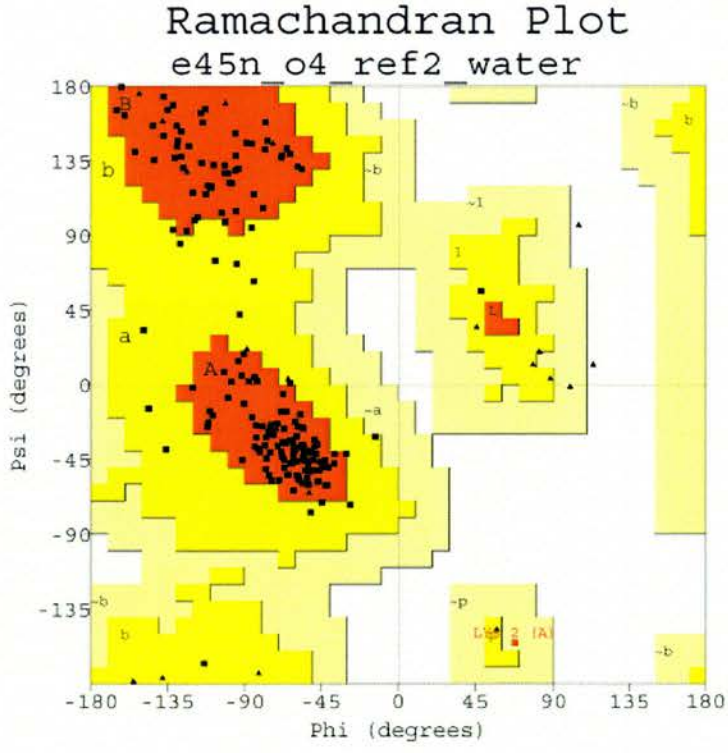


Figure 4.8 – Ramachandran Plot for the KDPG Aldolase E45N Mutant.

Table 4.4 - Refinement and Validation Statistics for the KDPG Aldolase E45N Mutant	
R_{factor}	17.9
R_{free}	24.6
Rmsd bonds (Å) / angles (°)	0.019 / 1.832
B factor deviation bonds/angles (Å^2/°)	
Main chain	0.741 / 1.342
Side chain	2.104 / 3.598
% Residues in Ramachandran core[a]	88.7
% Residues in additionally / generously allowed region	10.7 / 0.6
Protein atoms	1668
Water molecules	103
Sulfate molecules	1
Average B factor (Å^2)	63.54
PDB Accession code	1WAU

[a] Ramachandran core refers to the most favoured region in the ϕ/ψ Ramachandran plot as defined by Laskowski *et al.*, 1993.

4.3.8 Structure of E45N

The KDPG aldolase from *E. coli* is a homotrimer made up of three $(\alpha/\beta)_8$ monomers. Figure 4.9 shows the trimer as calculated from the symmetry related objects of the E45N mutant crystallised in space group P321. The trimeric unit resembles a ship's propeller, with identical monomers approximating elongated ellipsoids.

The mutant is unchanged from our earlier crystallisation and soaking attempts (Buchanan et al. 1999, Wymer et al., 2001, Louise Buchanan University of St. Andrews PhD Thesis 2000; Fullerton *et al.*, 2004, submitted) the active site is freely accessible to solvent, indicating why the presence of sulfate or phosphate ions in the crystallisation condition is detrimental to soaking experiments.

The KDPG aldolase is essentially identical to the ones from *P. putida* and *Thermotoga maritima*, the rms deviation for the Cα backbone was 0.96Å and 1.5Å respectively for the two enzymes and the active site Lys (Lys133) was located inside the barrel on β6. The rms deviation between wild-type and E45N mutant KDPG aldolase from *E. coli* was between 0.38Å and 0.78Å, depending on the wild-type monomer used, indicating that the mutant enzyme had not changed significantly compared to the native one. The density around the active site Lys was slightly disordered but did not contain convincing evidence for substrate, substrate intermediate or product (Figure 4.10).

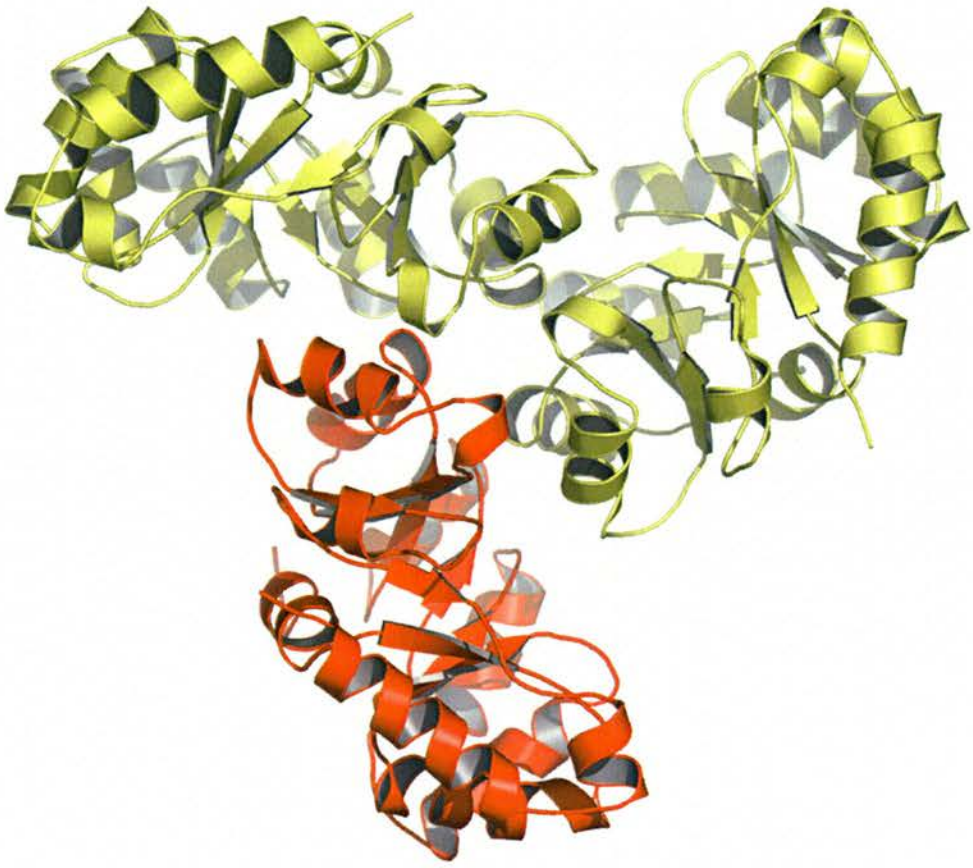


Figure 4.9 – KDPG Aldolase Trimer: the red monomer was the one determined from the crystals of the E45N mutant. The two yellow monomers were calculated as the symmetry related objects according to the P321 crystal symmetry.

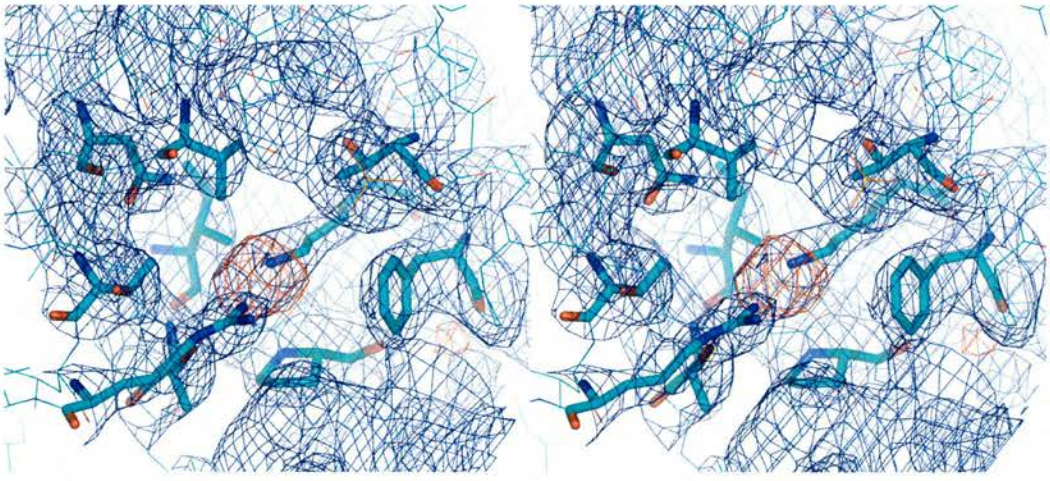


Figure 4.10 – Stereoimage of the Active Site: the single difference electron density (F_o-F_c) around Lys133 is contoured at 3.0 sigma and shown in red; the double difference electron density ($2F_o$-F_c) is contoured at 1.0 sigma and shown in blue.

One sulfate molecule (probably from the crystallisation condition) was found to be bound to Ser186 (Figure 4.11) which is however removed from the regions which is thought to be the active site.

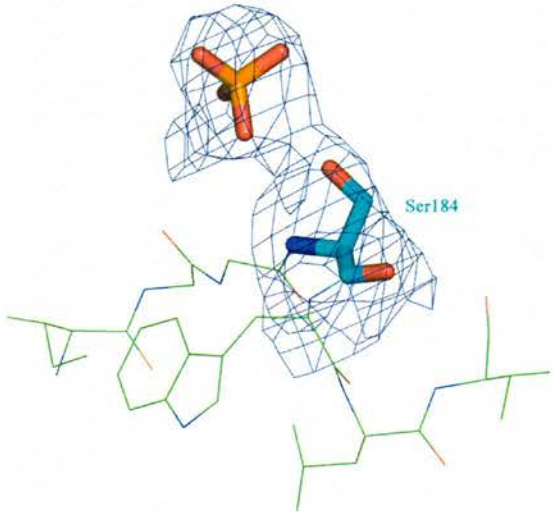


Figure 4.11 – Sulfate Molecule Bound to Ser184: the double difference electron density is contoured at 1.0 sigma. The surrounding residues are indicated to show the position of the main chain.

The ultra-high resolution structure of DERA was solved recently and a catalytic mechanism proposed. The active site Lys (133 in KDPG aldolase) is required to be uncharged to act as the nucleophile on the substrate, since the pKa of Lys in aqueous solution is 10.5 it needs to be significantly perturbed to be catalytically active. In DERA a second Lys (201) is in close proximity to the Schiff-base forming residue which decreases its pKa through electrostatic interactions, especially within a low dielectric environment (Schmidt and Westheimer, 1971; Westheimer, 1995). Heine *et al.*, 2004 proposed this mechanism to be conserved throughout class I aldolases, all of which contain either an Arg or a Lys in close proximity to the Schiff-base forming Lys. An anomaly seemed to be the KDPG Aldolase from *E. coli* since the distance between NH2 of Arg49 and NZ of Lys133 was more than 10Å. However, this could be explained by the presence of citrate in the active site which may have perturbed the side chains. The new crystal form, which does not contain any molecules in the active site, shows the distance between the NH2 of Arg49 and the NZ of Lys133 to be 4.0Å, which is reasonable for electrostatic interactions (Figure 4.12). This would promote an appropriate pKa for the Lys and cause it to be uncharged and act as the nucleophile in the reaction.

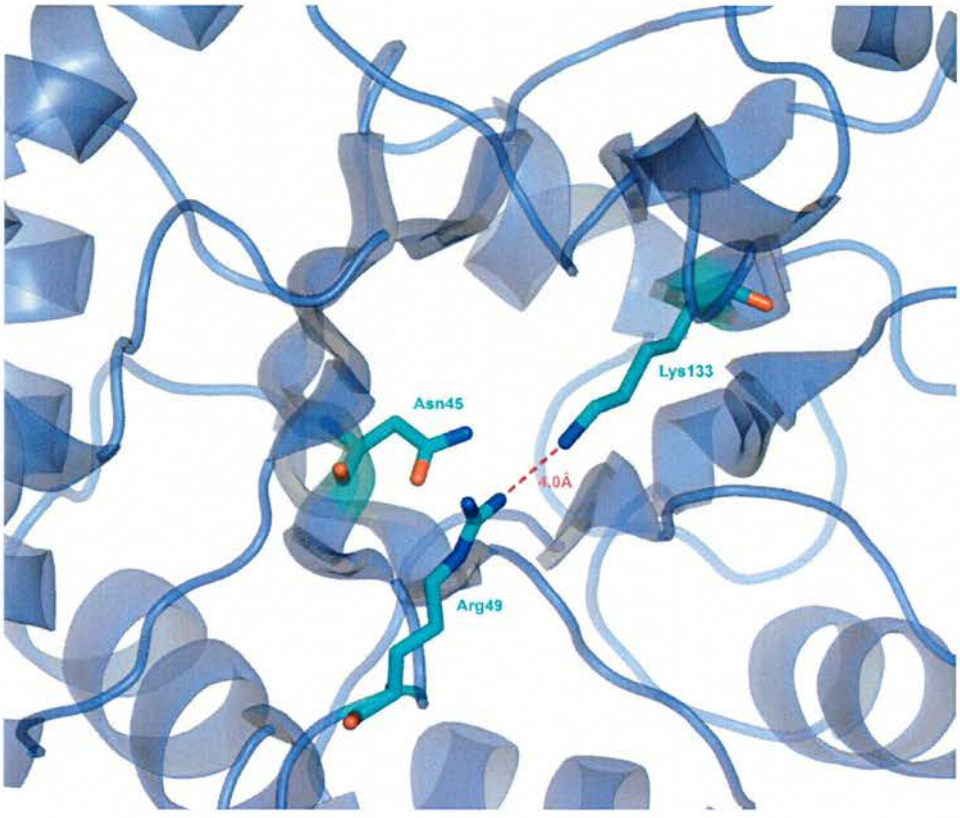


Figure 4.12 – Electrostatic Interactions: between NH2 of Arg49 and NZ of Lys133, which is buried within the beta barrel. In this structure the side chains are not perturbed by citrate present in the active site.

4.3.9 Data Collection and Processing of KA3-L2 with G3P

Apo-crystals of the KA3-L2 mutant had been obtained under the following conditions: 0.1M Ammonium sulfate, 15%w/v PEG3350 and 20mM Tris-HCl, pH7.5 using the sitting drop vapour diffusion method, a protein concentration of 15mg ml^{-1} and an incubation temperature of 20°C. Prior to data collection, the crystals were soaked in 10mM G3P for 20min, cryo-protected using 30%v/v MPD in mother liquor, mounted in a cryo-loop and flash-frozen in a nitrogen gas stream. X-ray data to 2.7Å resolution were collected as 170 non-overlapping 10min 1° oscillations. A total of 165,965 reflections, of which 15,425 were unique, were collected on the same in-house X-ray system as above, and the detector to crystal distance was 200mm. The data were

integrated, indexed and scaled in a primitive orthorhombic Bravais lattice, using MOSFLM, version 6.2.4 (Leslie, 1992) and Scala, version 3.2.1 (Evans, 1993). The space group was identified as $P2_122_1$ (space group no. 318) by the observation of n = 2 systematic absences at the h and l axes but not at the k-axis. The data were reindexed using Scala, version 3.1.2 to allow the expression of the space group in the more conventional format, $P2_12_12$ (space group no. 18). Table 4.5 details the data collection statistics for the individual resolution shells. Using the unit cell dimensions, a = 54.0Å, b = 73.0Å, c = 138.0Å, $\alpha = \beta = \gamma = 90°$, V_m was calculated to be 2.2 and the solvent content 42.6% assuming three molecules per au to allow the formation of the biologically relevant trimer.

Table 4.5 - Data Collection Details for Individual Resolution Shells for the KDPG Aldolase PKA3-L2 Mutant (G3P soak) the Primitive Orthorhombic Bravais Lattice

Resolution Shell (Å)	Measured Reflections	Completeness (%)	Multiplicity	R_{merge} [a]	I/sigma	Mn(I)/sd
39.07 – 12.07	201	94.7	2.7	0.069	7.9	25.1
12.07 – 8.54	357	99.3	3.0	0.065	7.6	28.1
8.54 – 6.97	435	99.9	3.1	0.068	7.2	23.7
6.97 – 6.04	505	99.8	3.1	0.072	8.2	18.3
6.04 – 5.40	578	100.0	3.1	0.079	7.8	16.7
5.40 – 4.93	627	99.9	3.2	0.082	7.2	16.8
4.93 – 4.56	679	100.0	3.2	0.081	7.4	18.2
4.56 – 4.27	733	99.9	3.1	0.087	6.7	17.2
4.27 – 4.02	764	99.9	3.2	0.094	6.9	16.6
4.02 – 3.82	798	99.9	3.1	0.106	6.4	14.4
3.82 – 3.64	835	99.9	3.1	0.118	5.8	12.5
3.64 – 3.49	901	99.9	3.1	0.133	5.0	10.4
3.49 – 3.35	900	99.7	3.0	0.148	4.7	8.4
3.35 – 3.23	955	99.6	3.0	0.163	4.2	7.5
3.23 – 3.12	971	99.4	3.0	0.185	3.9	6.0
3.12 – 3.02	997	98.6	2.9	0.212	3.4	5.0
3.02 – 2.93	1,025	97.9	2.8	0.228	3.2	4.1
2.93 – 2.85	1,021	96.9	2.8	0.253	2.9	3.4
2.85 – 2.77	1,092	96.2	2.8	0.287	2.6	2.8
2.77 – 2.70	1,051	95.2	2.8	0.328	2.3	2.3
38.93 – 2.70	**15,425**	**98.7**	**3.0**	**0.114**	**5.2**	**10.5**

[a] $R_{merge} = \Sigma\Sigma I(h)j - (I(h)) / \Sigma\Sigma I(h)j$ where I(h) is the measured diffraction intensity and the summation includes all observations

Molecular replacement was attempted in MolRep (Vagin and Teplyakov, 1997), AMoRe (Navaza, 1994) and Phaser (Read, 2003; Storoni *et al.*, 2004) using either a monomer or the trimer of the wild-type KDPG Aldolase structure from *E. coli* as search model. All data between 15.0 and 4.0Å were used, expecting a 0.99 fraction completeness of the model with 0.99 fraction similarity to the input structure. Despite the identification of the full space group all possible space groups, e.g. no. 16 to no. 19.

The atomic coordinates were calculated for all 'hits' and the output models refined by 10 cycles of rigid body refinement in Refmac, 5.0 to verify the molecular replacement output models. Since the R_{factor} failed to drop below 50% in the 10 cycles of rigid body refinement, the structures could not be assumed to be the correct ones. To check that this was not caused by local minima and Refmac's relatively small radius of convergence, the simulated annealing option (Adams *et al.*, 1997; Brünger *et al.*, 1990; Brünger *et al.*, 1987) in CNS (Crystallography and NMR System) (Brünger *et al.*, 1998) was employed, but did not solve the problem.

Molecular replacement of a mutant should have been straight forward; therefore, the problem encountered was attributed to identification of an incorrect space group. Consequently the data were re-processed in the lower symmetry Bravais lattice primitive monoclinic. A total of 171,331 reflections, of which 26,484 were unique were identified for the $P2_1$ space group identified as a possibility in MOSFLM, version 6.2.4. The unit cell dimensions were as follows: a = 54.2Å, b = 138.6Å, c = 73.1Å, $\alpha = \gamma =$ 90°, b = 90.01°. The R_{merge}, as calculated during scaling in Scala, version 3.2.1 was 0.081; Table 4.6 details the data collection statistics for the individual resolution shells, when processing, indexing, merging and scaling the data in $P2_1$.

Table 4.6 - Data Collection Details for Individual Resolution Shells for the KDPG Aldolase PKA3-L2 Mutant (G3P soak) in the Primitive Monoclinic Bravais Lattice:

Resolution Shell (Å)	Measured Reflections	Completeness (%)	Multiplicity	R_{merge}[a]	I/sigma	Mn(I)/sd
39.07 – 12.07	296	87.8	1.8	0.046	11.0	24.6
12.07 – 8.54	570	91.6	1.8	0.042	12.4	26.2
8.54 – 6.97	730	93.7	1.8	0.044	11.8	21.3
6.97 – 6.04	864	93.5	1.8	0.050	11.5	16.1
6.04 – 5.40	1,002	95.0	1.8	0.058	10.5	14.6
5.40 – 4.93	1,097	94.4	1.7	0.056	10.3	14.7
4.93 – 4.56	1,213	96.2	1.7	0.058	10.0	15.4
4.56 – 4.27	1,316	95.4	1.7	0.062	7.2	14.8
4.27 – 4.02	1,352	94.6	1.7	0.069	8.2	14.0
4.02 – 3.82	1,447	95.0	1.7	0.070	9.7	12.1
3.82 – 3.64	1,491	94.1	1.6	0.085	7.3	10.4
3.64 – 3.49	1,598	93.9	1.6	0.095	7.4	8.5
3.49 – 3.35	1,628	93.2	1.6	0.108	6.7	6.9
3.35 – 3.23	1,673	92.0	1.6	0.123	5.8	6.3
3.23 – 3.12	1,701	89.9	1.5	0.136	5.4	4.9
3.12 – 3.02	1,682	88.2	1.5	0.162	4.4	4.1
3.02 – 2.93	1,708	84.3	1.5	0.169	4.3	3.5
2.93 – 2.85	1,708	83.6	1.5	0.206	3.4	2.8
2.85 – 2.77	1,734	79.9	1.5	0.214	3.5	2.4
2.77 – 2.70	1,674	78.5	1.5	0.280	2.7	2.0
39.07 – 2.70	**26,484**	**89.8**	**1.6**	**0.081**	**7.2**	**9.1**

Molecular replacement in $P2_1$ was carried out as above searching for 1 or 2 trimers in the au. A good solution for two trimers in the au could be identified, its R_{factor} of and correlation coefficient were calculated to be 0.370 and 0.916 respectively; however, refining the new co-ordinates in Refmac, version 5.0 immediately increased the R_{factor} to 53% and the figure of merit dropped to 14%.

To check for symmetry peaks in the native Patterson map, the data was also processed in the primitive triclinic Laue group; however, the data was too incomplete to be phased by molecular replacement. Table 4.7 gives a summary of data collection statistics for all three Bravais lattice.

Table 4.7 – Data Collection Statistics for the KA3-L2 Mutant Data Set: the statistics are detailed for the data set scaled in three different space groups.

	Primitive triclinic	Primitive monoclinic	Primitive orthorhombic
Wavelength (Å)	1.54	1.54	1.54
Resolution (Å)	39.28 – 2.70	39.06 – 2.70	38.93 – 2.70
Space Group	P1 (no. 1)	$P2_1$ (no. 4)	P222 (no.16)
Unit-cell Parameters (Å, °)	a = 73.4, b = 54.6, c = 139.6, $\alpha = \gamma = 90$, $\beta = 90.02$	a = 54.2, b = 138.6, c = 73.1, $\alpha = \gamma = 90$, $\beta = 90.01$	a = 54.2, b = 73.0, c = 138.0, $\alpha = \beta = \gamma = 90$
V_m ($Å^3$ Da^{-1})[b]	3.0	4.4/2.2	2.2
Percentage solvent[b]	58.0	68.9/37.7	42.6
Total measurements	158,521	171,331	165,965
Unique reflections	32,150 (24,516)	26,484	15,425
I/σ	9.5 (3.6)	7.2 (2.7)	5.2 (2.3)
Mn(I)/sd	9.0 (2.1)	9.1 (2.0)	10.5 (2.3)
Average redundancy	1.4 (1.3)	1.6 (1.5)	3.0 (2.8)
Data completeness (%)	46.0 (38.9)	89.8 (89.8)	98.7 (95.2)
R_{merge}[a] (%)	0.065 (0.211)	0.081 (0.280)	0.114 (0.328)
Wilson B factor ($Å^2$)	53.8	53.8	53.9

[a] $R_{merge} = \Sigma\Sigma$ I(h)j – (I(h)) / $\Sigma\Sigma$ I(h)j where I(h) is the measured diffraction intensity and the summation includes all observations

[b] V_m and solvent content were calculated assuming 9 monomers per au for the triclinic, 3 or 6 monomers per au for the monoclinic and 3 monomers per au for the orthorhombic space group

Parentheses denote data for the highest resolution shell

4.3.10 Crystal Twinning

Morphologically, the twinned crystals are indistinguishable from non-twinned ones. Merohedral and pseudo-merohedral twinning arise when the rotational symmetry of the crystal lattice (the holohedry) is greater than the underlying Laue symmetry of the crystal. For monoclinic space groups, merohedral twinning can only occur if the

monoclinic angle is 90°. This kind of twinning affects all diffraction spots and causes them to overlap. Merohedral twinning is impossible to detect in the diffraction pattern of these also: the two twin domains align in three dimensions resulting in superimposable reciprocal lattices (Yeates and Fam, 1999).

Non-merohedral twinning can arise when extra symmetry occurs in a supercell; this means that only some zones are affected by overlapping diffraction spots, i.e. if the β-angle deviates from 90°. This means that the spots do not overlap exactly in reciprocal space, thus split and streaked reflection spots should be observed in the higher resolution shells.

Analysing the space group shows that the maximal angular difference for the monoclinic angle was only 0.012° which is insignificant, i.e. the assumption of merohedral twinning may be correct. This was also supported by the 'clean' diffraction pattern, an example of which is shown in Figure 4.13.

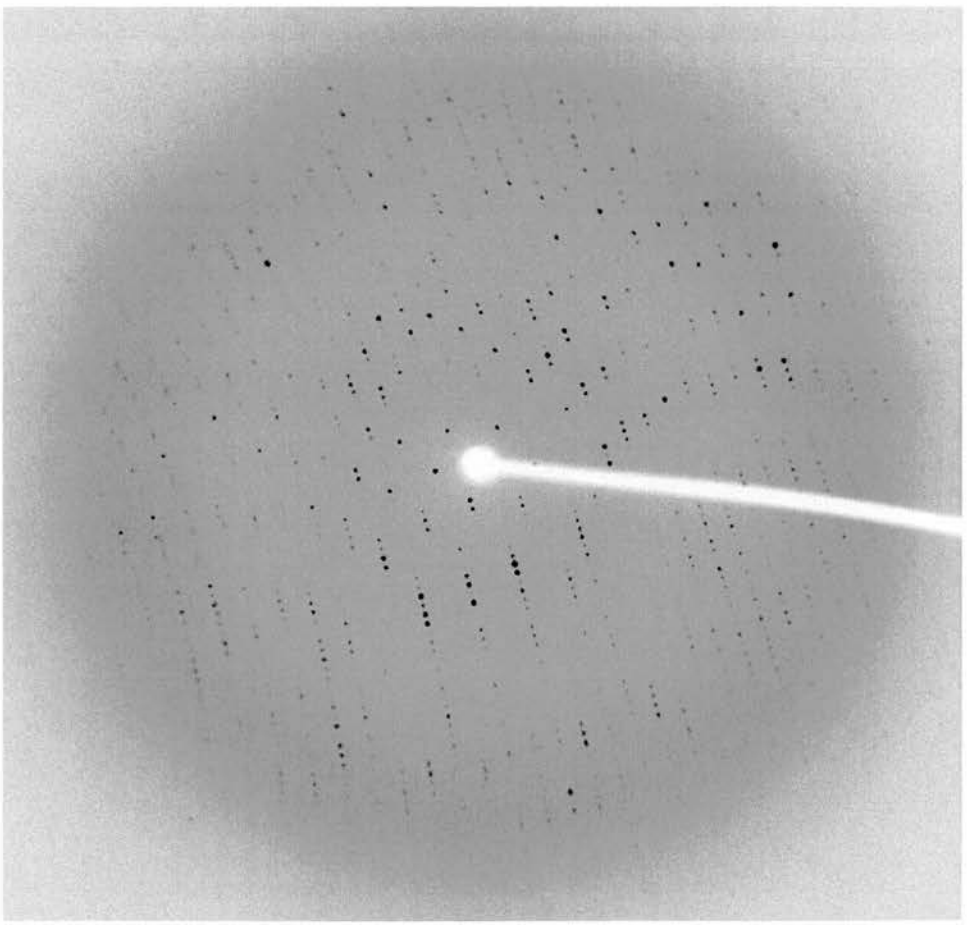


Figure 4.13 - Typical Diffraction Pattern of KA3-L2: collected to 2.4Å resolution (edge of detector); this particular image is number 50 and its phi angle is 54.5 - 55.0.

One of the most common signs of twinning is that the merging $R_{factors}$ for different Laue groups are very similar; secondly, the space group may be hard to determine as reflections from one domain can overlap with systematic absences from the other domain. Another common sign is a low value for $<|E^2 - 1|>$.

The degree of twinning is defined by the twin fraction α, which describes the volume of the twin taken up by the smaller of the two contributing twin domains (Rees, 1980). A perfect twin has a twin fraction of, or approaching 0.5; the perfect superimposition of twin-related reflections allows the data to obey and be processed to erroneously high symmetry (Yeates and Fam, 1999). Partial twins have a twin fraction smaller than 0.5 and may be successfully detwinned with the knowledge of α and the twinning operator (twin law) that aligns the two reciprocal lattices. The twin operation can be any symmetry operation in the holohedry but not in the true space group of the crystal.

There are several methods for detecting twinning and estimating the twinning ratio α. Twinning has profound effects on the statistics of reflection intensities; the cumulative intensity distribution, $N(z)$, is one of the simplest quantities to test. The theoretical Ny curve for a single crystal is exponential, whereas that of a twinned crystal is sigmoidal (Rees, 1980). This is caused by a lower percentage of very small and very large amplitudes in the twinned data, resulting from the low probability that in the two overlapped lattices both reflections are simultaneously either very weak or very strong. Yeates proposed a robust test for twinning based on the cumulative distribution of the ratio $H = |I_{t1} - I_{t2}|/(I_{t1} + I_{t2})$, which depends linearly on H (Yeates, 1988; Yeates, 1997).

The X-rays scattered from the two twin domains do not interfere and the reflection intensities from both domains (i.e. non-equivalent, overlapping reflections) contribute to the measured intensity proportionally to the twinning volume ratio α. The diffraction intensities of two reflections related by a twinning operation, measured for a crystal with $\alpha < 0.5$ are therefore

$$I_{t1} = (1 - \alpha)I_{s1} + \alpha I_{s2},$$

$$I_{t2} = \alpha I_{s1} + (1 - \alpha)I_{s2},$$

where I_{s1} and I_{s2} are the theoretical intensities contributed by the two single-crystal components (Yeates, 1988). I_{t1} and I_{t2} can then be determined as

$$I_{s1} = [(1 - \alpha)I_{t1} - \alpha I_{t2}]/(1 - 2\alpha),$$

$$I_{s2} = [(1 - \alpha)I_{t2} - \alpha I_{t1}]/(1 - 2\alpha).$$

Detwinning of intensities that assumes too high a value for α may lead to negative intensities for some reflections, which is unrealistic (Yang *et al.*, 2000).

Twinning was analysed using dataman (Rees, 1980; Stanley, 1972) and the Yeates Twinning Server located at UCLA (Yeates, 1997; Yeates and Fam, 1997) from the reflection file scaled and merged in space group $P2_1$. However, neither output was conclusive and a twin fraction could not be estimated in this way since the Yeates Twinning Server does not recognise twin laws for a monoclinic space group. This is caused by the fact that only pseudo-merohedral twinning is possible for a monoclinic space group and the Yeates Twinning Server 'expects' an input of a perfectly merohedrally twinned data set. For this reason the twin laws were determined theoretically: x, -y, -z (two-fold along x) and -x, -y, z (two-fold along z) are possibilities.

However, a second more quantitative indicator for merohedral twinning is the extraction of unusual intensity distribution from the diffraction data (Chandra *et al.*, 1999; Rees, 1982; Wilson, 1949; Yang *et al.*, 2000). From plot $N(z,\alpha)$ vs. z, as generated by the program truncate, (French and Wilson, 1978), it is apparent that there are fewer weaker reflections measured from the twinned crystal than would be expected. Twinning is especially apparent in the acentric reflections, where the observed and theoretical intensity values should be almost identical (Figure 4.14).

The diffraction data were subjected to the CCP4 program 'detwin' (Gomis-Ruth *et al.*, 1995; Redinbo and Yeates, 1993; Rees, 1980; Taylor and Leslie, 1998) in an attempt to deconvolute the overlapping lattices. Both possible twin operators were used in detwin. Analysis of the $S(H)$ vs. $<H>$ plot generated by the program suggested a twin fraction of 0.4 (Figure 4.15) when using the x, -y, -z twin operator.

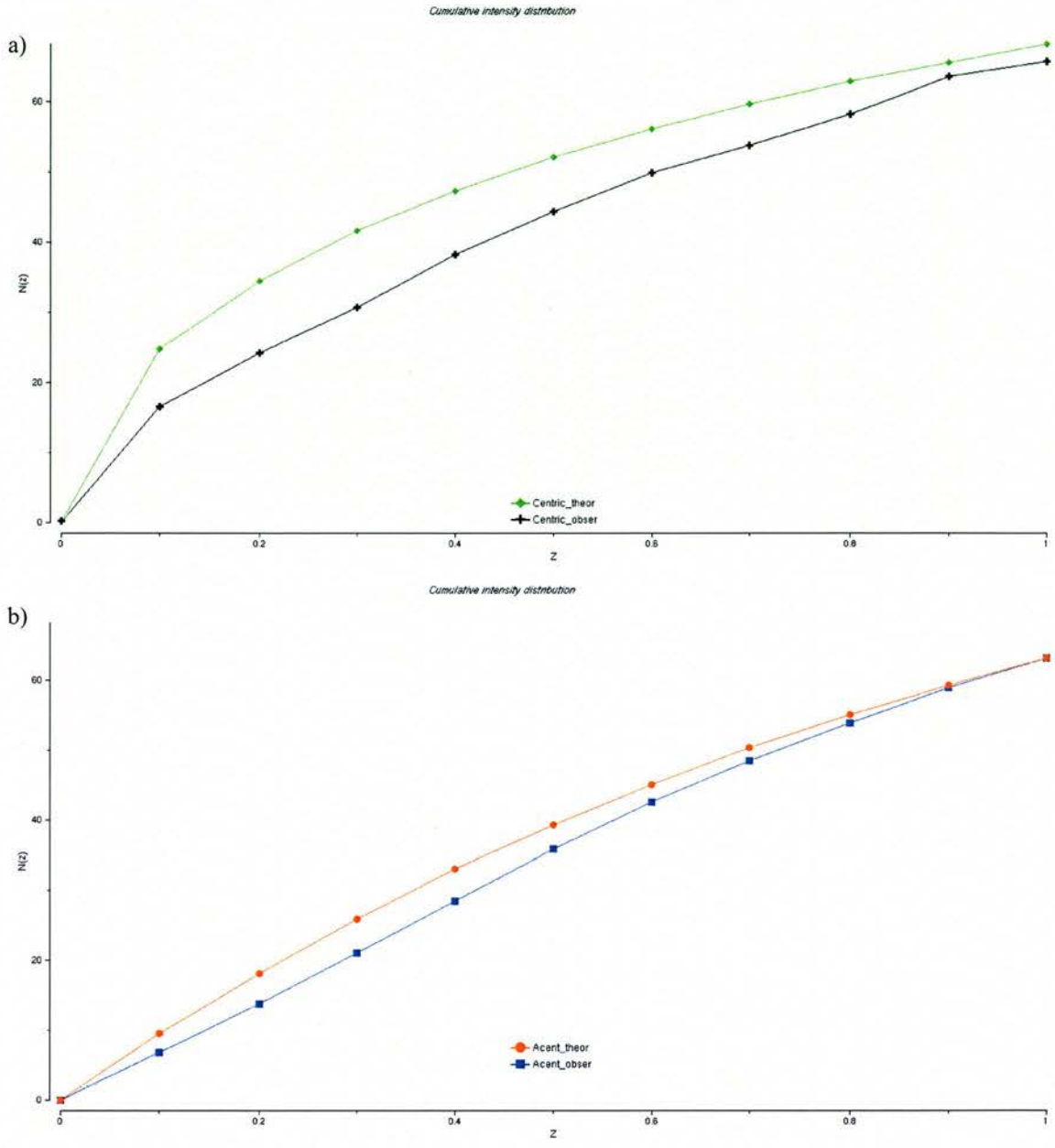


Figure 4.14 – Cumulative Intensity Distributions of the Twinned Crystal: a) for centric reflections and b) for acentric Reflections; the plot $N(z,\alpha)$ vs. z generated by truncate, as implemented in the CCP4 Program Suite.

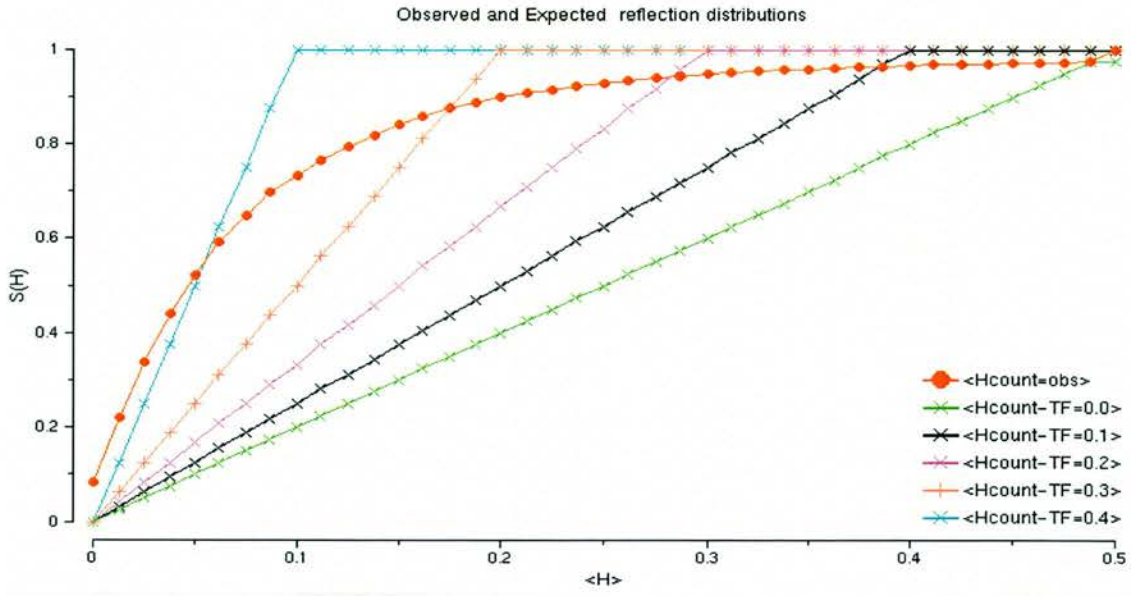


Figure 4.15 – Observed and Expected Reflection Distributions to Estimate the Twin Ratio α: plot $S(H)$ vs $\langle H \rangle$ generated by the CCP4 program detwin using x, -y, -z as twin operator.

A high twin fraction had already been indicated by the fact that the R_{symm}'s of the monoclinic and orthorhombic space groups were similar; thus the errors in the measured intensities were inflated by a factor of five, as determined by the equation which relates the error created by detwinnning to the twinning ratio $(1/(1-2\alpha)$ (Fischer and Sweet, 1980; Yeates and Fam, 1999). Nevertheless, data were detwinned using the twin operator x, -y, -z and the twin fraction $\alpha = 0.4$ in the CCP4 program detwin. The cumulative intensity distribution, calculated for the detwinned dataset using the CCP4 program truncate (French and Wilson, 1978), shows that the data is closer to a normal distribution (Figure 4.16), even though it is much noisier than the original, twinned data.

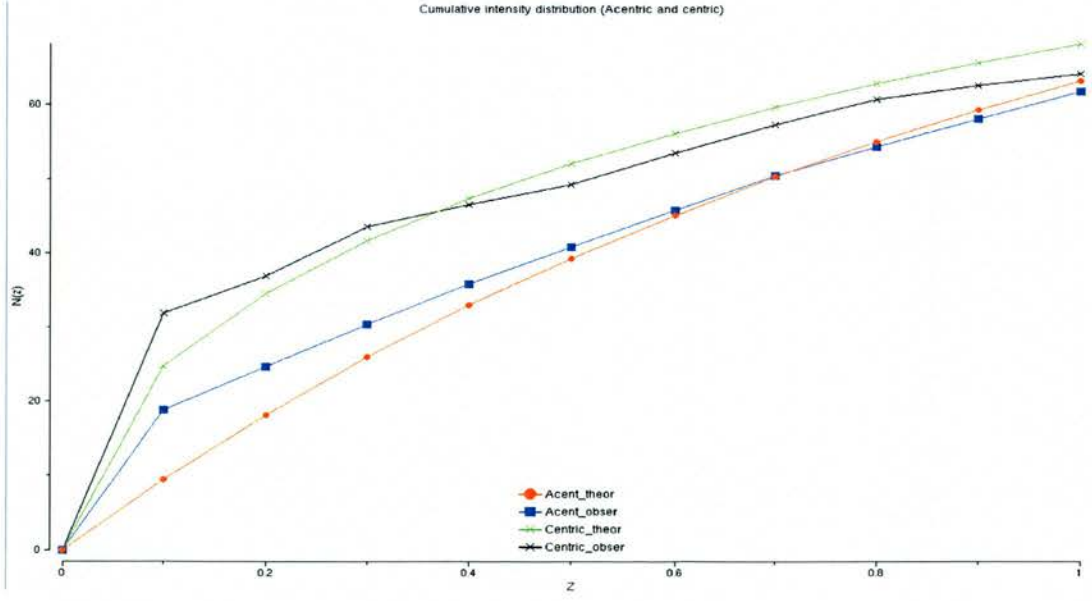


Figure 4.16 – Cumulative Intensity Distribution for the Detwinned Data Set: plot $N(z,\alpha)$ vs. z generated by truncate, as implemented in the CCP4 Program Suite.

Data was truncated to obtain structure factor amplitudes (French and Wilson, 1978) and used for molecular replacement. All molecular replacement attempts failed, as Table 4.8 shows, which presents the best solutions calculated by MolRep. Rigid body refining the new model coordinates against the detwinned crystal data did not decrease the R_{factor} below 45%. For this reason, the dataset was abandoned and new crystals, preferably under new crystallisation conditions, have to be grown in future.

Table 4.8 – Top Solutions for Molecular Replacement Using the Detwinned Data: no real solutions could be found for either one or two trimers per au.

No of trimers	alpha	beta	gamma	x_{frac}	y_{frac}	z_{frac}	Rf / sig	R_{fac}	Corr. Coef.
One	0.00	37.85	125.10	0.445	0.000	0.181	6.97	0.560	0.229
Two (1^{st})	0.00	37.85	125.10	0.445	0.000	0.181	6.97	-	-
Two (2^{nd})	179.20	-21.50	179.20	-0.391	0.197	-0.500	-	0.520	0.341

4.5 Conclusions

Enzymes are increasingly recognised as useful catalysts for organic synthesis. Their attractiveness arises from a number of important advantages over traditional methods. Enzymes are generally highly chemo-, regio-, diastereo- and enantioselective; as a result of this high selectivity and the mild reaction conditions employed, protecting group chemistry can be kept to a minimum.

However, enzymes in general tend to be very substrate specific; despite the relatively broad substrate profile of the KDPG aldolase, which makes it a useful catalyst for stereocontrolled carbon-carbon bond formation between pyruvate and a range of unnatural aldehydes (Allen *et al.*, 1992; Machajewski and Wong, 2000; Shelton *et al.*, 1996), it does fail to convert electrophilic substrates that lack polar functionality at C2′, C3′ or C4′ and nucleophilic substrates other than oxobutyrate or fluoropyruvate (Wymer *et al.*, 2001). To rationally engineer enzymes with an altered substrate profile or different catalytic properties, the catalytic mechanism, as well as the basis for substrate selectivity of the enzyme need to be determined in detail. The aim of this study was to further this current understanding of the aldolase mechanism. A number of crystallisation experiments using this enzyme have been carried out, which had lead to the identification of a number of crystallisation conditions. However, the usefulness of these conditions was limited, as we tried to crystallise the aldolase in the absence of citrate, to allow binding of inhibitors or substrate analogues to the active site residues of the crystals. Co-crystallisation experiments with substrate analogues and inhibitors failed to yield any useful, high-quality crystals. A number of crystallisation conditions had been identified which allowed the growth of apo-crystals, which could then be used for soaking experiments. Carrying out these did not result in anything bound to the active site of the enzyme. The active site Lys seems slightly disordered and a sulphate molecule from the crystallisation condition is bound to Ser184, but the structure only confirms the previously determined apo-enzyme structure with an rms deviation between the Cαs of between 0.3 and 0.7Å

A number of new crystallisation conditions have been identified, which do not contain sulfate, phosphate or citrate and may well allow subsequent soaking experiments. Nevertheless, most of the crystals obtained from these conditions diffracted to resolutions too low for any useful data to be collected. It may well be the

case that the crystals in these conditions are as variable in quality, as in the original crystallisation conditions and no high-quality crystals have been found as yet.

Some conditions yielded twinned crystals, whose phases could not be determined - even after detwinning - by molecular replacement, despite several attempts using different search models and a number of computer programs using different algorithms. This was most likely caused by the inflation of errors in measured intensities when detwinning a data set with a high twin ratio α.

4.6 Further Work

More crystallisation trials building on the identified KBr and KCSN conditions need to be carried out in order to gain higher-quality crystals that do not contain any hindering molecules in the active site. Secondly, crystallisation conditions that do not favour merohedral twinning of the crystals need to be identified, in order to solve the phases of further data sets collected on these crystals.

In addition to that, optimal soaking conditions need to be determined to successfully soak substrates, substrate analogues or inhibitors into the active site of the enzyme.

Bibliography

Adams, P.D., Pannu, N.S., Read, R.J. and Brünger, A.T. (1997) Cross-validated Maximum Likelihood Enhances Crystallographic Simulated Annealing Refinement. *Proc Natl Acad Sci U S A*, **94**, 5018-5023.

Aguirrezabalaga, I., Olano, C., Allende, N., Rodriguez, L., Brana, A.F., Mendez, C. and Salas, J.A. (2000) Identification and expression of genes involved in biosynthesis of L-oleandrose and its intermediate L-olivose in the oleandomycin producer *Streptomyces antibioticus*. *Antimicrob Agents Chemother*, **44**, 1266-1275.

Ahn, H.J., Yoon, H.J., Lee, B.n. and Suh, S. (2004) Crystal structure of chorismate synthase: a novel FMN binding protein fold and functional insights. *J Mol Biol*, **336**, 903-915.

Albermann, C., Soriano, A., Jiang, J., Vollmer, H., Biggins, J.B., Barton, W.A., Lesniak, J., Nikolov, D.B. and Thorson, J.S. (2003) Substrate specificity of NovM: implications for novobiocin biosynthesis and glycorandomization. *Org Lett*, **5**, 933-936.

Allard, J., Grochulski, P. and Sygusch, J. (2001) Covalent intermediate trapped in 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase structure at 1.95-A resolution. *Proc Natl Acad Sci U S A*, **98**, 3679-3684.

Allard, S.T. (2001) Carbohydrate Processing in Pathogenic Bacteria. *Chemistry*. University of St. Andrews, St. Andrews, p. 270.

Allard, S.T., Beis, K., Giraud, M.F., Hegeman, A.D., Gross, J.W., Wilmouth, R.C., Whitfield, C., Graninger, M., Messner, P., Allen, A.G., Maskell, D.J. and Naismith, J.H. (2002) Toward a structural understanding of the dehydratase mechanism. *Structure (Camb)*, **10**, 81-92.

Allard, S.T., Cleland, W.W. and Holden, H.M. (2004) High resolution X-ray structure of dTDP-glucose 4,6-dehydratase from *Streptomyces venezuelae*. *J Biol Chem*, **279**, 2211-2220.

Allen, S.T., Heintzelman, G.R. and Toone, E.J. (1992) Pyruvate Aldolases as Reagents for Stereospecific Aldol Condensations. *J. Org. Chem.*, **57**, 426-428.

Apic, G., Huber, W. and Teichmann, S.A. (2003) Multi-domain protein families and domain pairs: comparison with known structures and a random model of domain recombination. *J. Struct Funct Genomics*, **4**, 67-78.

Appelbaum, P.C. and Bozdogan, B. (2004) Vancomycin resistance in *Staphylococcus aureus*. *Clin Lab Med*, **24**, 381-402.

Apweiler, R. and Bairoch, A. (2004) Protein sequence databases. *Curr Opin Chem Biol*, **8**, 76-80.

Bailey, S. (1994) The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallogr D*, **50**, 760-763.

Bairoch, A., Boeckmann, B., Ferro, S. and Gasteiger, E. (2004) Swiss-Prot: Juggling between evolution and stability. *Brief. Bioinform.*, **5**, 39-55.

Baltz, R.H. (1986) Mutagenesis in Streptomyces spp. In Solomon, N.A. (ed.), *Manual of Industrial Microbiology and Biotechnology*. American Society for Microbiology, Washington, D.C., pp. 184-190.

Baltz, R.H. (1990) Gene expression using streptomycetes. *Curr Opin Biotechnol*, **1**, 12-20.

Baltz, R.H., McHenney, M.A., Cantwell, C.A., Queener, S.W. and Solenberg, P.J. (1997) Applications of transposition mutagenesis in antibiotic producing streptomycetes. *Antonie Van Leeuwenhoek*, **71**, 179-187.

Baltz, R.H. and Seno, E.T. (1988) Genetics of *Streptomyces fradiae* and tylosin biosynthesis. *Annu Rev Microbiol*, **42**, 547-574.

Barna, J.C. and Williams, D.H. (1984) The structure and mode of action of glycopeptide antibiotics of the vancomycin group. *Annu Rev Microbiol*, **38**, 339-357.

Barquera, B., Zhou, W., Morgan, J.E. and Gennis, R.B. (2002) Riboflavin is a component of the Na^+-pumping NADH-quinone oxidoreductase from *Vibrio cholerae*. *Proc Natl Acad Sci U S A*, **99**, 10322-10324.

Barran, L.R. and Wood, W. (1971) The Mechanism of 2-Keto-3-deoxy-6-phosphogluconate Aldolase. *J Biol Chem*, **246**, 4028-4035.

Barrett, C.T. and Barrett, J.F. (2003) Antibacterials: are the new entries enough to deal with the emerging resistance problems? *Curr Opin Biotechnol*, **14**, 621-626.

Bate, N. and Cundliffe, E. (1999) The mycinose-biosynthetic genes of *Streptomyces fradiae*, producer of tylosin. *J Ind Microbial Biotechnol*, **23**, 118-122.

Beebe, J.A. and Frey, P.A. (1998) Galactose mutarotase: purification, characterization, and investigations of two important histidine residues. *Biochem*, **37**, 14989-14997.

Bell, B.J., Watanabe, L., Rios-Steiner, J.L., Tulinsky, A., Lebioda, L. and Arni, R.K. (2003) Structure of 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase from *Pseudomonas putida*. *Acta Crystallogr D*, **59**, 1454-1458.

Berezovsky, I.N., Kilosanidze, G.T., Tumanyan, V.G. and Kisselev, L.L. (1999) Amino acid composition of protein termini are biased in different manners. *Protein Engineering*, **12**, 23-30.

Binnie, C., Cossar, J.D. and Stewart, D.I. (1997) Heterologous biopharmaceutical protein expression in Streptomyces. *Trends Biotechnol*, **15**, 315-320.

Blackwell, J.R. and Horgan, R. (1991) A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Letters*, **295**, 10-12.

Blankenfeldt, W., Kerr, I.D., Giraud, M.F., McMiken, H.J., Leonard, G., Whitfield, C., Messner, P., Graninger, M. and Naismith, J.H. (2002) Variation on a theme of SDR. dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD) shows a new Mg2+-dependent dimerization mode. *Structure (Camb)*, **10**, 773-786.

Blayer, S., Woodley, J.M., Lilly, M.D. and Dawson, M.J. (1996) Characterization of the chemoenzymatic synthesis of *N*-acetyl-$_D$-neuraminic acid (Neu5Ac). *Biotechnol. Prog.*, **12**, 758-763.

Borges-Walmsley, M.I. and Walmsley, A.R. (2001) The structure and function of drug pumps. *Trends Microbiol*, **9**, 71-79.

Bradford, M.M. (1976) A refined sensitive method for the quantification of microgram quantities of protein utilising the principle of protein - dye binding. *Anal. Biochem.*, **72**, 248.

Braga, R., Hecquet, L. and Blonski, C. (2004) Slow-binding inhibition of 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase. *Bioorg. Med. Chem.*, **12**, 2965-2972.

Brakoulias, A. and Jackson, R.M. (2004) Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: an automated all-against-all structural comparison using geometric matching. *Proteins*, **56**, 250-260.

Branden, C. and Tooze, J. (1998) Prediction, Engineering, and Design of Protein Structures. In *Introduction to Protein Structure*. Garland Publishing Inc., New York, p. 410.

Breuer, M. and Hauer, B. (2003) Carbon-carbon coupling in biotransformation. *Curr Opin Biotechnol*, **14**, 570-576.

Brückner, R. (1996) Chemie der Alkalimetall-Enolate. In Elvers, B. (ed.), *Reaktionsmechanismen*. Spektrum Akademischer Verlag GmbH, Heidelberg, pp. 332-385.

Brünger, A.T. (1992) Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.

Brünger, A.T. (1997) Free R Value: Cross-validation in crystallography. *Methods Enzymol*, **277**, 366-396.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Crystallography & NMR System*: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr D*, **54**, 905-921.

Brünger, A.T., Krukowski, A. and Erickson, J. (1990) Slow-Cooling Protocols for Crystallographic Refinement by Simulated Annealing. *Acta Crystallogr A*, **46**, 585-593.

Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) Crystallographic R factor Refinement by Molecular Dynamics. *Science*, **235**, 458-460.

Buchanan, L.V. (2000) Structural Investigations of *E. coli* KDPG Aldolase - a Carbohydrate Synthesising Protein. *School of Chemistry*. University of St. Andrews, St. Andrews, p. 212.

Buchanan, L.V., Mehta, N., Pocivavsek, L., Niranjanakumari, S., Toone, E.J. and Naismith, J.H. (1999) Initiating a structural study of 2-keto-3-deoxy-6-phosphogluconate aldolase from *Escherichia coli*. *Acta Crystallogr D*, **55**, 1946-1948.

Buchanan, R.E. (1917) Studies in the nomenclature and classification of bacteria. II. The primary subdivisions of the Schizomycetes. *J. Bacteriol.*, **2**, 155-164.

Burke, J.P. (1998) Antibiotic Resistance - Squeezing the Balloon? *JAMA*, **280**, 1270-1271.

Cassell, G.H. and Mekalanos, J. (2001) Development of Antimicrobial Agents in the Era of New and Reemerging Infectious Diseases and Increasing Antibiotic Resistance. *JAMA*, **285**, 601-605.

CDC. (2003) Nosocomial Infections caused by MRSA. CDC, **2003**

CDC. (2004) Nosocomial Infections caused by MRSA. CDC, **2004**

Cha, R. and Rybak, M.J. (2003) Daptomycin against multiple drug-resistant staphylococcus and enterococcus isolates in an in vitro pharmacodynamic model with simulated endocardial vegetations. *Diagnostic Microbiol Infectious Disease*, **47**, 539-546.

Chandra, N., Acharya, K.R. and Moody, P.C. (1999) Analysis and characterisation of data from twinned crystals. *Acta Crystallogr D*, **55**, 1750-1758.

Chater, K.F. (1990) The improving prospects for yield increase by genetic engineering in antibiotic-producing streptomycetes. *Bio/Technology*, **8**, 115-121.

Chater, K.F. and Bibb, M.J. (1997) Regulation of bacterial antibiotic production. In Doehren, H.v. (ed.), *Products of Secondary Metabolism*. VCH Press, Weinheim, Germany, Vol. 7, pp. 57-105.

Chater, K.F. and Merrick, M.J. (1979) Streptomycetes. In Parish, J.H. (ed.), *Developmental Biology of Prokaryotes*. Blackwell, Scientifc Publications, Oxford, pp. 93-114.

Chen, H., Thomas, M.G., Hubbard, B.K., Losey, H.C., Walsh, C.T. and Burkart, M.D. (2000) Deoxysugars in glycopeptide antibiotics: enzymatic synthesis of TDP-L-epivancosamine in chloroeremomycin biosynthesis. *Proc Natl Acad Sci U S A*, **97**, 11942-11947.

Chittum, H.S. and Champney, W.S. (1994) Ribosomal protein gene sequence changes in erythromycin-resistant mutants of *Escherichia coli*. *J Bacteriol*, **176**, 6192-6198.

Chou, W.-C., Fotsch, C. and Wong, C.-H. (1995) Synthesis of Nitrocyclitols based on enzymatic aldol reaction and intramolecular nitroaldol reaction. *J Org Chem*, **60**, 2916-2917.

Christendat, D., Saridakis, V., Dharamsi, A., Bochkarev, A., Pai, E.F., Arrowsmith, C.H. and Edwards, A.M. (2000) Crystal structure of dTDP-4-keto-6-deoxy-D-hexulose 3,5-epimerase from *Methanobacterium thermoautotrophicum* complexed with dTDP. *J Biol Chem*, **275**, 24608-24612.

Cleland, W.W. and Kreevoy, M.M. (1994) Low-Barrier Hydrogen-Bonds and Enzymatic Catalysis. *Science*, **264**, 1887-1890.

Cochrane, V.W. (1961) Physiology of actinomycetes. *Annu Rev Microbiol*, **15**, 1-26.

Combet, C., Blanchet, C., Geourjon, C. and Deléage, G. (2000) NPS@: Network Protein Sequence Analysis. *TIBS*, **25**, 147-150.

Conway, T. (1992) The Entner-Doudoroff pathway: history, physiology and molecular biology. *FEMS Microbiol Lett*, **103**, 1-27.

Cornish-Bowden, A. (1999) *Fundamentals in Enzyme Kinetics*. Portland Press, London, UK.

Cuff, J.A. and Barton, G.J. (2000) Application of Enhanced Multiple Sequence Alignment Profiles to Improve Protein Secondary Structure Prediction. *Proteins*, **40**, 502-511.

Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M. and Barton, G.J. (1998) Jpred: A Consensus Secondary Structure Prediction Server. *Bioinformatics*, **14**, 892-893.

Daffe, M., Brennan, P.J. and McNeil, M. (1990) Predominant structural features of the cell wall arabinogalactan of *Mycobacterium tuberculosis*as revealed through characterisation of oligoglycosyl alditol fragments by gas chromatography/mass spectrometry and by ^{1}H and ^{13}C NMR analyses. *J. Biol. Chem.*, **265**, 6734-6743.

Dalby, A., Dauter, Z. and Littlechild, J.A. (1999) Crystal structure of human muscle aldolase complexed with fructose 1,6-bisphosphate: Mechanistic implications. *Protein Sci*, **8**, 291-297.

Dalby, A., Dauter, Z. and Littlechild, J.A. (2001) The Structure of human liver fructose-1,6-bisphosphate aldolase. *Acta Crystallogr D*, **57**, 1526-1533.

Davis, G.D., Elisee, C., Newham, D.M. and Harrison, R.G. (1999) New fusion protein system designed to give soluble expression in *Escherichia coli*. *Biotechnol. Bioeng*, **65**, 382-388.

De Montigny, C. and Sygusch, J. (1996) Functional characterization of an extreme thermophilic class II fructose-1,6-bisphosphate aldolase. *Eur J Biochem*, **241**, 243-248.

Demain, A.L. (1992) Microbial secondary metabolism: a new theoretical frontier for academia, a new opportunity for industry. *Ciba Found Symp*, **171**, 3-16; discussion 16-23.

Demain, A.L. and Fang, A. (1995) Emerging concepts of secondary metabolism in actinomycetes. *Actinomycetol.*, **9**, 98-117.

Deupree, J. and Wood, W.A. (1975) L-ribulose-5-phosphate 4-epimerase from *Aerobacter aerogenes*. *Methods Enzymol*, **41**, 412-419.

Deupree, J.D. and Wood, W.A. (1972) L-Ribulose 5-phosphate 4-epimerase from Aerobacter aerogenes. Evidence for a role of divalent metal ions in the epimerization reaction. *J Biol Chem*, **247**, 3093-3097.

Diekema, D.J., BootsMiller, B.J., Vaughn, T.E., Woolson, R.F., Yankey, J.W., Ernst, E.J., Flach, S.D., Ward, M.M., Franciscus, C.L., Pfaller, M.A. and Doebbeling,

B.N. (2004) Antimicrobial resistance trends and outbreak frequency in United States hospitals. *Clin Infect Dis*, **38**, 78-85.

Dong, C. (2002) RmlC structure, mechanism and inhibitor design. *Centre for Biomolecular Sciences*. University of St. Andrews, St. Andrews, p. 376.

Dong, C., Major, L.L., Allen, A., Blankenfeldt, W., Maskell, D. and Naismith, J.H. (2003) High-resolution structures of RmlC from *Streptococcus suis* in complex with substrate analogs locate the active site of this class of enzyme. *Structure (Camb)*, **11**, 715-723.

Dunbrack, R.L., Jr. (1998) Protein Sidechain Conformational Analysis. Dunbrack, R.L., Jr., Vol. 2003.

Dunbrack, R.L., Jr. and Cohen, F.E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci*, **6**, 1661-1681.

Eckwall, E.C. and Schottel, J.L. (1997) Isolation and characterisation of an antibiotic produced by the scab disease-supressive *Streptomyces diastatochromogenes* strain. *J. Ind. Microbiol. Biotech.*, **19**, 220-225.

Effenberger, F. and Jäger, J. (1997) Synthesis of the adrenergic bronchodilators (R)-terbutaline and (R)-salbutamol from (R)-cyanohydrins. *Journal of Organic Chemistry*, **62**, 3867-3873.

Egan, S.E., Fliege, R., Tong, S., Shibata, A., Wolf, R.E., Jr. and Conway, T. (1992) Molecular characterization of the Entner-Doudoroff pathway in *Escherichia coli*: sequence analysis and localization of promoters for the edd-eda operon. *J Bacteriol*, **174**, 4638-4646.

Egan, S.E., Wiener, P., Kallifidas, D. and Wellington, E.M. (2001) Phylogeny of Streptomyces species and evidence for horizontal transfer of entire and partial antibiotic gene clusters. *Antonie Van Leeuwenhoek*, **79**, 127-133.

Ensign, J.C. (1978) Formation, properties, and germination of actinomycete spores. *Annu Rev Microbiol*, **32**, 185-219.

Eom, J.S., Hwang, I.S., Hwang, B.Y., Lee, J.G., Lee, Y.J., Cheong, H.J., Park, Y.H., Park, S.C. and Kim, W.J. (2004) Emergence of vanA Genotype Vancomycin-

Resistant Enterococci with Low or Moderate Levels of Teicoplanin Resistance in Korea. *J Clin Microbiol*, **42**, 1785-1786.

Errey, J.C. (2003) Mechanistic studies of novel anti-tuberculosis targets. *Chemistry*. University of East Anglia, Norwich, p. 207.

Escalante, A., Wacher-Rodarte, C., Garcia-Garibay, M. and Farres, A. (1998) Enzymes involved in carbohydrate metabolism and their role on exopolysaccharide production in *Streptococcus thermophilus*. *J Appl Microbiol.*, **84**, 108-114.

Evans, P.R. (1993) Data Reduction. *Data Collection & Processing*. CCP4, pp. 114-122.

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J., Hofmann, K. and Bairoch, A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res*, **30**, 235-238.

Feldman, W.H. and Hinshaw, H.C. (1944) Effects of streptomycin on experimental tuberculosis in guinea pigs: a preliminary report. *Proc. Staff Meetings Mayo Clinic*, **19**, 593-599.

Feldman, W.H. and Hinshaw, H.C. (1945) Effects of streptomycin on experimental tuberculosis in guinea pigs: a preliminary report. *Proc. Staff Meetings Mayo Clinic*, **20**, 313-318.

Felske, A., Vancanneyt, M., kersters, K. and Akkermans, A.D. (1999) Application of temperature-gradient gel electrophoresis in taxonomy of coryneform bacteria. *Int. J. Syst. Bacteriol.*, **49**, 113-121.

Fessner, W.D. (1998) Enzyme mediated C-C bond formation. *Curr Opin Chem Biol*, **2**, 85-97.

Fessner, W.D. and Walter, C. (1996) Enzymatic C-C bond formation in asymmetric synthesis. *Top. Curr. Chem.*, **184**, 97-194.

Fischer, E. (1890) *Ber.*, **23**, 2114.

Fischer, R.G. and Sweet, R.M. (1980) Treatment of diffraction data from protein crystals twinned by merohedry. *Acta Crystallogr B*, **36**, 755-760.

Fleming, A. (1929) On the bacterial action of cultures of a Penicillium, with special reference to their use in the isolation of *B. influenzae*. *Brithish Journal of Experimental Pathology*, **10**, 226-236.

Fleming, A. (1945) Penicillin's finder assays its future: Sir Alexander Fleming Says Improved Dosage Method Is Needed to Extend Use. *New York Times*, New York, p. 1.

Fluit, A.C. and Schmitz, F.-J. (2004) Resistance integrons and super-integrons. *Clin Microbiol Infect*, **10**, 272-288.

Foldes, M., Munro, R., Sorrell, T.C., Shanker, S. and Toohey, M. (1983) In-vitro effects of vancomycin, rifampicin, and fusidic acid, alone and in combination, against methicillin-resistant *Staphylococcus aureus*. *J Antimicrob Chemother*, **11**, 21-26.

Fong, S., Machajewski, T.D., Mal, C.C. and Wong, C.-H. (2000) Directed Evolution of D-2-keto-3-deoxy-6-phosphogluconate Aldolase to new Variants for the Efficient Synthesis of D- and L-Sugars. *Chem Biol*, **7**, 873-883.

Frankel, S., Sohn, R. and Leinwand, L. (1991) The use of sarkosyl in generating soluble protein after bacterial expression. *Proc Natl Acad Sci U S A*, **88**, 1192-1196.

Freel Meyers, C.L., Oberthur, M., Anderson, J.W., Kahne, D. and Walsh, C.T. (2003) Initial characterization of novobiocic acid noviosyl transferase activity of NovM in biosynthesis of the antibiotic novobiocin. *Biochem*, **42**, 4179-4189.

French, S. and Wilson, K. (1978) On the treatment of negative intensity observations. *Acta Crystallogr A*, **34**, 517-528.

Frey, P.A., Whitt, S.A. and Tobin, J.B. (1994) A Low-Barrier Hydrogen-Bond in the Catalytic Triad of Serine Proteases. *Science*, **264**, 1927-1930.

Fuhrman, L.K., Wanken, A., K.W., N. and Conway, T. (1998) Rapid accumulation of introcellular 2-keto-3deoxy-6-phosphogluconate in an Entner-Doudoroff aldolase mutant results in bacteriostasis. *FEMS Microbiol Lett*, **159**, 261-266.

Fullerton, S.W.F., Griffiths, J.S., Merkel, A.B., Wymer, N.J., Hutchins, M.J., Fierke, C.A., Toone, E.J., Naismith, J.H. (2004) Mechanism of the Class I KDPG Aldolase. *Biochem* (submitted)

Gaisser, S., Martin, C.J., Wilkinson, B., Sheridan, R.M., Lill, R.E., Weston, A.J., Ready, S.J., Waldron, C., Crouse, G.D., Leadlay, P.F. and Staunton, J. (2002) Engineered biosynthesis of novel spinosyns bearing altered deoxyhexose substituents. *Chem Commun (Camb)*, 618-619.

Gasteiger, E., Gattiker, A., Hoogland, C., I., I., Appel, R.D. and Bairoch, A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, **31**, 3784-3788.

Ge, M., Chen, Z., Onishi, H.R., Kohler, J., Silver, L.L., Kerns, R., Fukuzawa, S., Thompson, C. and Kahne, D. (1999) Vancomycin derivatives that inhibit peptidoglycan biosynthesis without binding D-Ala-D-Ala. *Science*, **284**, 507-511.

Gefflaut, T., Blonski, C. and Perie, J. (1996) Slow reversible inhibitions of rabbit muscle aldolase with substrate analogues: synthesis, enzymatic kinetics and UV difference spectroscopy studies. *Bioorg. Med. Chem.*, **4**, 2043-2054.

Georgiou, G. and Valax, P. (1996) Expression of correctly folded protein in *Escherichia coli*. *Curr Opin Biotech*, **7**, 190-197.

Gijsen, H.J.M., Qiao, L., Fitz, W. and Wong, C.-H. (1996) Recent advances in the chemoenzymatic synthesis of carbohydrates and carbohydrate mimics. *Chem Rev*, **96**, 48-51.

Giraud, M.F., Leonard, G.A., Field, R.A., Bernlind, C. and Naismith, J.H. (2000) RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase. *Nat Struct Biol*, **7**, 398-402.

Giraud, M.F. and Naismith, J.H. (2000) The rhamnose pathway. *Curr Opin Struc Biol*, **10**, 687-696.

Gomis-Ruth, F.X., Fita, I., Kiefersauer, R., Huber, R., Aviles, F.X. and Navaza, J. (1995) Determination of Hemihedral Twinning and Initial Structure Analysis of Crystals of the Procarboxypeptidase A Ternary Complex. *Acta Crystallogr D*, **51**, 819-823.

Gouet, P., Robert, X. and Courcelle, E. (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of protein. *Nucleic Acids Res*, **31**, 3320-3323.

Graninger, M., Nidetzky, B., Heinrichs, D.E., Whitfield, C. and Messner, P. (1999) Characterization of dTDP-4-dehydrorhamnose 3,5-epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-L-rhamnose biosynthesis in *Salmonella enterica* serovar Typhimurium LT2. *J Biol Chem*, **274**, 25069-25077.

Guillemot, D. (1999) Antibiotic use in humans and bacterial resistance. *Curr Opin Microbiol*, **2**, 494-498.

Hackl, E., Zechmeister-Boltenstern, S., Bodrossy, L. and Sessitsch, A. (2004) Comparison of diversities and compositions of bacterial populations inhabiting natural forest soils. *Appl Environ Microbiol*, **70**, 5057-5065

Hall, D.R., Bond, C.S., Leonard, G., Watt, C.I., Berry, A. and Hunter, W.N. (2002) Structure of tagatose-1,6-biphosphate aldolase. Insight into chiral discrimination, mechanism, and specificity of class II aldolases. *J Biol Chem*, **277**, 22018-22024.

Hall, D.R., Leonard, G.A., Reed, C.D., Watt, C.I., Berry, A. and Hunter, W.N. (1999) The Crystal Structure of *Escherichia coli* class II Fructose-1,6-bisphosphate Aldolase in Complex with Phosphoglycolohydroxamate Reveals Details of Mechanism and Specificity. *J Mol Biol*, **287**, 383-394.

Hamodrakas, S.J., Hoenger, A. and V.A., I. (2004) Amyloid fibrillogenesis of silkmoth chorion protein peptide-analogues via a liquid-crystalline intermediate phase. *J Struc Biol*, **145**.

Hancock, R. and Fitz-James, P.C. (1964) Some differences in the action of penicillin, bacitracin, and vancomycin on *Bacillus megaterium*. *J Bacteriol*, **87**, 1044-1050.

Hasslacher, M., Kratky, C., Griengl, H., Schwab, H. and Kohlwein, S.D. (1997) Hydroxynitrile lyase from *Hevea brasiliensis*: molecular characterization and mechanism of enzyme catalysis. *Proteins*, **27**, 438-449.

Hegeman, A.D., Gross, J.W. and Frey, P.A. (2001) Probing catalysis by *Escherichia coli* dTDP-glucose-4,6-dehydratase: identification and preliminary characterization of functional amino acid residues at the active site. *Biochem*, **40**, 6598-6610.

Heine, A., DeSantis, G., Luz, J.G., Mitchell, M., Wong, C.-H. and Wilson, I.A. (2001) Observationof Covalent Intermediates in an Enzyme Mechanism at Atomic Resolution. *Science*, **294**, 369-374.

Heine, A., Luz, J.G., Wong, C.-H. and Wilson, A.J.C. (2004) Analysis of the Class I Aldolase Binding Site Architecture Based on the Crystal Structure of 2-Deoxyribose-5-phosphate Aldolase at 0.99A Resolution. *J Mol Biol*, **343**, 1019-1034.

Heurgué-Hamard, V., Karimi, R., Mora, L., MacDougall, J., Leboeuf, C., Grentzmann, G., Ehrenberg, M. and Buckingham, R. (1998) Ribosome release factor RF4 and termination factor RF3 are involved in dissociation of peptidyl-tRNA from the ribosome. *EMBO J.*, **17**, 808-816.

Hood, D.W., Heidstra, R., Swoboda, U.K. and Hodgson, D.A. (1992) Molecular genetic analysis of proline and tryptophan biosynthesis in *Streptomyces coelicolor* A3(2): interaction between primary and secondary metabolism--a review. *Gene*, **115**, 5-12.

Hooft, R.W.W., Vriend, G., Sander, C. and Abola, E.E. (1996) Errors in protein structures. *Nature*, **381**, 272.

Horinouchi, S. and Beppu, T. (1994) A-factor as a microbial hormone that controls cellular differentiation and secondary metabolism in *Streptomyces griseus*. *Mol Microbiol*, **12**, 859-864.

Howlin, B., Moss, D.S. and Harris, G.W. (1989) Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model. *Acta Crystallogr A*, **45 (Pt 12)**, 851-861.

Hubbard, B.K. and Walsh, C.T. (2003) Vancomycin assembly: nature's way. *Angew Chem Int Ed Engl*, **42**, 730-765.

Hucho, F. and Wallenfels, K. (1971) The enzymatically catalyzed mutarotaton. The mechanism of action of mutarotase (aldose 1-epimerase) from Escherichia coli. *Eur J Biochem*, **23**, 489-496.

Hutchinson, C.R. and Fujii, I. (1995) Polyketide synthase gene manipulation: a structure-function approach in engineering novel antibiotics. *Annu Rev Microbiol*, **49**, 201-238.

Immirzi. (1966) Crystallographic Computing Techniques. In Ahmes, F.R. (ed.), *Crystallographic Computing Techniques*. Munksgaard.

Jaeger, K.E. and Reetz, M.T. (2000) Directed evolution of enantioselective enzymes for organic chemistry. *Curr Opin Chem Biol*, **4**, 68-73.

Jia, J., Schorken, U., Lindqvist, Y., Sprenger, G.A. and Schneider, G. (1997) Crystal structure of the reduced Schiff-base intermediate complex of transaldolase B from *Escherichia coli:* Mechanistic implications for class I aldolases. *Protein Sci*, **6**, 119-124.

Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models. *Acta Crystallogr A*, **47**, 110-119.

Jordan, D.C. and Inniss, W.E. (1959) Selective inhibition of ribonucleic acid synthesis in *Staphylococcus aureus* by vancomycin. *Nature*, **184(Suppl 24)**, 1894-1895.

Karagouni, A.D., Vionis, A.P., Baker, P.W. and Wellington, E.M.H. (1993) The effects of soil moisture content on spore germination, mycelium development and survival of a seeded streptomycete in soil. *Microbiol. Releases*, **2**, 47-51.

Karimi, R., Pavlov, M.Y., Heurgué-Hamard, V., Buckingham, R. and Ehrenberg, M. (1998) Initiation Factors IF1 and IF2 Synergistically Remove Peptidyl-tRNAs with Short Polypeptides from the P-site of Translating Escherichia coli ribosomes. *J Mol Biol*, **281**, 241-252.

Keller, U., Krengel, U. and Haese, A. (1985) Genetic analysis in *Streptomyces chrysomallus*. *J Gen Microbiol*, **131 (Pt 5)**, 1181-1191.

Kelley, L.A., MacCallum, R.M. and Sternberg, M.J.E. (2000) Enhanced Genome Annotation using Structural Profiles in the Program 3D-PSSM. *J Mol Biol*, **299**, 499-520.

Kerr, I.D. (2003) Molecular Recognition in Proteins. *School of Chemistry*. University of St. Andrews, St. Andrews, p. 194.

Kieser, T., Chater, K.F., Bibb, M.J. and Buttner, M.J. (2000) *Practical streptomyces genetics*. The John Innes Foundation, Norwich.

Kiliani, H. (1895) *Ber.*, **18**, 3066.

Kim, M.-J. and Lim, I.T. (1996) Synthesis of unsaturated C8-C9 sugars by enzymatic chain elongation. *Synlett*, 138-140.

Kirkpatrick, P.N., Scaife, W., Hallis, T.M., Liu, H., Spencer, J.B. and Williams, D.H. (2000) Characterisation of a sugar epimerase enzyme involved in the biosynthesis of a vancomycin-group antibiotic. *Chem Commun*, 1565-1566.

Kleywegt, G.J. and Brünger, A.T. (1996) Checking your imagination: applications of the free R value. *Structure*, **4**, 897-904.

Kleywegt, G.J. and Jones, T.A. (1993) Masks Made Easy. *ESF/CCP4 Newsletter*, **28**, 56-59.

Kleywegt, G.J. and Jones, T.A. (1994) Halloween ... Masks and Bones. In *From First Map to Final Model*. CCP4, pp. 59-66.

Kleywegt, G.J. and Jones, T.A. (1997a) Detecting folding motifs and similarities in protein structures. *Methods Enzymol*, **277**, 525-545.

Kleywegt, G.J. and Jones, T.A. (1997b) Model-building & refinement practice. *Methods Enzymol*, **277**, 208-230.

Kobes, R.D. and Dekker, E.E. (1971) Variant properties of bovine liver 2-keto-4-hydroxyglutarate aldolase; its decarboxylase activity, lack of substrate stereospecificity, and structural requirements for binding substrate analogs. *Biochim. Biophys. Acta*, **250**, 238-250.

Koonin, E.V., Makarova, K.S. and Aravind, L. (2001) HORIZONTAL GENE TRANSFER IN PROKARYOTES: Quantification and Classification. *Annu Rev Microbiol*, **55**, 709-742.

Laskowski, R.A. (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph*, **13**, 323-330, 307-328.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J App Cryst*, **26**, 548-558.

Lau, E.Y. and Bruice, T.C. (1999) Consequences of breaking the Asp-His hydrogen bond of the catalytic triad: effects on the structure and dynamics of the serine esterase cutinase. *Biophys J.*, **77**, 85-98.

Lechevalier, M.P., Prauser, H., Labeda, D.P. and Ruan, J.-S. (1986) Two New Genera of Nocardioform Actinomycetes: *Amycolata* gen. nov. and *Amycolatopsis* gen. nov. *J System Bacteriol*, **36**, 29-37.

Lee, B. and Richards, F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol*, **55**, 379-400.

Lee, L.V., Vu, M.V. and Cleland, W.W. (2000) 13C and deuterium isotope effects suggest an aldol cleavage mechanism for L-ribulose-5-phosphate 4-epimerase. *Biochem*, **39**, 4808-4820.

Leslie, A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EAMCB newsletter on protein crystallography*, **No 26**, 1-10.

Liu, H.W. and Thorson, J.S. (1994) Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria. *Annu Rev Microbiol*, **48**, 223-256.

Lloyd, A.B. (1969) Behaviour of streptomycetes in soil. *J.Gen. Microbiol.*, **56**, 165-170.

Long, M. (2001) Evolution of novel genes. *Curr Opin Genet Dev*, **11**, 673-680.

Low, D.E. and Scheld, W.M. (1998) Strategies for Stemming the Tide of Antimicrobial Resistance. *JAMA*, **279**, 394-395.

Maas, S. (1999) Efficient and rapid procedure for blue-white screening of recombinant bacterial clones. *BioTechniques*, **27**, 1126-1128.

Machajewski, T.D. and Wong, C.-H. (2000) The Catalytic Asymmetric Aldol Reaction. *Angew Chem Int Ed Engl*, **39**, 1352-1374.

Mackay, J.P., Gerhard, U., Beauregard, D.A., Maplestone, R.A. and Williams, D.H. (1994a) Dissection of the Contributions toward Dimerization of Glycopeptide Antibiotics. *J Am Chem Soc*, **116**, 4573-4580.

Mackay, J.P., Gerhard, U., Beauregard, D.A., Westwell, M.S., Searle, M.S. and Williams, D.H. (1994b) Glycopeptide Antibiotic-Activity and the Possible Role of Dimerization - a Model for Biological Signaling. *J Am Chem Soc*, **116**, 4581-4590.

Mahal, L.K., Yarema, K.J. and Bertozzi, C.R. (1997) Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis. *Science*, **276**, 1125-1128.

Mahmoudian, M., Noble, D., Drake, C.S., Middleton, R.F., Montgomery, D.S., Piercey, J.E., Ramlakhan, D., Todd, M. and Dawson, M.J. (1997) An efficient process for production of N-acetylneuraminic acid using N-acetylneuraminic acid aldolase. *Enzyme Microb Technol*, **20**, 393-400.

Major, L.L. (2001) Is the prokaryotic termination signal a simple triplet codon or an extended sequence element? *Biology*. University of Otago, Dunedin, p. 252.

Martin-Farmer, J. and Janssen, G.R. (1999) A downstream CA repeat sequence increases translation from leadered and unleadered mRNA in *Escherichia coli*. *Mol Microbiol*, **31**, 1025-1038.

Marumo, K., Lindqvist, L., Verma, N., Weintraub, A., Reeves, P. R. and Lindberg, A. A. (1992) Enzymatic synthesis and isolation of thymidine diphosphate-6-deoxy-D-*xylo*-4-hexulose and thymidine diphosphate-L-rhamnose. *Eur. J. Biochem.*, **204**, 539-545.

Marwick, C. (1999) Animal Feed Antibiotic Use Raises Drug Resistance Fear. *JAMA*, **282**, 120-122.

Massova, I. and Mobashery, S. (1998) Kinship and diversification of bacterial penicillin-binding proteins and beta-lactamases. *Antimicrob Agents Chemother*, **42**, 1-17.

Matthews, B.W. (1968) Solvent content of protein crystals. *J Mol Biol*, **33**, 491-497.

Mavridis, I.M., Hatada, M.H., Tulinsky, A. and Lebioda, L. (1982) Structure of 2-keto-3-deoxy-6-phosphogluconate aldolase at 2.8Å resolution. *J Mol Biol*, **162**, 419-444.

Mayfield, C.I., Williams, S.T., Ruddick, S.M. and Hatfield, H.L. (1972) Studies on the ecology of actinomycetes in soil. IV Observations on the form and growth of streptomycetes in soil. *Soil Biol. Biochem*, **4**, 79-91.

McAtee, J.J., Castle, S.L., Jin, Q. and Boger, D.L. (2002) Synthesis and evaluation of vancomycin and vancomycin aglycone analogues that bear modifications in the residue 3 asparagine. *Bioorganic & Med Chem Lett*, **12**, 1319-1322.

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

McIlroy, R.J. (1967) *Introduction to Carbohydrate Chemistry*. Butterworths & Co. Limited, London.

Meloche, H.P. (1967) Bromopyruvate inactivation of 2-keto-3-deoxy-6-phosphogluconic aldolase. I. Kinetic evidence for active site specificity. *Biochem*, **6**, 2273-2280.

Meloche, H.P., Monti, C.T. and Dekker, E.E. (1975) Enzyme stereoselectivity: the reversible reaction catalyzed by 2-keto-4-hydroxyglutarate aldolase of *Escherichia coli*. *Biochem Biophys Res Commun*, **65**, 1033-1039.

Menez, J., Heurgué-Hamard, V. and Buckingham, R. (2000) Sequestration of specific tRNA species cognate to the last sense codon of an overproduced gratuitous protein. *Nucleic Acids Res*, **28**, 4725-4732.

Meyers, C.L.F., Oberthur, M., Anderson, J.W., Kahne, D. and Walsh, C.T. (2003) Initial characterization of novobiocic acid noviosyl transferase activity of NovM in biosynthesis of the antibiotic novobiocin. *Biochem*, **42**, 4179-4189.

Miesel, L., Greene, J. and Black, T.A. (2003) *Nature Rev Genet*, **4**, 442-456.

Morgan, P.M., Sala, R.F. and Tanner, M.E. (1997) Eliminations in the reactions catalysed by UDP-N-Acetylglucosamine 2-Epimerase. *J Am Chem Soc*, **119**, 10269-10277.

Morita, R.Y. (1985) Starvation and miniaturisation of heterotrophs, with special emphasis on maintenance of the starved viable state. In Floodgate, G.D. (ed.), *Bacteria in their Natural Environments*. Academic Press, London, pp. 111-130.

Morris, A., Kellner, J.D. and Low, D.E. (1998) The superbugs: evolution, dissemination and fitness. *Curr Opin Microbiol*, **1**, 524-529.

Morris, R.J., Perrakis, A. and Lamzin, V.S. (2002) ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallogr D*, **58**, 968-975.

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

Navaza, J. (1994) *AMoRe*: an automated package for molecular replacement. *Acta Crystallogr A*, **50**, 157-163.

Nathan, C. (2004) Antibiotics at the crossroads. *Nature*, **431**, 899-902

NEB (2005-2006) *Catalogue and Technical Reference*, New England BioLabs Inc., Beverly, MA, USA: p.366

NEB (2004) *pMAL™ Protein Fusion and Purification System*, New England BioLabs Inc., Beverly, MA, USA: p.22

Nicolaou, K.C., Mitchell, H.J., Jain, N.F., Winssinger, N., Hughes, R. and Bando, T. (1999) Total synthesis of vancomycin. *Angew Chem Int Ed*, **38**, 240-244.

Nixon, A.E. and Firestine, S.M. (2000) Rational and 'Irrational' Design of Proteins and Their Use in Biotechnology. *IUBMB Life*, **49**, 181-187.

Noble, W.C. (1997) Antibiotic resistance in the staphylococci. *Sci Prog*, **80 (Pt 1)**, 5-20.

Noble, W.C., Virani, Z. and Cree, R.G. (1992) Co-transfer of vancomycin and other resistance genes from *Enterococcus faecalis* NCTC 12201 to *Staphylococcus aureus*. *FEMS Microbiol Lett*, **72**, 195-198.

Novagen Catalogue (2002-2003), Novagen, The GOLD standard in protein Expression, CN Biosciences, UK: p. 88.

Novotny, M., Madsen, D. and Kleywegt, G.J. (2004) Evaluation of protein fold comparison servers. *Proteins*, **54**, 260-270.

O'Donovan, C., Martin, M.J., Gattiker, A., Gasteiger, E., Bairoch, A. and Apweiler, R. (2002) High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief. Bioinform.*, **3**, 275-284.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. In Sweet, R.M. (ed.), *Methods in Enzymology, Volume 276: Macromolecular Crystallography, part A,*. Academic Press, New York, pp. 307-326.

Page, P., Blonski, C. and Perie, J. (1996) An improved chemical and enzymatic synthesis of new fructose derivatives for import studies by the glucose transporter in parasites. *Tetrahedron*, **52**, 1557-1572.

Pagni, M., Iseli, C., Junier, T., Falquet, L., Jongeneel, V. and Bucher, P. (2001) TrEST, trGEN and Hits: Access to databases of predicted protein sequences. *Nucleic Acids Res*, **29**.

Pootoolal, J., Neu, J. and Wright, G.D. (2002) Glycopeptide antibiotic resistance. *Annu Rev Pharmacol Toxicol*, **42**, 381-408.

Projan, S.J. (2003) Why is big Pharma getting out of antibacterial drug discovery. *Curr Opin Microbiol*, **6**, 427-430

Promega Catalogue (2004), Life Science Catalog, Promega, Madison, WI, USA: p.406

Qiagen (2003), *QIAprep® Miniprep Handbook*, Qiagen, Crawley, UK, p.46

Qiagen (2002), *QIAquick® Spin Handbook*, Qiagen, Crawley, UK, p.34

Ramachandran, G.N. and Sasisekharan, V. (1968) Conformation of polypeptides and proteins. *Adv. Prot Chem*, **23**, 283-438.

Read, R.J. (2003) Pushing the boundaries of molecular replacement with maximum-likelihood. *Acta Crystallogr D*, **57**, 1373-1382.

Read, R.J. and Schierbeek, A.J. (1988) A phased translation function. *J Appl Cryst*, **21**, 490-495.

Redinbo, M.R. and Yeates, T.O. (1993) Structure Determination of Plastocyanin from a Specimen with a Hemihedral Twinning Fraction of One-Half. *Acta Crystallogr D*, **49**, 375-380.

Rees, D.C. (1980) The Influence of Twinning by Merohedry on Intensity Statistics. *Acta Crystallogr A*, **36**, 578-580.

Rees, D.C. (1982) A general theory of x-ray intensity statistics for twins by merohedry. *Acta Crystallogr A*, **38**, 201-207.

Reeves, P.R., Hobbs, M., Valvano, M.A., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J., Maskell, D., Raetz, C.R. and Rick, P.D. (1996) Bacterial polysaccharide synthesis and gene nomenclature. *Trends Microbiol*, **4**, 495-503.

Riefler, R.G. and Smets, B.F. (2002) NAD(P)H:Flavin Mononucleotide Oxidoreductase Inactivation during 2,4,6-Trinitrotoluene Reduction. *Appl. Environ. Microbiol.*, **68**, 1690-1696.

Rodriguez, L., Aguirrezabalaga, I., Allende, N., Brana, A.F., Mendez, C. and Salas, J.A. (2002) Engineering deoxysugar biosynthetic pathways from antibiotic-producing microorganisms. A tool to produce novel glycosylated bioactive compounds. *Chem Biol*, **9**, 721-729.

Rodriguez, L., Oelkers, C., Aguirrezabalaga, I., Brana, A.F., Rohr, J., Mendez, C. and Salas, J.A. (2000) Generation of hybrid elloramycin analogs by combinatorial biosynthesis using genes from anthracycline-type and macrolide biosynthetic pathways. *J Mol Microbiol Biotechnol*, **2**, 271-276.

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

Sala, R.F., Morgan, P.M. and Tanner, M.E. (1996) Enzymatic formation and release of a stable glycal intermediate: the mechanism of the reaction catalysed by UDP-N-acetylglucosamine 2-epimerase. *J Am Chem Soc*, **118**, 3033-3034.

Salo, W.L., Fossitt, D.D., Bevill, R.D., 3rd, Kirkwood, S. and Wood, W. (1972) L-ribulose 5-phosphate 4-epimerase from *Aerobacter aerogenes*. Kinetic isotope effect with tritiated substrate. *J Biol Chem*, **247**, 3098-3100.

Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning, a laboratory manual*. Cold Spring Harbor Lanoratory Press, New York.

Schaberg, D.R., Rubens, C.E., Alford, R.H., Farrar, W.E., Schaffner, W. and McGee, Z.A. (1981) Evolution of antimicrobial resistance and nosocomial infection. Lessons from the Vanderbilt experience. *Am J Med*, **70**, 445-448.

Schaffner, J., Winter, J., Rudolph, R. and Schwarz, E. (2001) Cosecretion of chaperones and low-molecular size medium additives increases the yield of recombinant disulfide-bridged proteins. *Appl. Environ. Microbiol.*, **67**, 3994-4000.

Schein, C. (1990) Solubility as a function of protein structure and solvent components. *Bio/Technology*, **8**, 308-317.

Schenk, P.M., Baumann, S., Mattes, R. and Steinbiß, H.-H. (1995) Improved High-Level Expression System for Eukaryotic Genes in *Escherichia coli* Using T7 RNA Polymerase and Rare ^{Arg}tRNAs. *BioTechniques*, **19**, 196-198.

Schmid, A., Dordick, J.S., Hauer, B., Kiener, A., Wubbolts, M. and Witholt, B. (2001) Industrial biocatalysis today and tomorrow. *Nature*, **409**, 258-268.

Schmidt, D.E. and Westheimer, F.H. (1971) PK of the lysine amino group at the active site of acetoacetate decarboxylase. *Biochem*, **10**, 1249-1253.

Schneider, T.D., Stormo, G.D., Gold, L. and Ehrenfeucht, A. (1986) Information content of binding sites on nucleotide sequences. *J Mol Biol*, **188**, 415-531.

Schoolingin-Jordan, P.M., Spencer, P., Sarwar, M., Erskine, P.E., Cheung, K.-M.-. and Norton, E.B. (2002) 5-Aminolaevulinic acid dehydratase: metals, mutants and mechanism. *Biochem Soc Transactions*, **30**, 584-590.

Schurmann, M. and Sprenger, G.A. (2001) Fructose-6-phosphate aldolase is a novel class I aldolase from *Escherichia coli* and is related to a novel group of bacterial transaldolases. *J Biol Chem*, **276**, 11055-11061.

Sharman, G.J. and Williams, D.H. (1997) Common factors in the mode of action of vancomycin group antibiotics active against resistant bacteria. *Chem Comm*, 723-724.

Sheldrick, G.M. (1992) Shelx. In Moras, D., Podjarny, A.D. and Thierry, J.C. (eds.), *Crystallographic Computing*. Oxford University Press and I. U. Cr., Oxford, UK, pp. 145-157.

Shelton, M.C., Cotterill, I.C., Novak, S.T.A., Poonawala, R.M., Sudarshan, S. and Toone, E.J. (1996) 2-keto-3-deoxy-6-phosphogluconate Aldolases as Catalysts for Stereocontrolled Carbon-Carbon Bond Formation. *J Am Chem Soc*, **118**, 2117-2125.

Shibata, K., Shingu, K., Vassilev, V.P., Nishide, K., Fujita, T., Node, M., Kajimoto, T. and Wong, C.-H. (1996) Kinetics and Thermodynamic Control of L-threonine Aldolase Catalysed Reaction and its Application to the Synthesis of Mycestericin D. *Tetrahedron Lett.*, **37**, 2791-2794.

Sieradzki, K., Roberts, R.B., Haber, S.W. and Tomasz, A. (1999) The development of vancomycin resistance in a patient with methicillin-resistant St*aphylococcus aureus* infection. *N Engl J Med*, **340**, 517-523.

Skerman, V.B.D., McGowan, V. and Sneath, P.H.A. (1980) Approved lists of bacterial names. *Int. J. Syst. Bacteriol.*, **30**.

Smith, P.K. and Krohn, R.I. (1985) Measurement of protein using bicinchoninic acid. *Anal. Biochem.*, **150**, 76-85.

Sowden, J.C. and Fischer, H.O.L. (1947) The Condensation of Nitromethane with D- and L-Arabinose: Preparation of L-Glucose and L-Mannose. *J Am Chem Soc*, **69**, 1963-1965.

Stackebrandt, E., Rainey, F.A. and Ward-Rainey, N.L. (1997) Proposal for a new hierachic classification system, Actinobacteria classis nov. *Int. J. Syst. Bacteriol.*, **47**, 479-491.

Stanley, E. (1972) The Identification of Twins from Intensity Statistics. *J Appl Cryst*, **5**, 191-194.

Steitz, J.A. (1969) Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in baceriophage R17 RNA. *Nature*, **224**, 957-964.

Steitz, T.A. and Shulman, R.G. (1982) Crystallographic and NMR studies of the serine proteases. *Annu Rev Biophys Bioeng*, **11**, 419-444.

Stephenson, J. (2001) Researchers Describe Latest Strategies to Combat Antibiotic-Resistant Microbes. *JAMA*, **285**, 2317-2318.

Stephenson, J. (2002) Researchers Wrestle With Spread and Control of Emerging Infections. *JAMA*, **287**, 2061-2063.

Storoni, L.C., McCoy, A.J. and Read, R.J. (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr D*, **60**, 432-438.

Strenström, C.M., Jen, H., Major, L.L., Tate, W.P. and Isaksson, L.A. (2001) Codon bias at the 3'- side of the initiation codon is correlated with translational initiation efficiency in *Escherichia coli*. *Gene*, **263**.

Stryer, L. (1995a) *Biochemistry*. W. H. Freeman and Company, New York.

Stryer, L. (1995b) Carbohydrates. In *Biochemistry*. W. H. Freeman and Company, New York, p. 1000.

Stryer, L. (1995c) Glycolysis. In *Biochemistry*. W.H. Freeman and Company, New York, p. 1000.

Sutcliffe, J., Tait-Kamradt, A. and Wondrack, L. (1996) *Streptococcus pneumoniae* and *Streptococcus pyogenes* resistant to macrolides but sensitive to clindamycin: a common resistance pattern mediated by an efflux system. *Antimicrob Agents Chemother*, **40**, 1817-1824.

Taylor, H.O. and Leslie, A.G.W. (1998) A Program to Detwin Merohedrally Twinned Data. *CCP4 Newsletter*, **35**, 9.

Ten Eyck, L.F. (1973) Crystallographic fast Fourier transforms. *Acta Crystallogr A*, **29**, 183-191.

Tharp, R. (2004) *RxKinetics: software solutions for pharmacists*. RxKinetics, Vol. 2004.

Thoden, J.B., Frey, P.A. and Holden, H.M. (1996a) Crystal structures of the oxidized and reduced forms of UDP-galactose 4-epimerase isolated from *Escherichia coli*. *Biochem*, **35**, 2557-2566.

Thoden, J.B., Frey, P.A. and Holden, H.M. (1996b) High-resolution X-ray structure of UDP-galactose 4-epimerase complexed with UDP-phenol. *Protein Sci*, **5**, 2149-2161.

Thoden, J.B., Frey, P.A. and Holden, H.M. (1996c) Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from Escherichia coli: implications for the catalytic mechanism. *Biochem*, **35**, 5137-5144.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673-4680.

Thorell, S., Schurmann, M., Sprenger, G.A. and Schneider, G. (2002) Crystal structure of decameric fructose-6-phosphate ldolase from *Escherichia coli* reveals inter-subunit helix swapping as a structural basis for assembly differences in the transaldolase family. *J Mol Biol*, **319**, 161-171.

Toraya, T. (2000) Radical catalysis of B12 enzymes: structure, mechanism, inactivation, and reactivation of dial and glycerol dehydratases. *Cell Mol Life Sci*, **57**, 106-127.

Trefzer, A., Blanco, G., Remsing, L., Kunzel, E., Rix, U., Lipata, F., Brana, A.F., Mendez, C., Rohr, J., Bechthold, A. and Salas, J.A. (2002) Rationally designed

glycosylated premithramycins: hybrid aromatic polyketides using genes from three different biosynthetic pathways. *J Am Chem Soc*, **124**, 6056-6062.

Trias, J. and Gordon, E.M. (1997) Innovative approaches to novel antibacterial drug discovery. *Current Opinion in Biotechnology*, **8**, 757-762.

Tsoi, C.J. and Khosla, C. (1995) Combinatorial biosynthesis of 'unnatural' natural products: the polyketide example. *Chem Biol*, **2**, 355-362.

Vagin, A., Richelle, J. and Wodak, S.J. (1999) SFCHECK: a unified set of procedure for evaluating the quality of macromolecular structure-factor data and their agreement with atomic model. *Acta Crystallogr D*, **55**, 191-205.

Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J Appl Cryst*, **30**, 1022-1025.

van Duyne, G.D. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol*, **229**, 105-124.

Vastag, B. (2002) "Cipromania" and "Superclean" Homes Are Now Increasing Antibiotic Resistance. *JAMA*, **288**, 947-948.

Verlinde, C.L.M.J. and Quigley, P.M. (1999) Structure-based Re-evaluation of the Mechanism of Class I Fructose-1,6-bisphosphate Aldolase. *J Mol Biol*, **5**, 37-45.

Voet, D. and Voet, J.G. (1995) *Biochemistry*. John Wiley & Sons Inc., New York.

Vyas, N.K. (1991) Atomic features of protein-carbohydrate interactions. *Curr Opin Struc Biol*, **1**, 732-740.

Waksman, S.A. (1952) Streptomycin: background, isolation, properties, and utilization. *Nobel Lectures*, Stockholm, pp. 370-388.

Watanakunakorn, C. (1984) Mode of Action and Invitro Activity of Vancomycin. *J Antimicrob Chemotherapy*, **14**, 7-18.

Wellington, E.M., Cresswell, N. and Herron, P.R. (1992) Gene transfer between streptomycetes in soil. *Gene*, **115**, 193-198.

Wenzel, R.P.N. (2004) The antibiotic Pipeline - Cahallenges, Costs, and Value, *N Engl J Med*, **351**, 523-526

Westheimer, F.H. (1995) Coincidences, decarboxylation, and electrostatic effects. *Tetrahedron*, **51**, 3-20.

WHO. (2001) The World Health Report. WHO, **2001**.

WHO. (2003) The World Health Report. WHO. **2003**

WHO. (2004) The World Health Report. WHO, **2004**

Wiechelman, K.J. (1988) Investigation of the bicinchoninic acid protein assay: identification of the groups responsible for color formation. *Anal. Biochem.*, **175**, 231-237.

Wiener, P. (1996) Experimental studies on the ecological role of antibiotic production in bacteria. *Evol. Ecol.*, **10**, 405-421.

Wierenga, R.K., De Maeyer, C.H. and Hol, W.G.J. (1985) Interaction of pyrophosphate moeties with alpha-helices in dinucleotide binding proteins. *Biochemistry*, **24**, 1346-1357.

Wierenga, R.K. and Hol, W.G.J. (1983) Predicted nucleotide-binding properties of p21 protein and its cancer-associated variant. *Nature*, **302**, 842-844.

Wilson, A.J.C. (1949) The probability distribution of X-ray intensities. *Acta Crystallogr D*, **2**, 318-321.

Winn, M.D., Isupov, M.N. and Murshudov, G.N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D*, **57**, 122-133.

Wymer, N., Buchanan, L.V., Henderson, D., Mehta, N., Botting, C.H., Pocivavsek, L., Fierke, C.A., Toone, E.J. and Naismith, J.H. (2001) Directed Evolution of a New Catalytic Site in 2-Keto-3-Deoxy-6-Phosphogluconate Aldolase from *Escherichia coli*. *Structure*, **9**, 1-8.

Wymer, N. and Toone, E.J. (2000) Enzyme-catalyzed synthesis of carbohydrates. *Curr Opin Chem Biol*, **4**, 110-119.

Yang, F., Dauter, Z. and Wlodawer, A. (2000) Effects of crystal twinning on the ability to solve a macromolecular structure using multiwavelength anomalous diffraction. *Acta Crystallogr D*, **56**, 959-964.

Yang, K., Han, L. and Vining, L.C. (1995) Regulation of jadomycin B production in *Streptomyces venezuelae* ISP5230: involvement of a repressor gene, jadR2. *J Bacteriol*, **177**, 6111-6117.

Yeates, T.O. (1988) Simple statistics for intensity data from twinned specimens. *Acta Crystallogr A*, **44**, 142-144.

Yeates, T.O. (1997) Detecting and overcoming crystal twinning. *Methods Enzymol*, **276**, 344-358.

Yeates, T.O. and Fam, B. (1997) The Merohedral Crystal Twinning Server. UCLA, Vol. 2004.

Yeates, T.O. and Fam, B. (1999) Protein crystals and their evil twins. *Structure Fold Des*, **7**, 25-29.

Yuan, Y.P., Eulenstein, O., Vingron, M. and Bork, P. (1998) Towards detection of orthologues in sequence databases. *Bioinformatics*, **14**, 285-289.

Zdobnov, E.M. and Apweiler, R. (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, **17**, 847-848.

Zhang, L., Radziejewska-Lebrecht, J., Krajewska-Pietrasik, D., Toivanen, P. and Skurnik, M. (1997) Molecular and chemical characterization of the lipopolysaccharide O-antigen and its role in the virulence of *Yersinia enterolitica* serotype O:8. *Mol Microbiol*, **23**, 63-76.

Appendices

Appendix A – Primers

All primers were custom made by Oswel (now Eurogentec Ltd.).

EvaD - Site-Directed Mutagenesis

I50N:	F:	GTGTGGCACAGACGAACCACAGCATGTCC
	R:	GGACATGCTGTGGTTCGTCTGTGCCACAC
H63A:	F:	GTGCGGGGCATCGCCTACACCGTGACG
	R:	CGTCACGGTGTAGGCGATGCCCCGCAC
M131F:	F:	GAGGACGACACCGTCTTCTCGTACATGCTCTCC
	R:	GGAGAGCATGTACGAGAAGACGGTGTCGTCCTC
Y133F:	F:	GACACCGTCATGTCGTTCATGCTCTCCAG
	R:	CTGGAGAGCATGAACGACATGACGGTGTC
L135A:	F:	GTCATGTCGTACATGGCCTCCAGAAGCTATGTGACG
	R:	CGTCACATAGCTTCTGGAGGCCATGTACGACATGAC
M131F/L135A:	F:	GACGACACCGTCTTCTCGTACATGGCCTCCAGAAGCTATGTG
	R:	CACATAGCTTCTGGAGGCCATGTACGAGAAGACGGTGTCGTC

The bases introducing amino acid changes are highlighted in red.

EvaA - Amplification from pET22b+ Vector

For cloning into pEHISTEV and pLou3 vecotrs:

F: GGT CGT CCC ATG GCG TCC TTC GTC

R: CGG ATC GAA TTC TCA TGC ACC TCC C

The NcoI and EcoRI recognition sites are highlighted in red and green respectiveley.

For cloning into pMAL-c2X and pMAlp2X vecotrs:

F: GGT GAA TTC ATG TCG TCC TTC GTC

R: CGG ATC AAG CTT TCA TGC ACC TC

The EcoRI and HindIII restiction sites are highlighted in green and yellow respectively

EvaA Homologues

A. mediterranei F: GTCGTCCCCATGGTGCCTGACCTCGTTCCCCCGGTCGTGG

R: GTGACCGTGAATTCGTCAGGCACCGCCTCGAGGCTGGGCCG

S. antibioticus F: CCTGGATCATCATGATATGGGGAATTCCCGCGATGAGCGAAGCAATG GGATCGGTACCGAC

R: CATGGGGTAAGCTTCGGGGTGCTCAGCTCAGGGCCTGGATGCAGGCG ACG

S. avermitilis F: CGTGGGGTTCATGAGCGTACGAGCCGATGCCGACCACACCGAGC

R: GTACGGGCCGGATCCGGAGTCACAGCTCGACGGCGCGGGTGG

S. erythrea F: CTTACAGTTCATGAGGGTCTTGATCGACAACGCCCGGCGGCAGCAAG

R: AGATCGCCAGAATTCTCAGCGCGCGTTTCATCCGGCGGTG

S. fradiae F: GGGAGCCTTCCCATGGTTTCTGAAAGCCTTCTCGCAGTGGCCCAGCC GCGTGCGG

R: GATCCCCACGGATCCGCTCATCGGCTGTTCACCGCCATGGCGTTGAG GCAGGCCAG

S. nogalater F: AGTCAGCCTCATGACGAAACTGTCCGCCCACCCGGCCGCCCCCGCG

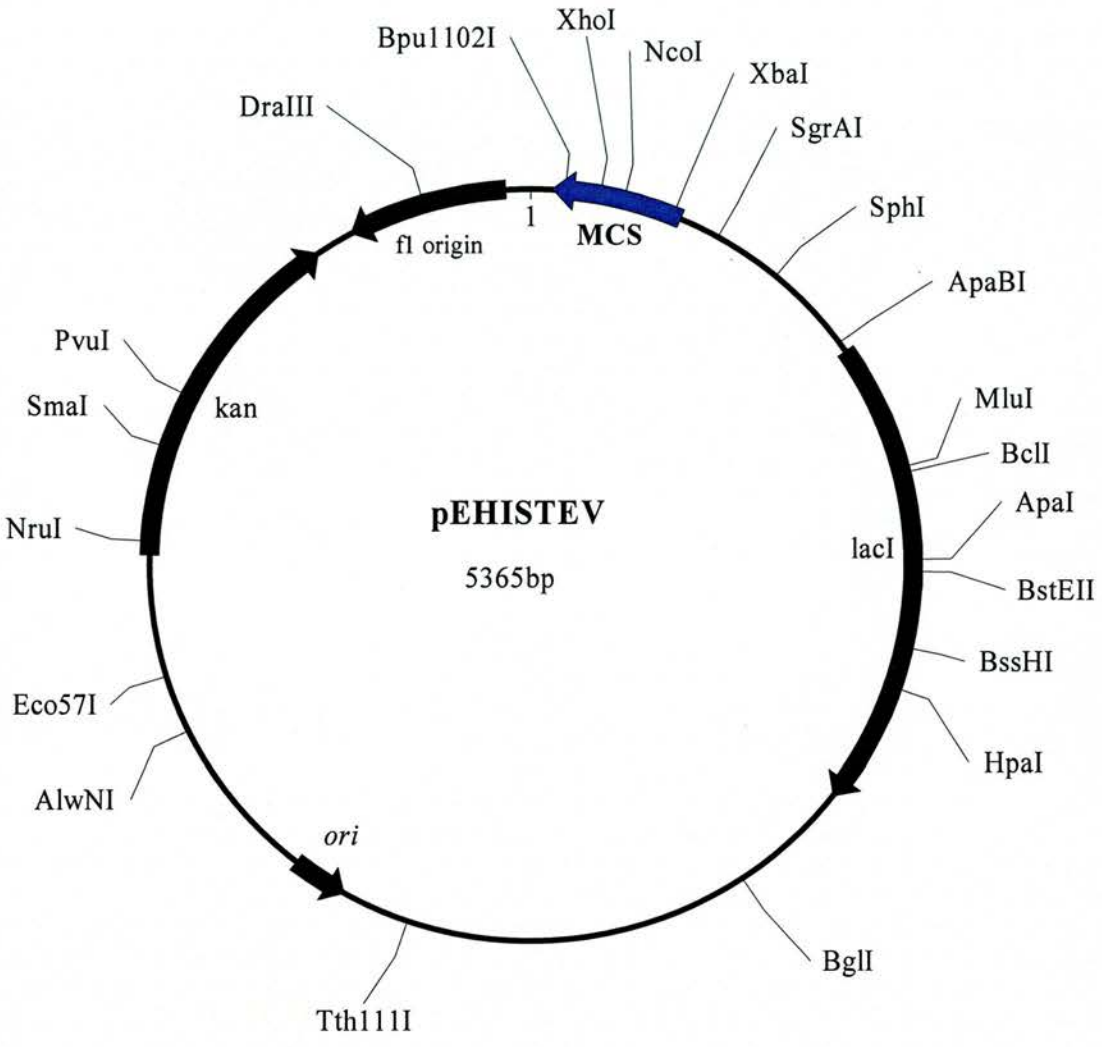
R: CTGGGCAGAATTCGCTCAGCGCCTCGGCGGCGCCGTGGCGAGGCC

S.violaceoruber F: GTGGCGAGCTCATGAGAATCACGGACACGGCCGGCTTCCATG

R: TCGGTGTGGAATTCGTTCACCGGGATGCTCCGTGGTCGGGCAAGG

The restriction sites are highlighted in red – NcoI, pink – BspHI, blue - BamHI, green – EcoRI, yellow – HindIII

Appendix B – Vector Maps: pEHISTEV



T7 Pro seq primer

T7 promoter lac operator *Xba* I

ATCGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTG

rbs *Nde* I 6xHis

TTTAACTTTAAGAAGGAGATATACAT ATG TCG TAC TAC CAT CAC CAT CAC CAT CAC GAT TAC GAT ATC CCA

M S Y Y H H H H H H D Y D I P

TEV protease site *Nco* I *BamH* I *EcoR* I *Sac* I

ACG ACC GAA AAC CTG TAT TTT CAG GGC **GCC ATG GCT GAT ATC GGA TCC GAA TTC GAG CTC CGT**

T T E N L Y F Q* G **A M A D I G S E F E L R**

Sal I *Hind* III *Not* I *Xho* I

CGA CAA GCT TGC GGC CGC ACT CGA GCA CCA CCA CCA CCA CCA CTG AGA TCC GGC TGC TAA CAA

R Q A C G R T R A P P P P P L R S G C end

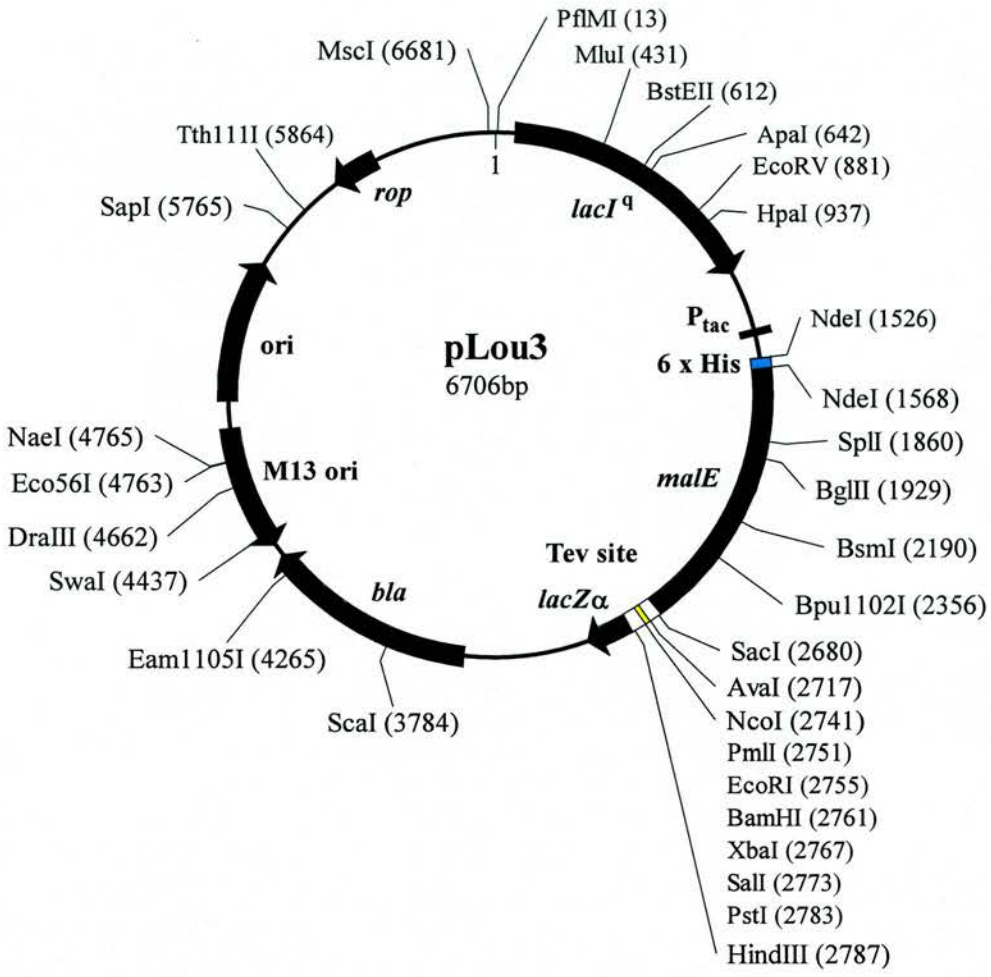
Bpu1102I T7 terminator

AGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGG

T7 Ter seq primer

GGTTTTTTGCTGAA

Appendix B – Vector Maps: pLou-3



pLou3 was derived from the NEB vector pMAL-c2X so has cytoplasmic expression of the fusion protein. New features are N-terminal 6xHis tag before MBP and a Tev protease cleavage site between MBP and the (modified) polylinker. Unique restriction sites are shown on the map (+ NdeI which cuts twice).

pLOU3 Tev cleavage site and cloning sites:

```
AvaI          TEV protease site          NcoI       PmlI    EcoRI   BamHI
CTC GGG GAA AAC CTG TAT TTT CAG GGC GCC ATG GAT CAC GTG GAA TTC GGA TCC
 L   G   E   N   L   Y   F   Q   G   A   M   D   H   V   E   F   G   S

 XbaI     SalI     PstI     HindIII
TCT AGA GTC GAC CTG CAG GCA AGC TTG GCA
 S   R   V   D   L   Q   A   S   L   A
```

Sequence for the 6xHis tag:

```
NdeI                    His-tag                          NdeI
CAT ATG AAA TAT TAC CAT CAC CAT CAC CAT CAC GAT TAC GAT CAT ATG
     M   K   Y   Y   H   H   H   H   H   H   S   Y   S   H   M
```

Appendix C – Rare Codon Results

Red = rare Arg codons **AGG, AGA, CGA**

Green = rare Leu codon **CTA**

Blue = rare Ile codon **ATA**

Orange = rare Pro codon **CCC**

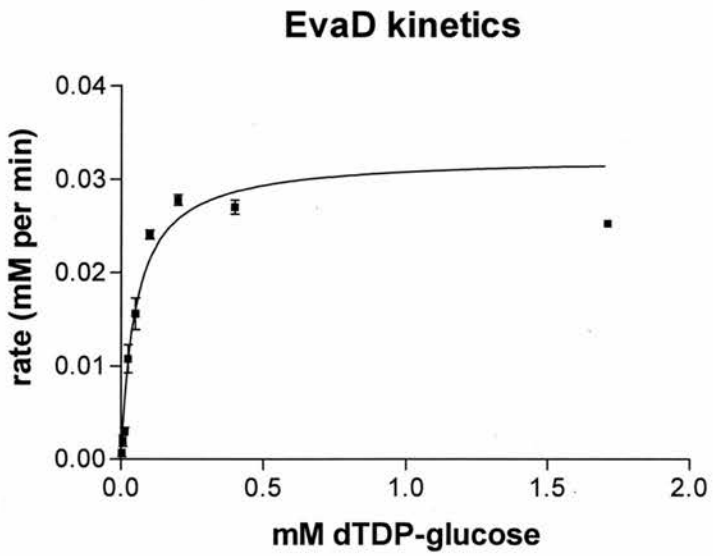
atg tcg tcc ttc gtc gtt cca tca ctc acc gcg gtg cgc cct cgt gac cac cac
gac tac gcc gac cgc atc gcc ttg tcg gcc gcg acc acc gat ggg gta cag atg
cgg acc gag gat gtc cgc gcc tgg atc gcc gaa cgc cgc gac gcc aac gtc ttc
cac gtc gaa CGA atc ccg ttc gcc gac ctg gac cag tgg tgg ttc gaa ggc gtg
acc ggc aat ctg gtg cac cgc agc ggg cgg ttc ttc acc atc gag ggc ctg cac
gtg atc gag cac gac ggc ccg cac ggc gac ggc ccg tac cgc gag tgg cag cag
ccg gtc atc AGG cag CCC gaa gtg ggc atc ctc ggc atc ctg gcc aag gag ttc
gac ggg gtg ctg cac ttc ctg atg cag gcc aag atg gag ccg ggc aac CCC aac
ctg gtg cag ctc tcg ccg acc gtg cag gcc acc cgg agc aac tac acc aag gcg
cac ggc ggt acg aac gtc aag ctg atc gag tac ttc gcg ccg ccg gac CCC gag
cgg gtc atc gtc gac gtt ctc cag gcg gag caa gga tcc tgg ttc ttc cgc aag
tcc aat cgc aac atg atc gtc gag acc gtc gac gac gtg ccg ttg tgg gac gac
ttc tgc tgg ctc acc ctc ggc cag atc gcc gaa ctg atg cac gag gac gag acg
atc aac atg aac tcc CGA agc gtg ttg tcg tgt ctg CCC tac cag gac ATA act
cct cgc gcg ctg ttc tcc gac gtt cag ctc ctg tcg tgg ttc acc aac gag cgt
tcg cgt cac gat gtg cgc gtc cgc cgc att ccg ctc gcg gac gtg tgc ggc tgg
aag caa ggc gcc gag gag atc gag cac gag gac ggc cgt tac ttc aag gtc ctc
gcg gtc gcc gtg aaa ggc agc aac cgc gag aag atc agt tgg acc cag ccg ttg
gtc gaa tcc gtc gac ttg ggt gtc gtc gcg ttc ctc gtg cgg aag atc gac ggt
gtg ccg cac gtt ctg gtg cag gcc cgc gtt gac ggt ggt ttc ctg gac acg gtc
gag ttg gcg ccg acc gtc cag tgc acg CCC ctc aac tac gca cac ctg CCC gcg
gag gag cgt ccg ccg ttc ctt gac ctc gtc cag aac gca CCC cgg tcg CGA atc
cgt tac gag gcg atc cat tcc gaa gaa ggc ggc cgc ttc ctc ggc gtc cgg gcc
CGA tat ctc gtg atc gac gcg gac gag gca atc gat cca ccg cct ggc tac gcc
tgg gtc acc ccg gcg cag ctc acc gcg ctc acc cgg cac ggg cac tac gtc aac
gtc gag gct cgc acg ctt ctc gcg tgc atc aac gcc gcg gcg gcc cag cct cgg
gga ggt gca tga

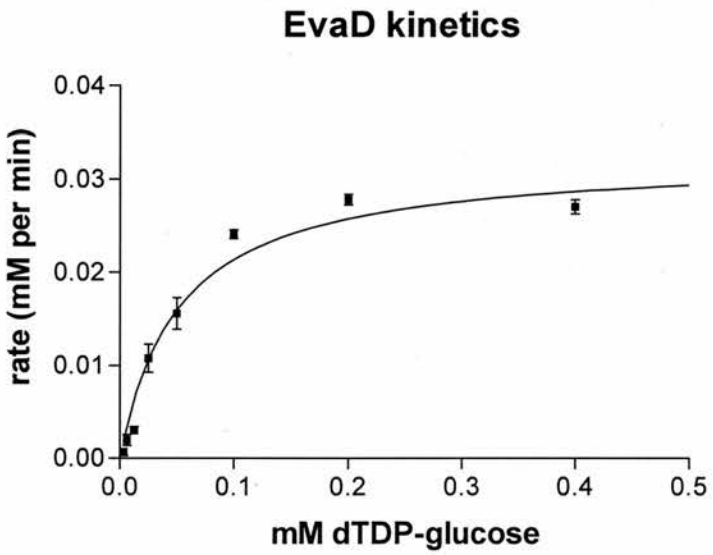
Number of total single rare Arg codons: **5**

Number of tandem rare Arg codon double repeats: **0**

Number of tandem rare Arg codon triple repeats: **0**

Appendix D - Kinetic Data EvaD





The program Graph Pad Prism 3 was used to analyse the data obtained for the biochemical studies. The non-linear regression was fitted to the curve resulting in 95% confidence in our results with good errors (presented at the appropriate place in the text).

Publications

Fullerton, S.W.F., Griffiths, J.S., **Merkel, A.B.**, Wymer, N.J., Hutchins, M.J., Fierke, C.A., Toone, E.J., Naismith, J.H. (2004) Mechanism of the Class I KDPG Aldolase. *Biochemistry* (submitted)

Merkel, A.B., Major, L.L., Errey, J.C., Burkhart, M.D., Field, R.A., Walsh, C.T., Naismith, J.H. (2004) The Position of a Key Tyrosine in dTDP-4-keto-6-deoxy-D-glucose-5-epimerase (EvaD) Alters the Substrate Profile for This RmlC-like Enzyme. *J Biol Chem*, **279**: 32683-32691

Merkel, A.B., Temple, G.K., Burkart, M.D., Losey, H.C., Beis, K., Walsh, C.T., Naismith, J.H. (2002) Purification, Crystallization and Preliminary Structural Studies of dTDP-4-keto-6-deoxy-glucose-5-epimerase (EvaD) from *Amycolatopsis orientalis*, the Fourth Enzyme in the dTDP-L-epivancoasmine Biosynthetic Pathway. *Acta Crystallogr D Biol Crystallogr*, **58**, 7: 126-128

Mechanism of the Class I KDPG aldolase

Stephen W.B. Fullerton[1,2], Jennifer S. Griffiths[3], Alexandra B. Merkel[1], Nathan J. Wymer[3,4], Michael J. Hutchins[1], Carol A Fierke[5], Eric J. Toone[3] & James H. Naismith[1]

[1]Centre for Biomolecular Sciences, The University, St Andrews, KY16 9ST UK

[2]Current address: Department of Biochemistry, Adrian building, University of Leicester, Leicester, LE17RH

[3]Department of Chemistry, Duke University, Durham, NC 27708 USA

[4]Current address: zuChem, Inc., Peoria, IL 61604

[5]Chemistry Department, University of Michigan, Ann Arbor, MI 48109

Abstract

2-Keto-3-deoxy-6-phosphogluconate (KDPG) aldolase catalyzes the reversible, stereospecific retro-aldol cleavage of KDPG into pyruvate and D-glyceraldehyde-3-phosphate. KDPG aldolase is a lysine-dependent (Class I) aldolase that functions through the intermediacy of a Schiff base. Here we report the structure of KDPG aldolase from the thermophile *Thermotoga maritima* determined to 1.9Å. The structure is the standard α/β barrel observed for all Class I aldolases with the active site Lys on β strand 6. At the active site Lys 129 we observe clear density for a pyruvate Schiff base. Density for a sulfate ion bound in a conserved cluster of residues close to the Schiff base is also observed. We have determined the structure of a mutant of *E. coli* KDPG aldolase in which the proposed general acid/base catalyst has been removed (E45N). One subunit of the trimer contains density suggesting a trapped pyruvate carbinolamine intermediate. All three subunits contain a bound phosphate ion. The location of the covalent intermediate and the anion are effectively identical in both enzymes. The sulfate and phosphate ions experimentally locate the putative phosphate binding site of the aldolase and on this basis we propose a model of the KDPG substrate complex. Our model requires only minimal positional adjustments of the experimentally determined covalent intermediate and bound anion. The model provides a plausible explanation why KDPG aldolase is a trimer. Furthermore, the model identifies the key catalytic residues and suggests that two water molecules play important roles in catalysis. The first water remains bound to the enzyme during the entire catalytic cycle, shuttling protons between Glu 40 (conserved putative base, *T. maritima* numbering) and substrate. The second water molecule arises from formation of the Schiff base and serves as the nucleophilic water during hydrolysis of the enzyme-product Schiff base. The second water may also mediate the base-catalyzed enolization required to form the carbon nucleophile, by bridging to Glu 40. Many aspects of this mechanism may be conserved throughout the Class I aldolases.

Introduction

2-Keto-3-deoxy-6-phosphogluconate (KDPG) aldolase, a class I aldolase of the Enter-Doudoroff glycolytic pathway (1), cleaves KDPG into the three-carbon units pyruvate and glyceraldehyde-3-phosphate. The mechanism proceeds through an imine, formed with an active site lysine, which undergoes retro-aldol carbon-carbon bond cleavage (Scheme 1). In the synthetic direction all aldolases (including KDPG aldolase) catalyze carbon-carbon bond formation, the central reaction of chemical synthesis. In both the synthetic and cleavage reactions, a series of proton transfers are facilitated by catalytically relevant acidic and basic sidechains.

Scheme 1 (KDPG aldolase reaction) Here

The enzymatic aldol reaction is highly efficient, regioselective and, in many cases, shows high facial stereoselectivity: for these reasons aldolases have long been of interest to the synthetic community (2). On the other hand, the limited substrate specificities of many aldolases limit their utility in chemical synthesis. Despite initial promise, rational redesign of enzyme active sites is not generally successful. Directed evolution has produced some notable successes and offers an alternative methodology for improving the utility of enzymes as synthetic catalysts (3-6) (Liu & greenblat). In addition to providing useful tools, directed evolution facilitates an understanding of structure-function relationships in protein catalysis. The interpretation of such experiments requires a detailed structural knowledge of the active site, including the substrate recognition regions. As part of a wider programme of directed evolution, we have investigated the structural basis of substrate specificity and the mechanism of catalysis of the *Thermotoga maritima* KDPG aldolase.

The first structures of class I aldolases located the active site lysine and identified catalytically important residues, including the presumed general base. More recently, substrate-bound co-complexes have been reported for aldolases from several sources (7-9). Nonetheless, several important mechanistic questions remained unresolved. The next important advance came with the

structure of a bound carbinolamine covalent intermediate at the active site of the *E. coli* KDPG aldolase (10). The structure identifies Glu 45 as the only plausible general base, although the orientation of the carbinolamine precludes direct deprotonation by the glutamate. Because the structure was trapped by crystallization at low pH, where protonation of the carbinolamine nitrogen inhibits imine formation, the relevance of this rotameric form to catalysis was unclear. Another major advance towards an understanding of aldolase mechanism came from a study of D-2-deoxyribose-5-phosphate aldolase (DERA)(11). DERA catalyzes a very similar reaction to KDPG aldolase; the principal difference is that DERA uses the aldehydic nucleophile acetaldehyde while KDPG aldolase uses the ketone pyruvate to react with glyceraldehydes-3-phosphate. The DERA study proposed a complex reaction mechanism involving the cyclic relay of protons (Scheme 2) between water and three residues during turnover. However, this mechanism cannot be entirely general since two of the three residues (Asp102 and Lys201) are not conserved in other aldolases. Studies on class II enzymes lagged behind although a comprehensive series of studies have now been reported(12-15).

Scheme 2 (DERA Mechanism) Here

We have previously reported the structure of the *E. coli* KDPG aldolase and an evolved mutant showing altered substrate specificity (16). Here we report the structure of *T. maritima* KDPG aldolase (17) covalently modified as the Schiff base complex with pyruvate and containing a tightly bound sulfate ion. We have also determined a high resolution pyruvyl carbinolamine complex with a catalytically incompetent mutant of *E. coli* KDPG aldolase. This protein contains a phosphate ion bound in the same locus as the *T. maritima* sulfate ion, even though the proteins were crystallised under different conditions. By assuming that these ions reveal the substrate phosphate binding site, we have created a model of KDPG-bound aldolase. The model suggests that proper recognition of the substrate requires the unusual trimeric structure of KDPG aldolases. The data also provide insight into the catalytic mechanism of KDPG aldolases. The proposed mechanism has some similarities with that proposed for the DERA, including a crucial role for an active site water. There are, however, differences in the location, identity, and role of the general base in the two enzymes.

Results and discussion

Overall structure

The *T. maritima* KDPG aldolase crystallises in spacegroup $P2_1$ with cell dimensions a = 42 Å b = 101 Å c = 124 Å with $\alpha = \gamma = 90°$ $\beta = 97°$ and data to 1.9Å was collected. The *E. coli* E45N mutant enzyme crystallises in space group $P2_12_12_1$ with cell dimensions a = 54.5Å b= 84.2 Å c= 132.4 Å and $\alpha = \beta = \gamma = 90$ to a diffraction resolution of 1.4Å. The overall fold is unchanged from our (16) and other's descriptions of the enzyme (*10, 18*). The two aldolase structures exhibit the classic α / β barrel structure. In looking at the monomeric structure, the *T. maritima* & *E. coli* enzymes are very similar; the rms deviation for 188 matching Cα atoms is 1.5Å. There are slight differences between the structures reflecting rigid body movement of the secondary structure elements. The most obvious differences are in the loop to the C-terminal helix and at the N-terminus, both of which are shorter in *T. maritima*. Looking at the quarternary structure, the *T. maritima* trimer is more compact. This is manifested in the closer contact in *T. maritima* between the open C-terminal end of the barrel from one monomer and the loop from residues 143 to 152 in the other. The enzyme from *Pseudomonas putida* (*18*) superimposes with the rms deviation for 183 matching Cα atoms of 1.5Å. Superposition of the various trimers increases the deviation up to 2.0Å, reflecting the subtly different arrangements seen for the trimer in all crystal structures.

Schiff base structure of *T. maritima enzyme*

There are two trimers (six monomers) in the unit cell structure of *T. maritima* KDPG aldolase: our description focuses on one of these trimers (labelled A). The trimers are essentially identical except for minor differences in side chain orientations and water structure. From the earliest electron density maps it was obvious that additional density was present on the Lys 129. As the structure was refined, the density clearly showed a pyruvate bound to Lys 129 as the Schiff base. The orientation of the pyruvate carboxylate group is the same as that described previously for the carbinolamine structure(10), forming a bidentate hydrogen bond to Thr 69 and Arg 17. This Thr, present in 76 of the 80 aligned KDPG sequences, follows a highly conserved GXG motif, found in

79 of 80 sequences. In the remaining 4 sequences Thr is replaced by Ser. The putative catalytic base, Glu 40, is quite remote from the key atoms of the Schiff base, including N (4.3Å), C2 (4.7) and C3 (5.8Å). The methyl group (C3) of pyruvate points towards the highly conserved Phe 131, in turn contained in the characteristic aldolase GxxxωKφFP motif that includes the key catalytic lysine. In the remainder of the sequence, ω is a hydrophobic residue (L,F,I,M,V,C) and φ is a bulky hydrophobic residue (F,I,V,L). Proline kinks the structure, ensuring that the aromatic side chain packs in position to cap the pyruvate methyl group. This hydrophobic cap likely limits the degree of substitution tolerated by the enzyme at the C3 position.

The nucleophilic C3 atom sits in a channel that reaches to the C-terminal end of the barrel. A sulfate ion is bound 7.7Å from C3 along this channel at the C-terminal end of the barrel, in an anion hole formed by the amide nitrogens of Gly 157, Gly 158, Val 159 and Ser 179. The Ser 179 hydroxyl group also donates a hydrogen bond. A sulfate is bound at an identical position in the *P. putida* structure. A Gly-Gly pair allows a very tight turn in the loop region connecting strands β7 and α8. This pair is found in well over 95% of all aldolase sequences and is part of a conserved PTGGω motif (where ω is either I or V) found in 70% of sequences. A water molecule (W1) is located less than 3.1Å from C2, consistent with a water eliminated from the pyruvyl carbinolamine. This water bridges Glu 40 and Lys 129. A second water molecule (W2) forms hydrogen bonds to W1 and Glu 40 in the well ordered A, B and C subunits. This water molecule makes a hydrogen bond with Thr 156.

E45N *E. coli* KDPG structure

The A subunit of the catalytically impaired *E. coli* E45N KDPG aldolase (Table 2) shows additional density consistent with a carbinolamine intermediate, although at less than full occupancy and there is some disorder at the Lys. The additional carbinolamine density in the mutant *E. coli* KDPG aldolase would appear essentially identical to that reported by Allard *et al.* (10) A phosphate ion in our structure replaces the sulfate in the native enzyme carbinolamine complex (10). This anion

binding site is located at the same Gly-Gly motif (162 and 163 in *E. coli*) described above for the *T. maritima* enzyme. The W1 molecule is replaced by the oxygen of the pyruvyl carbinolamine, consistent with our assertion that W1 arises from the elimination of water during Schiff base formation. Since the carbinolamine adduct of the E45N mutant is apparently formed, at least in part, the Glu side chain must not be essential for carbinolamine formation. On the other hand, a catalytic base is essential for formation of the Schiff base: this conclusion is similar to previous suggestions based on studies of the native protein (10). A water molecule is found in the E45N structure and in the native carbinolamine structure (10) close to W2 position described for the *T. maritima* structure. In this slightly shifted position, the water bridges the Glu and Lys residues (Glu 45 & Lys 133 in *E. coli*).

Table 2 (Catalytic data) Here

Model of the substrate in the *T. maritima* structure

KDPG was constructed using PRODG. The structure was then superimposed on the pyruvyl Schiff base and carbinolamine structures, with C1, C2 and C3 matching the experimentally determined positions. The remainder of the sugar was modelled into the protein by adjusting torsion angles along the carbohydrate to values very close to the energetic minima (±60° and 180°). In our model the sugar adopts an essentially fully staggered conformation, with carbon-carbon dihedral angles of -169° (C3-C4); -176° (C4-C5), and +177° (C5-C6). The C6-O6 bond adopts a gauche conformation, with a dihedral angle of -57°. In this orientation the phosphate of the substrate overlaps almost exactly with the experimental position of the sulfate anion (S to P distance is 1.0Å). The observation that a sulfate or phosphate ion is bound in all three enzymes at this highly conserved binding loop suggests that the region is indeed the recognition site for the phosphate moiety of KDPG. In contrast, other sugar phosphate aldolases utilize an Arg or Lys residue in the putative phosphate recognition motif: neither residue is found in a plausible position in KDPG aldolase.

In the model of the gluconyl Schiff base, both W1 and W2 are observed in equivalent positions to those of the pyruvyl Schiff base complex. In the carbinolamine structure W2 remains in position

while W1 is absent. We suggest that during retroaldol cleavage W1 adds across the imine to form the carbinolamine intermediate. Our model allows the substrate to make a number of favourable contacts with the protein. Although O4 of the sugar is in the vicinity of Thr 156, part of the highly conserved PTGGω motif, it is too far (4.7Å) from the hydroxyl to form a direct hydrogen bond. On the other hand, Thr156 could form a hydrogen bond to O4 of the sugar substrate, either by a 120° rotation around the sidechain dihedral or via the intermediacy of bridging water molecules. The equivalent threonines in the *E. coli* and *P. putida* structures adopt the same conformation about the sidechain bond and a bridging water network can be traced to O4 of the substrate. O5 of the sugar makes a hydrogen bond (3.1Å) with the backbone carbonyl of Ala 143 from the neighbouring subunit. Similar models of substrate binding are easily constructed for the *E. coli* and *P. putida* enzymes.

These models, however, do not predict the contact between O5 and the backbone of the neighboring subunit, seen in the *T. maritima* model. This difference may reflect the greater separation of monomers in the *E. coli* and *P. putida* structures relative to the *T. maritima* protein. As we have previously noted, the arrangement of monomers in the trimer is highly dependent on crystallisation conditions, even for iterative structures of the same protein. While these observations are not evidence that the *E. coli* enzyme can form the same inter-subunit contact as the *T. maritima* enzyme *per se*, they do offer plausible explanation for the unique trimeric arrangement of monomers of the KDPG aldolases.

Implications for catalysis

Aldol addition by KDPG aldolase involves sequential formation of a pyruvyl carbinolamine, dehydration to the imine and enolization to the nucleophilic enamine (Scheme 3). Addition to an aldehydic electrophile results in a gluconyl imine which is, in turn, hydrated to the corresponding carbinolamine and released as the ketoacid product. Catalysis is effected by the provision of a

nucleophilic amine and by a series of proton transfers to and from a highly conserved glutamate that serves, in turn, as both a general acid and a general base.

Scheme 3 (KDPG aldolase mechanism) Here

Formation of the initial carbinolamine requires a significant population of the deprotonated form of Lys133. The pKa of lysine ε-amino group is diminished by inclusion in a low dielectric media or the presence of a proximal positively charged residue. Here, both mechanisms may be operative. The central pore of the β-barrel is relatively hydrophobic, and the catalytic lysine is found among an unusual sequence of phenylalanyl residues. Additionally, inclusion of substrate within this pore should further lower the pKa of the lysine. KDPG aldolase also contains a conserved arginine residue (49 in *E. coli*, 17 in *T. maritima*). Although this sidechain is somewhat removed from the catalytic lysine (~5Å), alanine mutants at this position show significantly diminished activities (Toone & Fierke, unpublished results). Together, these effects presumably provide a lysine residue sufficiently nucleophilic to effect catalysis.

By comparison of both multiple structures and sequence alignments it seems clear that only the absolutely conserved glutamate is positioned to function as this general acid/general base. On the other hand, the glutamate forms direct contacts only to O2 of the carbinolamine intermediate; this same glutamate cannot form contacts to the catalytic lysine, to C3 of either the pyruvyl or gluconyl intermediates, or to O4 of the gluconyl carbinolamine or imine. This apparent separation of the putative general acid/general base from those sites requiring proton donation and abstraction represents a recurring challenge to the elucidation of aldolase mechanism.

A major step towards a solution to this problem came with the suggestion that water could act as a bridge to the catalytic base, based on studies of DERA (11). In the proposed mechanism, a triad of Lys201, Asp102 and water acts as the requisite base to shuttle protons off and on both the substrate

and the catalytic Lys during turnover (Scheme 2). However, this mechanism cannot be generalized to other aldolases, since DERA alone has a second Lys. Furthermore, KDPG (and other) aldolases lack a potential base near the position occupied by Asp 102 in DERA (Ile 92 in KDPG aldolase).

The mechanism proposed for DERA does, however, suggest that water mediated proton transfer should be considered for other aldolases. Our analysis of KDPG aldolase structures highlights two water molecules: W1, arising from elimination of water from pyruvate, and W2, which is always present. W2, which had previously been suggested as relevant to catalysis, bridges the catalytic Lys and Glu residues in both the apo and carbinolamine structures, allowing Glu to effect proton transfer during carbinolamine formation (Scheme 3). However, this Glu is not essential for carbinolamine formation; presumably proton transfer to solvent is sufficiently facile to support reaction even in the absence of a general base. In the second step W2 and Glu act in concert to transfer a proton to the carbinolamine hydroxyl, triggering the elimination of water (W1) and producing the imine. *In toto* then, the mechanism is conceptually similar to that proposed for DERA (11).

Prior to aldol addition, the Schiff base must be enolized to the nucleophilic enamine, a process that requires deprotonation of pyruvyl imine at C3. Both the Schiff base and carbinolamine structures place the conserved Glu some 6.0Å from this carbon, too far for direct proton transfer. Again, however, proton transfer could be mediated by an active-site water. In the Schiff base complex W1, which arises from elimination from the carbinolamine intermediate, bridges C3 and Glu with a C3-O1 distance of roughly 4Å. The distance between W1 and OE1 of Glu 45 (EC) is 2.5 – 2.6Å. Movement of less than 1.0Å brings this water into hydrogen bonding contact with both atoms, facilitating efficient proton transfer from pyruvyl C3 to Glu45. Given that *T. maritima* is a thermophile with an optimal growth temperature of roughly +80°C and that we have (by definition) crystallised a stable Schiff base adduct, this movement seems plausible. In contrast, rotation of the Lys Schiff base to bring C3 within hydrogen bonding distance of Glu45 dramatically alters the

recognition of caboxylate group and creates van der Waal clashes. Therefore based on the experimental locations of C3 in both the Schiff base and carbinolamine structures, we see no alternative to W1 fulfilling an intermediary role during proton transfer.

During aldol addition, C3 of the pyruvyl enamine attacks the aldehyde of glyceraldehyde-3-phosphate in what is presumably the rate-determining step. As the hybridization of the aldehydic carbon changes, the incipient oxyanion is protontated to form the alcohol. From our modelled substrate complex, a protonated Glu45 would be able to transfer a proton to this oxygen using the same bridging W1 (Scheme 3). In the reverse reaction, Glu would abstract a proton from O4, again via a bridging water, triggering retroaldol cleavage.

Contrast with DERA aldolase mechanism

The structures of native DERA bound as the carbinolamine intermediate and of a mutant variant trapped as the Schiff base intermediate first highlighted the role of bridging waters in the aldolase mechanism (11). The Schiff base-forming lysine is located in a similar position on the β-barrel of both DERA and KDPG aldolases. There are also similarities in the manner of recognition of the full length substrate by the two enzymes. Thus, for example, DERA binds the terminal phosphate group in a pocket composed primarily of main chain atoms, similar to the motif we propose in our substrate model for KDPG aldolase, and the substrate is in an extended conformation. Rather, the difference in substrate specificity and activity between the two enzymes appears to lie in the positions of the basic residues. The two residues implicated in proton transfer in DERA, Asp102 and Lys201, are occupied in space by Ile and Cys, respectively, in KDPG aldolase. Similarly, the glutamate residue proposed as the key catalytic residue of KDPG aldolases is replaced by a cysteine in DERA. While this cysteine could conceivably function in proton transfer, the DERA C47A mutant retains significant activity (k_{cat}/K_M = 43 s^{-1} mM^{-1}) (11), while the KDPG aldolase E45N mutant is much more significantly impared (k_{cat}/K_M = 0.003 s^{-1} mM^{-1}; Table 2).

It is tempting to speculate that the mechanism proposed here, one which utilizes two water molecules to shuttle protons between protein general acid/general base and substrate, is general among aldolases. On the other hand, one of the two water molecules proposed as catalytically relevant in KDPG aldolase (W1) is not visible in the mutant DERA Schiff base structure. Rather, this mutant (K201L) includes a water molecule in a position that would be occupied by the Nz atom of the wild-type enzyme. We suggest that this water molecule is, in fact, that expelled from the carbinolamine intermediate and corresponds to W1 in the KDPG aldolase mechanism. Its position, then, is an artifact of the mutant structure, which creates a pocket through replacement of K201 with the shorter leucine residue. Since dehydration and production of a water molecule is a necessary step in the aldol mechanism, the only alternative to this postulate is a water molecule eliminated and expelled from the active site, reentering the active site only when required to hydrate the Schiff base during retro-aldol cleavage.

We have examined the recently published co-complex structures of KDG aldolase. This aldolase belongs to a different class of aldolases (the neuraminate lyases) than KDPG and in this family a Tyr has been identified as the base. The mechanisms reported have not considered how the C3 carbon of the Schiff base is deprotonated during catalysis. The pyruvate Schiff base complex shows that the oxygen of the tyrosinate is over 4.5Å away from the C3 of pyruvate. However, a water molecule does bridge these positions. No carbinolamine complex is available to ascertain whether this water results from Schiff base formation, but its position is suggestive of link between the mechanism proposed here and the lyase aldolase mechanism (Theodossis et al., JBC epub).

Throughout, we have invoked a role for key water residues in the kinetic mechanism of the aldolase-catalyzed reaction. It is important to note that these same water molecules could play important roles in ligand binding and substrate recognition. Such water-mediated recognition would further facilitate the ready evolution of aldolases specific for a wide variety of substrates from a common precursor fold.

Conclusions

On the basis of our structural characterisation of a native Schiff base and a high resolution mutant carbinolamine structure and by comparison to other known structures, we have proposed a detailed mechanism for KDPG aldolase. This mechanism is general for aldolases that share a common fold and locate within that fold the two key residues required for catalysis: a nucleophilic Lys and a Glu/Asp to provide general acid/base catalysis. Two water molecules are crucial to the mechanism. The first is present throughout turnover, acting as a proton relay during both imine and carbinolamine formation. The second water arises from the dehydration of the intermediate carbinolamine and is consumed by the hydrolysis of the imine. The second water also relays protons during imine enolization and the formation of the carbon nucleophile. Water-mediated proton relay during carbanolimine formation was previously suggested during studies of DERA catalysis. Here, we hypothesise that DERA, like KDPG aldolase, uses a second water molecule (created during imine formation) to relay protons during imine/enamine interconversion. A series of structures of the unrelated 2-keto-3-deoxygluconate aldolase reveals that the base (Tyr) is too far from the C3 of the pyruvyl intermediate and that in some structures a water molecule may play an important role (*19*).

DERA and KDPG aldolase use different residues in different locations to achieve catalysis: the mechanisms are thus not completely equivalent. This observation is a potentially important one, since key residues are typically conserved in both space and sequence in such closely related enzyme activities. There may, in fact, be no completely general mechanism for the enzymatic aldol reaction. What does appear to be essential and conserved is general acid/base catalysis mediated by water. The location of the nucleophilic lysine within the barrel also appears to be conserved. On the other hand, the position and nature of the general acid/base with respect to the lysine is flexible since water mediated proton relays can support proton transfer to and from a variety of positions. Presumably, the precise positioning of these residues facilitates control over the stereochemical course of the reaction. Such control is non-trivial in a reaction with such a low activation barrier:

even the uncatalyzed additions of simple enamines to unfunctionalized aldehydes are predicted to show activation energies near 30 kcal mol^{-1}.

In support of this hypothesis we note that the catalytic 38C2 antibody is an efficient aldolase (*20*) despite having a totally different fold from the β-barrel aldolases. This efficiency stands in contrast to the general observation that catalytic antibodies are much less active than their enzyme counterparts. Similarly, a series of KDPG aldolase mutants in which the Lys residue is moved to various positions within the barrel retain activity but show altered substrate and stereoselectivities(16). Finally, mutation of the Schiff base-forming Lys 167 residue in DERA yields a protein that retains aldolase activity, presumably because a proximate lysine at position 201 can substitute for Lys 167 (11). We continue our evolutionary and mechanistic studies of the pyruvate aldolases and will report our results in due course.

Materials and methods

***T. maritima* KDPG aldolase crystallisation.** The protein was expressed from the previously described pTM-eda plasmid in BL21(DE3) cells(*17*). Protein production was induced by addition of 0.5mM IPTG at OD_{600} = 0.8. Cells were pelleted after 3 hours and lysed by sonication in the presence of 0.1mM PMSF. Cell debris was removed by centrifugation and the supernatant heated to 70°C for 3 minutes. The supernatant was further purified by anion exchange chromatography using POROS HQ media (Amersham Biosciences). Crystals of protein were obtained after 3 days by the hanging drop methodology at 22°C. 2μl of protein at 5$mgml^{-1}$ in 50mM Tris PH 7.0 was mixed with 2μl of a reservoir solution containing 0.075M sodium acetate, 0.1M ammonium sulfate and 27% w/v PEG 4000 at pH 4.6. Crystals were cryoprotected in a solution containing 50% PEG 4000K and identical salt concentrations to the reservoir. Data were collected at the ESRF ID14-2 using an ADSC Quantum-4 CCD detector. A data set to 1.9Å was collected at wavelength of 0.932Å. Data collection statistics are summarised in Table 1. Crystals belong to spacegroup $P2_1$

with cell dimensions a = 42 b= 101Å c = 124Å, $\alpha = \gamma = 90°$ β=97° and contains two trimers in the asymmetric unit. The structure was solved by molecular replacement with the *E. coli* KDPG aldolase structure(16). The structure was refined using REFMAC5 (*21, 22*) and rebuilt using O (23), the final statistics on the model are given in Table 1. The Fo-Fc density maps clearly showed that a Schiff base complex between Lys 129 and pyruvate was formed. Pyruvate was built using PRODRG and incorporated into the refinement.

Site-directed mutagenesis of *E. coli* KDPG aldolase. The E45N mutant of the *E. coli* enzyme was constructed in pET-30b with the following primers (mutated bases underlined): E45N For (5' – GCT GGT GGG GTG CGC GTT CTG AAC GTG ACT CTG CGT ACC GAG TG – 3') and E45N Rev (5' – CAC TCG GTA CGC AGA GTC ACG TTC AGA ACG CGC ACC CCA CCA GC – 3'). Details of PCR reaction conditions have been previously described (16). Presence of the mutation was confirmed by sequencing of the resulting plasmid.

***E. coli* E45N KDPG aldolase crystallisation.** The protein was purified and crystallised via the same method described for the native enzyme (16). A data set to 1.4Å was collected at wavelength of 0.932Å. Data collection statistics are summarised in Table 1. Crystals belong to spacegroup $P2_12_12_1$ with cell dimensions a = 54.5Å b= 84.2 Å c= 132.4 Å, $\alpha = \beta = \gamma = 90°$ and contain one trimer in the asymmetric unit. The structure was solved by molecular replacement with the *E. coli* KDPG aldolase structure(16). The structure was refined using REFMAC5 (*21, 22*) and rebuilt using O (23), the final statistics on the model are given in Table 1. The Fo-Fc density maps clearly indicate there may be a partial occupied carbinolamine bound at Lys 129.

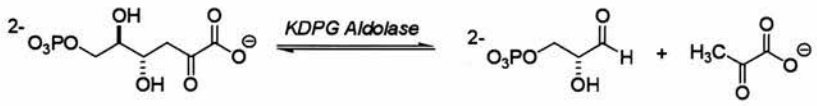
Kinetic characterization of *E. coli* E45N KDPG aldolase. In order to prevent wild-type KDPG aldolase contamination, the E45N mutant was expressed in a cell line lacking endogeneous KDPG aldolase. The DF71 cell line [*lacI22* LAM$^-$ e14$^-$ *eda-1 relA1 SpoT1 thi-1*] was obtained as a generous gift from the *E. coli* Genetic Stock Center at Yale University (24). The E45N mutation

was introduced into the previously described pUC-ECEDA2 using the same E45N For and E45N Rev primers described above(25). The resulting pUC-ECEDA2(E45N) plasmid was transformed into DF71 cells. A 50 mL TB/carbenicillin (60 μg mL^{-1}) culture was inoculated with a single colony and grown 12 hours at 37 °C and 220 rpm. A 1L TB/carbenicillin culture was subsequently inoculated with 10 mL of this overnight culture and grown to OD_{600} = 0.8 at 37 °C and 220 rpm. The culture was induced with 1 mM IPTG for 4 hours at 37 °C and then pelleted. The cells were disrupted by French Press and cell debris removed by centrifugation. The protein was purified via Ni^{2+} affinity chromatography (Novagen) and dialyzed against 3 x 2L 20 mM HEPES pH 7.5 to remove imidazole.

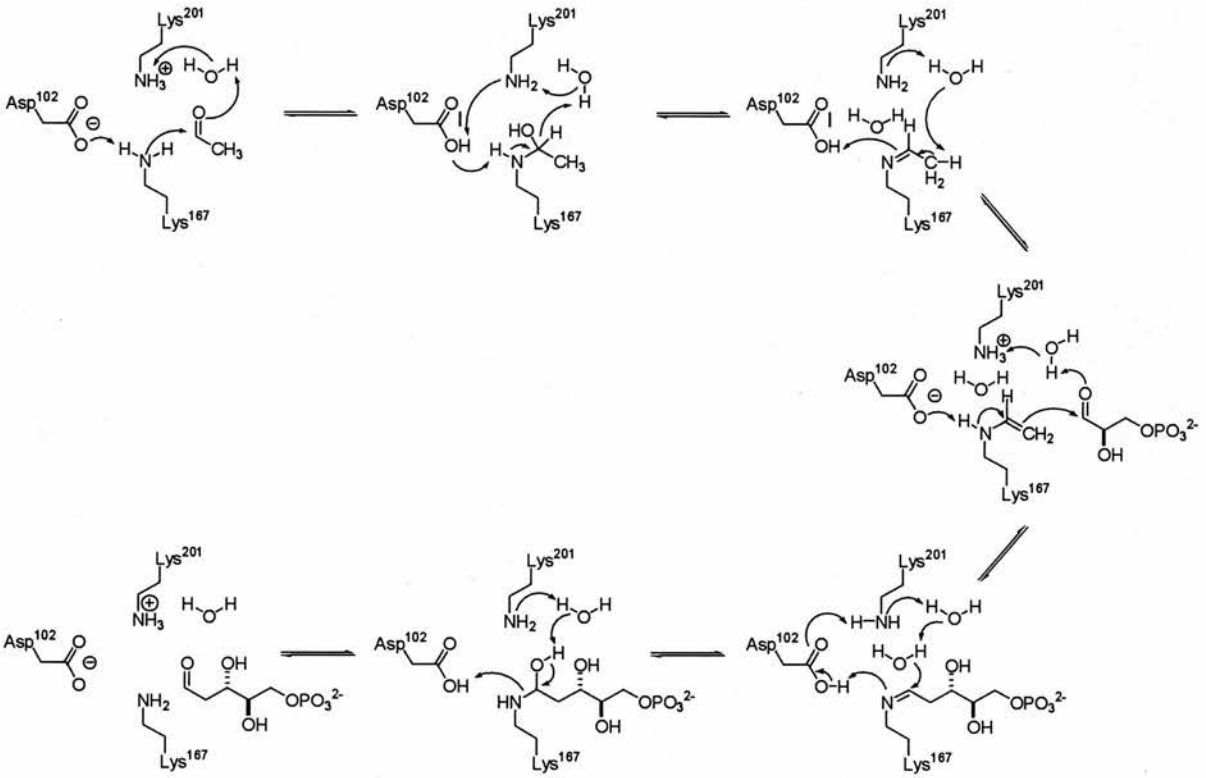
KDPG aldolase activity was determined by utilizing the standard coupled assay with L-lactic dehydrogenase (26). Briefly, HEPES (50mM, pH 7.5), NADH (0.1 mM), KDPG (0.4-16 mM) and L-lactic dehydrogenase (EC 1.1.1.27 Type II from rabbit muscle, 105 U) were combined in a total volume of 1 mL. KDPG aldolase (20 μM) was added to initiate the reaction and the dissappearance of reduced cofactor was monitored for 10 minutes by absorbance decrease at 340 nm. KDPG aldolase activity was determined from the initial slope of the absorbance versus time curve, and kinetic parameters were determined by fits of the Michaelis-Menten equation to velocity versus substrate concentration plot using Origin 5.0 (Microcal).

References

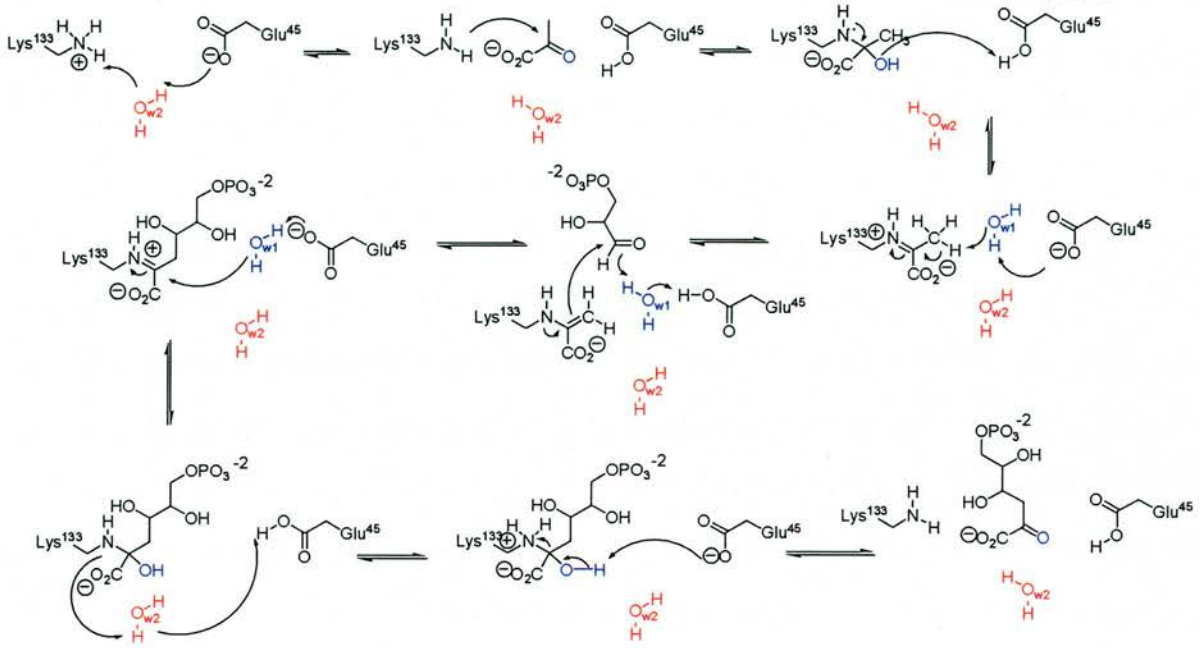
1. Conway, T. (1992) *Fems Microbiology Reviews 103*, 1-28.
2. Fessner, W. D., and Walter, C. (1997) *Bioorganic Chemistry 184*, 97-194.
3. Joerger, A. C., Mayer, S., and Fersht, A. R. (2003) *Proceedings of the National Academy of Sciences of the United States of America 100*, 5694-5699.
4. Hult, K., and Berglund, P. (2003) *Current Opinion in Biotechnology 14*, 395-400.
5. Williams, G. J., Domann, S., Nelson, A., and Berry, A. (2003) *Proceedings of the National Academy of Sciences of the United States of America 100*, 3143-3148.
6. Fong, S., Machajewski, T. D., Mak, C. C., and Wong, C. H. (2000) *Chemistry & Biology 7*, 873-883.
7. Jia, J., Schorken, U., Lindqvist, Y., Sprenger, G. A., and Schneider, G. (1997) *Protein Science 6*, 119-124.
8. Dalby, A., Dauter, Z., and Littlechild, J. A. (1999) *Protein Science 8*, 291-297.
9. Thorell, S., Schurmann, M., Sprenger, G. A., and Schneider, G. (2002) *Journal of Molecular Biology 319*, 161-171.
10. Allard, J., Grochulski, P., and Sygusch, J. (2001) *Proceedings of the National Academy of Sciences of the United States of America 98*, 3679-3684.
11. Heine, A., DeSantis, G., Luz, J. G., Mitchell, M., Wong, C. H., and Wilson, I. A. (2001) *Science 294*, 369-374.
12. Joerger, A. C., Mueller-Dieckmann, C., and Schulz, G. E. (2000) *Journal of Molecular Biology 303*, 531-543.
13. Kroemer, M., Merkel, I., and Schulz, G. E. (2003) *Biochemistry 42*, 10560-10568.
14. Cooper, S. J., Leonard, G. A., McSweeney, S. M., Thompson, A. W., Naismith, J. H., Qamar, S., Plater, A., Berry, A., and Hunter, W. N. (1996) *Structure 4*, 1303-1315.
15. Plater, A. R., Zgiby, S. M., Thomson, G. J., Qamar, S., Wharton, C. W., and Berry, A. (1999) *Journal of Molecular Biology 285*, 843-855.
16. Wymer, N., Buchanan, L. V., Henderson, D., Mehta, N., Botting, C. H., Pocivavsek, L., Fierke, C. A., Toone, E. J., and Naismith, J. H. (2001) *Structure (Camb) 9*, 1-9.
17. Griffiths, J. S., Wymer, N. J., Njolito, E., Niranjanakumari, S., Fierke, C. A., and Toone, E. J. (2002) *Bioorganic & Medicinal Chemistry 10*, 545-550.
18. Bell, B. J., Watanabe, L., Rios-Steiner, J. L., Tulinsky, A., Lebioda, L., and Arni, R. K. (2003) *Acta Crystallographica Section D-Biological Crystallography 59*, 1454-1458.
19. Theodossis, A., Walden, H., Westwick, E. J., Connaris, H., Lamble, H. J., Hough, D. W., Danson, M. J., and Taylor, G. L. (2004) *Journal of Biological Chemistry Epub ahead of publication.*
20. Hertweck, C. (2000) *Journal Fur Praktische Chemie-Practical Applications and Applied Chemistry 342*, 832-835.
21. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallographica Section D-Biological Crystallography 53*, 240-255.
22. Bailey, S. (1994) *Acta Crystallographica Section D-Biological Crystallography 50*, 760-763.
23. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallographica Section A 47*, 110-119.
24. Fradkin, J. E., and Fraenkel , D. G. (1971) *Journal of Bacteriology 108*, 1277-1283.
25. Griffiths, J. S., Cheriyan, M., Corbell, J. B., Pocivavsek, L., Fierke, C. A., and Toone, E. J. (2004) *Bioorganic & Medicinal Chemistry 12*, 4067-4074.
26. Meloche, H. P., and Wood, W. A. (1964) *Journal of Biological Chemistry 239*, 3511-3514.



Scheme 1. KDPG aldolase reaction



Scheme 2 The mechanism of DERA



Scheme 3. The mechanism of KDPG Aldolase

Table 1 Crystallographic data

	T. maratime	*E. coli* E45N
Resolution (Å)	33 - 1.9	40-1.55
High resolution shell	2.0 – 1.9	1.69-1.55
Cell dimensions	a = 42.6 b= 101.1Å c = 124.6Å α = γ = 90°, β=97.2°	a = 54.5 b= 84.2Å c = 132.4Å α = β=γ = 90°
Spacegroup	$P2_1$	
R-merge	7.2% (29.6)	7.5% (35.6%)
Completeness	94.3% (94.2%)	99.4% (98.8%)
Redundancy	4.0 (3.9)	3.4(3.2)
Refinement		
R-factor	18.8% (20.0%)	17.0% (21.2%)
R-free	23.7% (28.2%)	21.0% (28.5%)
RMS deviation		
Bonds	0.016	0.016
Angles	1.53	1.5
Core Ramachandran (%)	95.3	96%

Table 2. Kinetic data for *E. coli* KDPG aldolase E45N mutant with KDPG

Enzyme	**k_{cat} (s^{-1})**	**K_M (mM)**	**k_{cat}/K_M (s^{-1} mM^{-1})**
Wild-type	100	3.9	26
E45N	0.02	3.2	0.003

Figure 1 The Structure of the KDPG aldolase is a trimer.

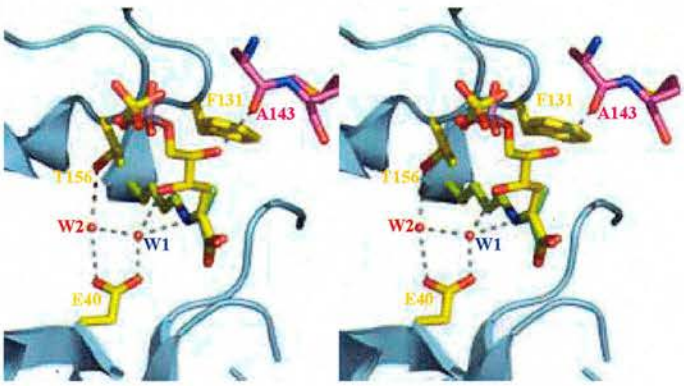


Figure 2 The location of the Schiff base with the active site Lys and the key catalytic and recognition residues. Oxygen atoms are coloured red and nitrogen blue for all molecules. In the protein the carbons are coloured yellow. The Schiff base carbon atoms are coloured green. The experimentally located sulphate and phosphate ions are shown yellow and red. The model for KDPG is shown with carbon atoms colored grey. The contact with the neighbouring subunit is shown clearly.

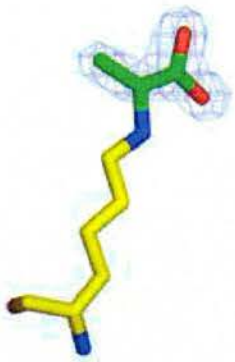


Figure 3 Omit density Fo-Fc for the Schiff base of *T. maritime* KDPG aldolase. The map is countered at 025e$Å^3$ (3.5σ).