CARBOHYDRATE PROCESSING IN PATHOGENIC BACTERIA

Simon Timothy Morgan Allard

A thesis submitted for the degree of

Doctor of Philosophy

19th November 2001

Contents

List of Figures

List of Tables

List of abbreviations

ADP	Adenosine diphosphate
AGME	ADP-L-glycero-D-mannoheptose 6-epimerase
BPGD	Bacterial Polysaccharide Gene Database
CCD	Charged coupled device
CDP	Cytidine diphosphate
CNS	Crystallography and NMR System
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
dTDP	deoxythymidine 5'-diphosphate
DTT	Dithiothreitol
EC	Enzyme Commission
EDTA	Ethylenediaminetetraacetic acid
EGALE	UDP-galactose 4-epimerase from *Escherichia coli*
GALE	UDP-galactose 4-epimerase
GDP	Guanosine diphosphate
GFS	GDP-fucose synthetase
GMD	GDP-mannose 4,6-dehydratase
HEPES	(N-[2-Hydroxyethyl] piperazine-N'-[2-ethanesulphonic acid])
HGALE	UDP-galactose 4-epimerase from *Homo sapiens*
HPLC	High-performance liquid chromatography
IPTG	Isopropyl β-D-thiogalactopyranoside
K_a	The equilibrium constant for acid dissociation
k_{cat}	The enzyme turnover number

K_m	Michaelis constant
LB	Luria-Bertani medium
LPS	Lipopolysaccharide
LUMO	Lowest unoccupied molecular orbital
MALDI-TOF	Matrix assisted laser desorption/ionisation time-of-flight mass spectrometry
MES	2-[N-Morpholino] ethanesulphonic acid
NAD^+	Nicotinamide adenine dinucleotide (oxidised)
NADH	Nicotinamide adenine dinucleotide (reduced)
NADP	Nicotinamide adenine dinucleotide phosphate (oxidised)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
NCS	Non crystallographic symmetry
PEG	Polyethylene glycol
PDB	Protein Data Bank
pK_a	$-\log(K_a)$
PMSF	Phenylmethylsulphonyl flouride
RmlB	dTDP-D-glucose 4,6-dehydratase
rms	Root mean square
SDR	Short chain dehydrogenase/reductase
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SQD1	Sulfolipid biosynthesis protein
Tris	Tris (hydroxymethyl) aminomethane
UDP	Uridine diphosphate
V_M	Matthews number

Declarations

I, Simon T.M. Allard, hereby certify that this thesis, which is approximately 38,000 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 19/Nov/01. Signature of candidate

I was admitted as a research student in September 1998 and as a candidate for the degree of Ph.D. in September 1999; the higher study for which this is a record was carried out in the University of St Andrews between 1998 and 2001.

Date 19/Nov/01. Signature of candida

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 that the candidate is qualified to submit this thesis in application for that degree.

Date 19/Nov/01. Signature of supervisor

In submitting this thesis to the University of St Andrews I 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 worker.

Date 19/09/15 Signature of candidat

Abstract

The enzyme dTDP-D-glucose 4,6-dehydratase (RmlB) along with other 4,6-dehydratases catalyses the first committed step in all 6-deoxysugar biosynthetic pathways. Numerous 6-deoxysugars are found in bacterial cell surface polysaccharides and a diverse array of secondary metabolites. RmlB is involved in the L-rhamnose biosynthetic pathway where it catalyses the dehydration of the nucleotide sugar dTDP-D-glucose to dTDP-4-keto-6-deoxy-D-glucose. L-rhamnose is of particular interest as it is found in the cell walls and envelopes of many pathogenic bacteria and in many cases has been shown to be essential for the organism's virulence. The fact that the L-rhamnose biosynthetic pathway is not found in humans makes RmlB a potential therapeutic target.

The mechanism of RmlB has been studied for over 30 years. RmlB is a member of the short-chain dehydrogenase/reductase (SDR) superfamily and contains the characteristic catalytic triad of Thr-Tyr-Lys. The work described here includes the first reported structures of the RmlB holoenzyme, substrate complexes and a substrate analogue complex. These structures have been solved for both *Salmonella enterica* Typhimurium and *Streptococcus suis* serotype 2 allowing comparisons to be made. The substrate complexes reveal important conformational changes in the enzyme that take place on binding substrate. The structural movements enable the conserved active site tyrosine to act directly as the active site base. The complexes disclose how a conserved aspartic acid/glutamic acid pairing is involved in the dehydration step of catalysis. The complexes also suggest how water elimination drives movement of the substrate to allow the enzyme to complete the reaction by reducing the glucosene

intermediate. As many of the residues involved in the RmlB mechanism are found in other members of the SDR family these results have important implications for the study of all dehydratases, in particular hydride transfer

Acknowledgements

There are two people without whom I would never have completed this Ph.D. Firstly, I am very grateful to my supervisor Prof. Jim Naismith. I have had a great project to work on, and with his guidance, experience and imagination I have achieved more than I ever believed possible three years ago. Secondly, I am also grateful to Dr. Marie-France Giraud, who with a certain Gallic charm took a wet behind the ears postgraduate student and helped make him a scientist (some may disagree with the scientist bit!).

Special thanks also go to Kostas Beis, an undergraduate student with almost as much enthusiasm for crystallography as my boss Jim. His belief that RmlB from *S. suis* would eventually produce diffracting crystals kept him, myself and Hampton Screens in business!

Other people who deserve a special mention include Helen Walden whose ability to understand and explain to me all the necessary crystallographic programs made my life much easier. I am also grateful to Genevieve Temple, Lesley Thomson, Stephen McMahon, 'Jimmy' Dong, Miryam Asuncion, Wulf Blankenfeldt and Deans Court for making my time at St Andrews great fun.

And finally, to all those in the Taylor/Naismith groups – Cheers!

Gratias tibi ago

"It is not enough to have a good mind; the main thing is to use it well"

Rene Descartes

"When something can be read without effort, great effort has gone into its writing"

Enrique Jardiel Poncela

Lege Feliciter

To my parents, Paul and Anne Allard, who have supported me since the very beginning, and to my brother and sister-in-law, Martin and Su Allard, who indulged me when times were hard.

Chapter One

Introduction

1.1. *Overview*

Carbohydrates are essential to life in all its forms; their synthesis and modification common to every organism. They fill many roles in biology, including such vital tasks as molecular recognition markers, structural elements and energy sources. They are also used as convenient precursors for the biosynthesis of important building blocks.

Deoxysugars have long been recognised as an important class of carbohydrate, being found in lipopolysaccharides (LPS), glycoproteins, glycolipids and many secondary metabolites. L-rhamnose is a 6-deoxysugar synthesised only in bacteria. In many Gram-negative pathogens, such as *Pseudomonas aeruginosa* and *Salmonella enterica* serovar Typhimurium, L-rhamnose occupies a key position in the LPS molecule that decorates the cell surface. Loss of the ability to synthesise L-rhamnose leads to a truncation of the LPS molecule. Organisms with incomplete LPS molecules are usually cleared by the host and are consequently avirulent. L-rhamnose is not restricted to Gram-negative bacteria, but is also found in *Mycobacterium tuberculosis* where an L-rhamnosyl residue plays an essential structural role in the cell wall, linking the inner peptidoglycan layer to the arabinogalactan polysaccharides. In other Gram-positive bacteria such as *Streptococcus mutans* and *Streptococcus suis* serotype 2, rhamnose is found in the capsule, which is a key virulence determinant.

These observations confirm the importance in bacterial pathogenicity of enzymes implicated in the biosynthesis of LPS and capsular sugar precursors. Attention has therefore turned to pathways such as L-rhamnose biosynthesis, as the enzymes are

highly conserved amongst pathogenic bacteria and therefore represent possible novel therapeutic targets.

1.2. *Carbohydrates*

Carbohydrate diversity

Carbohydrates form the most abundant group of natural products and are found in all classes of living organisms (He *et al.*, 2000; Kirschning *et al.*, 1997). They play a central role in metabolism, with organisms capable of photosynthesis harvesting the solar energy in order to drive reactions in which glucose is anabolised from carbon dioxide and water. In addition carbohydrates play an essential structural role in many organisms; examples include: cellulose, chitin, peptidoglycan and LPS all of which are derived from repeating sugar units. Carbohydrates and carbohydrate containing structures are also involved in many more active biochemical and bioinorganic processes. Cell surface oligosaccharides such as tumour associated antigens, lectins, glycoproteins, glycolipids and immunodeterminants are important for recognition and specificity in cell-cell interactions (Kirschning *et al.*, 1997). Furthermore, many bioactive secondary metabolites such as cardioglycosides, macrolide antibiotics and aminoglycoside antibiotics rely on sugar components for solubility and activity. Carbohydrates are also used as building blocks for the biosynthesis of aromatic amino acids (He *et al.*, 2000).

Deoxysugars

Although carbohydrates are recognised as a whole to be important, it is the role of deoxysugars that have come under much closer scrutiny in recent years. Since carbohydrates possess a carbon skeleton that is decorated with hydroxyl groups, a variety of mechanisms have evolved for the deoxygenation of these molecules to generate modified sugar structures. Deoxysugars are derived from common sugars by

the replacement of at least one hydroxyl group with a hydrogen or a non-O-linked functional group (Williams and Wander, 1980) (Figure 1.1). Such a substitution generally induces a fundamental change in the chemical properties of the resulting sugar, for instance enhancement of its thermal stability and hydrophobicity (Hallis and Liu, 1999). These properties contribute to the wide range of activities exhibited by deoxysugars as a class, particularly in LPS and glycoproteins. Deoxysugars are very common in secondary metabolism. Indeed as a prime component of deoxyribonucleic acid (DNA), 2-deoxy-D-ribose is probably the best-known deoxysugar. The most ubiquitous form of deoxysugar is deoxygenated at the C6 position (He *et al.*, 2000), of which L-rhamnose and L-fucose are the most frequently encountered (Tonetti *et al.*, 1998). Monodeoxygenation of a sugar at other positions is rare, except in the cases of some cardiac glycosides. C6 deoxygenation is also the first committed step in almost all deoxysugar biosynthetic pathways leading to the formation of di-, tri-, and tetra-deoxysugars.

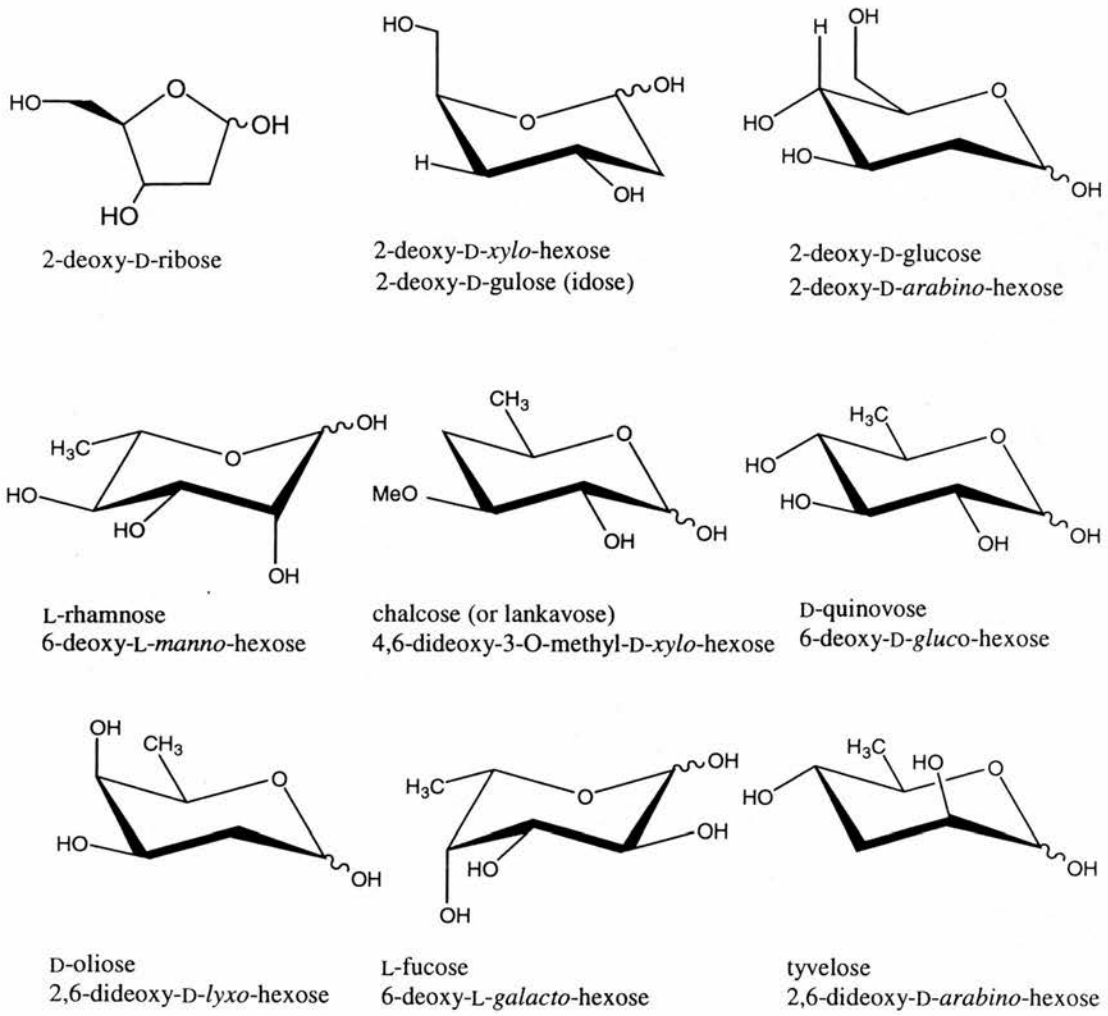


Figure 1.1. Examples of deoxysugars.

L-Rhamnose

L-rhamnose and L-fucose are two 6-deoxysugars which are characterised by the presence of a methyl group bound to carbon atom C5. Another important feature is that in all organisms, both prokaryotes and eukaryotes, deoxysugars are mainly used in the L-configuration instead of the D-configuration that is usually observed in all other monosaccharides (Tonetti *et al.*, 1998). L-rhamnose is involved in the structure of the cell wall, LPS and capsule (should one exist) of many bacteria, and hence is of significant functional importance in bacterial cells. It is an essential component of the surface antigens of many microorganisms. In particular it occurs in the O-antigen of LPS of Gram-negative bacteria, which is an important virulence factor. The differences in LPS composition are the basis for variations among the different serotypes of bacterial species such as the *Shigellae* and *Salmonellae*, where the O-antigen is formed by repeating units extending outside the cell surface, mainly constituted by L-rhamnose (Jiang *et al.*, 1991; Macpherson *et al.*, 1994). In *Escherichia coli* K12, L-rhamnose is found in the inner core of the LPS as well as being a vital component of the side-chain polysaccharide (O-antigen) of LPS. In *Salmonella*, the O-antigens form hydrophilic surface layers that protect the bacterium from complement-mediated cell lysis (Joiner, 1988), and in the case of *Shigella* the absence of rhamnose-containing O-antigen alters the distribution of the IcsA protein and in so doing affects intracellular spread (Sandlin *et al.*, 1996; Van den Bosch *et al.*, 1997). Consequently the pathways involved in the formation of O-antigens are critical to virulence and represent potential novel targets for therapeutic intervention. In the pathogenically important Gram-negative *P. aeruginosa*, L-rhamnose is a key component of the LPS core oligosaccharides (Sadovskaya *et al.*, 2000) and its absence precludes formation of a complete and protective LPS structure. Amongst the

Gram-positive bacteria, rhamnose is often found in the cell wall polysaccharides (McNeil *et al.*, 1990; Michalek *et al.*, 1984). In *S. suis* serotype 2 rhamnose has been found in the capsule. *S. suis* serotype 2 is an important zoonotic pathogen causing a wide spectrum of porcine diseases of considerable economic cost (Staats *et al.*, 1997), as well as being one of the major causes of human meningitis in Hong Kong (Lutticken *et al.*, 1986); crucially, loss of the capsule correlates with loss of virulence (Smith *et al.*, 1999). In *S. mutans*, which has been strongly implicated as one of the major causes of dental caries and is also associated with systemic diseases such as rheumatic fever (Tsuda *et al.*, 2000), the cell wall polysaccharide consists of a backbone structure of 1,2- and 1,3-linked rhamnosyl polymers with glucose side chains (Pritchard *et al.*, 1986). In *S. mutans* the serotype-specific rhamnose-containing polysaccharide antigens of its cell wall have been proposed as putative mediators for the colonisation of tooth surfaces (Tsukioka *et al.*, 1997) and its ability to bind to heart, kidney and muscle tissues (Stinson *et al.*, 1980; Stinson *et al.*, 1984). Recent work also suggests that these serotype-specific polysaccharide antigens play an essential role in resistance to phagocytosis and complement killing by human polymorphonuclear leukocytes (Tsuda *et al.*, 2000).

The cell wall of *M. tuberculosis*, the causative agent of tuberculosis, is unique. It has an outer layer of mycolic acids connected to peptidoglycan, the major structural component of Gram-positive cell walls, via an arabinogalactan polysaccharide. This connection is essential for the viability of the bacterium, and growth is inhibited by ethambutol, which prevents the attachment of the mycolic acids to the inner layer of peptidoglycan (Ma *et al.*, 1997). A single L-rhamnosyl residue has a vital structural role in linking the inner peptidoglycan layer to the arabinogalactan polysaccharides

(McNeil *et al.*, 1990) (Figure 1.2). Consequently inhibition of rhamnose biosynthesis in *M. tuberculosis* is an attractive proposition for the design of anti-mycobacterial drugs.

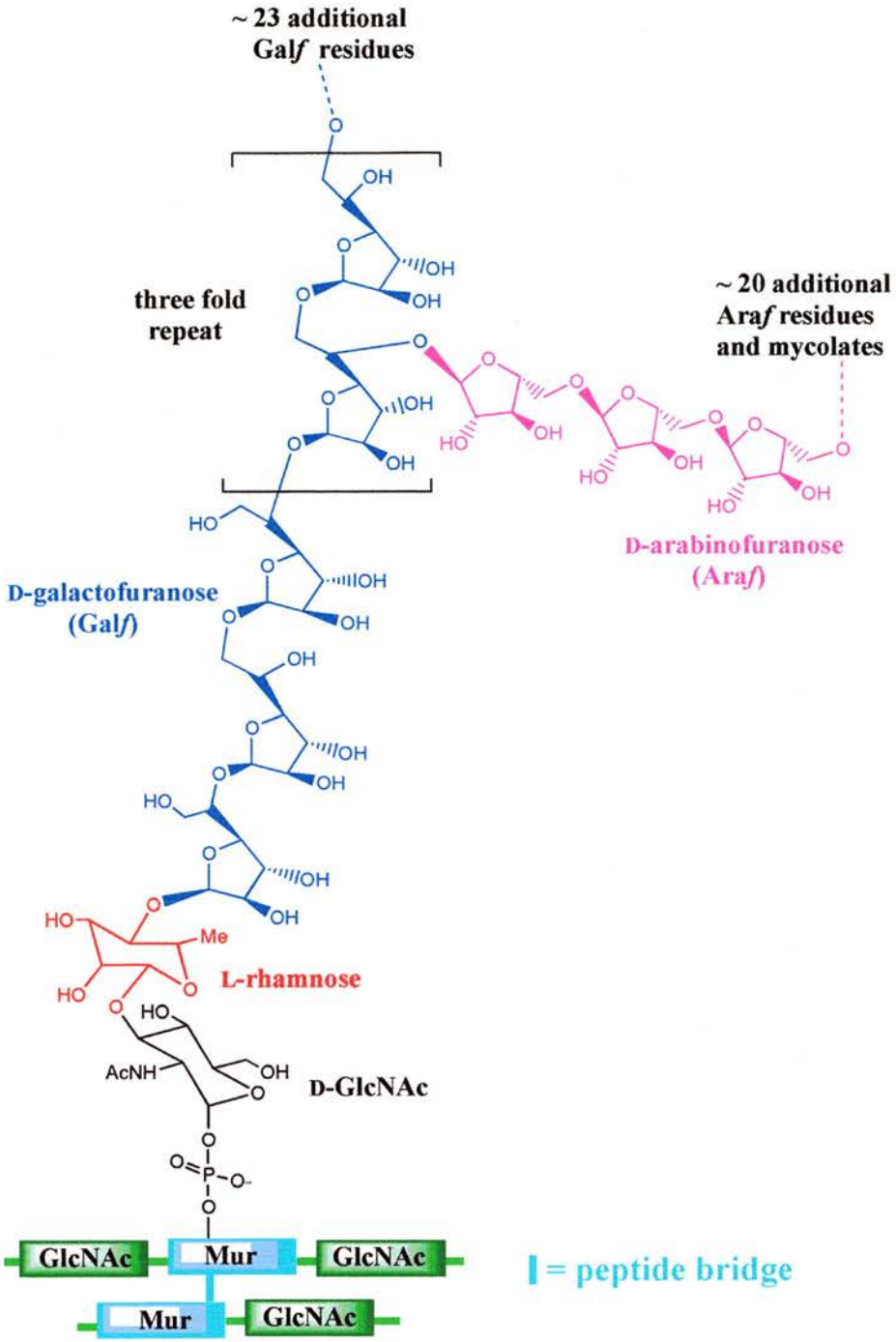


Figure 1.2. The mycobacterial cell wall; the single L-rhamnosyl residue has a vital structural role in linking the inner peptidoglycan layer to the arabinogalactan polysaccharides and is shown in red. Abbreviations: Mur, murein; GlcNAc, N-acetylglucosamine (Sanders et al., *2001).*

1.3. Bacteria

Infections due to bacteria are responsible for diseases ranging from mild tonsillitis to epidemics of cholera and plague. Bacteria are remarkably adaptable microorganisms. They may cause severe disease or innocently colonise the skin. Some bacteria, such as *S. suis,* are primarily pathogens of animals, only infecting humans as a chance event whereas others, such as *M. tuberculosis,* are once again causing untold misery and death around the world. Whilst most bacteria can replicate in hours or days, some are much slower growers, leading to chronic infections and difficulties in treatment (Hart and Shears, 1996; Ingraham and Ingraham, 1995).

The fight against bacterial disease is an ongoing one with organisms constantly developing new forms of resistance. The appearance of multi-drug resistant tuberculosis and vancomyocin resistant enterococci underlie the necessity for the development of new types of antimicrobial agents. The part played by protein crystallography in both drug design and the understanding of bacterial metabolic enzymes will be important.

The bacterial cell wall and associated structures

Bacteria that inhabit aquatic environments (most of the prokaryotic kingdom) need mechanically strong cell walls to protect their fragile plasma membranes from the very high cytoplasmic turgor pressures that can be generated by the osmotic gradient across the membrane. Therefore many bacteria (other than mycoplasma) are surrounded by a complex cell wall. Based on their cell-wall structure bacteria may be divided into two main groups depending on their reaction to Gram's stain; Gram-

positive and Gram-negative (Hart and Shears, 1996) (Figure 1.3). In most cases, the component of the cell wall that provides the necessary strength is peptidoglycan, which constitutes a three-dimensional covalently cross-linked net-like layer completely enveloping the bacterial protoplast (Hancock, 1997) (Figure 1.4).

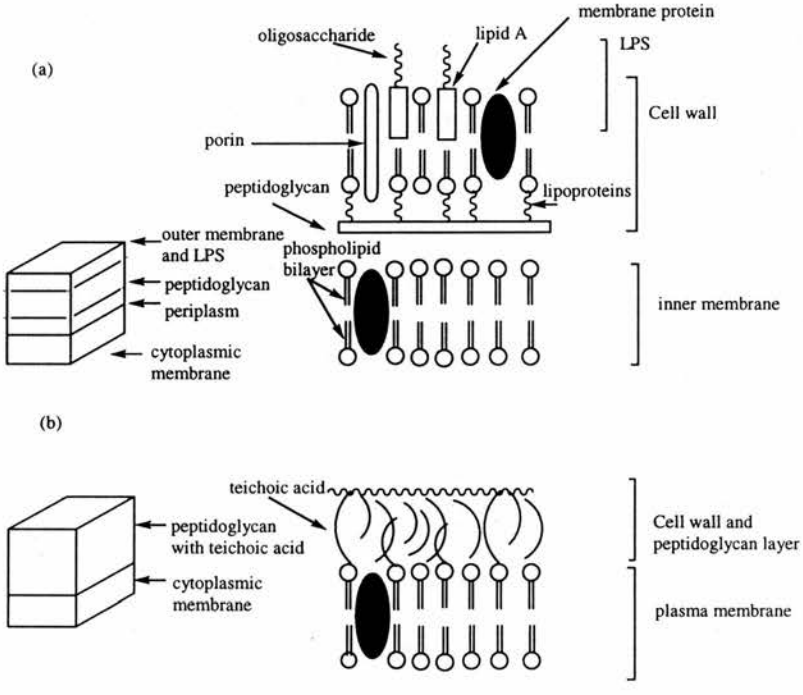


Figure 1.3. Diagram showing the structure of Gram-negative and Gram-positive bacterial cell walls.

Peptidoglycan is unique to bacteria, and is composed of long chains of polysaccharides (glycan) cross-linked by peptides (short proteins). Peptidoglycan is a biopolymer consisting of alternating units of *N*-acetylmuramic acid (NAM) and *N*-acetylglucosamine (NAG) with a short peptide linked to the lactyl moiety of the NAM residues. The peptide component varies among bacterial species but the most common type in both Gram-negative and Gram-positive bacteria is four amino acids long and contains L-isomers, which are normally found in proteins, and D-isomers,

which are otherwise rare in nature. The sugars are bound by β(1,4) glycosidic linkages. The dibasic amino acid in Gram-positive bacteria is usually D-lysine and in Gram-negative bacteria *meso*-diaminopimelic acid. Covalent peptide linkages or cross-bridges lead to the formation of the three-dimensional rigid meshlike molecule that forms the bacterial cell wall (Baron *et al.*, 1993).

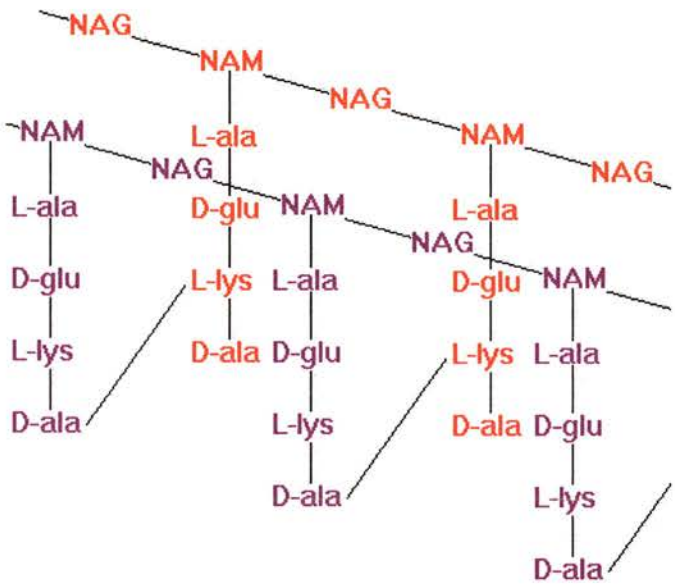


Figure 1.4. Peptidoglycan consists of a heteropolymer of amino sugars, alternating N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) residues. These are linked with β1,4 bonds to give chains of 10-65 disaccharide units in length.

In Gram-positive bacteria the peptidoglycan comprises 40-80% of the dry weight of the cell wall. The layer is typically 20-30nm thick and there is extensive cross-linking in three dimensions. Teichoic acids are also commonly present and in *Mycobacteria, Nocardia* and some other genera there may be long chain fatty acids. The cell wall may also contain polysaccharide antigens such as the Lancefield antigens of Streptococci (Lancefield, 1933) and Griffith proteins (Griffith, 1928). In contrast peptidoglycan comprises only 5-10% of the dry weight of the cell wall of

Gram-negative bacteria. It is only about 3-5nm thick and has fewer cross bridges than in the Gram-positive wall and hence is less rigid. Surrounding the cell wall is the outer membrane, which contains LPS and other specific outer membrane proteins such as pore-forming proteins and receptor sites for bacteriophages and bacteriocins.

Bacterial polysaccharides

Bacterial polysaccharides include LPS, lipo-oligosaccharide and extracellular polysaccharide (EPS). As the name implies lipopolysaccharide is a complex molecule, with lipid on one end and polysaccharide on the other. LPS is present in most Gram-negative bacteria and characteristically comprises three components: the lipid A structure required for insertion into the outer leaflet of the outer membrane bilayer; a covalently attached core composed of 2-keto-3-deoxyoctonic acid (KDO), heptose, ethanolamine, *N*-acetylglucosamine, glucose and galactose; and polysaccharide chains linked to the core, Figure 1.5. The polysaccharide chains constitute the O-antigens of the Gram-negative bacteria and extend from the cell surface, consisting of repeating oligosaccharide units generally composed of 3-6 sugars. The structural diversity of the O-antigens stems from variations in sugar composition, the sequences of sugars and linkages and the substitution of monomers with either sugars or non-sugar residues. These differences give rise to O-serotype specificity within bacterial species. The O-antigens form hydrophilic surface layers that often protect the cell from complement-mediated serum killing, and they are therefore essential virulence determinants (Whitfield, 1995).

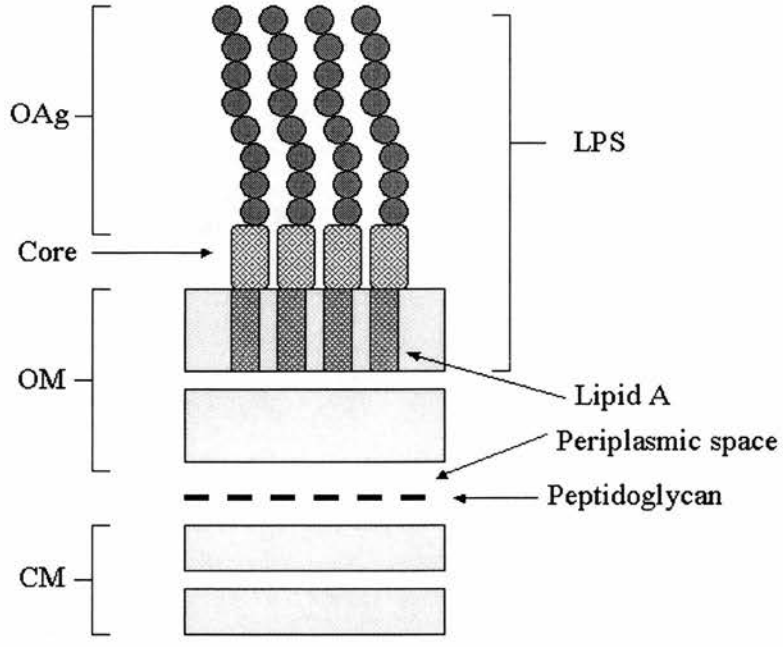


Figure 1.5. Schematic diagram showing the basic components of lipopolysaccharide found in most Gram-negative bacteria; lipid A, core oligosaccharide and O-antigen (OAg). Abbreviations: OM, outer membrane; CM, cytoplasmic membrane. Adapted from Reeves et al. *1996.*

Table 1.1. The structure of LPS (adapted from http://gsbs.utmb.edu/microbook/ch002.htm).

Lipid A	Core	O-antigen
Glucosamine	Ketodeoxyoctonate	Polysaccharide chains:
β-hydroxymyristate	Phosphoethanolamine	Repeating units of species-specific monosaccharides, e.g., galactose, rhamnose, mannose and abequose in *S. typhimurium* LPS
Fatty acids	Heptose	
	Glucose, galactose	
	N-acetylglucosamine	

EPS may be present in both Gram-negative and Gram-positive bacteria and, like O-antigens, is made of repeating units. The EPS can be released from the cell into the environment as a slime or remain attached to the cell to form a capsule as in *S. suis* (Reeves *et al.*, 1996).

Lipopolysaccharide

LPS is also called endotoxin and on its release by Gram-negative bacteria can exert a truly astonishing variety of toxic effects on the human body ranging from fever and diarrhoea to death. It should be noted however, that endotoxin is generally of low toxicity and is fatal only under conditions of overwhelming infection. The toxic effects are expressed when endotoxin stimulates human cells to secrete particular messenger proteins. For example, septic shock occurs when endotoxin causes phagocytes to secrete a protein called tumour necrosis factor (TNF). This can lead to

loss of fluid from the circulation, which results in insufficient blood reaching vital organs and as a consequence they shut down causing a potentially life threatening condition (Ingraham and Ingraham, 1995).

LPS assembly

Sugar nucleotides are the activated precursors for cell-surface polysaccharides. To assemble O-antigens, monomers are not transferred directly to a growing LPS molecule. Instead, O-antigens are synthesised separately on a lipid carrier by enzymes encoded by the *rfb* gene cluster (Figure 1.6). This gene cluster encodes all the enzymes required for formation of the O-antigens. These enzymes include those that catalyse the synthesis of the sugar-nucleotide precursors that are unique to O-antigens such as L-rhamnose, the glycosyltransferases and polymerases needed for assembly of the O-polysaccharides. In addition, the cluster also encodes the components required for the transfer of O-antigen polymers, or O-units, across the plasma membrane (Whitfield, 1995). As with the synthesis of cell-wall peptidoglycan and capsular polysaccharides, the lipid carrier is the C_{55}-polyisoprenoid derivative, undecaprenol phosphate. Once complete the O-antigen is transferred and covalently linked (ligated) to a preformed lipid-A-core acceptor at the periplasmic face (McGrath and Osborn, 1991). Following ligation the completed LPS molecule is translocated to the cell surface by unknown mechanisms (Whitfield and Valvano, 1993).

The Capsule

Surrounding both Gram-positive and negative bacteria may be a capsule. This is usually polysaccharide in nature and can be thin or several times as thick as the rest

of the cell. The principle function of the capsule is protection. Its sticky nature also helps a bacterium to adhere to a surface where conditions are favourable for growth. For example the capsule of *S. mutans*, the bacterium principally involved in tooth decay, helps anchor it to the surface of the tooth (Tsukioka *et al.*, 1997). Capsules also provide protection against phagocytosis and destruction by another cell (Klein, 1990). Many organisms rely on the presence of a capsule to aid evasion of the host's immune system and thus aid their ability to cause diseases. The capsule can therefore be viewed in many cases as an important virulence factor of many bacteria; for example only encapsulated strains of *Streptococcus pneumoniae* cause disease in man.

1.4. The Salmonellae

Salmonellae belong to the Enterobacteriaceae, which are a heterogeneous group of Gram-negative aerobic bacilli that are commensals of the intestinal tract of mammals. They are named after the scientist who discovered them, Dr. Daniel Salmon. The genus *Salmonella* contains a large number of species (more correctly, serotypes); the classification of which is complex, with over 2000 serotypes being distinguished (Elliott *et al.*, 1997; Hess and Kaufmann, 1996). Many of these have been given binomial names such as *S. typhimurium* and *S. enteritidis*, although they are not separate species, merely different serotypes. The majority of the components of these bacteria are identical, and at the DNA level, they are between 95% and 99% identical. They are commensals of many animals including poultry, domestic pets, birds and humans. Transmission is via the faecal oral route (Elliott *et al.*, 1997).

Structure and morphology

Salmonellae possess three major antigens; H (flagellar), O (somatic) which occur on the surface of the outer membrane and are determined by specific sugar sequences on the cell surface, and Vi (virulence), which are possessed by only a few serovars such as *S. typhi* where they are associated with invasiveness. There are over 60 different O-antigens, and individual strains may possess several O-antigens and H-antigens. The classification of the serotypes is often undertaken by identifying these antigens. Unfortunately, identification is complex, depending on the detection of several antigens. A single strain can possess two different sets of H antigens at different times (phase variation), and thus both sets must be analysed for identification (Sleigh and Timbury, 1998).

In common with other Gram-negative bacilli, the cell envelope of salmonellae contains a complex LPS structure. The nature of the repeating sugar units in the outer O-polysaccharide chains of the LPS is responsible for O-antigen specificity and may also help determine the virulence of the organism. Indeed, salmonellae lacking the complete sequence of O-sugar repeat units are called 'rough' due to the rough appearance of their colonies and are usually avirulent or less virulent than the smooth strains which posses a full complement of O-sugar repeat units. In *S. typhimurium*, the enzymes of O-antigen biosynthesis are encoded in the *rfb* gene cluster, which has been cloned and sequenced and comprises about 20 kb and 16 genes (Xiang *et al.*, 1993), Figure 1.6. The genes encoded at this locus include those for nucleotide-sugar biosynthesis enzymes, for example, the *man* genes for GDP-mannose biosynthesis, the *rml* genes for dTDP-rhamnose biosynthesis and the *ddh* genes for NPD-

dideoxyhexose biosynthesis (Reeves *et al.*, 1996). Mutation of any genes in the *rfb* locus leads to loss of O-antigen structures from the LPS (Maskell and Allen, 1997).

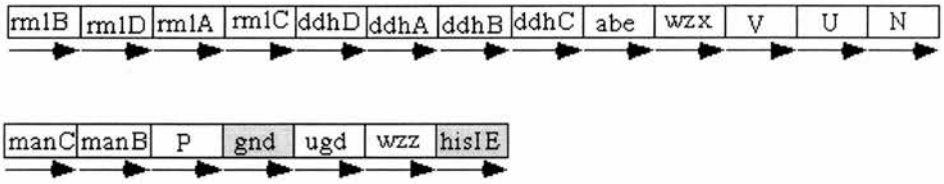


Figure 1.6. The genetic map of the S. typhimurium rfb *locus; intergenic gaps are not shown. Boxes represent genes; arrows underneath show their transcriptional direction. Genes not involved in polysaccharide production are shaded. The majority of genes in a cluster will have the same first three letters in their names. Taken from the Bacterial Polysaccharide Gene Database (BPGD) (http://www.microbio.usyd.edu.au/BPGD/default.htm).*

S. typhimurium has an O-antigen with a repeat unit of four sugars: abequose, mannose, rhamnose and galactose (Figure 1.7). Its biosynthesis begins with the synthesis of nucleotide sugars and is followed by assembly of O-units by sequential transfer of sugars onto undecaprenol phosphate by transferases (Jiang *et al.*, 1991). The polymerisation of O-units is catalysed by O-antigen polymerase. Finally, the O-polysaccharides are ligated to lipid A-core to form LPS (Curd *et al.*, 1998).

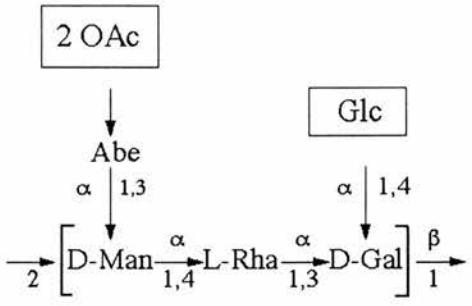


Figure 1.7. The polysaccharide structure of S. typhimurium. *Abbreviations: Abe, abequose; D-Man, D-mannose; L-Rha, L-rhamnose; D-Gal, D-galactose. Taken from the BPGD (http://www.microbio.usyd.edu.au/BPGD/default.htm).*

Pathogenesis

Salmonellae enter the body when contaminated food is ingested. The intestine represents both the port of entry as well as the major site of replication; the organisms colonise the ileum and colon, invade the intestinal epithelium (often via specialised epithelia known as M cells) and proliferate within the epithelium and lymphoid follicles (Finlay and Falkow, 1989). The mechanism by which salmonellae invade the epithelium is incompletely understood but involves an initial binding to specific receptors on the epithelial cell surface followed by invasion (Donnenberg, 2000; Vazquez-Torres *et al.*, 1999). Invasion occurs with the bacterium inducing the enterocyte membrane to undergo "ruffling" and thereby to stimulate their pinocytosis (Finlay *et al.*, 1989) (http://gsbs.utmb.edu/microbook/ch021.htm). Invasion is dependent on rearrangement of the cell cytoskeleton and probably involves increases

in cellular phosphate and calcium. Attachment and invasion are under distinct genetic control and involve multiple genes in both chromosomes and plasmids.

Having invaded the intestine, most salmonellae induce an acute inflammatory response. They may invoke cytokines that inhibit protein synthesis, although it is not known whether these cytokines contribute to the inflammatory response or ulceration (Giannella *et al.*, 1973a; Giannella *et al.*, 1973b; Giannella *et al.*, 1975). Invasion of the mucosa causes the epithelial cells to synthesise and release various proinflammatory cytokines that in turn lead to an acute inflammatory response and may be responsible for damage to the intestine. Due to this intestinal inflammatory reaction, symptoms of inflammation such as abdominal pain and diarrhoea are common (Figure 1.8).

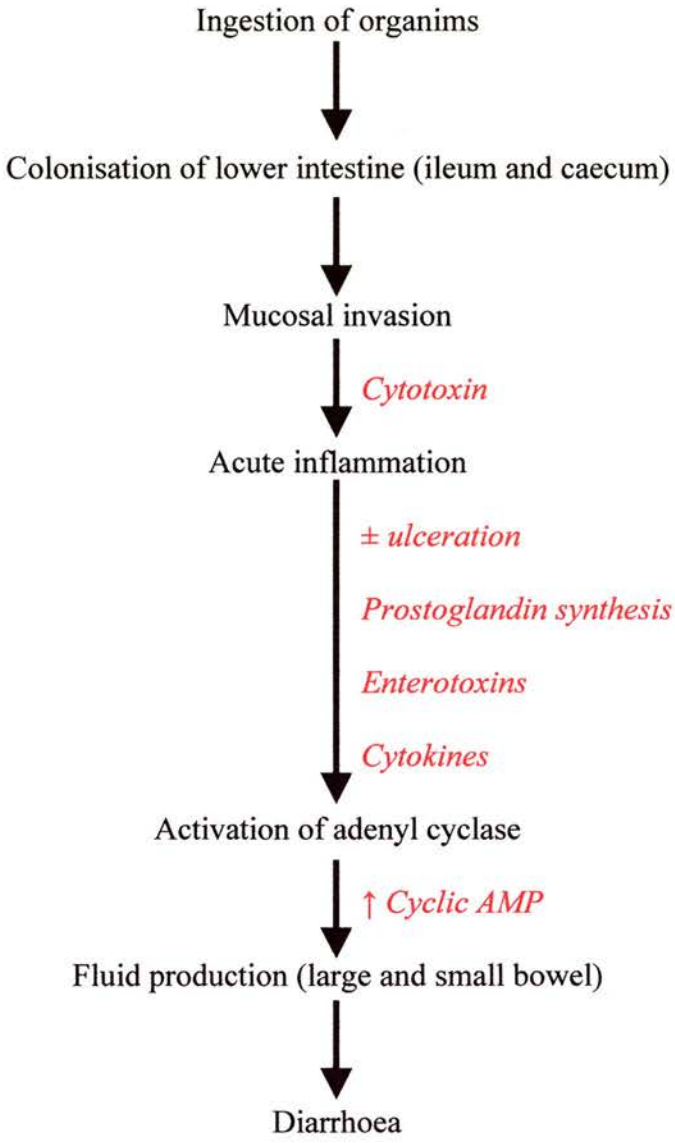


Figure 1.8. Scheme of the pathogenesis of Salmonella entercolitis and diarrhoea (adapted from http://gsbs.utmb.edu/microbook/ch021.htm).

In the susceptible host, salmonellae tend to cause disease ranging from mild entercolitis to severe diarrhoea. Nevertheless, some salmonellae are not restricted to the intestinal tract and can cause bacteraemia, sometimes resulting in focal infection or enteric fever (*S. typhi, S. paratyphi*) – the most severe form of salmonellosis (Eisenstein *et al.*, 1996; Elliott *et al.*, 1997). In humans this disease is known as typhoid fever and is caused by *S. enterica* serovar Typhi. In mice a similar disease evolves after infection with *S. enterica* serovar Typhimurium. Although human typhoid fever is now well controlled in many countries, the entercolitis and diarrhoea caused by various serovars represent a significant threat to public health and economy.

Salmonella enterica *serovar* Typhimurium

Salmonella enterica serovar Typhimurium (hereafter referred to as *S. typhimurium*) was until recently the most common cause of food poisoning by *Salmonella* species. As its name indicates, it causes typhoid-like disease in mice. In humans however, it causes food poisoning (that is not normally fatal), characterised by diarrhoea, abdominal cramps, vomiting and nausea, and generally lasts up to seven days (http://www.salmonella.org). Unfortunately, in immunocompromised people such as the elderly, the young, or people with depressed immune systems, these infections are often fatal if they are not treated with antibiotics.

Antimicrobial resistance

Investigations of *S. typhimurium* in many countries have demonstrated high levels of resistance to antimicrobial agents in humans as well as animal isolates. Furthermore, these investigations have shown that the occurrence of antimicrobial-resistant *S.*

typhimurium in domestic animals is associated with the use of antimicrobial agents in animal herds (Seyfarth *et al.*, 1997). Studies in the United Kingdom have indicated that salmonellae in pig and cattle herds develop resistance to antimicrobial agents more frequently than similar serotypes in poultry flocks (Threlfall *et al.*, 1993). Threlfall *et al.* (1997) have shown that around 20% of human isolates of *S. typhimurium* were resistant to one or more antimicrobial agents, a figure that had increased from less than 2% in August 1988. Almost half of all resistant isolates were multi-resistant.

1.5. The Streptococci

Classification

Streptococcus suis serotype 2 belongs to the family Streptococcaceae. Streptococci are capable of causing a variety of suppurative diseases in both humans and animals. Classification of the Streptococcaceae is complex and continually under review and as yet no single system of classification suffices to differentiate this heterogeneous group of organisms. Currently there are approximately forty described species of streptococci of which the majority have been shown to form six species groups by 16s RNA comparative analysis. Two species, *S. suis* and *S. acidominimus* have failed to cluster with any other streptococcal species so far tested although they exhibit a loose mutual association on the basis of rRNA sequence analysis (Hardie and Whiley, 1997).

More commonly the streptococci can be classified by their phenotypic, serologic and biochemical characteristics e.g. type of haemolysis produced around colonies

growing on blood agar, the Lancefield antigen in the bacterial cell wall (Lancefield, 1933), fermentation and other biochemical tests. *S. suis* is α or β haemolytic on sheep or horse blood agar, respectively (Table 1.2).

More precise identification of the streptococci is often accomplished by serogrouping the organisms. This is based on antigenic differences in cell wall carbohydrates and was pioneered by Lancefield (Lancefield, 1933). To date, serogroups A to H and K to V have been designated. Groups A, B, C, D and G are those most commonly found in humans (Table 1.3).

Table 1.2. Classification of streptococci based on haemolysis (adapted from Sleigh and Timbury, 1998).

Haemolysis	**Appearance**	**Designation**
Complete	Colourless, clear, sharply-defined zone	β
Partial	Greenish discolouration	α
None	No change	γ or non-haemolytic

Table 1.3. Streptococcal serogroups most frequently involved in human disease (adapted from Bisno and van de Rijn, 2000).

Serogroup	**Group Specific Cell Wall Antigen**	**Usual Clinical Features**
A	Rhamnose-*N*-Acetylglucosamine Polysaccharide	Pharyngitis, tonsillitis, otitis media, sinusitis, scarlet fever, pneumonia, septicaemia
B	Rhamnose-Glucosamine Polysaccharide	Chorioamnionitis, puerperal sepsis, meningitis
C	Rhamnose-*N*-Acetylglucosamine Polysaccharide	Upper respiratory infections
D	Glycerol teichoic acid	Genitourinary tract infections, wound infections, endocarditis
G	Rhamnose-Galactosamine Polysaccharide	Upper respiratory infections, cellulitis, septicaemia, deep tissue infections

Laboratory characteristics of Streptococci

Streptococci are Gram-positive spherical or oval cocci in pairs or chains with a diameter in the range of 0.7 to 0.9 μm. They grow well on blood agar aerobically but sometimes better anaerobically with 10% carbon dioxide added. Their major fermentative product is lactic acid that can accumulate in cultures and rapidly terminate growth (Sleigh and Timbury, 1998).

The importance of streptococci as both human and animal pathogens cannot be overestimated. Their disease-causing spectrum ranges from mild tonsillitis in humans to fatal meningitis and septicaemia in both humans and animals.

Streptococcus suis

S. suis was first described in 1966 and is a Gram-positive, haemolytic and facultatively anaerobic coccus. Based on its polysaccharide capsular antigens, at least 35 serotypes (types 1 to 34 and type 1/2) of *S. suis* exist (Staats *et al.*, 1997). Originally described strains were thought to be new Lancefield groups R, S, RS and T (de Moor, 1963). It was then demonstrated that the material described earlier as group antigen was in fact capsular material and that all strains were part of Lancefield group D. However, hybridisation studies have shown that *S. suis* is a genetically homogeneous species that is not closely related to group D streptococci. Groups S, R and RS were replaced by serotypes 1, 2 and 1/2, respectively (Elliott, 1966; Windsor and Elliott, 1975). Serotype 2 is the serotype most commonly associated with disease and also the one that is most frequently isolated (Higgins and Gottschalk, 1995).

S. suis serotype 2 (hereafter referred to as *S. suis*) is mainly a porcine pathogen and can cause a wide spectrum of disease syndromes including arthritis, meningitis, pneumonia, septicaemia, endocarditis, encephalitis and abscesses (Staats *et al.*, 1997). In the United States alone the economic loss due to *S. suis* infection is estimated to be in the region of 300 million dollars. *S. suis* is also a zoonotic pathogen and can occasionally cause disease in humans. In general, the disease is rare in man; however, it has been reported to be one of the major causes of meningitis in adults in Hong Kong (Lutticken *et al.*, 1986). Deaths from *S. suis* infection have been reported in slaughterhouse workers and butchers handling infected pork and *S. suis* infection may be viewed as an occupational hazard for those workers who come into contact with pigs or raw pork (Tayoro *et al.*, 1996). Control of the disease by vaccines and antimicrobials has generally been ineffective, partly because of increased resistance by *S. suis* to antimicrobials and a paucity of information regarding its virulence factors (Staats *et al.*, 1997). Recent work has demonstrated that the synthesis of two proteins, muramidase-released protein (MRP) and extracellular protein factor (EF) along with the possession of a thick capsule are required for *S. suis* to exert its full pathogenicity (Vecht *et al.*, 1991).

***Epidemiology of* S. suis**

In the United Kingdom studies have shown that with swine infections morbidity due to *S. suis* can range from less than 1% to more than 50% although it rarely exceeds 5% (Heard, 1984). Higher rates of infection can usually be attributed to poor hygiene and/or concurrent disease. Most outbreaks only affect only a small number of pigs and last a few weeks (Power, 1978). With prompt and appropriate treatment, mortality in swineherds is usually low although without treatment mortality can

approach 20%. Pigs of any age can be affected, but susceptibility generally decreases with age following weaning (Staats *et al.*, 1997). Many pig herds are now endemically infected, often with more than one serotype, and disease is not strictly related to the presence of serotype 2 (Chanter *et al.*, 1993).

S. suis can be carried in the tonsillar crypts of pigs for many months in the absence of apparent disease and may sometimes be cultured from the nose (Chanter *et al.*, 1993). Carriers of *S. suis* are infectious to other pigs and are significant in transmission of the disease; *S. suis* may persist in the tonsils of carrier pigs for over one year (Staats *et al.*, 1997).

Carriage of *S. suis* is not affected by circulating opsonic antibodies or administration of antibiotics with the intention of reducing bacterial shedding. Often episodes of disease seem to correlate with the appearance of strains of increased virulence and changes in host susceptibility particularly related to age, overcrowding and poor ventilation (Chanter *et al.*, 1993).

Vectors such as houseflies can also play a role in the transmission of *S. suis* and have been shown to transmit the organism between farms (Staats *et al.*, 1997). *S. suis* is also an important contaminant of faeces, dust and water. In water the organism can survive for ten minutes at 60°C while in faeces it can survive for three months at 0°C (Clifton-Hadley and Enright, 1984). Formites, such as manure covered work boots and needles have also been shown to be culpable in the transmission of the organism (Staats *et al.*, 1997).

***Pathogenesis and virulence factors of* S. suis**

Many factors play a role in the pathogenesis of *S. suis*. Different bacterial structures, such as capsule, fimbriae (pili), haemagglutinins, extracellular and cell wall associated proteins and haemolysis have all been put forward as possible virulence factors (Brazeau *et al.*, 1996). Several pathogenic mechanisms have been suggested for *S. suis* infections and it is generally accepted that the roles these play in meningitis include:

- Entry of *S. suis* into the blood from tonsils
- Uptake of the bacteria by monocytes
- Transport of bacteria to the cerebrospinal fluid *via* the choroid plexus
- Stimulation of cytokine production by monocytes/macrophages, which leads to an inflammatory infiltrate from the blood to the cerebrospinal fluid.

This increase in cells in the cerebrospinal fluid blocks sites of fluid efflux, increases intracranial pressure and produces the neural damage typical of the clinical signs of meningitis (Lutticken *et al.*, 1986).

Virulence differs among various strains of *S. suis*. As already mentioned it has been shown in *S. suis* serotype 2 that virulence correlates with the possession of a 136kDa protein on the bacterial cell surface released by muramidase (MRP) and with a secreted protein of 110kDa termed extracellular factor (EF) (Vecht *et al.*, 1991). Strains lacking both these proteins tend to be largely isolated from healthy pigs. In experimental infections of pigs these strains were found to be avirulent. Of current interest is the role of the capsule in *S. suis* virulence. Capsules have long been

recognised as being an important component of successful invasion and disease production by many bacteria. The serotype 2 polysaccharide is composed of five sugars: rhamnose, galactose, glucose, *N*-acetyl glucosamine and sialic acid in a molar ratio of 1.07:3.17:1.00:0.94:1.00 (type 1 substitutes the rhamnose with *N*-acetyl galactosamine) (Elliott and Tai, 1978). Work by Smith *et al.* (1999) has shown that the capsulate of *S. suis* serotype 2 is a surface component with antiphagocytic activity. This work also showed that wild-type encapsulated bacteria were ingested by phagocytes at very low frequency whereas mutant unencapsulated bacteria were efficiently ingested by porcine macrophages. This suggested that the loss of capsular material is associated with loss of the capacity to resist uptake by macrophages. The loss of resistance to *in vitro* phagocytosis was associated with an almost complete attenuation of the virulence of the mutant strains in germ free pigs (Smith *et al.*, 1999). Importantly, the work showed that all pigs inoculated with the mutant strains survived the experiment and did not show any specific clinical signs of disease. This supports the proposal that in *S. suis*, as in other pathogenic streptococci, the capsule acts as an important virulence factor.

One of the best-known mechanisms by which capsules protect bacteria is by impairing their interaction with phagocytic cells. Recent work has shown that the capsule of *S. suis* does play an antiphagocytic role; even the presence of compliment did not have a significant influence on phagocytosis of *S. suis* (Brazeau *et al.*, 1996). It has been shown that phagocytosed, non-pathogenic, capsulated *S. suis* organisms are killed whereas intracellular, capsulated, pathogenic *S. suis* organisms are able to survive and replicate within the phagosome for up to three hours (Williams, 1990). The mechanism by which virulent capsulated strains of *S. suis* are able to do this

remains unknown although this intracellular survival does not seem to be due to superoxide dismutase synthesis (Langford *et al.*, 1991). It is now generally accepted that in the pathogenesis of *S. suis* meningitis, bacteria enter into the cerebrospinal fluid compartment in association with monocytes.

S. suis *as a zoonotic infection*

In humans, *S. suis* can cause meningitis and/or septicaemia followed by deafness (Staats *et al.*, 1997). In nearly all cases of *S. suis* disease published so far, meningitis is the main clinical feature. Septicaemia without involvement of the meninges is rare (Lutticken *et al.*, 1986). In man the complications from *S. suis* infection are due to bacteraemia for which skin lesions may be the portal of entry. Permanent deafness is the most frequently occurring neurologic sequela typically due to the affection of the eighth cranial nerve occurring early in the course of the disease (Lutticken *et al.*, 1986). Between 1968 and 1984, the estimated annual risk of developing meningitis due to *S. suis* among Dutch abattoir workers compared to the general population was 3.5 versus 0.002 / 100,000, respectively (Arends and Zanen, 1988).

Although the disease is sporadic it may be under diagnosed because *S. suis* can often be mistaken for other organisms with a similar appearance (Lutticken *et al.*, 1986). Because many hospital laboratories are not aware of this streptococcus, it may be primarily identified as enterococcus or pneumococcus. In general *S. suis* disease in man has a favourable prognosis if the diagnosis is established early and treatment is started without delay. Generally high dose intravenous penicillin G is the first choice of drug although ampicillin appears to act just as well (Lutticken *et al.*, 1986).

***Treatment and control of* S. suis**

The treatment of choice for pigs exhibiting signs of *S. suis* infection is injectable ampicillin or penicillin with supportive nursing care (Staats *et al.*, 1997). Medication of feed with antibiotics at therapeutic levels may suppress clinical disease while it is being done but does not eliminate carriers (Amass *et al.*, 1996). Research has shown that the most active drugs against *S. suis* are the β-lactams; unfortunately because of high levels of antimicrobial resistance, antimicrobial susceptibility testing of *S. suis* isolates should be performed prior to initiation of antibiotic therapy (Turgeon *et al.*, 1994). Plasmids may carry antimicrobial resistance genes and there is evidence that further types of resistance in *S. suis* may be transposon mediated (Staats *et al.*, 1997).

Control of *S. suis* in pigs using vaccines has generally been unsuccessful (too many capsule types), as has been the use of antimicrobials for prophylaxis and treatment. With the advent of intensive management in swine production, *S. suis* has become a major problem in swine producing countries. Along with the problems of increased antimicrobial resistance and lack of knowledge concerning all the organism's virulence factors, there is now a real need to produce a drug capable of combating and controlling this important pathogen.

1.6. The Rhamnose Pathway

The immediate source of rhamnose in carbohydrate polymers in bacteria is L-rhamnose. L-rhamnose is formed from its precursor dTDP-L-rhamnose, which is in turn synthesised from glucose-1-phosphate and deoxythymidine triphosphate (dTTP). The biosynthesis of L-rhamnose has been studied for over three decades; however it is only relatively recently that there has been interest in it as a potential therapeutic target (Giraud and Naismith, 2000). The pathway itself involves a number of interesting reactions, involving four distinct enzymes that are highly conserved amongst microorganisms (Graninger *et al.*, 1999; Ma *et al.*, 1997) (Figure 1.9): α-D-glucose-1-phosphate thymidylyl transferase (RmlA, EC: 2.7.7.24), dTDP-D-glucose 4,6-dehydratase (RmlB, EC: 4.2.1.46), dTDP-6-deoxy-D-xylo-4 hexulose 3,5-epimerase (RmlC, EC: 5.1.3.13) and dTDP-6-deoxy-L-*lyxo*-4-hexulose-4-reductase (RmlD, EC: 1.1.1.133) (Allard *et al.*, 2001a; Blankenfeldt *et al.*, 2000; Giraud *et al.*, 1999; Giraud *et al.*, 2000). The fact that the L-rhamnose biosynthetic pathway is not found in humans makes all four enzymes potential targets for antibacterials.

The first enzyme RmlA, catalyses the transfer of a thymidylmonophosphate nucleotide to glucose-1-phosphate. This enzyme is important in that it is allosterically regulated by dTDP-L-rhamnose and hence provides a feedback mechanism that regulates the production of L-rhamnose in bacteria. The enzyme is a homotetramer (Figure1.10a) with each monomer consisting of three functional domains (Blankenfeldt *et al.*, 2000).

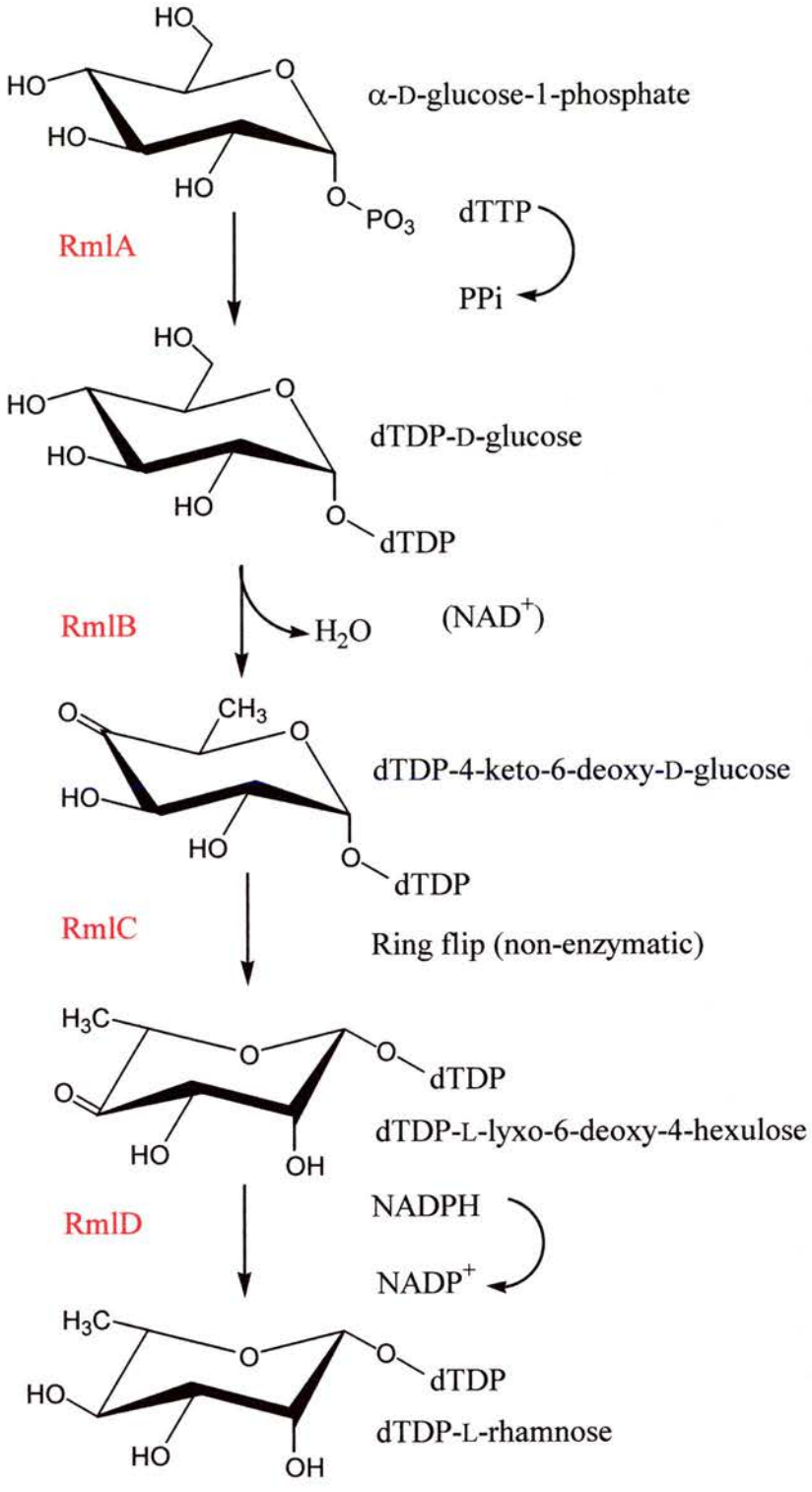


Figure 1.9. The dTDP-L-rhamnose pathway.

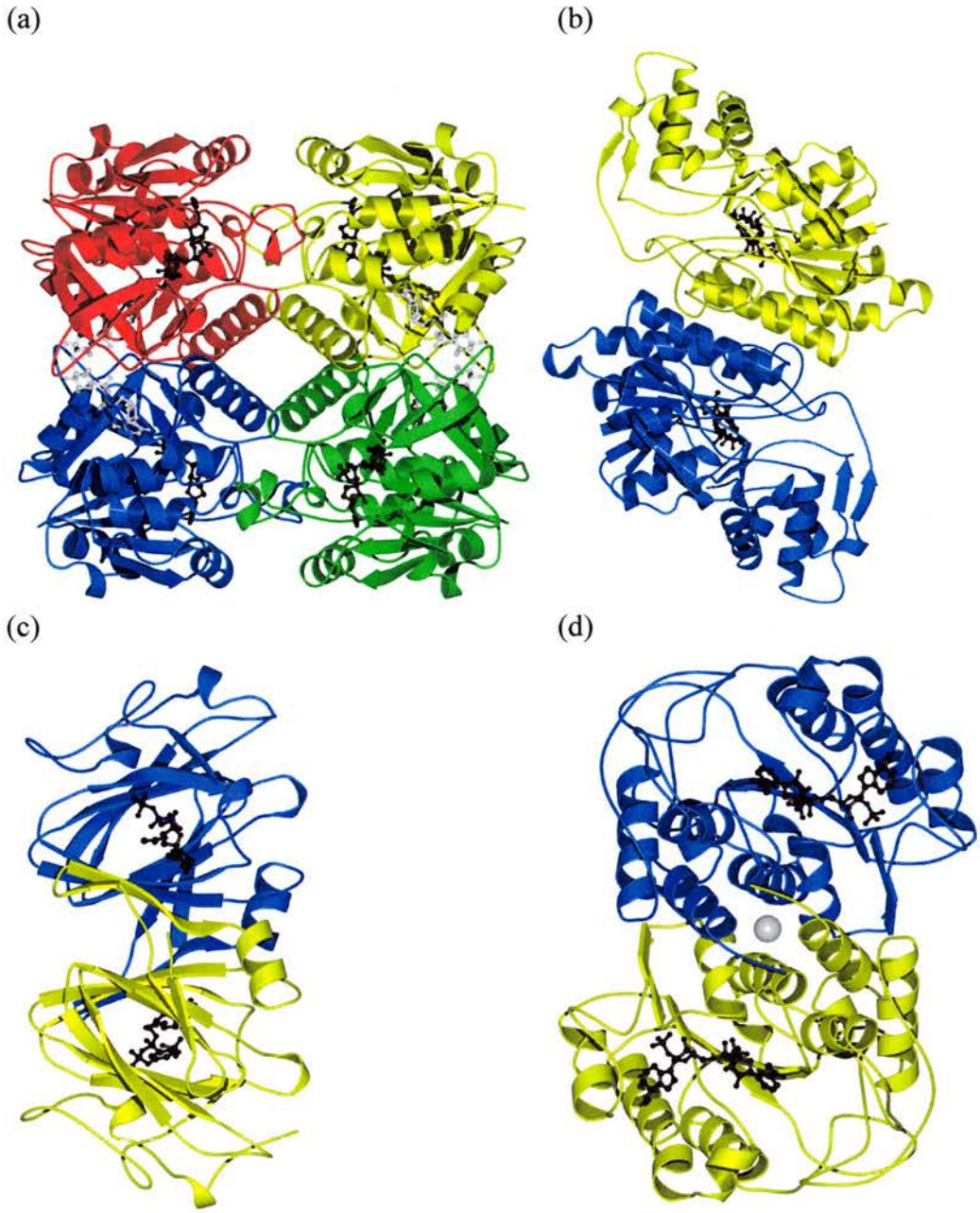


Figure 1.10. The four enzymes of the dTDP-L-rhamnose pathway. (a) α-D-glucose-1-phosphate thymidylyl transferase (RmlA, EC: 2.7.7.24) (Blankenfeldt et al., *2000), (b) dTDP-D-glucose 4,6-dehydratase (RmlB, EC: 4.2.1.46) (Allard* et al., *2001a), (c) dTDP-6-deoxy-D-xylo-4 hexulose 3,5-epimerase (RmlC, EC: 5.1.3.13) (Giraud* et al., *2000) and (d) dTDP-6-deoxy-L-lyxo-4-hexulose-4-reductase (RmlD, EC: 1.1.1.133) (Blankenfeldt, personal communication). All molecular representations in this thesis were prepared with BOBSCRIPT (Esnouf, 1999).*

RmlB catalyses the second step in the pathway; the dehydration of dTDP-D-glucose to dTDP-4-keto-6-deoxy-D-glucose. The enzyme functions as a homodimer with each monomer consisting of two domains (Figure 1.10b) (Allard *et al.*, 2001a). The mechanism of RmlB has been intensively studied for 40 years (Frey, 1987; Gabriel and Lindquist, 1968; Glaser, 1963; Gross *et al.*, 2000; He *et al.*, 2000; Hegeman, 2001; Hegeman *et al.*, 2001; Melo *et al.*, 1968; Snipes *et al.*, 1977; Tanner, 2001; Wang and Gabriel, 1970; Zarkowsky *et al.*, 1970), first being proposed back in 1963 [Glaser, 1963 #4978}. The mechanism involves oxidation of dTDP-D-glucose at glucosyl-C4' by nicotinamide adenine dinucleotide (NAD^+) to generate the intermediate dTDP-4-ketoglucose and NADH. There is then dehydration across the C5 and C6 bond of the dTDP-4-ketoglucose intermediate to produce a second intermediate dTDP-4-keto-glucose-5,6-ene which is then reduced by NADH at glucosyl-C6' forming the product dTDP-4-keto-6-deoxy-D-glucose and regenerating NAD^+.

RmlC, the third enzyme in the pathway catalyses a double epimerisation at positions C3 and C5 to give dTDP-L-lyxo-6-deoxy-4-hexulose. The structure of RmlC from *S. typhimurium* was reported last year (Giraud *et al.*, 2000) closely followed by the enzyme from *Methanobacterium thermoautotrophicum* (Christendat *et al.*, 2000). RmlC is a mainly beta class protein and functions as a dimer with the active site being formed from both subunits (Figure 1.10c). The epimerisation of the C3 and C5 positions requires that the hydrogens attached to the C3 and C5 carbons are removed and replaced at the opposite face. The mechanism of RmlC involves a series of at least four sequential steps and further work is being done in order to establish and

understand the full mechanism of this epimerase (Asuncion & Naismith, unpublished results, Dong & Naismith, personal communication) (Giraud *et al.*, 2000).

The final reaction in the dTDP-L-rhamnose pathway is catalysed by RmlD, which reduces the C4 keto group of dTDP-L-lyxo-6-deoxy-4-hexulose to hydroxyl giving the product dTDP-L-rhamnose (Figure 1.10d). During the catalysis, RmlD utilises nicotinamide adenine dinucleotide phosphate (NADPH) as the reducing cofactor. Like RmlB, the enzyme is a homodimer and a member of the short-chain dehydrogenase/reductase family. However, in contrast to other well-known family members the monomer/monomer interface is in an entirely different location in RmlD and contains a magnesium-binding site (Blankenfeldt, personal communication). Like RmlB, each monomer can be subdivided into two domains, an N-terminal nucleotide-binding Rossmann-fold possessing a glycine rich motif (GlyXXGlyXXGly in RmlD) and a smaller C-terminal substrate-binding module that contains the commonly found Thr-Tyr-Lys catalytic triad (Giraud and Naismith, 2000; Giraud *et al.*, 1999).

1.7. Short and Medium Chain Dehydrogenases/Reductases

The short-chain dehydrogenases/reductases (SDR) family constitutes a group of over 60 characterised proteins representing a wide variety of enzyme activities (Persson *et al.*, 1995; Persson *et al.*, 1999). The first characterised member of the SDR family was the fruit-fly alcohol dehydrogenase in 1976 (Schwartz and Jornvall, 1976). In addition to the SDR enzymes which typically have subunits of 250-odd residues there are also the slightly larger medium-chain dehydrogenases/reductases (MDR) which have subunits of around 350 residues. Originally both the SDR and MDR families were regarded as special, with SDR enzymes found only in prokaryotes and insects (Jornvall *et al.*, 1981) and with MDR proteins representing only enzymes with an active-site zinc in a widely spaced ligand pattern (Eklund *et al.*, 1985). However, this distinction has now been removed with evidence showing both SDR and MDR enzymes to be of wide occurrence with diverse activities (Jornvall *et al.*, 1999).

Multiple alignments of SDR enzymes show that only 23 residues are conserved in more than 70% of the structures (Persson *et al.*, 1995). Amongst these conserved residues one finds the highly conserved N-terminal coenzyme-binding motif containing three glycines, GlyXGlyXXGly (as in RmlB). The glycines are positioned so as to participate as part of a Rossmann fold (βαβ motif common to nucleotide binding enzymes) (Rossmann *et al.*, 1975; Thatcher and Sawyer, 1980). There is also a highly conserved TyrXXXLys couple as part of the enzyme's active site. Despite their grouping in a family, the SDR enzymes are highly divergent with a typical pairwise identity of 15-30%, covering a wide range of substrate specificity, including steroids, alcohols and aromatic compounds (Jornvall *et al.*, 1999), (Table 1.4).

Importantly, there also exists SDR enzymes with around 350 residues that are more distantly related which exhibit dehydrogenase, dehydratase, epimerase or isomerase activity; for example RmlB and UDP-galactose 4-epimerase (GALE) which are considered as part of the extended SDR family (Allard *et al.*, 2001a; Jornvall *et al.*, 1995; Persson *et al.*, 1995; Thoden *et al.*, 1996a). The main difference between the extended SDR family and the shorter classical SDR family (apart from the subunit length) is that the extended SDR enzymes tend to have subunits that consist of two domains (as do MDR enzymes) as compared to the one domain SDR subunit (Jornvall, 1999; Persson *et al.*, 1995). An alignment of UDP-galactose 4-epimerase from *Escherichia coli* (EGALE) with 3α/20β-hydroxysteroid dehydrogenase (the first SDR structure solved by X-ray crystallography (Ghosh *et al.*, 1991)) and dihydropteridin reductase (both SDR enzymes) shows that the enzymes have a block of about 180 residues in common encompassing the important conserved residues, indicating similar catalytic mechanisms (Persson *et al.*, 1995).

Table 1.4. The enzyme activity types represented by members of the extended SDR family (adapted from Persson et al., *1995).*

Enzyme activity type	EC class
Dehydrogenases	EC 1
Reductases	EC 1
Dehydratases	EC 4.2
Epimerases	EC 5.1
Isomerases	EC 5.3

As pointed out by Persson *et al.* (1995) most of the enzymes of the SDR family are equidistantly related, with enzymes like GALE and RmlB forming a separate sub-

family. However, it is now common to include enzymes like GALE and RmlB as part of an SDR 'superfamily'. Since the SDR family of enzymes shares a rather low degree of sequence identity/homology but a remarkably similar tertiary fold, it can help serve in the understanding of protein folding and evolution (Grundy *et al.*, 1997).

1.8. SDR Enzymes similar to RmlB

A number of SDR enzymes are both structurally and mechanistically similar to RmlB; the most studied being UDP-galactose 4-epimerase from both *E. coli* and humans (GALE, interconverts UDP-glucose and UDP-galactose, Figure 1.12) (Berger *et al.*, 2001; Thoden *et al.*, 1996a; Thoden *et al.*, 1996b; Thoden *et al.*, 1996c; Thoden *et al.*, 1997a; Thoden *et al.*, 1997b; Thoden and Holden, 1998; Thoden *et al.*, 2000; Thoden *et al.*, 2001). Other key enzymes include GDP-mannose 4,6-dehydratase (GMD, converts GDP-D-mannose to GDP-4-keto-6-deoxy-D-mannose, Figure 1.13b) (Somoza *et al.*, 2000), SQD1 (involved in the biosynthesis of UDP-sulfoquinovose, Figure 1.14) (Mulichak *et al.*, 1999) and ADP-L-glycero-D-mannoheptose 6-epimerase (AGME, catalyses the interconversion of ADP-D-glycero-D-mannoheptose and ADP-L-glycero-D-mannoheptose, Figure 1.15) (Deacon *et al.*, 2000). All of these enzymes contain a catalytic triad that includes a TyrXXXLys couple and a characteristic glycine rich (GlyXGlyXXGly, as seen in RmlB) motif at the N-terminus. The initial member of the catalytic triad is either a serine as in GALE and AGME or a threonine as found in RmlB, SQD1 and GMD. GALE, RmlB, GMD and SQD1 function as homodimers but AGME is a homopentamer (Figure1.11).

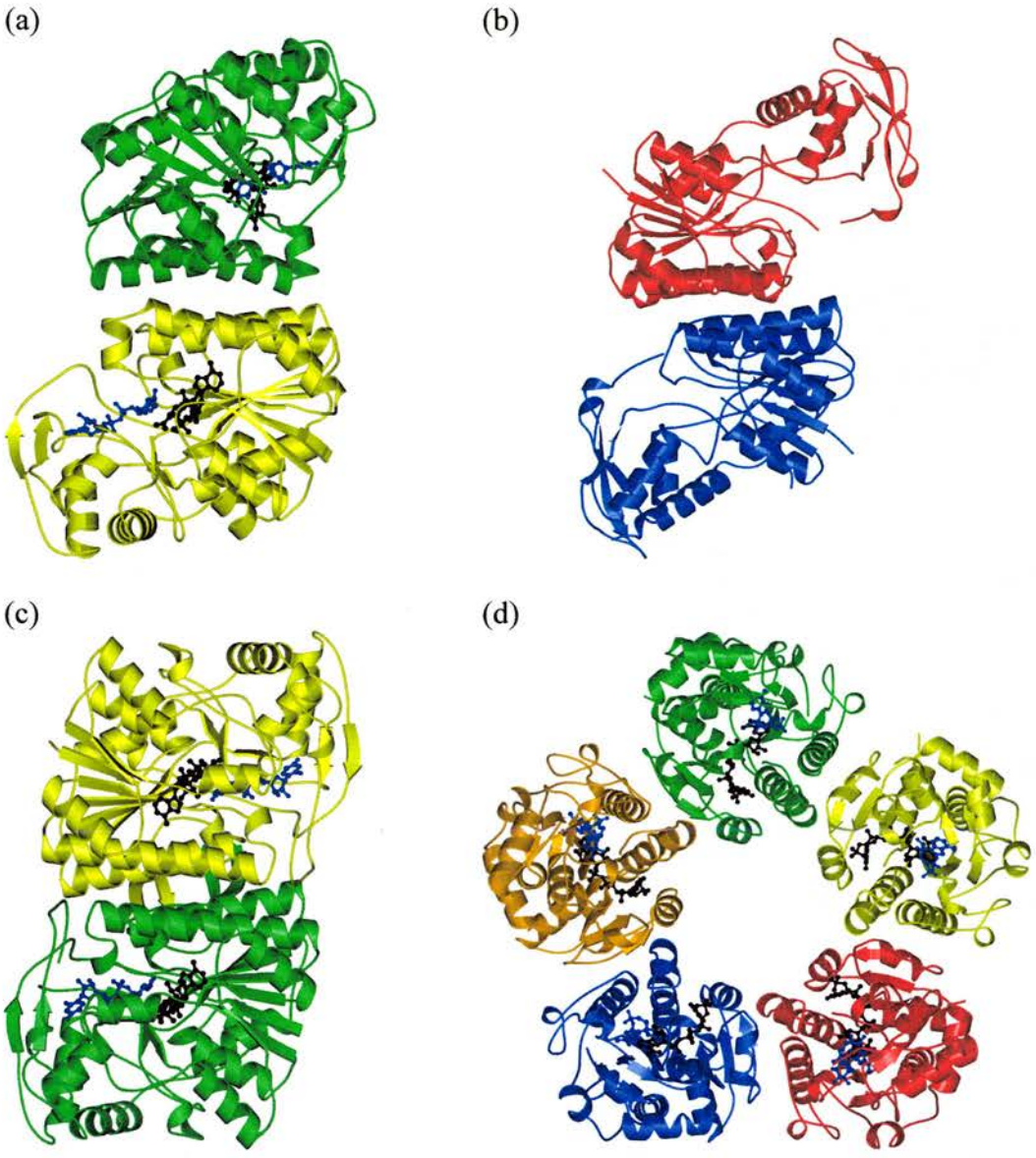


Figure 1.11. (a) The abortive complex of E. coli *UDP-galactose 4-epimerase with NADH (black) and UDP-glucose (blue) bound, PDB: 1XEL (Thoden* et al., *1996b); (b) the apoenzyme form of GDP-mannose 4,6-dehydratase from* E. coli, *PDB: 1DB3 (Somoza* et al., *2000); (c) SQD1 from* Arabidopsis thaliana *with* NAD^+ *(black) and UPD-glucose (blue) bound, PDB: 1QRR (Mulichak* et al., *1999); (d) the homopentameric ADP-L-glycero-D-mannoheptose 6-epimerase from* E. coli *with* $NADP^+$ *(black) and ADP-glucose (blue) bound, PDB: 1EQ2, (Deacon* et al., *2000).*

UDP-galactose 4-epimerase

GALE catalyses the interconversion of UDP-glucose and UDP-galactose by inverting the stereochemistry at the C4 position (Figure 1.12). The enzyme thus obviates the need for the *de novo* synthesis of both sugars (Carnell, 1999). The biological interconversion of galactose and glucose is known as the Leloir pathway, and in addition requires the enzymes galactokinase and galactose-1-P uridylyltransferase (Frey, 1996; Leloir, 1951). In humans, the role of GALE in galactose metabolism is critical. Two forms of GALE deficiency have been described: one is benign and involves only red and white blood cells and not other tissues (the 'peripheral' form). The other is very rare and presents with symptoms resembling transferase deficiency (the 'general' form) (Gitzelmann *et al.*, 1977; Henderson *et al.*, 1983; Petry and Reichardt, 1998).

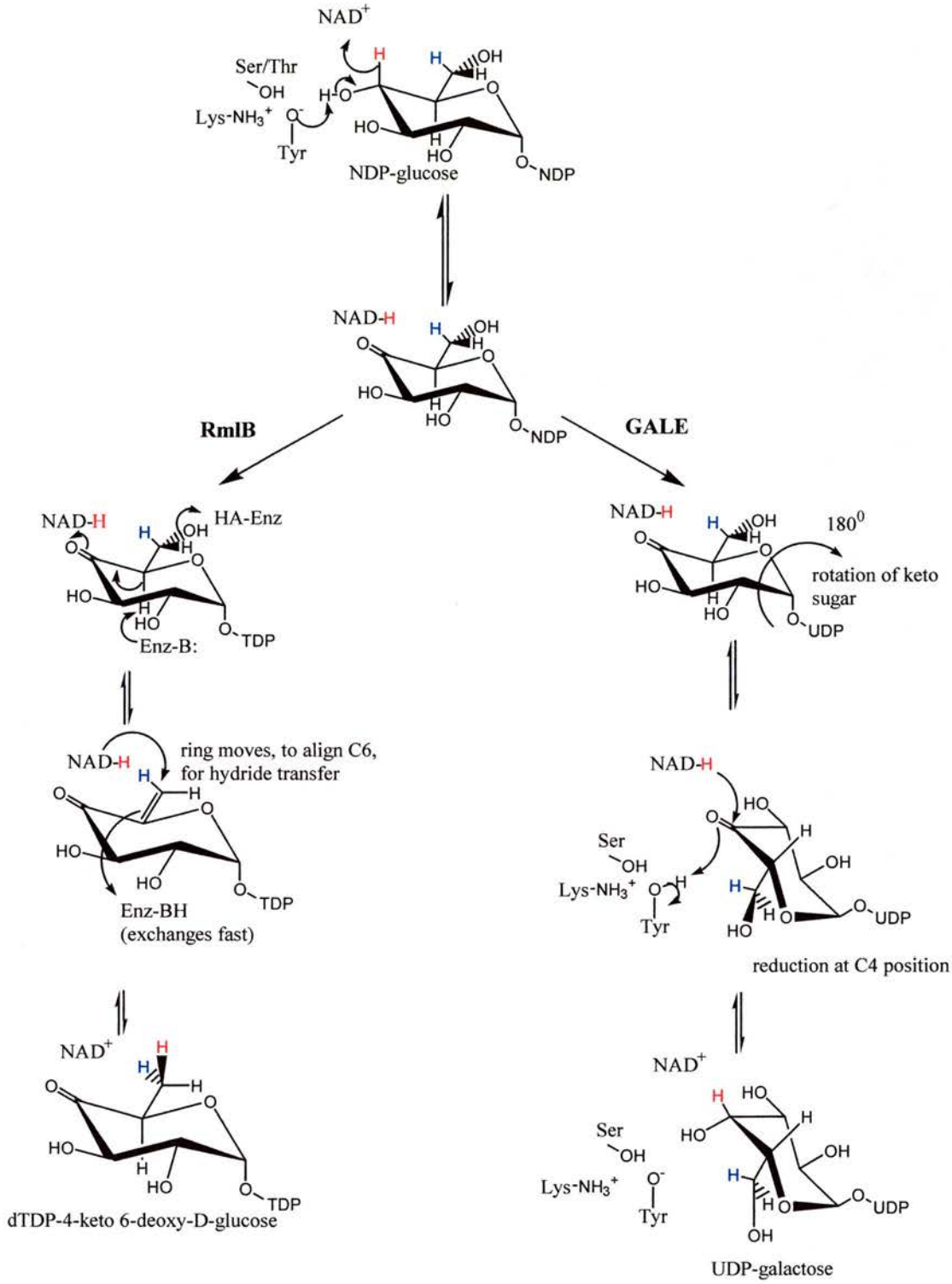


Figure 1.12. The mechanisms of RmlB and GALE compared. These enzymes have similar structures and mechanisms although the fate of the keto sugar is different.

GALE has been extensively studied from both *E. coli* (EGALE) (Bauer *et al.*, 1992; Thoden *et al.*, 1996a; Thoden *et al.*, 1996b; Thoden *et al.*, 1996c; Thoden *et al.*, 1997a; Thoden *et al.*, 1997b; Thoden and Holden, 1998) and more recently *Homo sapiens* (HGALE) (Thoden *et al.*, 2000; Thoden *et al.*, 2001). Structurally, HGALE is 55% identical to the homologous *E. coli* enzyme (Thoden *et al.*, 2000). Both HGALE and EGALE are homodimers with molecular masses of 38.3 and 37.3 kDa for their respective monomers. Each monomer consists of 348 (HGALE) and 338 (EGALE) amino acids with one molecule of NAD^+ in the *syn*-conformation with its *si*-face oriented towards the sugar substrate. The monomers fold into two distinct domains; an N-terminal or nucleotide binding domain dominated by a seven-stranded parallel β-pleated sheet flanked on either side by α-helices, and a smaller C-terminal domain responsible for the proper positioning of the UDP-sugar substrates. The crevice created by the two domains contains the active site (Thoden *et al.*, 1996a; Thoden *et al.*, 1996b; Thoden *et al.*, 1996c). The structure of EGALE is shown in Figure 1.11a. The structures of both EGALE and HGALE are virtually identical, as they can be superimposed with a root mean square deviation of 0.98Å for 299 structurally equivalent α-carbon atoms (Thoden *et al.*, 2000). A subtle difference is that the human enzyme has an active site volume ~15% larger than that observed for EGALE and the substitution of an active site amino acid Tyr299 (EGALE) for the less bulky Cys307 (HGALE) (Thoden *et al.*, 2001). This is may be related to the secondary role of HGALE in catalysing the interconversion of UDP-*N*-acetylgalactosamine and UDP-*N*-acetylglucosamine.

GALE contains three amino acids that have been identified by site directed mutagenesis as important in the catalytic process: Lys153, Tyr149 and Ser124

(numbers refer to EGALE enzyme) (Swanson and Frey, 1993; Thoden *et al.*, 1996b). The conservation of Lys and Tyr has already been highlighted across the SDR superfamily, with many of the other members also containing either a Ser or Thr residue as the third member of the catalytic triad.

The broad details of the GALE mechanism were worked out in the 1970s (Wee *et al.*, 1972; Wee and Frey, 1973). The first step is an abstraction of the 4-hydroxyl proton by an enzymatic base and an abstraction of a hydride from the C4' position of the sugar to the C4' position on NAD^+ to form NADH. Studies have shown that Tyr149 is in the phenolic form and is stabilised in this form by Lys153 (Liu *et al.*, 1997). It therefore seemed obvious that Tyr149 acts as the main base deprotonating the 4-hydroxyl group. However, with the exception of the HGALE structure, Tyr149 is too far from the substrate to directly abstract this proton (4.3Å) (Thoden *et al.*, 1996b). To solve this conundrum, a proton shuttle mechanism had been advanced, in which the hydroxyl side chain of the conserved serine (Ser124) abstracts the proton from the substrate and loses its own proton to Tyr149 (Thoden *et al.*, 1996a). At the active site a transient keto sugar is formed which of course has no chirality at the C4' position. The keto sugar is not released by the enzyme and remains bound (Frey, 1996). The UDP-group serves a vital role in the mechanism by anchoring the pyranosyl groups of the galactosyl/glucosyl and 4-ketosugar intermediate in the active site. Binding studies have shown that in EGALE most of the binding free energy between the enzyme and these molecules is directed towards binding the nucleotide portion (Kang *et al.*, 1975; Wong and Frey, 1977). In the final step of the process NADH transfers the hydride back to the C4' of the sugar, but this time to the opposite face, with

inversion of configuration at C4' of the sugar. The proton extracted by Tyr149 (or Ser124) is transferred back to the sugar.

A key question is how the hydride is abstracted from one face and transferred to the other. The NAD^+/NADH transformation in GALE is itself stereospecific, with hydrogen transfer exclusively to and from the *si*-face. This would seem to indicate that hydride transfer in GALE must follow a stereospecific and fairly well defined trajectory, even if the overall result is nonstereospecific. Cursory examination of the substrate complex confirms that it is not possible to abstract and add hydride to opposite faces without a change in the orientation of sugar substrate relative to the cofactor. Weak binding of the 4-ketosugar intermediate in the active site suggests that it may be conformationally mobile in the active site. This has led to the proposition that there is a rotation of the 4-ketosugar intermediate within the active site around the β-phosphate of UDP. The α-phosphate and uridine remain fixed in position during this rotation. This would place the 4-keto group in a position relative to the NADH that would allow it to accept hydride transfer from either face to form UDP-glucose or UDP-galactose. Indeed, space filling models show that this rotation can allow virtually equivalent positioning of the hydroxyl groups at glycosyl-C3' and C4' and the hydrogen at C4' in the substrate epimers (Thoden and Holden, 1998). This rotation of the carbohydrate portion of the substrate allows the NADH to transfer the hydride to the opposite face while conserving an optimum trajectory for hydride transfer. Such a revolving door mechanism is unusual in biology.

At a more subtle level, it is known that the uridine nucleotide also induces a conformational change in GALE. The conformational change activates the NAD^+ to

reduction (Frey, 1987; Frey, 1996; Nelsestuen and Kirkwood, 1971; Wong and Frey, 1977). The structure of the enzyme in its active conformation reveals that the 6-ammonium group of Lys153 is hydrogen bonded to both the 2'- and 3'-hydroxyl groups of the nicotinamide riboside in NAD^+ (Thoden *et al.*, 1996b). The 6-ammonium group is seen to be much closer to the nicotinamide-N1 (5.3Å) than to the nicotinamide C-4 (>7Å). The positive electrostatic field between Lys153 and nicotinamide-N-1 in the active conformation should polarise the π-electrons in the nicotinamide ring, decreasing the positive charge on N-1 and increasing it on C4, thus activating it for hydride abstraction. The effect is brought about through electrostatic repulsion between the nicotinamide ring and a residue of the enzyme, Lys153, and not through enzymatic binding of the transition state (Frey, 1996).

Recently the structure of HGALE has been solved which has led to some revision of the epimerase mechanism (Thoden *et al.*, 2000). Upon binding of the UDP-glucose, there is a structural movement of one of the HGALE subunits relative to the other causing the C-terminal domain of one monomer to clamp down more tightly over the active site. The major significance of this is that the active site tyrosine (Tyr157) is now within hydrogen-bonding distance not only of the nicotinamide ribose 2'-hydroxyl group and the 3'-hydroxy group of the glucose moiety, but also the 4'-hydroxyl group of the UDP-sugar. The net result of this would be to enable the tyrosine to act directly as the active site base, without the need for the serine (Ser132) to act as a proton shuttle. This subunit movement has not yet been observed for EGALE and the possibility that EGALE and HGALE have slightly different mechanisms cannot be ruled out. The function of Ser132 in HGALE could be to form a low-barrier hydrogen bond between the O^{γ} of Ser132 and the 4'-hydroxyl group of

the UDP-glucose substrate. This would lead to both the pK_a values of the alcohol groups being similar resulting in a proton being shared between two oxygen atoms. This close interaction would facilitate the removal of the 4-hydroxyl hydrogen of the sugar by the phenolic side chain of Tyr157 and the transfer of the hydride from C4' of the sugar to C4' of the dinucleotide cofactor. Interestingly, these new data on the HGALE enzyme imply that the knowledge of one, or even a few members of an enzyme superfamily may be not sufficient to fully characterize either the structures or functions of all the members (Thoden *et al.*, 2000). Further studies are probably required to fully appreciate the exact role of Ser124.

GDP-mannose 4,6 dehydratase

GDP-mannose 4,6-dehydratase (GMD) catalyses the conversion of GDP-D-mannose to GDP-4-keto-6-deoxy-D-mannose. This is the first step in the biosynthesis of GDP-L-fucose (Figure 1.13). In bacteria fucose is found as part of the cell wall and has been shown to play a role in pathogenicity (Flowers, 1981) and nodulation (Mergaert *et al.*, 1997). In humans, fucose is widely distributed in many glycoconjugates, perhaps the best-known example being that of the human blood group antigens. More recently, fucose has been shown to comprise part of the epitope recognised by the selectin family of cell adhesion molecules and play a role in immune function, inflammation, and metastasis (McEver, 1997). In humans, defects in the reaction catalysed by GMD have been linked to the immune disorder leukocyte adhesion deficiency type II (LADII) (Etzioni *et al.*, 1992; Karsan *et al.*, 1998). Patients with this defect fail to fucosylate their glycoconjugate and as a consequence show the rare Bombay blood group. They also fail to synthesise fucosylated ligands for selectins,

are deficient in selectin-mediated cell adhesion and have defects in leukocyte extravasation (Sturla *et al.*, 1998).

In all species looked at so far the primary biosynthetic route to GDP-fucose is from GDP-mannose (Somoza *et al.*, 2000). This process is accomplished by two enzymes, GMD and GDP-fucose synthetase (GFS). These two enzymes convert GDP-mannose to GDP-fucose in a three-step reaction (Figure 1.13a). The first step of the reaction is catalysed by GMD and involves the oxidation of the C4 position of the mannose ring to a keto functionality followed by the reduction of the C6 carbon to a methyl group (Figure 1.13b). This concerted oxidation and reduction involves a intramolecular hydride transfer from the C4 to the C6 on the mannose ring by a tightly bound $NADP^+$ coenzyme, which is transiently reduced during the course of the reaction.

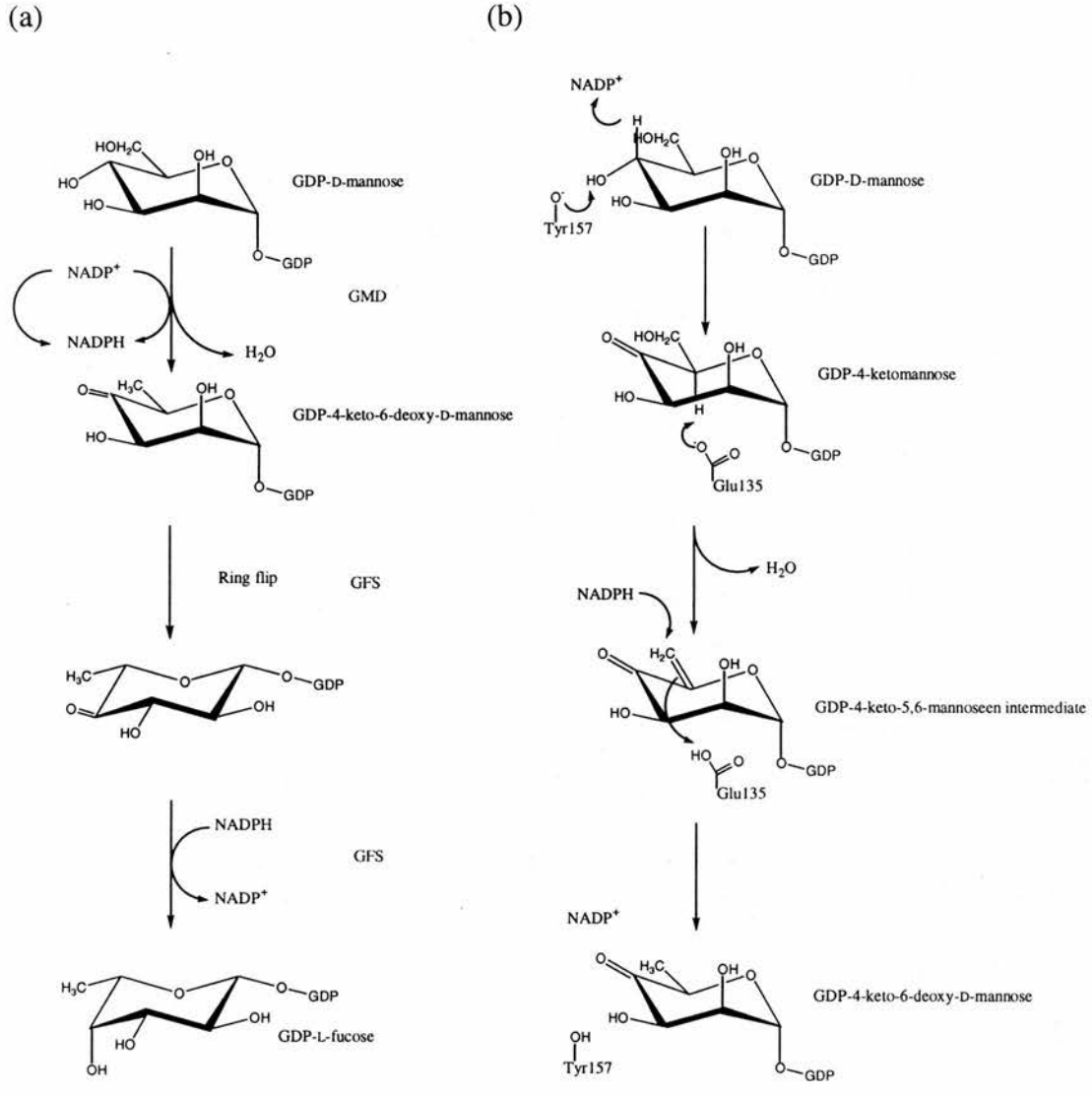


Figure 1.13. (a) GDP-fucose biosynthesis. $NADP^+$ bound to GMD is reduced and then oxidised during the course of the reaction. GFS catalyses two further distinct reactions: the epimerisation of GDP-4-keto-6-deoxy-D-mannose at C3 and C5 followed by the subsequent reduction at C4 to give the product GDP-fucose; (b) the proposed mechanism of GMD by Somoza et al. *(2000). Here Tyr157 plays the role of active site base deprotonating the 4-hydroxyl and Glu135 functions as a general base to remove the C5 proton during the dehydration step.*

The structure of the apoenzyme form of GMD from *E. coli* was solved in 2000 (Somoza *et al.*, 2000) (Figure 1.11b). The structure reveals a homodimer, with each monomer being composed of two domains. The larger N-terminal domain binds the NADP(H) coenzyme and contains a Rossmann fold consisting of seven-stranded β-sheet interwoven by seven α-helices. The slightly smaller C-terminal domain is responsible for binding the substrate. Somoza *et al.* (2000) modelled in the $NADP^+$ coenzyme and substrate, and performed site-directed mutagenesis studies on the active site catalytic triad, Thr133, Tyr157 and Lys161 as well as Glu135. The results obtained suggested the mechanism shown in Figure 1.13b. Their proposal that Glu135 plays the role of an active site base in the deprotonation/reprotonation of C5 during catalysis is important, in that this residue is totally conserved in all RmlB enzymes.

Sulfolipid Biosynthesis Protein (SQD1)

SQD1 is believed to be involved in the biosynthesis of the sulfoquinovosyl headgroup of plant sulfolipids, catalysing the transfer of SO_3^- to UDP-glucose (Figure 1.14). SQD1 binds NAD^+ and contains the Thr-Tyr-Lys catalytic triad (Thr145, Tyr182, Lys186 in *A. thaliana*) and glycine rich motif as part of a Rossmann fold found in RmlB and GMD. The structure of SQD1 from *A. thaliana* was solved in 1999 (Mulichak *et al.*, 1999) and shows the enzyme to be a homodimer (Figure 1.11c) with each monomer exhibiting a bidomain structure. The larger N-terminus binds the NAD^+ and consists of a seven-stranded parallel β-sheet with six-connecting α-helices. The C-terminal, as with the other enzymes discussed here, is responsible for binding the substrate. In terms of the mechanism (Figure 1.14), the first step involves proton abstraction from the UDP-glucose 4'-hydroxyl, with concomitant hydride

transfer from C4' to NAD^+, to produce the 4'-keto intermediate. The conserved tyrosine (Tyr182), in the form of a negatively charged tyrosinate, acts as the general base that abstracts the proton. This tyrosinate catalytic base hypothesis was originally suggested for *Drosophila* alcohol dehydrogenase (Chen *et al.*, 1993) and is now a generally accepted mechanistic principle among many SDR enzymes including GALE, GMD and RmlB (Berger *et al.*, 2001; Gerratana *et al.*, 2001; Liu *et al.*, 1997; Somoza *et al.*, 2000). Interestingly, in SQD1 the Tyr182 makes a close hydrogen bond (2.5Å) with the O4' hydroxyl of the bound UDP-glucose and may act directly as the catalytic base without the need for a proton shuttle mechanism (with Thr145) (Mulichak *et al.*, 1999) as proposed for EGALE (Thoden *et al.*, 1996a). A general base is also required to abstract a proton from the acidic glucosyl C5' atom during the formation of intermediate **II** (Figure 1.14); a role that Mulichak *et al.* (1999) propose for His183. Significantly, His183 occupies almost the same position as Glu135/Glu127 in the *S. typhimurium* and *S. suis* RmlB structures, respectively.

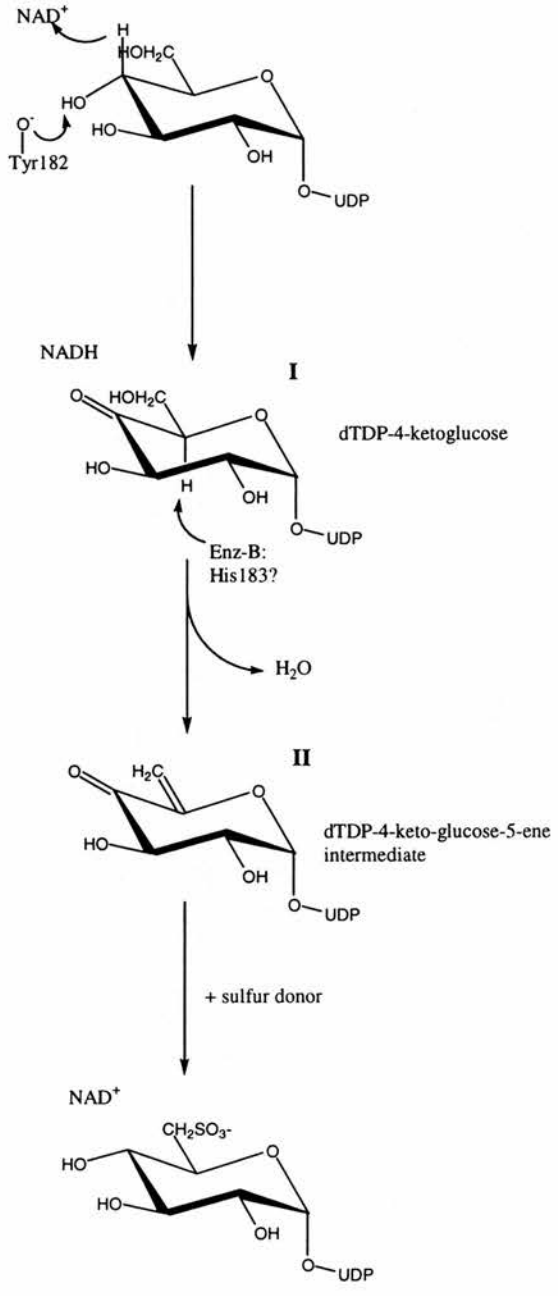


Figure 1.14. The proposed mechanism of SQD1 (Essigmann et al., *1999; Mulichak* et al., *1999). Tyr182 acting as an active site base abstracts a proton from the UDP-glucose 4'-hydroxyl, with His183 acting as a general base abstracting a proton from the acidic glucosyl C5' atom during formation of the 4-keto-glucose-5-ene intermediate (II).*

ADP-L-glycero-D-mannoheptose 6-epimerase

ADP-L-glycero-D-mannoheptose 6-epimerase (AGME) is a 34 kDa enzyme required in Gram-negative bacteria for ADP-L-glycero-D-mannoheptose (heptose) synthesis, a component of the core domain of the LPS portion of the bacterial cell wall (Adams *et al.*, 1967). The substrate for this enzyme (ADP-D-glycero-D-mannoheptose), is a sugar with seven carbons, and the epimerisation occurs at the C6' position (Figure 1.15). The structure of AGME in complex with $NADP^+$ and ADP-glucose, a catalytic inhibitor of the enzyme, has been solved at a resolution of 2Å (Deacon *et al.*, 2000). The protein has two domains. The N-terminal domain consists of a modified 7-stranded Rossmann fold, which contains the $NADP^+$ binding site. The smaller C-terminal domain is an α/β fold and provides residues that create the specificity for the substrate. The two domains interact to form a crevice where the substrate is brought close to the $NADP^+$ cofactor. The three key catalytic residues identified in AGME (Ser116, Tyr140, Lys144) are found to be located in structurally similar positions to GALE. This and other evidence confirms the epimerisation proceeds through a transient keto intermediate. However, the quaternary structure of AGME diverges from the other proteins of the SDR family since AGME forms homopentamers as opposed to the dimers or tetramers of most SDR proteins (Figure 1.11d) (Rizzi *et al.*, 1998). The five monomers are arranged in a ring, with the crevices containing the catalytic site oriented towards the centre. This position of the active site crevice varies between monomers and reflects the two binding modes of ADP-glucose. Although this may be a result of crystal packing, there are similarities to the domain movements in HGALE and in the different UDP-galactose/UDP-glucose binding modes discussed above. Depending on the C-terminal conformation AGME ("closed" or "open"), the torsion angle, defined by the oxygen linking the two

α and β phosphates, the β-phosphate, the glycosyl oxygen, and the C1 atom of ADP-glucose, changes by 125°. Such a 'rotation' of substrate would permit epimerisation to take place and provides indirect evidence that such a rotation also occurs in GALE. Significantly, in this mechanism, oxidation and reduction occurs at the C6' position. In RmlB, the final step of the mechanism involves reduction of the dTDP-4-keto-5,6-glucosene intermediate at C6' by NADH (Figure 1.12).

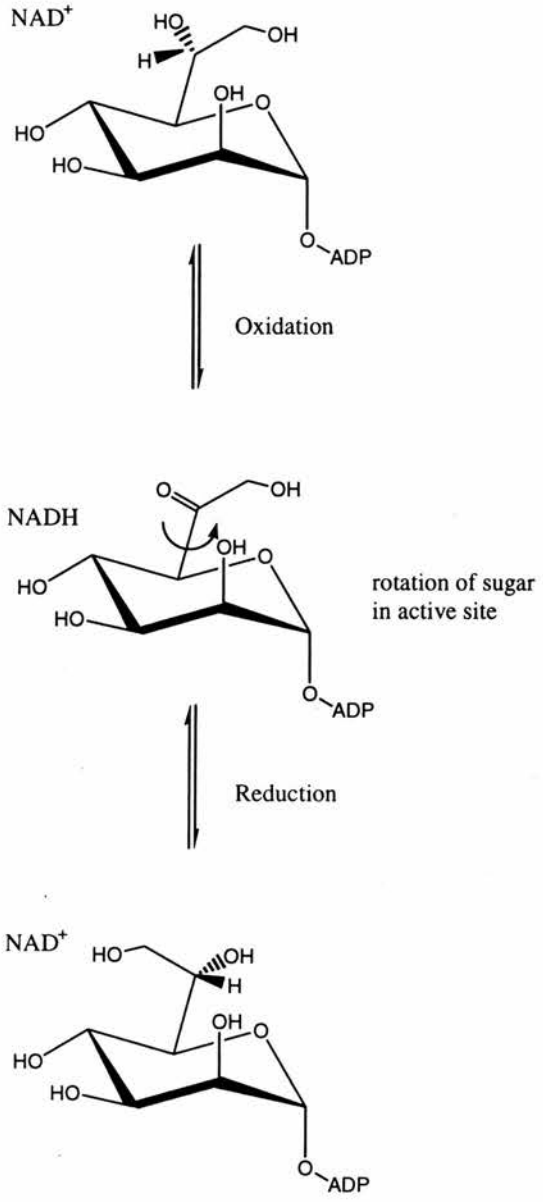


Figure 1.15. The C6 epimerisation catalysed by ADP-L-glycero-D-mannoheptose 6-epimerase, adapted from (Allard et al., *2001b).*

1.9. Experimental Aims

In the 1960s 6-deoxy dTDP-linked sugars were identified in various bacteria, for example dTDP-L-rhamnose was found in *P. aeruginosa* (Glaser and Kornfeld, 1961). dTDP-D-glucose-4,6-dehydratase (RmlB) is the second enzyme of the dTDP-L-rhamnose biosynthesis pathway, and catalyses the irreversible dehydration of dTDP-D-glucose to dTDP-4-keto-6-deoxy-D-glucose. This C6-deoxygenation catalysed by RmlB and other NDP-hexose 4,6-dehydratases is the first committed step in almost all deoxyhexose biosynthetic pathways leading to the formation of di-, tri- and tetra-deoxysugars (He *et al.*, 2000; Johnson and Liu, 1998; Liu and Thorson, 1994). For this reason, the mechanism of RmlB has been the subject of extensive investigation (Gerratana *et al.*, 2001; Gross *et al.*, 2000; Hegeman, 2001; Hegeman *et al.*, 2001). However, no structure has been reported and how the enzyme carries out catalysis remains unknown. The aim of this research was to both structurally and mechanistically characterise RmlB from more than one bacterial species. The following chapters reveal the structure of RmlB from both *S. typhimurium* and *S. suis* as well as the first complete understanding of the dehydratase mechanism.

Chapter Two

The over-expression and purification of dTDP-D-glucose 4,6-dehydratase (RmlB) from Salmonella enterica *serovar* Typhimurium *and* Streptococcus suis *serotype 2*

2.1. Summary

Plasmids containing the DNA encoding dTDP-D-glucose 4,6-dehydratase (RmlB) from *S. typhimurium* and *S. suis* were initially received as frozen glycerol stocks. In order to over-express the RmlB proteins the plasmids were transformed into *E. coli* BL21(DE3) cells. The samples were purified with a two-step HPLC purification protocol using anion exchange and hydrophobic interaction columns. The samples were verified as pure RmlB protein using silver stained SDS-PAGE gel analysis and N-terminal sequencing. Before crystallisation trails were set up the protein sample was assayed to verify activity and checked for monodispersity as revealed by dynamic light scattering.

2.2. Introduction

dTDP-D-glucose 4,6-dehydratase (RmlB) was first identified in the L-rhamnose biosynthetic pathway where it catalyses the conversion of dTDP-D-glucose into dTDP-4-keto-6-deoxy-D-glucose. L-rhamnose is a 6-deoxysugar, which occurs in the cell walls and envelopes of many pathogenic bacteria. As L-rhamnose is not found in humans, the enzymes involved in its biosynthesis represent attractive potential therapeutic targets. RmlB and other 4,6-dehydratases also catalyse the first committed step in all 6-deoxysugar biosynthetic pathways. Numerous 6-deoxysugars are used in bacterial lipopolysaccharide production and in the biosynthesis of a diverse array of secondary metabolites including antibiotics.

The enzymes from *S. typhimurium* and *S. suis* share 38.5% identity. The genes for both enzymes were supplied in pET plasmids by our collaborators Chris Whitfield (University of Guelph) and Duncan Maskell and Andrew Allen (Cambridge University). These pET plasmids were incorporated into *E. coli* BL21(DE3) cells. The pET (**p**lasmid for **E**xpression by **T**7 RNA polymerase) system is a powerful expression tool; over-expression of the protein being induced by providing a source of T7 RNA polymerase for the host cell which directs the high level expression of the cloned RmlB target genes.

Following over-production of the required proteins, high-pressure liquid chromatography (HPLC) was used to purify the RmlB enzymes to a purity necessary for crystallisation trails to be started. Dr Marie-France Giraud, a postdoctoral worker in our laboratory, had already completed some preliminary work on *S. typhimurium*

RmlB, including ascertaining the conditions needed to purify the protein and an initial crystallisation condition. It is at this point that I took over the project.

2.3. S. typhimurium *Experimental*

Any material requiring centrifugation was undertaken using a Beckman Avanti J-25 centrifuge employing either a JA-25.50 or JLA-9.1000 fixed angle rotor. Optical density readings were recorded using a Pharmacia Biotech Ultraspec 1000 spectrophotometer. Cell suspensions were sonicated with a Heat Systems-Ultrasonics W-220F cell disrupter. All chromatographic separations were performed using a BioCAD 700E Perfusion Chromatography Workstation fitted with an Advantec SF-2120 Super Fraction Collector. Dynamic Light Scattering was measured with a Protein Solutions DynaPro-801 Molecular Sizing Instrument and protein purity was assessed by running SDS polyacrylamide gels with a silver stain method on a Pharmacia LKB-PhastSystem. All chemicals were obtained from Sigma (Poole, UK) unless otherwise stated.

Our collaborator Chris Whitfield (University of Guelph) amplified the open reading frame of the gene encoding RmlB using the polymerase chain reaction (PCR) with primers designed to introduce a unique 5′ *Nco*I site overlapping the initiation ATG codon. A downstream 3′ *Sst*I site was introduced to facilitate cloning into the expression vector pET-28a(+). Expression involves the IPTG (isopropyl-β-D-thiogalactoside)-inducible T7 promoter and ribosome-binding sites conferred by the vector, but uses the natural *rmlB* initiation ATG codon. The sequence of the amplified and cloned gene was confirmed to be identical to the chromosomal copy. The expressed protein is therefore identical in amino acid sequence to the authentic product and carries no extensions or mutations. The pET-28a(+) plasmid was

transformed into *E. coli* DH5α cells and transported to the University of St Andrews on dry ice.

In order to over-express RmlB, *E. coli* BL21(DE3) cells were transformed with the pET-28a(+) plasmid. The transformed BL21(DE3) cells were grown on a Luria agar plate containing 50μg ml^{-1} kanamycin. From this, a single colony was picked and used to inoculate an overnight culture of 10ml Luria-Bertani medium (LB) plus 80μg ml^{-1} kanamycin, incubated at 310K, 180rpm shake. This starter culture was then used to inoculate 500ml of Terrific Broth (Sambrook and Russell, 2001) in a 2l shaker flask that contained 80μg ml^{-1} kanamycin and was incubated at 310K, 180rpm shake until an OD_{600} of 0.6 – 0.8; over-expression was then induced by the addition of 2mM IPTG. Novagen (Novagen, 1999) recommend induction of cells when their OD_{600} reaches 0.6-1.0 as prolonged incubation of the culture at stationary phase may deplete the media of antibiotic and allow overgrowth of non-recombinants or non-expressing mutants.

To determine the best time and temperature in which to grow and harvest the induced cells, time course experiments were performed at 303K and 310K. It is known that growth at 310K causes some proteins to accumulate as inclusion bodies, while incubation at 303K may lead to soluble, active proteins (Schein and Noteborn, 1988). 1ml samples were taken every 15 mins for 4 hours and centrifuged at 13000*g* for 3 mins. The resultant pellets were resuspended and lysed to check for soluble protein using Coomassie stained SDS-PAGE analysis. The most favourable incubation condition was an incubation temperature of 310K for 3 hours.

After 3 hours of incubation at 310K the cells were harvested by centrifugation at 8000*g* (277K) for 15 mins. The pellets were stored at 253K until required. Immediately prior to purification the pellets were defrosted on ice and resuspended in ice-cold lysis buffer; 20mM Tris Base pH 8.0, 100mM NaCl, 5mM PMSF (phenylmethysulphonyl fluoride), 2mM DTT (dithiothreitol) and 2mM EDTA (ethylenediaminetetraacetic acid). Lysozyme was then added to give a final concentration of 100μg ml^{-1}. The viscosity of the cell suspension was decreased by the addition of DNAse I (20μg ml^{-1}) and 2mM EDTA. The cell suspension was incubated on ice and at room temperature for alternating 5 min periods totalling 30 mins and then sonicated for four cycles of 30 s interrupted by 1 min periods on ice. Centrifugation at 18000g (277K) for 15 mins removed the cell debris, the resulting supernatant being carefully pipetted into a small glass beaker on ice.

The supernatant was carefully brought to 20% $(NH_4)_2SO_4$ saturation and incubated for 1 hour at 277K. After a second centrifugation at 18000*g* (277K) for 15 mins, 277K, the supernatant was dialysed overnight against three changes of 2l 50mM NaCl, 20 mM Tris-Base, pH 8.3 at 277K.

DTT and EDTA were added to the dialysed supernatant to give final concentrations of 2mM and 1mM respectively and the supernatant 0.2μm filtered. The supernatant was passed through a POROS-1760 HQ/M HPLC column equilibrated with 50mM NaCl, 20mM Tris-Base, pH 8.3. The protein was eluted with a 50-2000mM NaCl, 20mM Tris-Base, pH 8.3 gradient. Fractions containing *S. typhimurium* RmlB were identified by SDS-PAGE analysis and were typically eluted at 370mM NaCl, however all fractions were collected and analysed (Figure 2.1). A typical POROS-

1760 HQ/M HPLC column profile is shown in Figure 2.2. Coomassie stained SDS-PAGE analysis showed the fraction eluted at 370mM NaCl to contain a protein with a molecular mass corresponding to RmlB, 41 kDa.

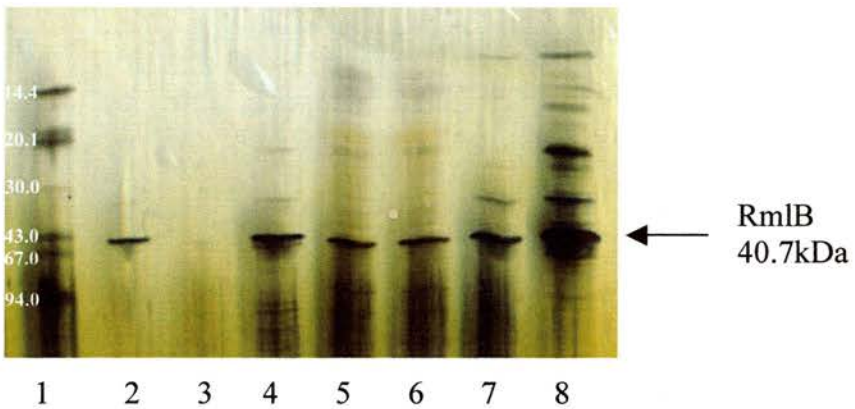


Figure 2.1. A silver nitrate stained SDS-PAGE gel of RmlB from S. typhimurium *showing the purification process. Lane 1 shows the molecular weight markers (kDa), lane 2 pure protein after HP, lane 3 is empty, lanes 4 to 7 protein after HQ and lane 8 over-expressed* S. typhimurium *RmlB.*

Fractions corresponding to this peak at 370mM NaCl were pooled, concentrated with an Amicon ultrafiltration unit, and DTT and EDTA added to give final concentrations of 2mM and 1mM, respectively. The sample was then dialysed overnight against three changes of 1l 20mM Na_2HPO_4, pH 7.6 at 277K. $(NH_4)_2SO_4$ was added gradually to 35% whilst stirring at 277K. DTT was added to give a final concentration of 2mM and the protein sample 0.2μm filtered.

The filtered protein sample was loaded onto a POROS HP2/M HPLC column equilibrated with 35% $(NH_4)_2SO_4$), 20mM Na_2HPO_4, pH 7.6. Elution of the protein was achieved with a decreasing gradient of 35 to 0% $(NH_4)_2SO_4$, 20mM Na_2HPO_4,

pH 7.6; the fractions were collected and analysed by silver stained SDS-PAGE, which gave a single band (Figure 2.3). A typical POROS HP2/M HPLC column profile for RmlB is shown in Figure 2.4 with RmlB being eluted between 46% and 36% buffer A (35% $(NH_4)_2SO_4$, 20mM Na_2HPO_4, pH 7.6). Fractions corresponding to *S. typhimurium* RmlB were concentrated to 4mg ml^{-1} with an Amicon ultrafiltration unit, and DTT and EDTA were added to give final concentrations of 2mM and 1mM, respectively. The protein solution was then dialysed overnight against two changes of 2l 25mM Tris-Base pH 7.6. Pure protein was stored as 500µl aliquots in 50% glycerol at 253K.

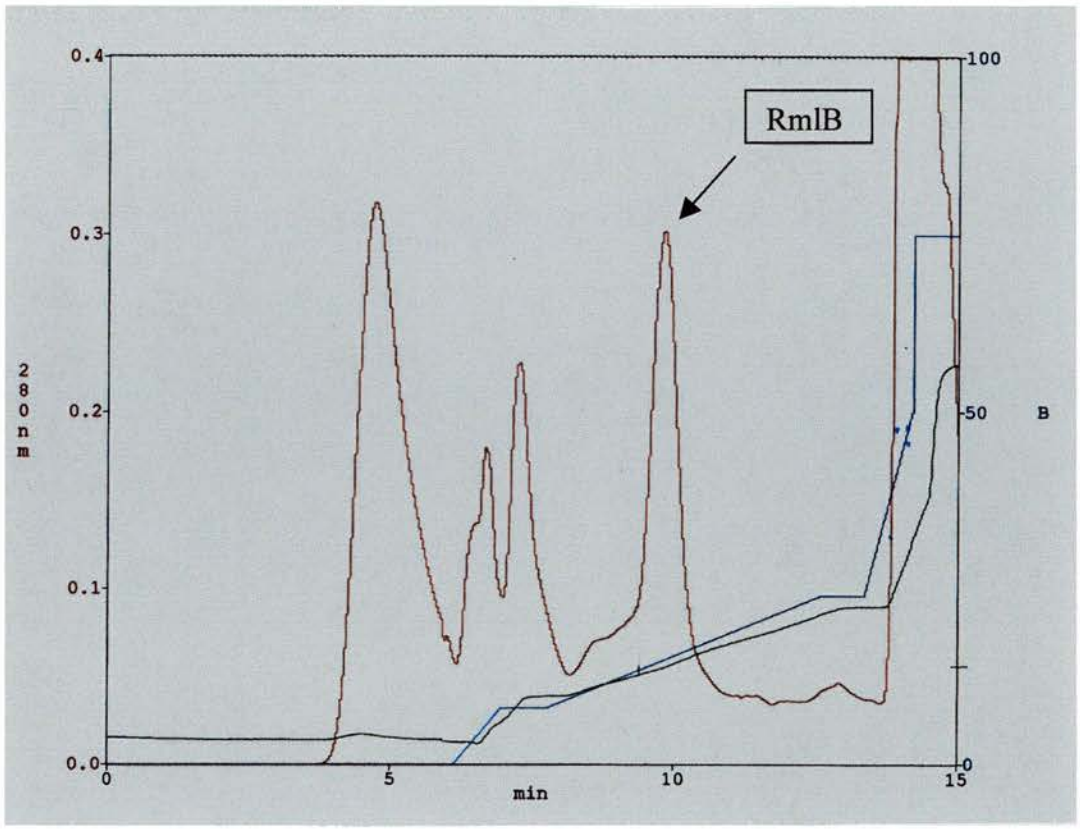


Figure 2.2. A UV_{280} trace of the protein fractions eluted during anion exchange chromatography (HQ). Time is on the x-axis, left y-axis is 280nm reading, right y-axis increasing % of buffer B; 2000mM NaCl, 20mM Tris-Base, pH 8.3, decreasing % buffer A; 50mM NaCl, 20mM Tris-Base, pH 8.3.

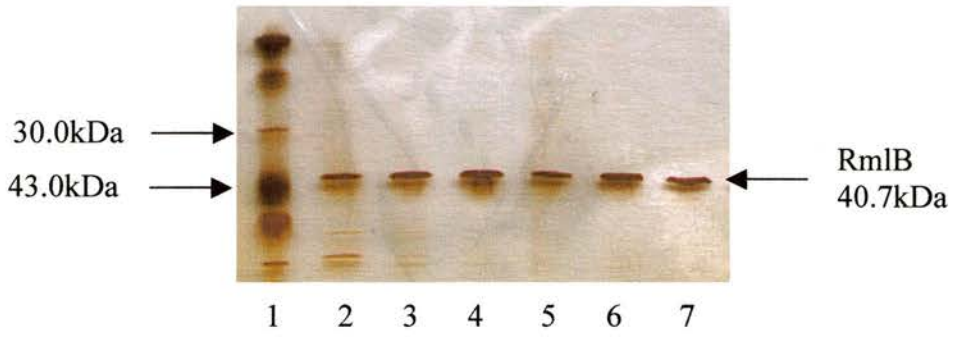


Figure 2.3. A silver nitrate stained SDS-PAGE gel of RmlB from S. typhimurium *after hydrophobic exchange chromatography. Lane 1 shows the molecular weight markers and lanes 2 to 7* S. typhimurium *RmlB after HP. Apparent band duplication in lanes 2 to 7 of the gel is due to band reflection from the scanner surface.*

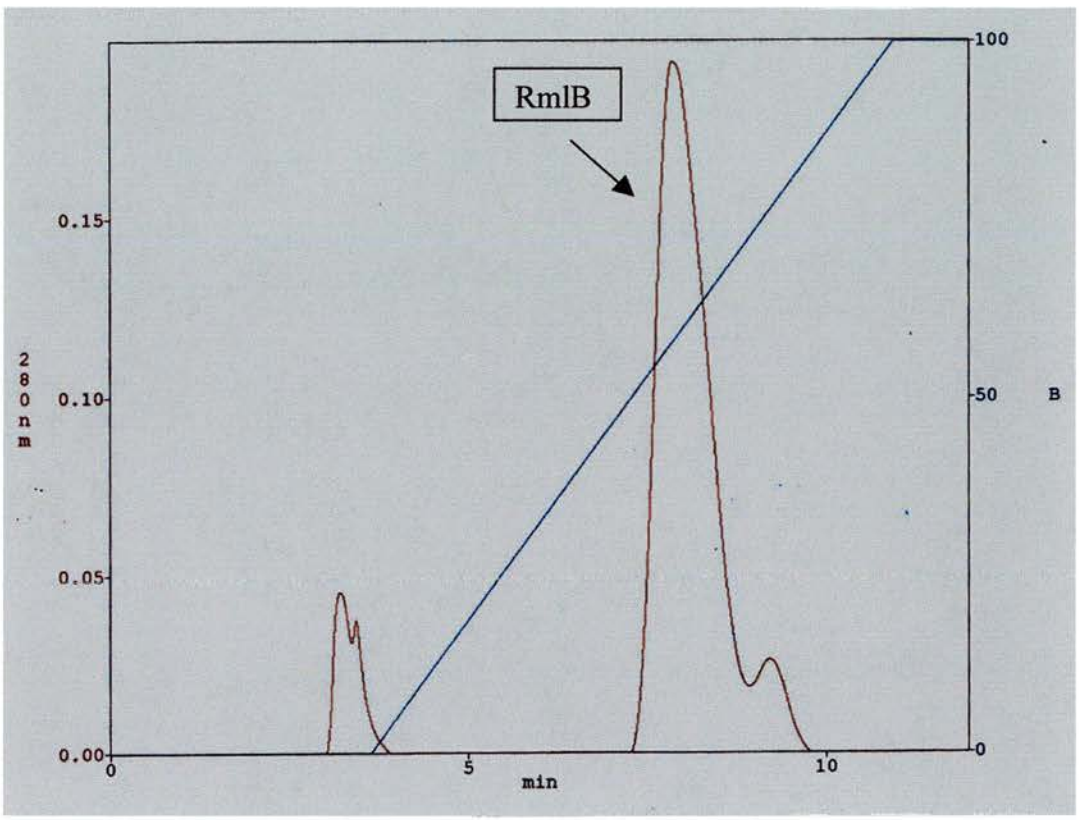


Figure 2.4. A UV_{280} trace of the protein fractions eluted during hydrophobic exchange chromatography (HP). Time is on the x-axis, left y-axis is 280nm reading, right y-axis is increasing % buffer B; 20mM Na_2HPO_4, pH 7.6, decreasing % buffer A; 35% $(NH_4)_2SO_4$, 20mM Na_2HPO_4, pH 7.6.

2.4. S. typhimurium *protein analysis*

Following the two HPLC steps, the protein appeared to be pure as judged by a SDS silver nitrate stained gel, giving a single band at an apparent molecular weight of 41 kDa (Figure 2.3); the calculated molecular mass based on the sequence being 40722 Da. Matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry was also performed on the pure protein samples, producing a single clean peak at the correct molecular mass (data not shown). Dynamic light-scattering results (DLS) indicated the protein to be monodisperse, shown by a base line of 1.000, with a molecular mass indicative of a dimer (Table 2.1). Purified protein was N-terminal sequenced using the Edman degradation reaction (Edman and Begg, 1967); 5 cycles were performed and the resulting sequence unambiguous. The first 6 residues were found to be Met-Lys-Ile-Leu-Ile-Thr, which exactly match residues 1 to 6 of *S. typhimurium* RmlB. The final yield of protein following purification was approximately 18mg of RmlB per litre of Terrific Broth as estimated at UV_{280}, given that 1AU corresponds to 0.62mg ml^{-1} for *S. typhimurium* RmlB.

All purified protein was also assayed to check for activity. The assay is based upon that developed by Okaskai *et al.* (1962) and follows the formation of a unique, broad UV chromophore centred at 318nm following incubation of dTDP-4-keto-6-deoxy-D-glucose in 0.1M NaOH at 310K for 20 min. The structure of the species responsible for this chromophore has not yet been determined. In addition, several different values for the extinction coefficient for this species have been utilised. The value used here is that of Wang and Gabriel (1969) of $4500M^{-1}$ cm^{-1}.

All readings were done in triplicate. 10µl of protein (between 10-50µg of RmlB) were carefully pipetted into a sterile 1ml Eppendorf containing 86µl of 50mM HEPES, pH 7.6 and 2.5µl of 40mM NAD^+. 1µl of 100mM dTDP-D-glucose was added and the sample incubated at 310K for 20 mins. 1ml of 0.1M NaOH was then added and the sample incubated for a further 20 min at 310K. The spectrophotometer was zeroed against 0.1M NaOH and the absorbance of the sample read at 318nm in a quartz cuvette. Negative controls were also set up containing no dTDP-D-glucose. The amount of dTDP-D-glucose used was 100nmoles. After the addition of 1ml NaOH, the concentration of dTDP-D-glucose is at a little less than 100µM. Using an extinction coefficient of $4500M^{-1}\ cm^{-1}$, an optical density of 0.45 would be obtained if the reaction went to completion (1000mM/4500 = 0.1mM/optical density). Values of around 0.3 indicated strong activity. Only protein with good activity was used for crystallisation trials.

Table 2.1. DLS results for pure S. typhimurium *RmlB*

Amplitiude	Differential Coefficient (nm)	Radius (nm)	Polydispersity (nm)	Estimated Weight (kDa)	Count Rate	Base Line	SOS Error
0.773	572	3.7	---	79	34	1.012	2.898
0.799	585	3.6	---	75	34	1.000	2.262
0.808	584	3.6	---	86	34	1.000	2.602
0.813	592	3.5	---	83	34	1.002	2.252
0.802	561	3.7	---	82	34	1.004	2.932
0.797	605	3.4	---	80	34	0.999	3.087
0.809	591	3.5	---	83	34	1.000	3.421
0.801	591	3.5	---	74	34	1.000	2.926
0.801	597	3.5	---	82	35	1.000	3.798
0.788	600	3.5	---	83	35	1.003	2.339

Polydispersity: <15% of Radius →Negligible, < 30% of Radius →Moderate Amount of Polydispersity, > 30% of Radius →Significant Amount of Polydispersity. Base Line Parameter: 0.997-1.001 →Narrow Monomodal Size Distribution, 1.002-1.005 →Broad Monomodal Size Distribution, > 1.005 →Multimodal Size Distribution, cannot be resolved. SOS Error: 1.000-5.000 →Negligible Error, Low Noise, 5.000-20.000 →Somewhat Significant Background Noise, > 20.000 →High Noise/Error.

2.5. S. suis *serotype 2 Experimental*

Our collaborators Duncan Maskell and Andrew Allen at Cambridge University cloned the open reading frame of the gene encoding *S. suis* serotype 2 RmlB and incorporated it into the expression vector pET-21a(+). The pET-21a(+) plasmid was transformed into *E. coli* DH5α cells and transported to the University of St Andrews on dry ice.

In order to over-express *S. suis* RmlB, *E. coli* BL21(DE3) cells were transformed with the pET-21a(+) plasmid. The transformed BL21(DE3) cells were grown on a Luria agar plate containing 200μg ml^{-1} ampicillin. From this a single colony was picked and used to inoculate an overnight culture of 10ml LB plus 200μg ml^{-1} ampicillin, which was incubated at 310K, 180rpm shake. This starter culture was then used to inoculate 500ml of Tryptone Yeast Extract (TYE) medium (Sambrook and Russell, 2001) in a 2l shaker flask containing 200μg ml^{-1} ampicillin. Incubation was carried out at 310K, 180rpm shake until an OD_{600} of 0.6 – 0.8; over-expression was then induced by the addition of 1.5mM IPTG.

As with *S. typhimurium* RmlB, time course experiments were performed at 303K and 310K to determine the best time and temperature in which to grow and harvest the induced cells. 1ml samples were taken every 15 mins for 4 hours and centrifuged at 13000*g* for 3 mins. The resultant pellets were resuspended and lysed to check for soluble protein using Coomassie stained SDS-PAGE analysis. The most favourable incubation condition was an incubation temperature of 310K for 3 hours.

After 3 hours of incubation at 310K the cells were harvested by centrifugation at 8000*g* (277K) for 15 mins. The pellets were stored at 253K until further use. Immediately prior to purification the pellets were defrosted on ice and resuspended in ice-cold lysis buffer; 20mM Tris Base pH 8.0, 100mM NaCl, 5mM PMSF, 2mM DTT and 2mM EDTA. Lysozyme was added to give a final concentration of 100μg ml^{-1}. The viscosity of the cell suspension was decreased by the addition of DNAse I (20μg ml^{-1}) and 2mM EDTA. The cell suspension was then incubated on ice and at room temperature for alternating 5 min periods totalling 30 mins and then sonicated for four cycles of 30 s interrupted by 1 min periods on ice. Centrifugation at 18000g for 15 mins at 277K removed the cell debris and the resulting supernatant was carefully pipetted into a small glass beaker on ice.

The supernatant was carefully brought to 20% $(NH_4)_2SO_4$ saturation and incubated for 1 hour at 277K. After a second centrifugation at 18000*g* (277K) for 15 mins, the supernatant was dialysed overnight against three changes of 2l 50mM NaCl, 20 mM Tris-Base, pH 8.3 at 277K.

DTT and EDTA were added to the dialysed supernatant to give final concentrations of 2mM and 1mM, respectively, and the supernatant 0.2μm filtered. The supernatant was passed through a POROS-1760 HQ/M HPLC column equilibrated with 50mM NaCl, 20mM Tris-Base, pH 8.3. The protein was eluted with a 50-2000mM NaCl, 20mM Tris-Base, pH 8.3 gradient. *S. suis* RmlB fractions were typically eluted at 275mM NaCl, however all fractions were collected and analysed (Figure 2.5). A typical POROS-1760 HQ/M HPLC column profile is shown in Figure 2.6.

Coomassie stained SDS-PAGE analysis showed the fraction eluted at 275mM NaCl to contain a protein with a molecular mass of 39kDa corresponding to *S. suis* RmlB.

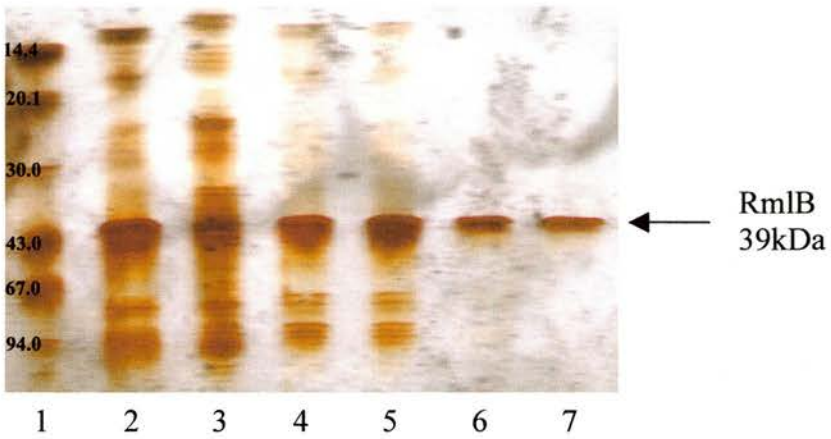


Figure 2.5. A silver nitrate stained SDS-PAGE gel of RmlB from S. suis *showing the purification process. Lane 1 shows the molecular weight markers (kDa), lanes 2 and 3 over-expressed* S. suis *RmlB, lanes 4 and 5 protein after HQ and lanes 6 and 7 pure protein after HP.*

Fractions corresponding to this peak at 275mM NaCl were pooled, concentrated with an Amicon ultrafiltration unit, and DTT and EDTA added to give final concentrations of 2mM and 1mM, respectively. The sample was then dialysed overnight against three changes of 1l 20mM Na_2HPO_4, pH 7.6 at 277K. $(NH_4)_2SO_4$ was added gradually to 40% whilst stirring at 277K. DTT was added to give a final concentration of 2mM and the protein sample 0.2μm filtered.

The filtered protein sample was loaded onto a POROS HP2/M HPLC column equilibrated with 40% $(NH_4)_2SO_4$), 20mM Na_2HPO_4, pH 7.6. Elution of the protein was achieved with a decreasing gradient of 40 to 0% $(NH_4)_2SO_4$, 20mM Na_2HPO_4, pH 7.6; the fractions were collected and analysed by silver stained SDS-PAGE,

which gave a single band (Figure 2.7). A typical POROS HP2/M HPLC column profile for RmlB is shown in Figure 2.8 with RmlB being eluted between 66% and 59% buffer A (40% $(NH_4)_2SO_4$, 20mM Na_2HPO_4, pH 7.6). Fractions corresponding to *S. suis* RmlB were concentrated to 4mg ml^{-1} with an Amicon ultrafiltration unit, and DTT and EDTA were added to give final concentrations of 2mM and 1mM, respectively. The protein solution was then dialysed overnight against two changes of 2l 25 mM Tris-Base pH 7.6. Pure protein was stored as 500µl aliquots in 50% glycerol at 253K.

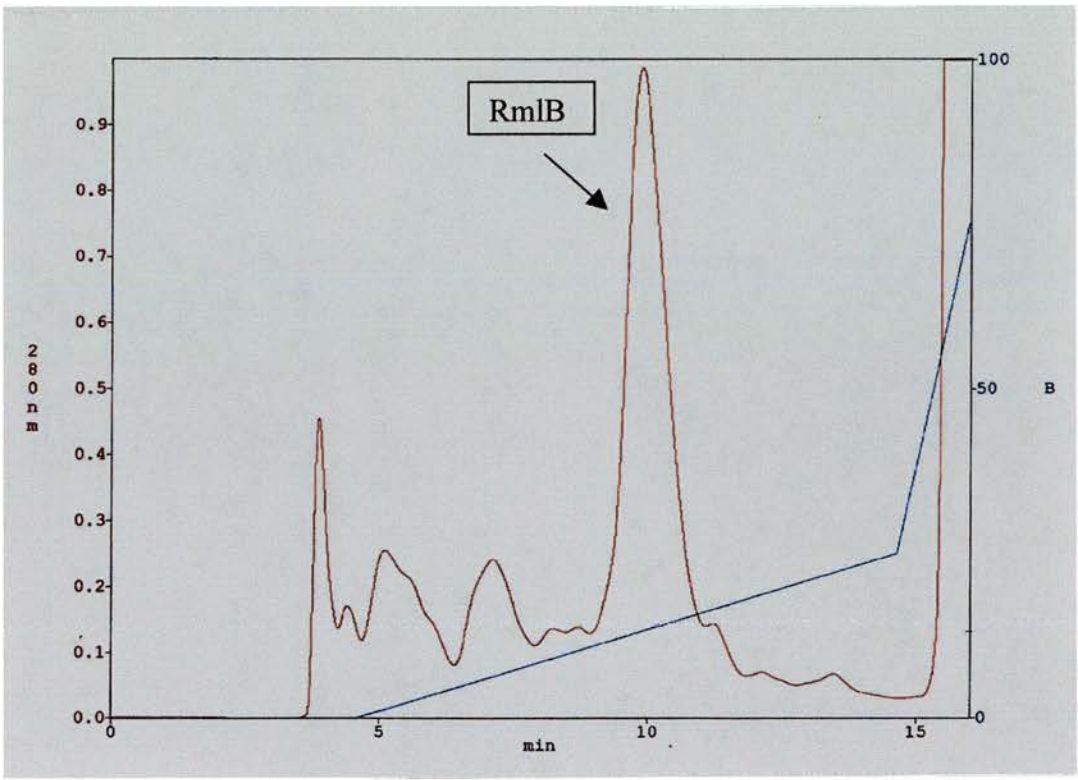


Figure 2.6. A UV_{280} trace of the protein fractions eluted during anion exchange chromatography (HQ). Time is on the x-axis, left y-axis is 280nm reading, right y-axis increasing % of buffer B; 2000mM NaCl, 20mM Tris-Base, pH 8.3, decreasing % buffer A; 50mM NaCl, 20mM Tris-Base, pH 8.3.

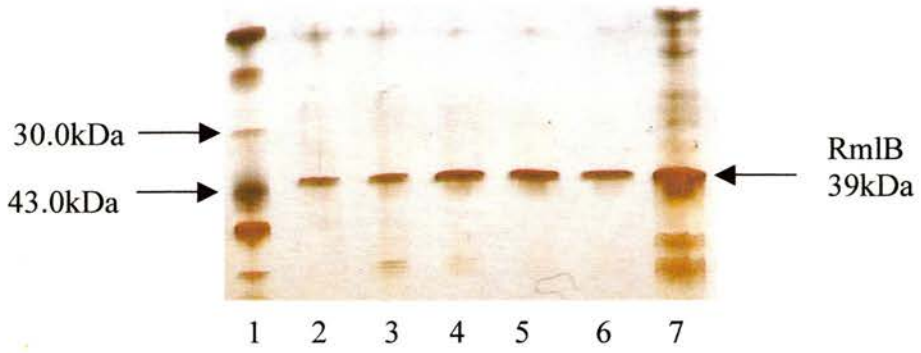


Figure 2.7. A silver nitrate stained SDS-PAGE gel of RmlB from S. suis *after hydrophobic exchange chromatography. Lane 1 shows the molecular weight markers, lanes 2 to 6* S. suis *RmlB and lane 7 over-expressed protein before HPLC.*

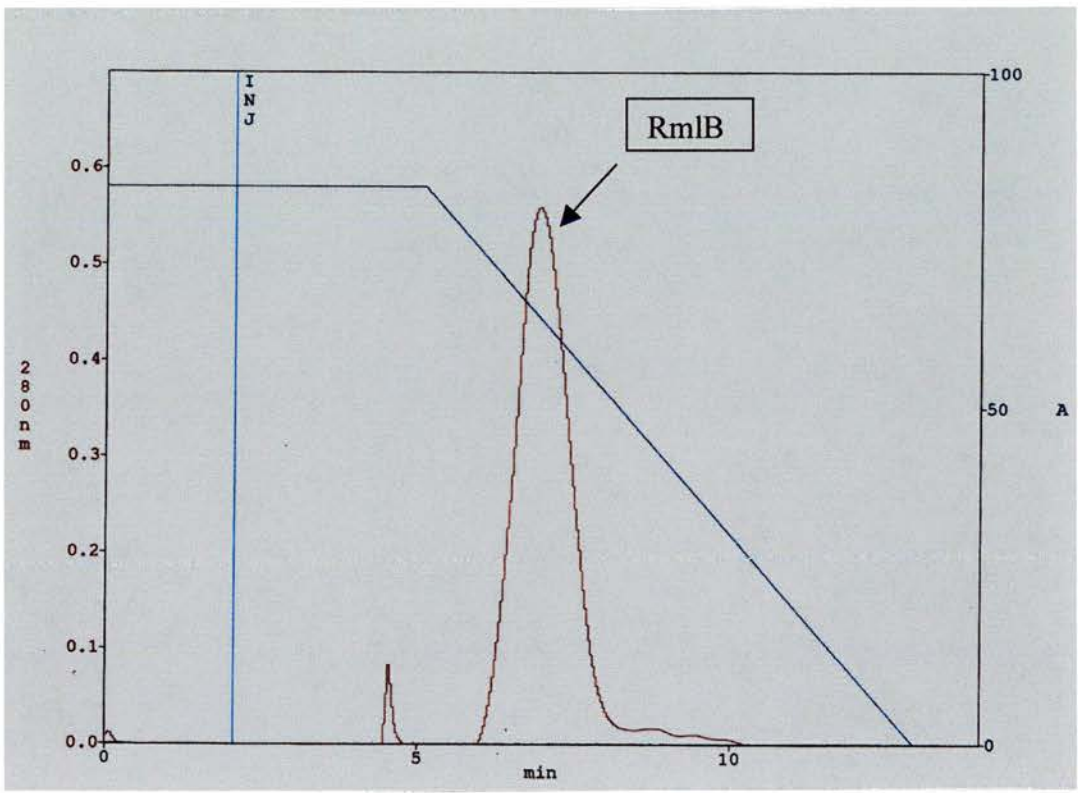


Figure 2.8. A UV_{280} trace of the protein fractions eluted during hydrophobic exchange chromatography (HP). Time is on the x-axis, left y-axis is 280nm reading, right y-axis is decreasing % buffer A; 40% $(NH_4)_2SO_4$, 20mM Na_2HPO_4, pH 7.6, increasing % buffer B; 20mM Na_2HPO_4, pH 7.6.

2.6. S. suis *serotype 2 protein analysis*

Following the two HPLC steps, the protein appeared to be pure as judged by a SDS silver nitrate stained gel (Figure 2.7) giving a single band at an apparent molecular mass of 39 kDa, the calculated molecular mass based on sequence being 38936 Da. MALDI-TOF mass spectrometry produced a single clean peak at the correct molecular mass. DLS indicated the protein to be monodisperse, shown by a base line of 1.000, with a molecular mass indicative of a dimer (Table 2.2). N-terminal sequencing was also performed. After 5 cycles of the Edman degradation reaction the results confirmed the protein as *S. suis* RmlB. The first 6 residues were: Ser-Gln-Phe-Lys-Asn-Ile, exactly matching residues 2 to 7 of *S. suis*; the initial Met is often lost during over-expression in *E. coli* BL21(DE3) cells. The final yield of protein following purification was approximately 15 mg of RmlB per litre of TYE as estimated at UV_{280}, given that 1AU corresponds to 0.63 mg ml^{-1} for *S. suis* RmlB. As with *S. typhimurium* RmlB, the *S. suis* protein was analysed for activity using the assay based on that of Okaskai *et al.* (1962).

Table 2.2. DLS results for pure S. suis *RmlB. The average polydispersity value as a function of the hydrodynamic radius is 15.4%.*

Amplitiude	Differential Coefficient (nm)	Radius (nm)	Polydispersity (nm)	Estimated Weight (kDa)	Count Rate	Base Line	SOS Error
0.868	664	4.0	0.7	74	105	1.002	2.277
0.858	655	4.0	1.2	82	117	1.003	3.449
0.858	659	4.0	1.1	74	102	1.004	4.302
0.836	664	3.9	0.6	91	102	1.004	3.252
0.845	633	4.1	---	75	117	1.002	3.912
0.850	683	3.8	0.8	81	101	1.005	2.057
0.857	660	4.0	0.5	81	109	1.002	3.241
0.857	657	4.0	0.6	74	102	0.999	4.112
0.845	649	4.0	---	86	101	1.003	1.541

Polydispersity: <15% of Radius →Negligible, < 30% of Radius →Moderate Amount of Polydispersity, > 30% of Radius →Significant Amount of Polydispersity. Base Line Parameter: 0.997-1.001 →Narrow Monomodal Size Distribution, 1.002-1.005 →Broad Monomodal Size Distribution, > 1.005 →Multimodal Size Distribution, cannot be resolved. SOS Error: 1.000-5.000 →Negligible Error, Low Noise, 5.000-20.000 →Somewhat Significant Background Noise, > 20.000 →High Noise/Error.

Chapter three

Crystallisation, data collection and structural determination of RmlB holoprotein from S. typhimurium *and* S. suis *serotype 2*

3.1. Summary

The structures of *S. typhimurium* and *S. suis* RmlB crystals with coenzyme NAD^+ bound have been determined. The *S. typhimurium* crystals were grown using the sitting-drop vapour-diffusion technique with lithium sulphate as precipitant. Two sets of synchrotron diffraction data were obtained to resolutions of 2.8Å and 2.47Å using single frozen crystals. Processing of the data revealed that the crystals belong to space group $P2_1$, with unit cell parameters of a=110.6Å, b=87.5Å, c=111.3Å and β=98.0°. The asymmetric unit contains four monomers in the form of two RmlB homodimers with a solvent content of 62%. Molecular replacement was used to solve the structure and the model refined to give a final crystallographic R-factor of 20.4% and an R_{free} value of 24.9% with good stereochemistry.

The *S. suis* RmlB crystals were also grown using the sitting-drop vapour-diffusion technique but with polyethylene glycol (PEG) 4000 as precipitant. A set of data to 2.9Å was collected at the European Synchrotron Radiation Facility (ESRF) Grenoble using a single frozen crystal. Processing of the data revealed that the crystals belong to space group $P2_12_12_1$ with unit cell parameters of a=61.5Å, b=97.8Å, c=181.3Å and $\alpha=\beta=\gamma=90.0°$. The asymmetric unit contains one homodimer with a solvent content of 63%. Molecular replacement was used to solve the structure and the model refined to give a final crystallographic R-factor of 21.5% and an R_{free} value of 26.9% with good stereochemistry. The active site of each monomer also appears to contain one molecule of dTDP bound, although none was used in the crystallisation trials.

The structure reveals that RmlB functions as a homodimer with monomer association occurring principally through hydrophobic interactions *via* a four-helix bundle. Each monomer exhibits an α/β structure that can be divided into two domains. The larger N-terminal domain binds the nucleotide cofactor NAD^+ and consist of a seven stranded β-sheet surrounded by α-helices. The smaller C-terminal domain is responsible for binding the sugar substrate dTDP-D-glucose and contains four β-strands and six α-helices. The two domains meet to form a cavity in the enzyme. The highly conserved active site TyrXXXLys catalytic couple and the GlyXGlyXXGly motif at the N-terminus characterise RmlB as a member of the SDR extended family.

The substrate dTDP-D-glucose was modelled into the active site of *S. typhimurium* RmlB enabling the identification of key residues.

3.2. Introduction

RmlB catalyses the second step in the dTDP-L-rhamnose pathway; the dehydration of the nucleotide sugar dTDP-D-glucose to dTDP-4-keto-6-deoxy-D-glucose (Figure 3.17). This reaction has been shown to be the first unique step in the bacterial biosynthesis of several naturally occurring sugars (Wang and Gabriel, 1969).

RmlB has been purified from several bacterial sources (Romana *et al.*, 1991; Sohng and Yoo, 1996; Thompson *et al.*, 1992; Vara and Hutchinson, 1988). The enzyme functions as a homodimer; the *S. typhimurium* monomer consisting of 361 amino-acid residues and the *S. suis* monomer 348 amino-acid residues. In both the RmlB proteins, one molecule of the cofactor NAD^+ is bound to each monomer in the Rossmann fold located at the N-terminus. The sequence of RmlB indicates it to be a tyrosine-dependent oxidoreductase (also known as the SDR family), containing the highly conserved TyrXXXLys catalytic couple (Jornvall *et al.*, 1995; Jornvall, 1999; Jornvall *et al.*, 1999).

The structures of RmlB from *Salmonella enterica* serovar Typhimurium at 2.47Å resolution and *S. suis* serotype 2 at 2.9Å resolution with cofactor NAD^+ bound are discussed in this chapter. The structures identify the key residues involved in catalysis and confirm RmlB belongs to the SDR superfamily. On the basis of homology and this work, a mechanism for RmlB is proposed.

3.3. The crystallisation of S. typhimurium *RmlB holoprotein*

Before crystallisation trials were carried out the protein was incubated overnight at 277K with 2.5mM DTT and 4mM NAD^+ to try and obtain crystals with NAD^+ coenzyme bound in the active site. Initial crystallisation trials were carried out using Hampton Crystal screens I and II™. The sitting-drop vapour-diffusion technique (Ducruix and Giege, 1992) was employed throughout the initial screening process, with 4µl of sample being mixed with 4µl reservoir solution. Preliminary protein concentrations were varied between 2mg ml^{-1} and 5mg ml^{-1} at temperatures of 277K and 293.5K. All the crystal plates were regularly examined over a period of one month. Initial results suggested using protein at a concentration of around 4mg ml^{-1} at 293.5K. Poor needle-shaped crystals were obtained in only two conditions of Hampton Crystal screens I and II™; screen I, condition 6: (30% PEG 4000, 0.1M Tris-HCl pH 8.5 and 0.2M sodium citrate); screen II, condition 25: (1.8M ammonium sulphate, 0.1M MES pH 6.5 and 0.01M cobalt chloride). A further 5 conditions gave either birefringent microcrystals or rosettes; screen I, conditions: 4, 22, 28, 40; screen II: condition 32 (see appendix I). Optimisation of Hampton Crystal screen I condition 6 and Hampton Crystal screen II condition 25 followed. This involved varying buffer pH, precipitant concentration (and the precipitant itself), the salt added and protein concentration.

Following the optimisation trials the best crystals were grown at 293K from a solution containing 4µl protein sample at 3.8mg ml^{-1} and 4µl precipitant (0.1M MES pH 6.3, 1.5M lithium sulphate). Large needle shaped crystals grew within 14 days to a maximum size of 1.5 x 0.15 x 0.15 mm (Figure 3.1), (Allard *et al.*, 2000).

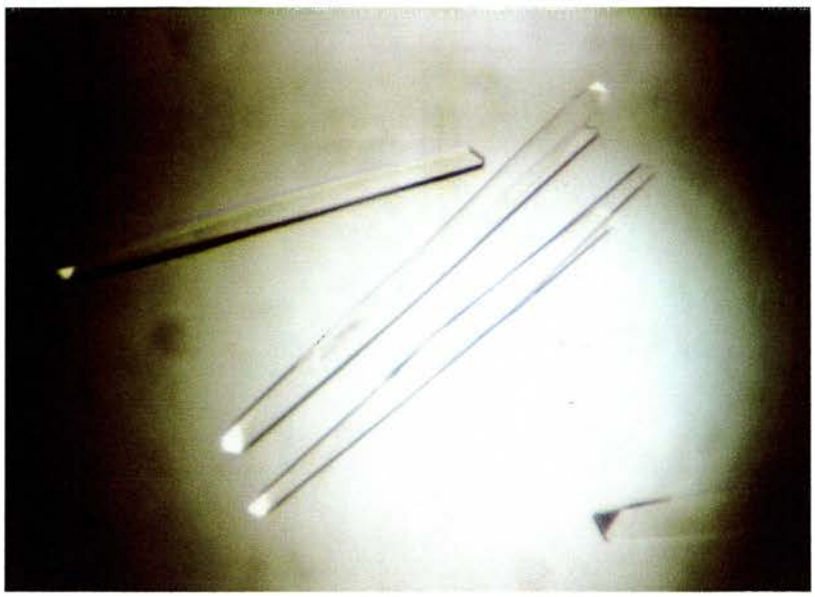


Figure 3.1. Photograph of S. typhimurium *RmlB crystals, with dimensions of 1.5 x 0.15 x 0.15mm.*

3.4. Data Collection of S. typhimurium *holoprotein*

Initial attempts to collect processable data in house failed due to the weak diffracting nature of the crystals, although confirmation of the crystals as protein was achieved. In house studies also enabled the optimisation of cryoprotection conditions. This required the crystals to be brought up through fresh solutions of mother liquor containing 5, 10, 15 and 20% glycerol with a 30 s soak in each.

Data were eventually collected from a single frozen crystal on station 7.2 Daresbury SRS at a wavelength of 1.488Å using a *mar345* Image Plate Detector (Figure 3.2, Table 3.1). The data were collected at 110K to a resolution of 2.8Å with 100% completeness. The crystal to detector distance was 220mm with data being recorded as 214 non-overlapping 90 s 1° rotation exposures. Data were indexed and integrated

with the program DENZO (Otwinowski and Minor, 1997) and merged with SCALEPACK (Otwinowski and Minor, 1997). A total of 210905 measurements were recorded, with 51729 unique reflections. The RmlB crystals were assigned to space group $P2_1$ following the observation of a 2n condition along the *k* axis. A summary of the data collection statistics is given in Table 3.2.

The solvent content of the crystals was estimated using the Matthews equation (Matthews, 1968):

$$V_M = \frac{V}{M_r \, n \, nmol}$$

V = unit cell volume ($Å^3$, abc sinβ for a monoclinic cell)

M_r = relative molecular mass

N = number of asymmetric units per unit cell – is space group dependent

nmol = number of molecules per asymmetric unit (unknown)

solvent fraction = $1 - (1.23/V_M)$

The estimation of the solvent content indicated the presence of two dimers in the asymmetric unit, leading to a Matthews number (V_M) of 3.3 $Å^3$ Da^{-1}, with a corresponding solvent content of 62%.

For 2 and 6 molecules in the asymmetric unit the V_M values were 6.5$Å^3Da^{-1}$ and 2.2 $Å^3Da^{-1}$, respectively. Given these V_M values, it seemed conceivable that either 4 or 6 molecules could exist in the asymmetric unit.

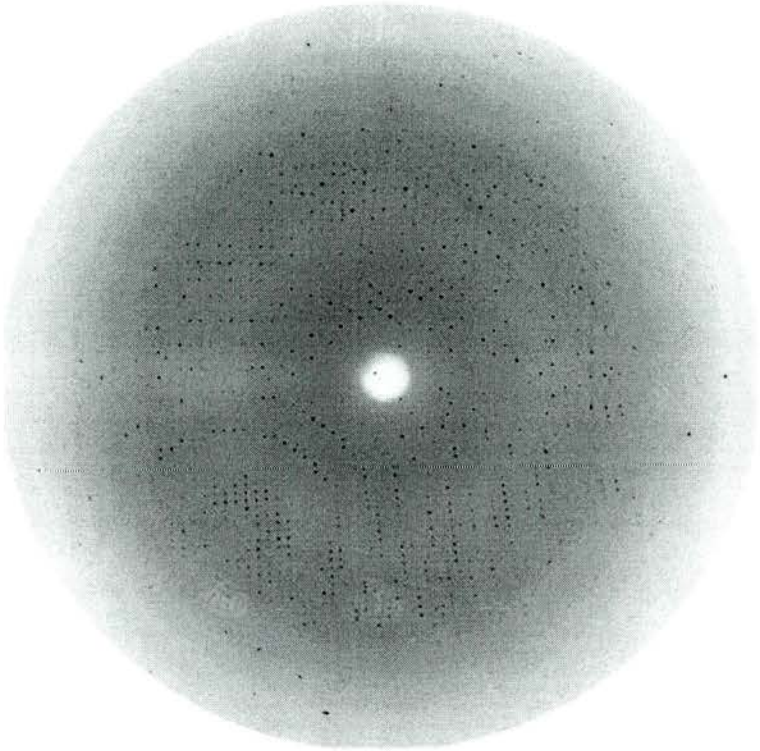


Figure 3.2. A single diffraction image of S. typhimurium *RmlB to 2.8Å from station 7.2 Daresbury SRS.*

Table 3.1. The 2.8Å data statistics for S. typhimurium *RmlB, station 7.2 Daresbury.*

Resolution (Å)	No. Reflections	Completeness (%)	†R_{merge} (%)	Redundancy
30.0 – 6.02	5315	100.0	2.8	4.2
6.02 – 4.78	5201	100.0	4.2	4.3
4.78 – 4.18	5180	100.0	4.3	4.3
4.18 – 3.80	5145	99.9	6.8	4.3
3.80 – 3.53	5160	99.9	9.4	4.3
3.53 – 3.32	5159	99.9	11.9	4.3
3.32 – 3.15	5137	99.9	16.4	4.3
3.15 – 3.02	5123	99.9	25.3	4.2
3.02 – 2.90	5159	99.9	34.5	4.2
2.90 – 2.80	5150	99.9	39.3	4.1
30.0 – 2.80	51729	99.9	7.4	4.2

† $R_{merge} = \sum_i [I_i - (I_i)] / \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

Table 3.2. Summary of x-ray data collection statistics for RmlB from S. typhimurium *(values in parentheses refer to the highest resolution shell). The two data sets are isomorphous.*

	Data set 1: Daresbury SRS	Data set 2: ESRF Grenoble
Wavelength (Å)	1.488	0.934
Space group	$P2_1$	$P2_1$
Resolution (Å)	30-2.8 (2.9-2.8)	40.6-2.47 (2.6-2.47)
Cell parameters	a=110.6Å, b=87.5Å, c=111.3Å β=98.0°	
V_M (four molecules per Asymmetric unit) ($Å^3$ Da^{-1})	3.27	
Percentage solvent	62	
Total Measurements	210905 (20036)	201143 (15645)
Unique reflections	51729 (5150)	67669 (7903)
$\langle I/\sigma \rangle$	10.1 (4.0)	9.5 (2.1)
Redundancy	4.2 (4.1)	3.0 (2.0)
Data completeness (%)	99.9 (99.9)	90.4 (72.9)
R_{merge}† (%)	7.4 (39.3)	7.0 (35.8)

† $R_{merge} = \sum_i [I_i - (I_i)] / \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

3.5. Structure determination of S. typhimurium *holoprotein*

The structure of RmlB from *S. typhimurium* was solved by molecular replacement using *AMoRe* (Navaza, 1994) as implemented in the CCP4 program suite (Collaborative Computational Project, Number 4, (1994). The phasing model used was the 1.9Å structure of RmlB from *E. coli*: Protein Data Bank (PDB, Berman *et al.*, 2000) ref: 1BXK, unpublished results. The amino acid sequence shows a 69.7% identity with that of *S. typhimurium* RmlB. All residues in *S. typhimurium* RmlB that differed from that of *E. coli* RmlB were mutated to alanine, and insertions in 2 loop regions were removed.

Using a dimer search model and data in the resolution range 10-3.3Å, clear solutions for two dimers in the asymmetric unit were obtained. α, β, γ, are the Eulerian angles and x, y, z, define the molecule's position in the cell.

A cross rotation function gave eight solutions, two of which were clearly higher than the rest: (i) α=355.34, β=46.10, γ=235.46 with a correlation coefficient of 27.8, equal to 16 standard deviations, (ii) α=179.27, β=39.05, γ=238.11 with a correlation coefficient of 26.2, 15 standard deviations.

All eight solutions were put into the translation function, which produced one clear solution for one dimer. This solution was fixed in the translational search and the second translation function (the second dimer) searched for (Table 3.3). At this point the space group was confirmed as $P2_1$. Following rigid body refinement in *AMoRe* the final two sets of coordinates for two-dimers in the asymmetric unit gave R-factors

and correlation coefficients of 41.2% and 62.6%, respectively (Table 3.4). The transformations were applied to the dimer search model using MOLEMAN2 (Kleywegt, 1999) and the crystal packing checked using the program O (Jones *et al.*, 1991).

Table 3.3. Peaks from the translation function.

Peak	α	β	γ	x	y	z	c.c*	R_{factor}	#Height/σ
1	355.39	46.12	235.40	0.0495	0.2305	0.2803	53.6	42.1	38
2	179.20	39.12	238.13	0.4914	0.7709	0.3148	52.8	42.6	37

**c.c refers to the correlation coefficient. # Refers to height of peak at this position divided by the r.m.s of the map.*

Table 3.4. Molecular replacement solution for two dimers following rigid body refinement.

Dimer	α	β	γ	x	y	z	c.c*	R_{factor}
1	355.65	45.98	235.14	0.0494	-0.0004	0.2804	62.6	41.2
2	179.11	39.23	238.40	0.4918	0.7745	0.3149	62.6	41.2

**c.c refers to the correlation coefficient.*

3.6. Model building and refinement of S. typhimurium *holoprotein*

The model was refined using *CNSsolve* (Brunger *et al.*, 1998). A random subset of data (5%) was omitted from all refinement calculations to provide an assessment of the refinement process (R_{free}). After 50 cycles of rigid body refinement the R-factor had decreased from 40.7% to 38.9% and R_{free} from 41.0% to 39.1%. SIGMAA (Read, 1986) weighted $2F_o$-F_c and F_o-F_c electron density maps were generated and viewed using the program O (Jones *et al.*, 1991). The $2F_o$-F_c and F_o-F_c maps revealed strong density for the protein chains and one molecule of NAD^+ bound tightly in the active site of each monomer. The model was then refined further using three rounds of individual *B*-factor refinement, one round of positional refinement and one round of simulated annealing, reducing the R-factor to 31.8% and R_{free} to 35.1%. Positional non-crystallographic symmetry restraints were applied throughout the refinement process. Real space electron density averaging was carried out using the program RAVE (Jones, 1992). At this stage a number of changes were made in the model protein structure, with most of the correct *S. typhimurium* residues being incorporated into the electron density maps. The ligand geometry was defined using the Engh and Huber stereochemical parameter dictionary (Engh and Huber, 1991) and was included in the model for further refinement. Subsequent refinement proceeded smoothly by alternating rounds of energy-minimisation, individual *B*-factor refinement and manual model building in O. Overall anisotropic thermal parameters and bulk solvent corrections were applied to the data using standard *CNSsolve* protocols (Brunger *et al.*, 1998). The model was refined to give an R-factor of 24.2% and R_{free} of 28.8% when a second data set to 2.47Å became available.

3.7. The collection of data to 2.47Å on S. typhimurium RmlB

Towards the end of the refinement of the 2.8Å model a second set of data to 2.47Å were collected at beam line 14.1 ESRF, Grenoble at a wavelength of 0.934Å using a MARCCD image plate detector. The crystal to detector distance was 200mm with data being recorded as 180 non-overlapping 10 s 1° rotation exposures (Table 3.5). The data were processed using MOSFLM (Leslie, 1992) and merged and scaled with SCALA from the CCP4 program suite (Table 3.2).

Table 3.5. 2.47Å data statistics for S. typhimurium *RmlB, station 14.1 ESRF Grenoble.*

Resolution (Å)	No. Reflections	Completeness (%)	R_{merge} (%)	Redundancy
40.6 – 7.81	2427	98.1	2.9	3.4
7.81 – 5.52	4410	100.2	4.0	3.7
5.52 – 4.51	5663	100.2	4.3	3.7
4.51 – 3.91	6634	99.6	5.2	3.7
3.91 – 3.49	7534	100.0	7.3	3.4
3.49 – 3.19	8070	97.3	9.7	3.0
3.19 – 2.95	8282	92.1	14.9	2.8
2.95 – 2.76	8342	86.4	21.6	2.6
2.76 – 2.60	8404	81.9	31.1	2.5
2.60 – 2.47	7903	72.9	35.8	2.0
40.8 – 2.47	67669	90.4	7.0	3.0

The partially refined 2.8Å data set was then used as the starting model against which to refine the 2.47Å data. The UNIQUEFY procedure from the CCP4 suite (Collaborative Computational Project, Number 4, (1994) was implemented in order to flag the same reflections for the R_{free} set as were used in the refinement of the 2.8Å structure. Refinement using *CNSsolve* was carried out, with further alternating rounds of energy-minimisation, individual *B*-factor refinement and manual model building in O. In addition, water molecules were added in batches using the water-pick procedure from *CNSsolve* (R-factor and R_{free} before water addition were 22.5% and 27.6%, respectively) provided that they satisfied the following criteria: (i) they corresponded to a peak > 3.5σ in the F_o-F_c map, (ii) they made hydrogen bonds with reasonable stereo chemistry, (iii) they reappeared in at least 1σ in subsequently calculated $2F_o$-F_c maps, and (iv) a drop in R_{free} was observed. Following multiple rounds of model building and refinement the model gave a final R-factor of 20.2% and an R_{free} of 24.9%. The final refinement statistics are summarised in Table 3.6.

3.8. Analysis of the final S. typhimurium *RmlB model*

The final structure contains two dimers in the asymmetric unit, each dimer comprising two identical subunits, 4 molecules of NAD^+, 9 sulphate molecules and 512 solvent molecules. Each monomer consists of 352 residues with 2800 non-hydrogen atoms. Of the 361 residues making up a complete monomer, residues 353-361 in all four monomers were omitted from the final model as no clear electron density could be observed for these residues. The final model had good overall stereochemistry as revealed by PROCHECK (Laskowski *et al.*, 1993). WHATIF (Vriend, 1990) was used to check the best orientation of O* and N* atoms in residues

such as Asn, Arg, Asp, Gln and His. PROMOTIF (Hutchinson and Thornton, 1996) aided in the evaluation of secondary structure elements. VOIDOO (Kleywegt and Jones, 1994) and *CNSsolve* (Brunger *et al.*, 1998) were used to calculate protein surface and volumes. Hydrogen bonds and salt bridges were analysed using *CNSSolve* (Brunger *et al.*, 1998) and the Protein-Protein Interactions Server (http://www.biochem.ucl.ac.uk/bsm/PP/server/).

The overall geometry of the model was good, with root mean square (rms) deviations of the model from ideal geometry of 0.009Å and 1.44° for bond lengths and bond angles, respectively (Engh and Huber, 1991). A Ramachandran plot generated using PROCHECK (Laskowski *et al.*, 1993) showed 89.5% of residues lay within the most favoured regions and 10.5% in further allowed regions; no residues were found in disallowed regions (Ramakrishnan and Ramachandran, 1965) (Figure 3.3). The average temperature factor of the refined model for all atoms was 40.4Å^2. The mean errors in the atomic coordinates estimated from a Luzzati plot (Luzzati, 1952) were 0.30Å (working set) and 0.37Å (test set).

Table 3.6. Refinement statistics for the final model of the 2.47Å S. typhimurium *RmlB holoenzyme structure.*

Resolution Range (Å)	40.6 – 2.47
Number of protein atoms in a.s.u	11200 (4 x 2800)
Number of solvent molecules	512
R-factor (working)[a] (%)	20.2
R_{free} (test)[b] (%)	24.9
R.m.s. deviations from ideal values:	
Bond lengths (Å)	0.009
Bond angles (deg)	1.44
Dihedrals (deg)	22.43
Improper dihedrals (deg)	0.97
Mean *B*-value ($Å^2$)	40
B-factor deviation bonds ($Å^2$)	1.4
B-factor deviation angles ($Å^2$)	2.4
Ramachandran:	
Residues in most favoured region (%)	89.5
Residues in allowed regions (%)	10.5

[a] R-factor = $\sum | F_{obs} - F_{calc} | / \sum | F_{obs} |$.
[b] 5% of reflection data.

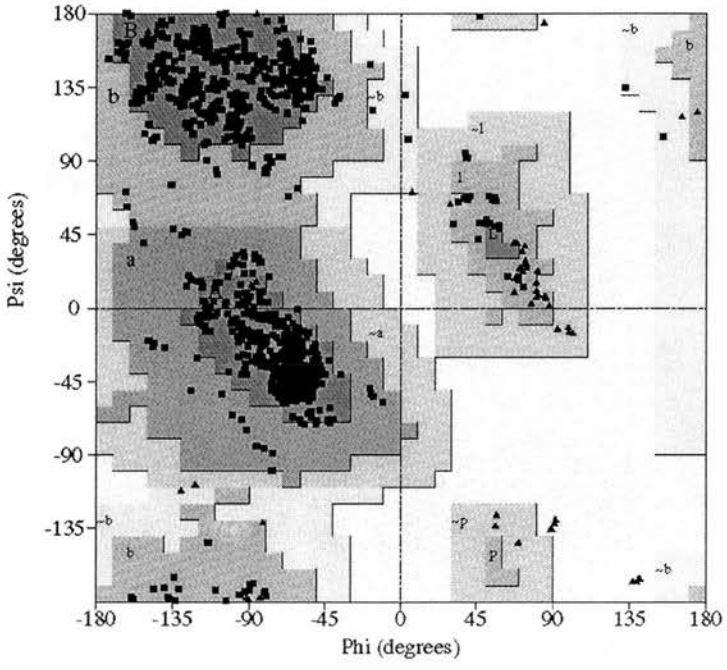


Figure 3.3. Ramachandran plot for the 2.47Å S. typhimurium *RmlB holoenzyme structure. Glycine residues (not restricted to any particular region of the plot) are represented by a ▲ and non-glycine residues by a ■. The darker the shaded region, the more favourable the φ, ψ combination.*

3.9. The crystallisation of S. suis *RmlB holoprotein*

Before crystallisation trials were carried out the protein was incubated overnight at 277K with 2.5mM DTT and 4mM NAD^+. Initial crystallisation trials were carried out using Hampton Crystal screens I and II™. The sitting-drop vapour-diffusion technique (Ducruix and Giege, 1992) was employed throughout the initial screening process, with 4µl sample being mixed with 4µl reservoir volume. Preliminary protein concentrations were varied between 3mg ml^{-1} and 5mg ml^{-1} at a temperature of 293.5K. All the crystal plates were regularly examined over a period of one month. Initial results suggested using protein at a concentration of around 4mg ml^{-1} at 293.5K. Crystals were obtained in only one condition, Hampton screen I, condition 9: 0.2M ammonium acetate, 0.1M *tri*-sodium citrate dihydrate pH 5.6 and 30% w/v PEG 4000. This condition produced both flat diamond crystals in large clusters and thin needles. Optimisation around this condition followed, including the use of Hampton additive screens I and II. Large single diamond shaped crystals were eventually obtained using the sitting drop method with protein at a concentration of 3mg ml^{-1} and a mother liquor containing: 0.2M ammonium acetate, 0.1M *tri*-sodium citrate dihydrate pH 5.4, 35% w/v PEG 4000 and the additive 3% w/v 1,6-hexanediol. Unfortunately, these crystals diffracted poorly both in-house and at the synchrotron, exhibiting high mosaicity. At this point an undergraduate student, Konstantinos Beis joined the laboratory. He was assigned to work with me and has proved invaluable in helping to set up and grow *S. suis* crystals. After more optimisation trials around the condition, some needle shaped crystals were grown using the hanging drop method with protein at a concentration of 2.3mg ml^{-1} and a mother liquor containing: 0.3M ammonium acetate, 0.1M *tri*-sodium citrate

dihydrate pH 5.4, 35% w/v PEG 4000 and the additive 3% w/v 1,6-hexanediol. These crystals were very fine needles (Figure 3.4), but were protein and diffracted, albeit to only a medium resolution.

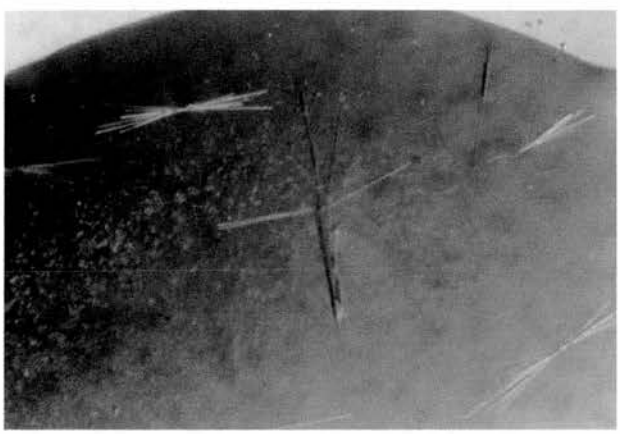


Figure 3.4. Photograph of S. suis *RmlB crystals.*

3.10. The collection of data to 2.9Å on S. suis *RmlB*

Due to the weak diffraction produced by the *S. suis* RmlB crystals, data had to be collected at a synchrotron radiation source. The crystals could be cryoprotected by being soaked for 30 s in mother liquor containing 15% glycerol.

Data were collected from a single frozen crystal on station 14.1 ESRF Grenoble at a wavelength of 0.934Å using an ADSC Quantum 4R CCD detector (Table 3.7). The data were collected at 110K to a resolution of 2.9Å with 92.4% completeness. The crystal to detector distance was 310mm with data being recorded as 180 non-overlapping 120 s 1° rotation exposures. Data were indexed and integrated in space group $P2_12_12_1$ with the program MOSFLM (Leslie, 1992) and merged and scaled using SCALA from the CCP4 program suite. A total of 148438 measurements were

recorded, with 23025 unique reflections. Table 3.8 gives a summary of the final data collection statistics.

Table 3.7. 2.9Å data statistics for S. suis *RmlB holoprotein, station 14.1, ESRF Grenoble.*

Resolution (Å)	No. Reflections	Completeness (%)	†R_{merge} (%)	Redundancy
22.36 – 9.17	906	100.0	7.7	5.9
9.17 – 6.48	1508	99.3	8.5	6.8
6.48 – 5.29	1927	100.0	11.2	7.0
5.29 – 4.59	2228	99.5	10.5	7.0
4.59 – 4.10	2493	99.2	11.6	6.9
4.10 – 3.74	2763	100.0	13.7	6.7
3.74 – 3.47	2977	99.5	17.7	6.6
3.47 – 3.24	3143	98.6	23.0	6.5
3.24 – 3.06	2871	86.3	31.6	5.7
3.06 – 2.90	2209	63.4	43.0	5.3
22.36 – 2.90	23025	92.4	15.2	6.4

† $R_{merge} = \sum_i [I_i - (I_i)] / \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

Table 3.8. Summary of x-ray data collection statistics for S. suis *RmlB holoprotein (values in parentheses refer to the highest resolution shell).*

	ESRF Grenoble, station 14.1
Wavelength (Å)	0.934
Space group	$P2_12_12_1$
Resolution (Å)	2.9
Cell parameters	a=61.5Å, b=97.8Å, c=181.3Å $\alpha=\beta=\gamma=90^o$
V_M (two molecules per Asymmetric unit) ($Å^3$ Da^{-1})	3.50
Percentage solvent	63%
Total Measurements	148438 (11648)
Unique reflections	23025 (2209)
$\langle I/\sigma \rangle$	4.2 (1.5)
Redundancy	6.4 (5.3)
Data completeness (%)	92.4 (63.4)
R_{merge}† (%)	15.2 (43.0)

† $R_{merge} = \sum_i [I_i - (I_i)]/ \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

3.11. Structure determination of S. suis *RmlB*

The structure of the RmlB holoprotein from *S. suis* was solved by molecular replacement using *AMoRe* (Navaza, 1994) as implemented in the CCP4 program suite (Collaborative Computational Project, Number 4, (1994). The phasing model used was the 2.47Å structure of RmlB from *S. typhimurium* (Allard *et al.*, 2001) (PDB ref:

1G1A). The amino acid sequence shows a 38.5% identity with that of *S. suis* RmlB. Residues in *S. suis* RmlB that differed from that of *S. typhimurium* RmlB were mutated to alanine. The V_M suggested one dimer in the asymmetric unit. A monomer search model was used in *AMoRe* with data in the resolution range 10-4.0Å.

The cross rotation function gave 56 solutions; the two highest solutions are shown in Table 3.9.

Table 3.9. The top two solutions from the rotation function.

Peak	α	β	γ	c.c*	R_{factor}	#Height/σ
1	29.00	38.68	126.00	15.0	57.7	5.69
2	111.25	28.82	138.77	15.2	57.5	5.20

**c.c refers to the correlation coefficient. # Refers to the height of peak at this position divided by the r.m.s of the map.*

The top forty-five solutions were put into the translation function. This produced a clear solution for one monomer, which was then fixed in the translation search whilst a second monomer solution searched for (Table 3.10):

Table 3.10. The solutions for two monomers from the translation function.

Monomer	α	β	γ	x	y	z	c.c*	R_{factor}	#Height/σ
1	111.25	28.82	138.77	0.1751	0.2007	0.4318	31.0	54.3	7.98
2	29.00	38.68	126.00	0.2276	0.4725	0.1855	39.0	51.5	11.17

**c.c refers to the correlation coefficient. # Refers to the height of peak at this position divided by the r.m.s of the map.*

Following rigid body refinement in *AMoRe,* the final two sets of coordinates for a dimer in the asymmetric unit gave R-factors and correlation coefficients of 50.8% 44.5%, respectively (Table 3.11). The solutions were applied to the dimer search model using MOLEMAN2 (Kleywegt, 1999) and the crystal packing checked using the program O (Jones *et al.*, 1991).

Table 3.11. Molecular replacement solution for one dimer following rigid body refinement.

Solution	α	β	γ	x	y	z	c.c*	R_{factor}
1	110.54	28.22	139.32	0.1746	0.2011	0.4318	44.5	50.8
2	28.70	38.38	127.27	0.2279	0.4727	0.1855	44.5	50.8

**c.c refers to the correlation coefficient.*

3.12. Model building and refinement of S. suis *RmlB*

The model was refined using *CNSsolve* (Brunger *et al.*, 1998); positional non-crystallographic symmetry restraints were applied throughout the refinement process. A random subset of data (5%) was omitted from all refinement calculations to provide an assessment of the refinement process (R_{free}). Initially rigid body refinement was performed to minimise any bad contacts, however after 50 cycles the R-factor had not decreased. Simulated annealing was then undertaken which reduced the R-factor to 46.0% and produced an R_{free} of 52.1%. SIGMAA (Read, 1986) weighted $2F_o$-F_c and F_o-F_c electron density maps were then generated and viewed using the program O (Jones *et al.*, 1991). The $2F_o$-F_c map was of poor quality, but did

allow the building in of some of the aromatic residues. One of the monomers in the asymmetric unit was built as far as the $2F_o$-F_c and F_o-F_c maps could be interpreted and then related to the position of the second monomer by using the program LSQKAB (Kabsch, 1976). This dimer was then put back into annealing which brought the R-factor down to 43.8 and R_{free} to 50.6%. The generated $2F_o$-F_c and F_o-F_c were of slightly better quality enabling more residues to be built in. One further round of annealing followed by energy minimisation refinement and group *B*-factor refinement reduced the R-factor to 38.7% and R_{free} to 44.2%. The superposition of the *E. coli* and *S. typhimurium* RmlB proteins onto the *S. suis* model enabled additional model building in O. Subsequent cycles of refinement continued with alternating rounds of energy-minimisation, group *B*-factor refinement and manual model building in O. The model was refined to give an R-factor of 21.5% and R_{free} of 26.9%. The final refinement statistics are summarised in Table 3.12.

Table 3.12. Refinement statistics for S. suis *RmlB holoprotein.*

Resolution Range (Å)	22.36 – 2.9
Number of protein atoms in a.s.u	5479 (2741+2738)
Number of solvent molecules	0
R-factor (working)[a] (%)	21.5
R_{free} (test)[b] (%)	26.9
R.m.s. deviations from ideal values:	
Bond lengths (Å)	0.007
Bond angles (deg)	1.27
Dihedrals (deg)	22.37
Improper dihedrals (deg)	0.778
Mean *B*-value ($Å^2$)	25
B-factor deviation bonds ($Å^2$)	5.2
B-factor deviation angles ($Å^2$)	8.0
Ramachandran:	
Residues in most favoured region (%)	86.7
Residues in allowed regions (%)	13.1

[a] R-factor = $\sum | F_{obs} - F_{calc} | / \sum | F_{obs} |$.
[b] 5% of reflection data.

3.13. Analysis of the S. suis *holoprotein model*

The final structure contains one homodimer in the asymmetric unit, comprising two identical subunits, and two molecules of NAD^+. Monomer A consists of 346 residues (3 to 348) with 2741 non-hydrogen atoms and monomer B also consists of 346 residues (2 to 347) with 2738 non-hydrogen atoms. The first two residues of monomer A and the first and last residues of monomer B were not built in because no clear density at 1σ was observed. The final stereochemistry of the structure was checked using the same programs that were used for the *S. typhimurium* structures.

The overall geometry of the model was good, with rms deviations from ideal geometry of 0.007Å and 1.27° for bond lengths and bond angles, respectively (Engh and Huber, 1991). A Ramachandran plot generated using PROCHECK (Laskowski *et al.*, 1993) showed 86.7% of residues lay within the most favoured regions and no residues were found in disallowed regions (Ramakrishnan and Ramachandran, 1965) (Figure 3.5). The average temperature factor of the refined model for all atoms was 25Å^2. The mean errors in the atomic coordinates estimated from a Luzzati plot (Luzzati, 1952) were 0.32Å (working set) and 0.41Å (test set).

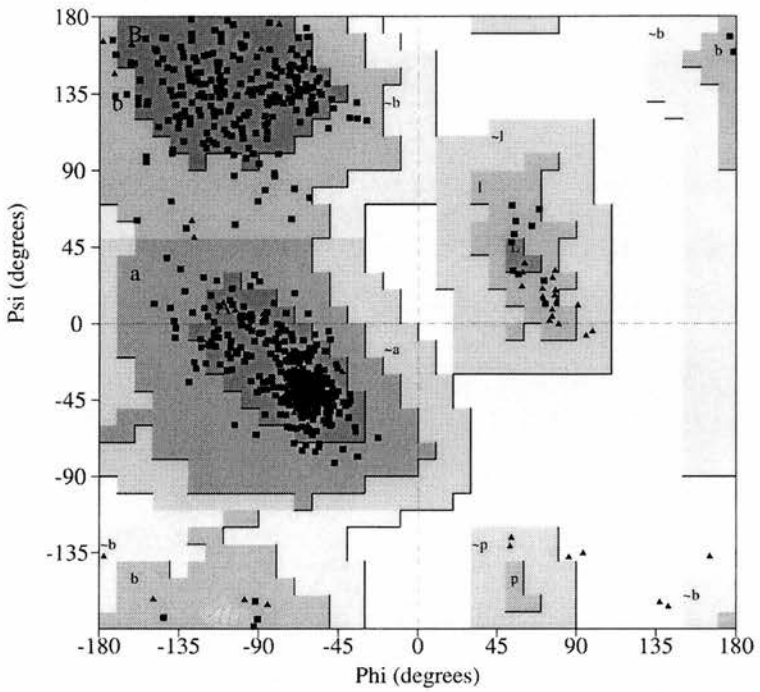


Figure 3.5. Ramachandran plot for the 2.9Å S. suis *RmlB holoenzyme structure. Glycine residues (not restricted to any particular region of the plot) are represented by a ▲, non-glycine residues by a ■. The darker the shaded region, the more favourable the φ, ψ combination.*

3.14. Results and discussion

Structural overview

The monomer from *S. typhimurium* consists of 361 amino acids with a molecular mass of 40.7 kDa, and a surface area and volume of 15100Å^2 and 46400Å^3, respectively. The *S. suis* monomer is slightly smaller, being made up of 348 amino acids with a molecular mass of 38.9 kDa, and a respective surface area and volume of 14400Å^2 and 45000Å^3. The monomers of both proteins belong to the α/β class of proteins. The complete *S. typhimurium* monomer consists of 43% α-helices and 14% β-strands (Table 3.13, Figure 3.6) compared to 39% α-helices and 16% β-strands for the *S. suis* monomer (Figure 3.7).

S. suis shares 38.5% sequence identity with *S. typhimurium* RmlB. The monomers from the two enzymes can be superimposed to give an rms deviation of 2.4Å over 315 Cα atoms (Figure 3.8). The reason for this high rms deviation between the two monomers is the apparent incorporation of dTDP into the active site of the *S. suis* monomers (Figure 3.9). Although the protein was not co-crystallised with substrate/substrate analogue, the density at 3σ clearly resembles dTDP, which was possibly a contaminant from *E. coli* metabolism during protein over-expression. The binding of dTDP results in major conformational changes in the subunits of RmlB, with the C-terminal clamping down over the active site (see chapter 4 for discussion of enzyme complex structures). This would explain the more ordered C-terminal region of the protein as compared to the *S. typhimurium* holoprotein and the lower average *B*-factor.

An alignment of the two proteins is shown in Figure 3.10. The homodimers from both organisms have monomers that interact *via* a four-helix bundle (Figure 3.11). Monomer-monomer interaction is mediated principally through hydrophobic interactions.

The structures of the monomers from *S. typhimurium* and *S. suis* are almost identical. As the *S. typhimurium* model is of a higher quality, the following discussion will principally focus on this structure. Important residues mentioned in the text have the corresponding conserved residues for *S. suis* shown in brackets.

For both proteins the main peptide chain is well defined in the $2F_o$-F_c electron density maps. However, in *S. typhimurium* RmlB there are two loop regions in all four monomers, residues His85 to Ala95 and Val136 to Ala160 that have small areas of high average *B*-factor. Residues Arg88 to Gly92 and Glu145 to Thr151 have average *B*-factors of 73Å^2 and 75Å^2, respectively, resulting in diffuse electron density. In the *S. suis* structure there were no missing loop regions or areas of particularly high *B*-factors, the *S. suis* structure having a much lower average temperature factor than the *S. typhimurium* model (25Å^2 as opposed to 40Å^2). Any atoms that were not clearly located in the $2F_o$-F_c map at 1σ were set to zero occupancy.

Table 3.13. The secondary structural elements of S. typhimurium *RmlB.*

Element[a]	Residues	Turns (β,γ)-Type[b]	Sequence
βA	2-6		
	5-8	β-IV	ITGG
	7-10	β-I	GGAG
αA	11-23		
βB	27-31		
	34-37	β-IV	LTYA
	35-38	β-I	TYAG
αB*	40-43		
	43-46	β-I	LSDI
	44-47	β-IV	SDIS
	46-49	β-I	ISES
	49-52	β-I	SNRY
βC	52-55		
	55-57	γ-inv	EHA
	58-61	β-I	DICD
αC	62-72		
βD	76-79		
	88-92	β-VIII	SITG
αD	95-100		
αE	102-116		
αF	120-125		
βE	127-133		
αG*	134-137		
	142-145	β-I	HPDE
	143-146	β-IV	PDEV
	144-147	β-IV	DEVE
	145-147	γ-inv	EVE
	156-159	β-I	TETT
αH	166-185		
	187-189	γ-inv	LPT
βF	189-194		
βG	196-198		
	199-202	β-II	GPYH
	203-206	β-I	FPEK
αI	207-216		
βH	221-223		
	224-227	β-IV	YGKG
	225-228	β-IV	GKGD
βI	233-235		
αJ	236-250		
	251-254	β-II	KAGE
βJ	255-258		
αK	266-280		
	279-282	β-IV	EIVP
	283-286	β-IV	KATS
αL*	287-290		
βK	291-293		
	297-300	β-IV	RPGH
αM	309-315		
αN	323-335		
	335-338	β-IV	LANT
αO	338-346		
αP*	348-351		

* Indicates a 3_{10} helix

[a] Secondary structure elements were assigned using the program DSSP (Kabsch and Sander, 1983)

[b] b-turns were classified using the program PROMOTIF v2.0 (Hutchinson and Thornton, 1996)

The structure of RmlB

For functional purposes the RmlB monomer can be divided into two domains, an N-terminal NAD^+ cofactor-binding domain and a C-terminal sugar-nucleotide binding domain (Figures 3.6 and 3.7). The N-terminal domain contains the characteristic dinucleotide-binding motif, the Rossmann fold (Rossmann *et al.*, 1975). In *S. typhimurium* this domain encompasses a central β-sheet composed of seven parallel β-strands (βA, βB, βC, βD, βE, βF, βJ), ranging in length from 4 to 7 amino acids. Surrounding the β-sheet are ten α-helices (αA, αB, αC, αD, αE, αF, αG, αH, αJ, αM) ranging from 4 to 20 amino acids in length. In *S. suis* the N-terminal domain also includes a central β-sheet composed of the same seven parallel β-strands. Surrounding the β-sheet are eleven α-helices (αA, αB, αC, αD, αE, αF, αG, αH, αJ, αL, αN). In *S. typhimurium*, the smaller C-terminal domain responsible for binding dTDP-D-glucose is composed of four β-strands: βG, βH, βI and βK (five β-strands in *S. suis*: βH, βK, βL, βI, βG) and six α-helices: αI, αK, αL, αN, αO and αP (five α-helices in *S. suis*: αK, αM, αO, αP, αQ). The two domains meet to form a cavity in the enzyme.

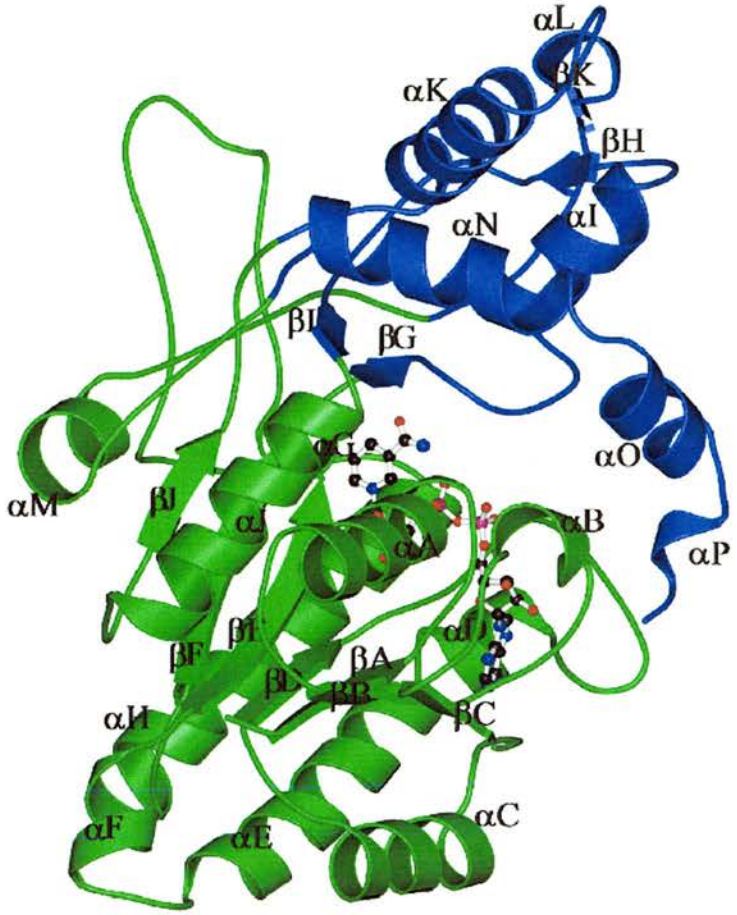


Figure 3.6. A ribbon representation of the S. typhimurium *RmlB monomer. The secondary structural elements are labelled proceeding from the N to C terminus. The* NAD^+ *coenzyme is shown bound in ball and stick representation. The N-terminal and C-terminal domains are colour-coded in green and blue, respectively.*

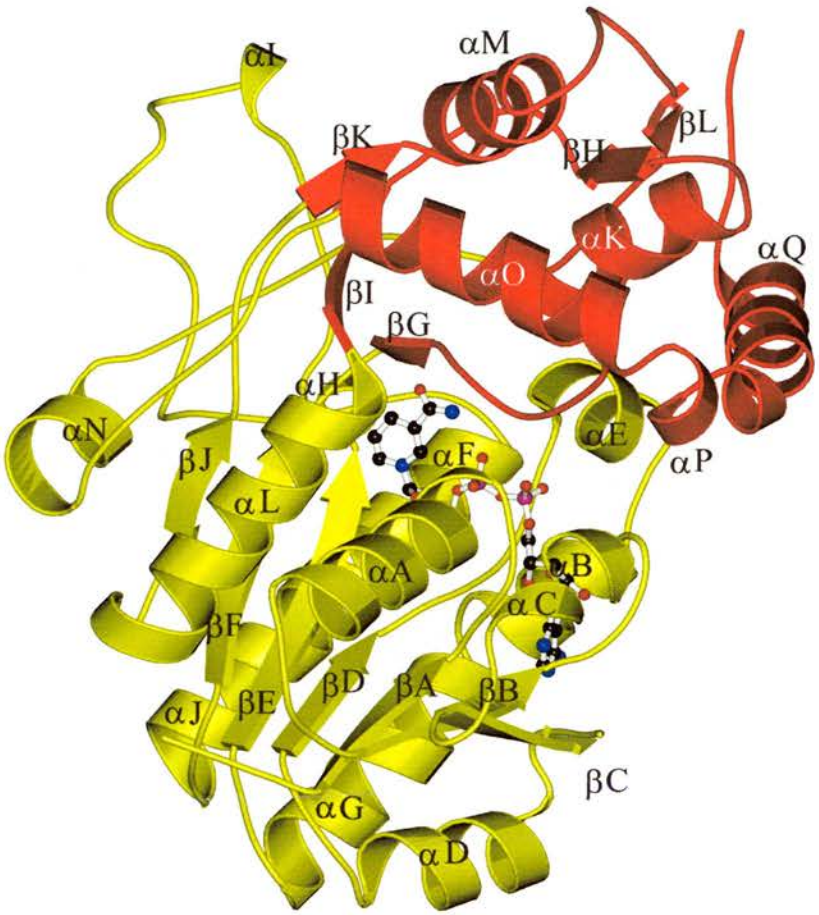


Figure 3.7. A ribbon representation of the S. suis *RmlB monomer. The secondary structural elements are labelled proceeding from the N to C terminus. The* NAD^+ *coenzyme is shown bound in ball and stick representation. The N-terminal and C-terminal domains are colour-coded in yellow and red, respectively.*

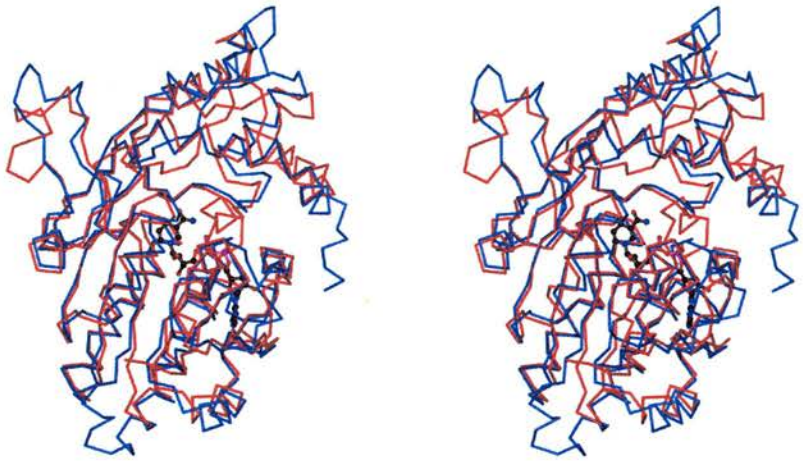


Figure 3.8. Stereo Cα trace superposition of the monomers of RmlB from S. typhimurium *(blue) and* S. suis *(red). The two monomers can be superimposed with an rmsd of 2.4Å over 315 cα atoms. The NAD^+ coenzyme is shown bound in ball and stick.*

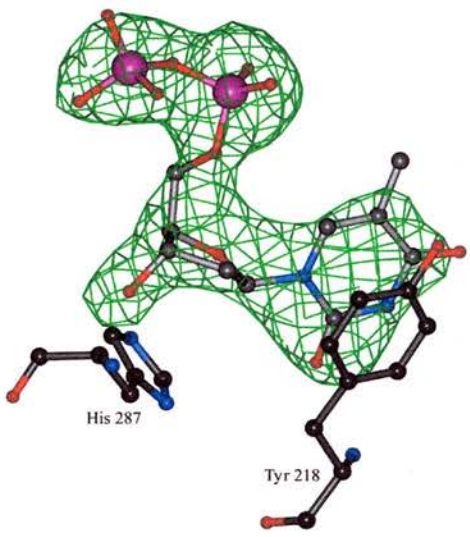


Figure 3.9. The density at 3σ from the active site of the S. suis *'holoprotein' monomer. As can clearly be seen this resembles dTDP. For clarity two active site residues are also shown (for a complete discussion of dTDP binding see chapter 4).*

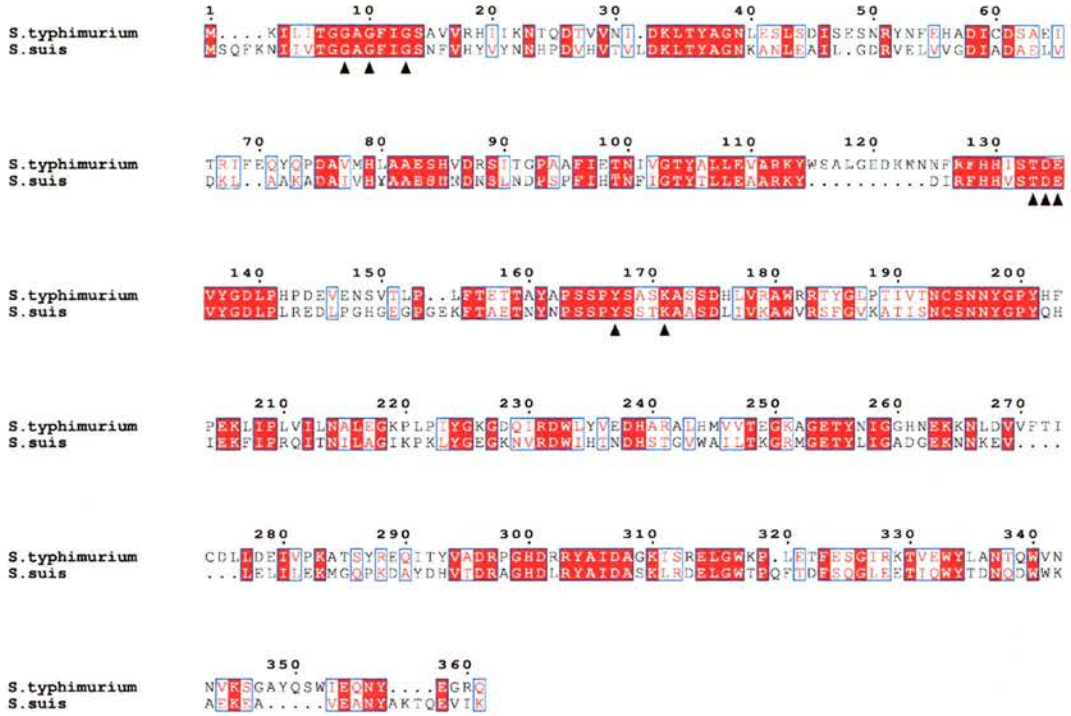
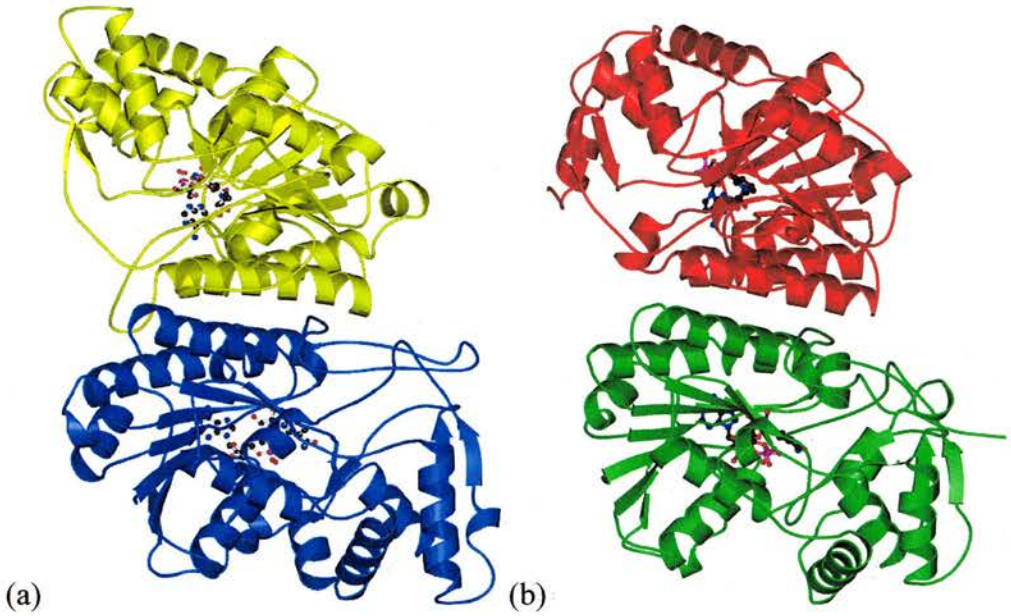


Figure 3.10. Sequence based alignment of the RmlB proteins from S. typhimurium *and* S. suis. *The numbering scheme refers to* S. typhimurium. *Totally conserved residues are highlighted in solid red. Black triangles below the alignment indicate important residues involved in catalysis. Alignment generated using ESPript (Gouet* et al., *1999).*

Subunit association

The active form of the enzyme is a homodimer (Figure 3.11), with each monomer binding one molecule of NAD^+ cofactor in the open conformation. RmlB monomers interact *via* a four-helix bundle, which is a common theme in SDR enzymes (Jornvall *et al.*, 1995). In *S. typhimurium* this utilises helices αE and αH (residues 102-116 and residues 166-185, respectively) of both monomers and in *S. suis* helices αG and αJ (residues 104-116 and 160-179, respectively). The major molecular forces holding the monomers together are hydrophobic interactions, which is evidenced by the high percentage of hydrophobic residues at the dimer surface interface (*S. typhimurium*, 63% and *S. suis*, 58%). Two salt bridges are also seen to be involved in the dimerisation process. Each salt bridge involves an interaction between Arg183 (Arg177) and Asp301 (Asp288) from the two monomers; both residues are highly conserved amongst RmlB enzymes (see Figure 3.12). In addition, there are a further four hydrogen bonds between the monomers in *S. typhimurium*: Ser164, His176 (2 bonds) and Thr184; in *S. suis* there are an additional seven hydrogen bonds: Ser96 (2), Asp131, Glu152, Ser158, Ser178 and Gly286. A total of 1263Å^2 (1302Å^2 in *S. suis*) of each monomer, representing 8.2% (8.9% in *S. suis*) of the monomer surface area, is used in subunit association. This figure is similar to that of RmlB and GMD from *E. coli* (1300Å^2 and 1150Å^2, respectively), although it is substantially larger than that of EGALE (900Å^2), whose subunit interactions involve only 6.4% of its monomer surface area.



(a) (b)

Figure 3.11. Ribbon representations of the homodimers from (a) S. typhimurium *and (b)* S. suis *RmlB. The NAD^+ coenzyme is shown bound in ball and stick representation. The monomers are coloured separately.*

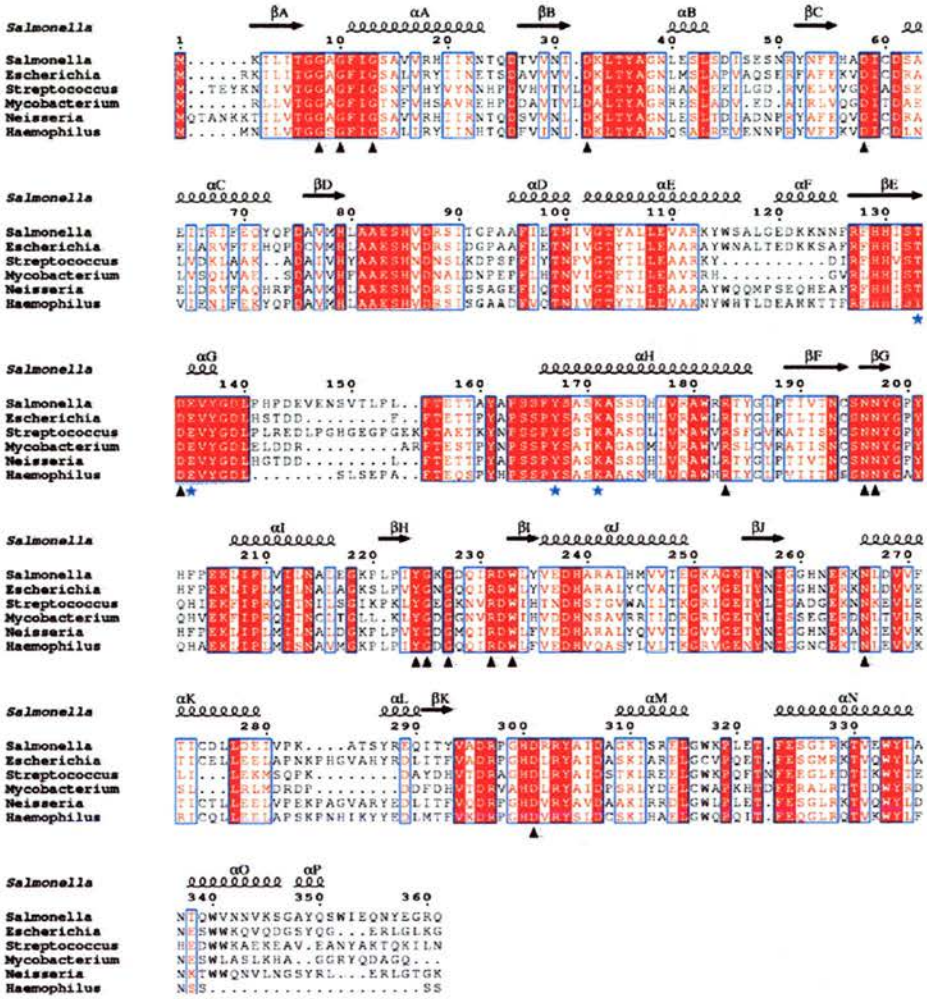


Figure 3.12. Sequence alignment of RmlB proteins from S. typhimurium *(Genbank accession number: P26391),* E. coli *(P27830),* S. mutans *(P95780),* M. tuberculosis *(Q50556),* Neisseria meningitidis *(P55294) and* Haemophilus influenzae *(P44914) produced using ESPript (Gouet* et al., *1999); the Genus names are shown next to the amino acid sequences. The secondary structural elements of RmlB from* S. typhimurium *are shown above the alignment. Totally conserved residues amongst all six RmlB proteins are highlighted in solid red. Blue stars below the alignment indicate important catalytic residues discussed in the text. Black triangles below the alignment represent important conserved residues mentioned in the text.*

The active site

The active site of RmlB consists of the cavity created by the NAD^+ and dTDP-D-glucose binding regions.

NAD^+ binding region

The enzyme is a class B dehydrogenase (Wang and Gabriel, 1970), in that the nicotinamide ring of the cofactor adopts the *syn* conformation, presenting the proper *si* face for pro-S hydride transfer from C4 of the dTDP-D-glucose to C4 of the dinucleotide cofactor (Figure 3.13). Like GALE, the NAD^+ is tightly bound in the active site and does not dissociate from the enzyme either in the course of catalysis or between catalytic cycles, which is in contrast to other SDR enzymes such as dihydropterine reductase where the NAD^+ dissociates from the enzyme after the reaction has occurred (Varughese *et al.*, 1994). Figure 3.14a shows that the NAD^+ is also bound in the base-base destacked open conformation, with a distance of 15.8Å between C6 of the adenine ring and C6 of the nicotinamide ring. The key residues involved in interactions with the NAD^+ cofactor are shown schematically in Figure 3.14b. This region contains the characteristic βαβ motif, which is highly conserved in SDR enzymes. The motif includes the conserved SDR sequence GlyXGlyXXGly (residues 8-13 in *S. typhimurium* and 12-17 in *S. suis*, (Figure 3.10)) that is responsible for binding the nucleotide phosphate. The first glycine, Gly8 (Gly12), a strictly conserved residue, allows a tight turn of the main chain, which is important for positioning the second glycine. The second glycine, Gly10 (Gly14), facilitates the close contact of the main chain to the diphosphate of the NAD^+ (Bellamacina, 1996).

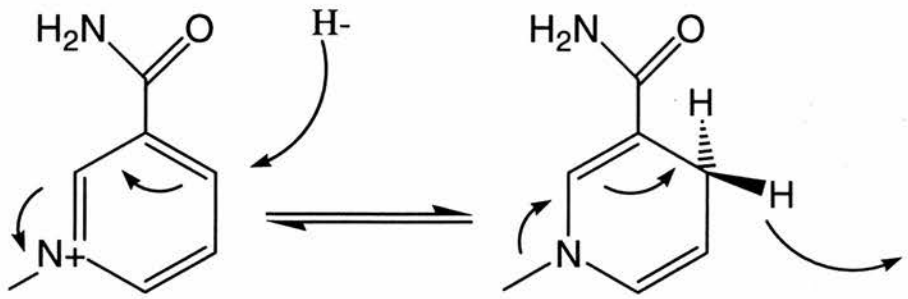
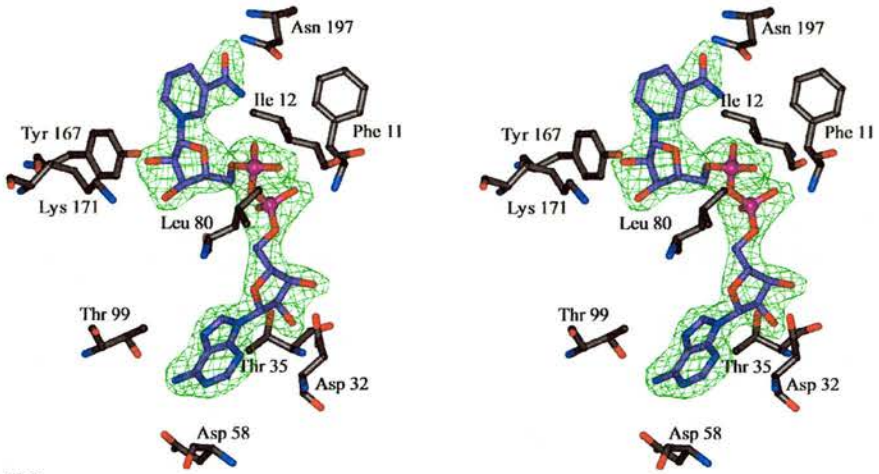


Figure 3.13. The stereo chemical nature of hydride transfer in RmlB. Here the NAD(H) adopts the syn *conformation presenting the si face for pro-S hydride transfer. The si face faces out of the paper, the substrate dTDP-D-glucose would be above the plane of the paper, with the glucose ring aligned almost on top of the nicotinamide ring. It is the pro-S hydrogen (facing out) that is transferred back from the C4' position of the NADH to C6' of the sugar.*

(a)



(b)

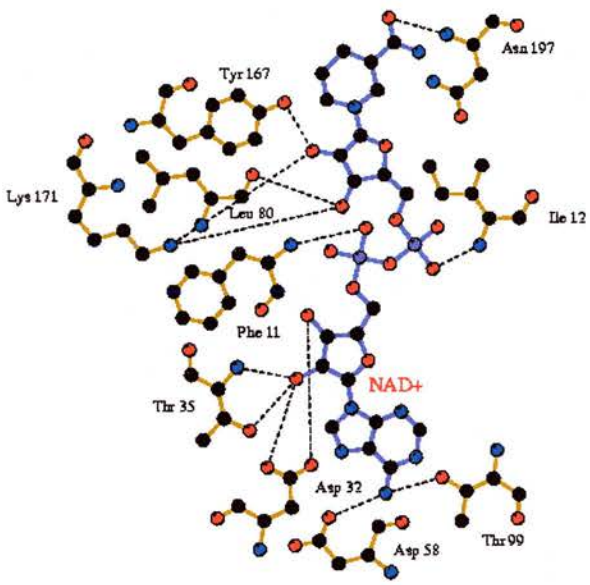


Figure 3.14 (a) Stereo view of the residues interacting with the cofactor NAD^+ from S. typhimurium *RmlB with the F_o-F_c map surrounding the NAD^+ contoured at 3.0σ. (b) Schematic representation of the binding of the cofactor NAD^+ in the active site of RmlB from* S. typhimurium *generated by LIGPLOT (Wallace* et al., *1995). Hydrogen bonds are indicated as black dashed lines. Atoms are colour-coded as follows: black, carbon; blue, nitrogen; red, oxygen. Ligand bonds are shown in purple and non-ligand bonds in orange.*

In both *S. typhimurium* and *S. suis* RmlB, the adenine segment of NAD^+ binds to a hydrophobic crevice. In *S. typhimurium* this pocket includes residues: Ile21, Ala57, Ile59, Val77, Ala81 and Leu107. The nicotinamide ring of the cofactor has an intramolecular hydrogen bond of 2.9Å between the nitrogen of its carboxamide group and its phosphoryl oxygen; this is also seen in *S.suis* RmlB, EGALE and HGALE. Six side chains interact directly with the nucleotide: Asp32 (Asp37), Thr35 (Thr40), Asp58 (Asp62), Thr99 (Thr101), Tyr167 (Tyr161) and Lys171 (Lys165). Asp32 (Asp37) discriminates between the nucleotides NAD^+ and NADP. In RmlB Asp32 forms hydrogen bonds with the 2'-hydroxyl and 3'-hydroxyl of the ribose sugar (2.7Å and 2.8Å, respectively); this conserved aspartate is present in both EGALE and HGALE, which like RmlB do not use NADP. Asp32 would clash both sterically and electrostatically with the extra phosphate group attached to the 2'-hydroxyl group of the adenosine ribose of NADP. In GMD, which binds NADP, Gly31 replaces the negatively charged Asp in this position, creating space for the extra phosphate group of NADP. Both positively charged residues nearby, Arg34 and Arg35, facilitate binding of the negatively charged phosphate groups. This GlyXXArgArg sequence is conserved in other GMD proteins that also bind NADP.

Asp58 (Asp62) in RmlB is an important residue in that it hydrogen bonds with the N1 atom of the adenine ring holding it in position, a role that it also plays in both EGALE and HGALE. Two other highly conserved residues, Tyr167 (Tyr161) and Lys171 (Lys165), both hydrogen bond to the ribose moiety of the nicotinamide ring; the tyrosine to the 2'-hydroxyl group (2.7Å) and the lysine to the 2' and 3'-hydroxyl groups (3.3Å & 2.8Å). Mutation of the corresponding lysine in EGALE which also hydrogen bonds with the ribose hydroxyl groups, resulted in a partial loss of NAD^+

during purification (Swanson and Frey, 1993), confirming its significance in binding NAD^+. In *S. typhimurium* a further five hydrogen bonds are contributed by Phe11 (Phe15), Ile12 (Ile16), Thr35 (Thr40), Leu80 and Asn197 (Asn191); all but the Leu are highly conserved in the other RmlB sequences studied. In *S. suis,* there is in addition to the four residues noted above, Ser86 and Tyr82, which also supply two more hydrogen bonds. Figure 3.12 shows a sequence alignment of six RmlB proteins.

Modelling of substrate dTDP-D-glucose into the active site of* S. typhimurium *RmlB

Due to the binding of dTDP in the *S. suis* active site (see Figure 3.9), plus the higher quality of the *S. typhimurium* structure, and that at this time no co-crystals had been obtained with substrate or substrate analogue, substrate was modelled in to the active site of the *S. typhimurium* holoprotein.

A monomer from EGALE (Thoden *et al.*, 1996b), (PDB code 1XEL), can be superimposed onto *S. typhimurium* RmlB with an rms deviation of 1.55Å over 282 Cα atoms (Figure 3.15). Using the superposition of the two monomers, UDP-glucose from the EGALE substrate complex was positioned in the RmlB active site. Trivial modification of UDP to dTDP completed the modelling of dTDP-D-glucose in the RmlB active site (Figure 3.16).

The thiamine ring sits in a hydrophobic pocket, surrounded by Pro204, Leu207, Leu210, Val 11, Pro222, Ile223, Val269, Val270 with Tyr224 positioned just above the thiamine ring. The thiamine ring of dTDP-D-glucose π-stacks against the side chain of Tyr224 (Tyr218); a β-IV turn, containing two glycines at positions 225 (219)

and 227 (221), enables Tyr224 to make this interaction. This is similar to HGALE where Phe226, the structural equivalent of Tyr224 in RmlB, stacks against the uridine ring (Thoden *et al.*, 2000). An alignment of different bacterial RmlB enzymes (Figure 3.12) confirms this region is conserved.

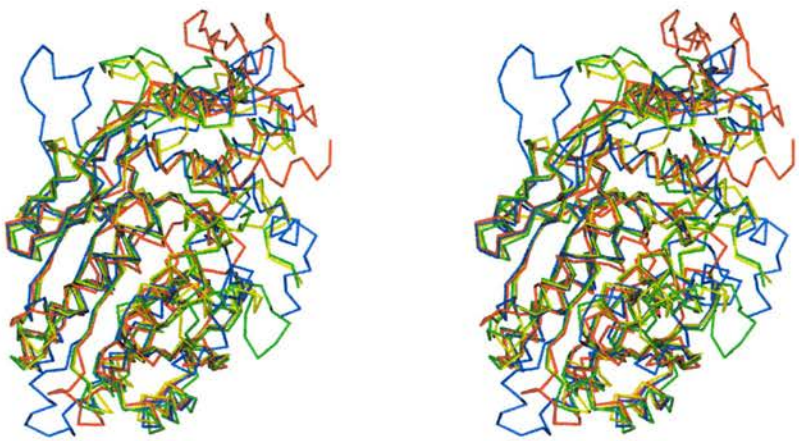
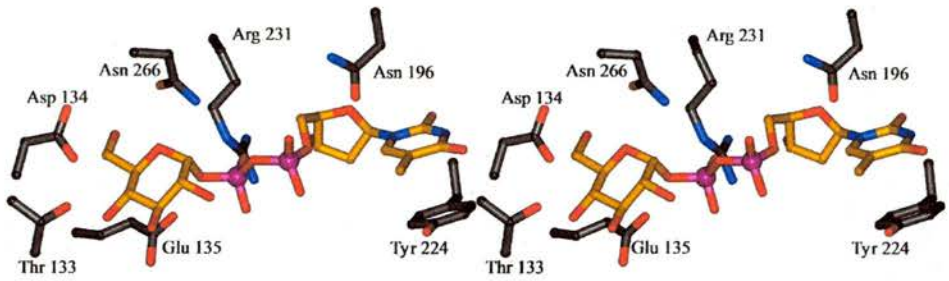


Figure 3.15. Stereo Cα trace superposition of the monomer from S. typhimurium *RmlB (blue) with GMD (red), EGALE (yellow) and HGALE (green). The monomers are in the same orientation as that of the RmlB monomer from* S. typhimurium *in Figure 3.6.*

A number of residues make contacts with the nucleotide sugar when it is modelled into the active site: Thr133 (Thr125), Asp134 (Asp126), Glu135 (Glu127), Asn196 (Asn190), Arg231 (Arg225) and Asn266 (Asn260) (Figure 3.16). All these residues are highly conserved amongst RmlB sequences. Residues Thr133, Asp134 and Glu135 make key electrostatic interactions with the glucose moiety of the substrate. Thr133 and Glu135 hydrogen bond to the 4'-hydroxyl of the glucose ring, with Asp134 hydrogen bonding to the 6'-hydroxyl group. Asn196 and Arg231 interact with the phosphoryl oxygens of the dTDP-D-glucose. Asn266 hydrogen bonds to the 3'-hydroxyl group of the ribose sugar, suggesting it may also control the selection for the deoxy-nucleotide sugar substrate in the binding site. An extra hydroxyl group at the 2' position of the ribose ring of the oxy-nucleotide sugar would sterically clash with the Asn266 side chain as the sugar moves to stack with Tyr224. All the RmlB sequences studied show an Asn at this position, however EGALE and HGALE, both of which bind oxy-nucleotide substrates, have a Val at this position with an Asp located close by for binding to the ribose hydroxyls.

(a)



(b)

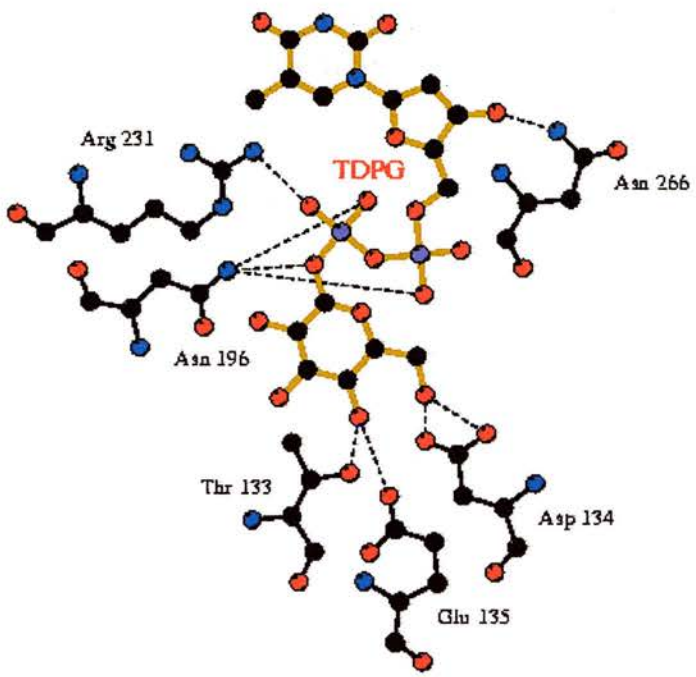


Figure 3.16. (a) Stereo view of the residues interacting with the sugar nucleotide dTDP-D-glucose (orange) when modelled into the active site of RmlB from S. typhimurium. *(b) Schematic representation of the binding of the sugar nucleotide dTDP-D-glucose (TDPG) when modelled into the active site of RmlB from* S. typhimurium *generated by LIGPLOT. Hydrogen bonds are indicated as black dashed lines. Atoms are colour-coded as follows: black, carbon; blue, nitrogen; red, oxygen. Ligand bonds are shown in orange and non-ligand bonds in black.*

What the* S. typhimurium *2.47Å structure reveals about the mechanism of RmlB

As already discussed in chapter 1, the mechanism of RmlB has been the subject of extensive investigation. The C6' dehydration catalysed by RmlB is a prototypical reaction necessary for the biosynthesis of all 6-deoxyhexoses (Hallis and Liu, 1999).

In the first step the NAD^+ cofactor abstracts a hydride from the C4' position of the glucose ring. The nicotinamide ring is planar, with a distance of 3.8Å between C4' of the nicotinamide ring and C4' of the glucose in the modelled dTDP-D-glucose complex. The C4' atom of the glucose ring forms an angle with the N1-C4 atoms of the nicotinamide ring of 90°. An analysis of redox enzyme substrate complexes in the PDB, (including flavin adenine dinucleotide and flavin mononucleotide dependent proteins) reveals the site that donates or accepts hydride to or from co-factor is at a distance of between 3.5Å and 4.0Å from C4 of the nicotinamide or N5 of the flavin ring (Fraaije and Mattevi, 2000) (see appendix II). The angle that the site of oxidative or reductive attack defines with the nicotinamide or flavin rings is between 80° and 120°.

Tyr167 (Tyr161) acting as an active site base, deprotonates the 4'-hydroxyl of the glucose ring, leading to the 4-keto-sugar intermediate (Figure 3.16). Lys171 (Lys165) stabilises the phenolic form of Tyr by hydrogen bonding to the 2' and 3'-hydroxyl groups of the nicotinamide ribose moiety of NAD^+ (Figure 3.14) creating a positive electrostatic field around itself and the nicotinamide ring, lowering the pK_a of Tyr167. This is confirmed by measurement of the pK_a of the corresponding tyrosine in *E. coli* RmlB (Tyr160) that gives a value of 6.41 (Gerratana *et al.*, 2001).

In this modelled structure, as with EGALE, it is seen that the phenolic oxygen of Tyr167 is over 4.0Å from the 4'-hydroxyl group of the glucose moiety, which in this case would preclude the Tyr acting directly as the active site base. To overcome this in EGALE, the conserved Ser (Thr133 in RmlB) has been proposed to act as proton shuttle (Thoden *et al.*, 1996a). This would be consistent with this model as Thr133 sits between the glucose 4'-hydoxyl group and Tyr167. However, recent work on HGALE (Thoden *et al.*, 2000) revealed that on binding substrate (UDP-glucose), there is a conformational change of the C-terminal sugar-binding domain that moved the Tyr into a position enabling it to act directly as the active site base.

At this point the catalytic mechanisms of RmlB and GALE diverge as a consequence of the fate of their 4-keto sugar intermediates (Figure 3.17). In GALE the keto intermediate rotates in the active site. NADH then reduces the 4-keto group by hydride transfer from the opposite face from abstraction, thus inverting the configuration at C4.

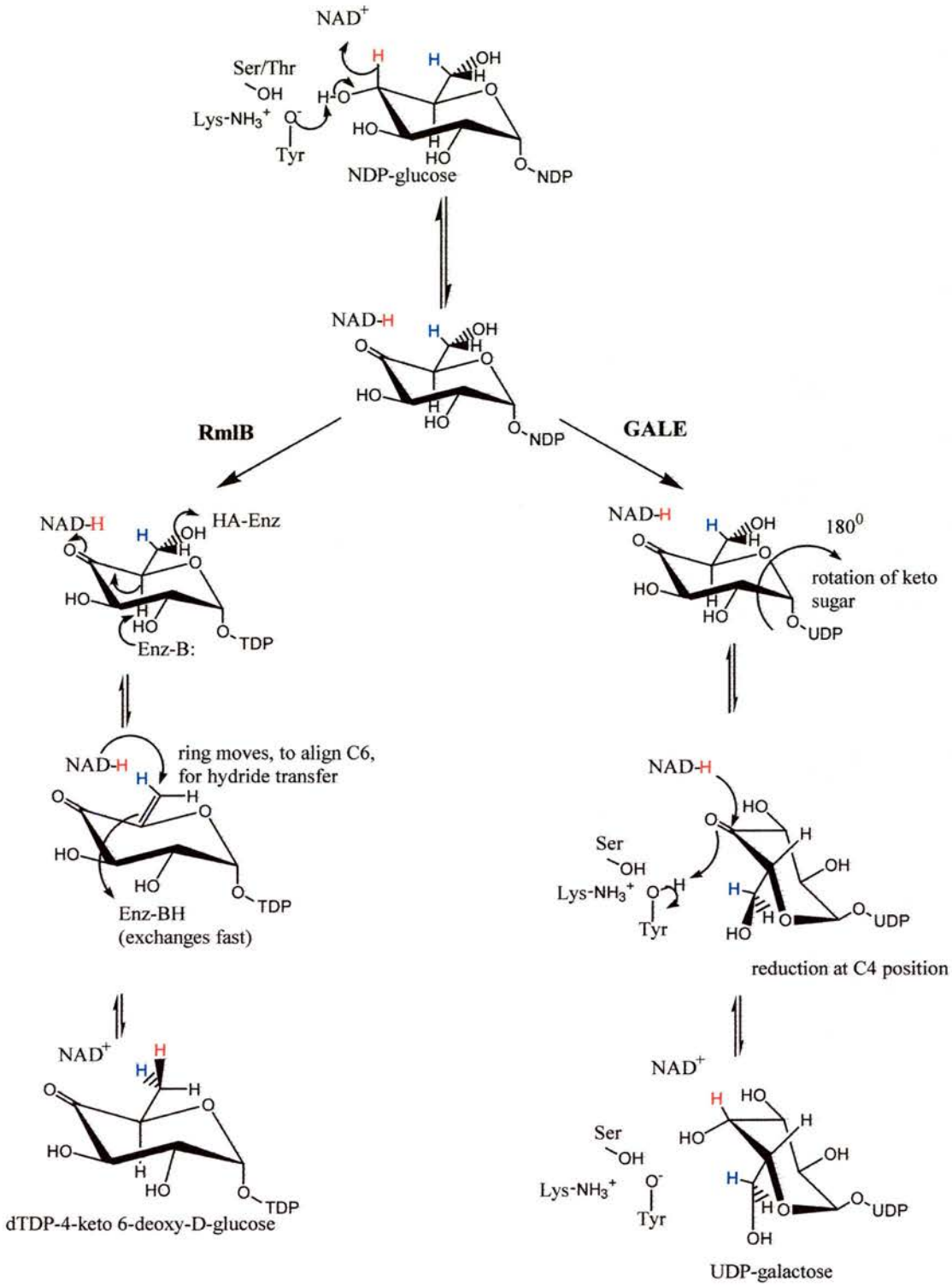


Figure 3.17. The mechanisms of RmlB and GALE compared. These enzymes have similar structures and mechanisms although the fate of the keto sugar is different.

In RmlB, following the abstraction of the hydride from the glucose ring, Glu135 (Glu127) is in a position to act as a base and remove a proton from the C5' of the sugar ring, a role that Glu135 may also play in GMD (Somoza *et al.*, 2000). Mutation of this corresponding Glu in GMD to a Gln results in a 1500 fold reduction in k_{cat}, confirming that the group abstracts the proton (Somoza *et al.*, 2000). The C5 proton is acidified by the keto group, which stabilises the formal negative charge through enolate delocalisation of charge (Figure 3.18). The enolate collapses by formation of a double bond between C5' and C6'. The formation of the dTDP-4-keto-5,6-glucosene intermediate was recently confirmed for *E. coli* RmlB (Gross *et al.*, 2000).

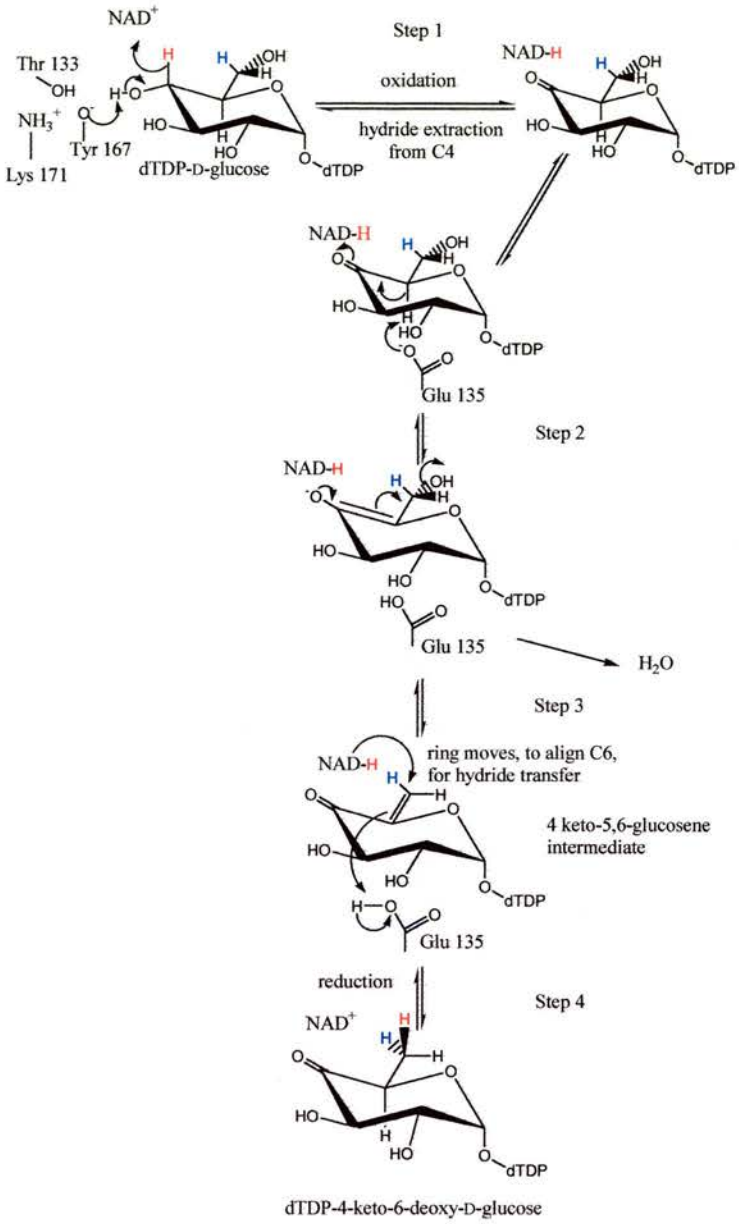


Figure 3.18. The proposed catalytic mechanism of dTDP-D-glucose 4,6-dehydratase (RmlB) based on dTDP-D-glucose being modelled into the active site of S. typhimurium *RmlB. Step 1: NAD^+ extracts a hydride from the C4 position of the glucose ring. Step 2: Glu135 functioning as a general base removes a C5 proton. Step 3: The syn elimination of a water molecule between C5 and C6 to give the 4-keto-5,6-glucosene intermediate. Step 4: Transfer of hydride from NADH to C6 of the glucose ring with inversion of configuration; Glu135 acting as a general acid protonates C5.*

The final step is transfer of hydride to C6' resulting in inversion of configuration at C6' of the glucose (Snipes *et al.*, 1977) and protonation of C5' by Glu135. The distance in the model between C6' glucose and C4' of the nicotinamide ring is 4.8Å, forming an angle of 131° with the N1-C4 atoms of the nicotinamide ring. These values are inconsistent with hydride transfer. It was not possible to position the dTDP-D-glucose within the active site in such a way that the distances and angles involved in hydride transfer are within the known parameters for such processes without there being either steric clashes with amino acid side chains or making proton abstraction by Glu135 sub-optimal. These results suggest therefore, that there would be some rearrangement in the active site to allow movement of the substrate in order to optimise its position for hydride abstraction/addition and proton abstraction /addition.

3.15. Conclusions

RmlB is an α/β protein that can be divided into two domains; a nucleotide cofactor binding domain consisting of a seven parallel stranded β-sheet at the N-terminal and a smaller sugar binding C-terminal domain. The enzyme functions as a homodimer principally through hydrophobic interactions *via* a four-helix bundle. Its catalytic mechanism involves the highly conserved SDR catalytic couple of Tyr167 (Tyr161) and Lys171 (Lys165) with an associated Thr133 (Thr125) and Glu 135 (Glu127).

Chapter four

The co-crystallisation, data collection and refinement of the structures of RmlB from S. typhimurium *with dTDP-D-glucose and dTDP bound and RmlB from* S. suis *with dTDP-D-glucose, dTDP and dTDP-xylose bound*

4.1. Summary

The structures of RmlB from *S. typhimurium* in complex with substrate (dTDP-D-glucose) and dTDP, and RmlB from *S. suis* in complex with dTDP-D-glucose, dTDP and dTDP-xylose have all been solved at resolutions between 1.8Å and 2.4Å. The structures show that the active sites are highly conserved, both binding coenzyme (NAD^+) and substrate in identical manners. Comparing the polypeptide chains for the previously solved resting enzyme of *S. typhimurium* RmlB (RmlB with NAD^+ bound) with the complexes reveals there to be substantial conformational changes in three main areas of the protein. Importantly, the substrate bound proteins show that the active site tyrosine functions directly as the active site base, and reveals an Asp/Glu pair required for dehydration.

4.2. Introduction

Deoxysugars have long been recognised as an important class of carbohydrate, being found in lipopolysaccharides, glycoproteins, glycolipids and many secondary metabolites (Johnson and Liu, 1998). The C6-deoxygenation catalysed by RmlB and other NDP-hexose 4,6-dehydratases is the first committed step in almost all deoxyhexose biosynthetic pathways leading to the formation of di-, tri- and tetra-deoxysugars (He *et al.*, 2000; Johnson and Liu, 1998; Liu and Thorson, 1994). For this reason, the mechanism of RmlB has been the subject of extensive investigation (Gerratana *et al.*, 2001; Gross *et al.*, 2000; Hegeman, 2001; Hegeman *et al.*, 2001). Catalysis proceeds through three distinct chemical steps (Gabriel and Lindquist, 1968). The substrate dTDP-D-glucose is initially oxidised at glucosyl-C4' by the tightly bound NAD^+ coenzyme. This is followed by the concerted elimination of water from glucosyl-C5' and C6' to form the 4 keto-5,6-glucosene intermediate (Gross *et al.*, 2000). The final step is the reduction of this intermediate by NADH at glucose C6', giving both the product dTDP-4-keto-6-deoxy-D-glucose and the regenerated NAD^+ coenzyme. The existence of the dTDP-4-keto-5,6-glucosene intermediate, as revealed by rapid mix-quench mass spectroscopy, and characterisation of the dehydration step of the mechanism from *E. coli* RmlB have provided further corroborative evidence for this mechanism (Gross *et al.*, 2000; Hegeman, 2001).

4.3. The co-crystallisation of S. typhimurium *with dTDP-D-glucose and dTDP*

Having already obtained good holoprotein crystals at 293K from a solution containing 4µl protein sample at 3.8mg ml^{-1} and 4µl precipitant (0.1M MES pH 6.3, 1.5M lithium sulphate), it was originally attempted to co-crystallise using these conditions. Unfortunately, there was no immediate success; new trials were therefore carried out using Hampton screens I and II. Before any trails were set up the protein (at concentrations between 3 and 5mg ml^{-1}) was incubated overnight at 277K with 2.5mM DTT and 4mM NAD^+ and either 10mM dTDP-D-glucose or 10mM dTDP. The Hampton screen trials were carried out using the sitting-drop vapour-diffusion technique (Ducruix and Giege, 1992), with 4µl sample being mixed with 4µl reservoir volume. For both the dTDP-D-glucose and dTDP co-crystallised protein, the same two conditions in the Hampton screens gave crystals. These were Hampton screen I, condition 41: 1.0M HEPES pH 7.5, 10% v/v iso-Propanol, 20% w/v PEG 4000, and Hampton screen II, condition 24: 0.05M Caesium chloride, 0.1M MES pH 6.5 and 30% v/v jeffamine M-600. Both conditions used protein at a concentration of 3mg ml^{-1}. Although the two conditions produced protein crystals of a size that were large enough to shoot, none of the crystals diffracted effectively and optimisation trials around these conditions gave no better results.

At the same time as setting up the Hampton screens, further co-crystallisation experiments were undertaken using conditions based around the condition that had been successfully used for the *S. typhimurium* holoprotein. To save on protein, the hanging-drop rather than sitting-drop vapour-diffusion technique was used. Having

seen nothing of interest in any of these trials for a month, these plates were left in the incubator at 293K undisturbed. After being left for just under 3 months, the plates were rescanned revealing two large crystals, one with protein co-crystallised with dTDP-D-glucose and the other with dTDP. For the dTDP-D-glucose RmlB crystal the condition was: 2μl protein at 5.32mg ml^{-1} mixed with 2μl mother liquor (0.1M MES pH 6.32 and 1.5M ammonium sulphate), with a well volume of 1ml mother liquor incubated at 293K (Figure 4.1).

Figure 4.1. S. typhimurium *RmlB co-crystallised with substrate, dTDP-D-glucose. The crystal is nearly 1mm in length.*

For the dTDP RmlB crystal the condition was slightly different in that the precipitant used was 1.5M lithium sulphate.

4.4. Data collection of S. typhimurium *with dTDP-D-glucose bound*

Data were collected from the single frozen crystal at Daresbury SRS, station 9.6 using a wavelength of 0.870Å with a ADSC Quantum 4 CCD detector. Cryoprotection of the crystal was achieved by bringing the crystal up through mother liquor containing increasing concentrations of glycerol (5, 10, 15 and 20%) with a 30s soak in each. The data were collected at 110K to a resolution of 1.8Å with 100% completeness. The crystal to detector distance was 180mm with data being recorded as 200 90 s 1° rotation exposures. Data were processed with the programs MOSFLM (Leslie, 1992) and merged and scaled using SCALA from the CCP4 program suite (Collaborative Computational Project, Number 4 (1994). A total of 2555543 measurements were recorded, with 428551 unique reflections. Unfortunately, because the data were collected at 1° rotation exposures, processing proved difficult due to a high percentage of overlaps. The data could be indexed in space group $P2_1$ but only with an unacceptable high amount of overlaps. In addition, there was a strong indication of hexagonal space group symmetry. After some experimentation, the data could be successfully indexed in spacegroup P3 and then scaled in spacegroup $P6_1$. This gave satisfactory results, and removed all but a small percentage of overlaps (<5%). Unfortunately, to achieve this the data had to be truncated from a reasonably high resolution of 1.8Å to 2.4Å (Table 4.1). The salutatory lesson learnt from this was to always process one or two frames of data at the beginning of a collection using the 'testgen' option in MOSFLM in order to determine the correct oscillation angle for data collection. When this was done post-collection, the angle that would have given less than 5% overlaps was 0.3° rotation exposures, something that would come in

useful when collecting the dTDP data. Table 4.2 gives a summary of the final data collection statistics.

Table 4.1. The 2.4Å data statistics for S. typhimurium *RmlB with dTDP-D-glucose bound, Daresbury SRS, station 9.6.*

Resolution (Å)	No. Reflections	Completeness (%)	†R_{merge} (%)	Redundancy
34.2 – 7.59	2024	99.9	9.3	5.9
7.59 – 5.37	3630	99.7	9.4	6.7
5.37 – 4.38	4672	99.8	9.2	6.8
4.38 – 3.79	5498	99.9	9.2	6.7
3.79 – 3.39	6232	100.0	9.4	6.4
3.39 – 3.10	6886	99.9	9.4	6.2
3.10 – 2.87	7461	99.7	9.8	6.1
2.87 – 2.68	7932	98.9	10.3	6.0
2.68 – 2.53	8438	98.6	10.7	5.9
2.53 – 2.40	8886	98.5	11.3	5.8
34.2 – 2.40	61659	99.3	9.5	6.2

† $R_{merge} = \sum_i [I_i - (I_i)] / \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

Table 4.2. Summary of x-ray data collection statistics for RmlB from S. typhimurium *with dTDP-D-glucose bound (values in parentheses refer to the highest resolution shell).*

	Daresbury SRS, station 9.6
Wavelength (Å)	0.870
Space group	$P6_1$
Resolution (Å)	2.4
Cell parameters	a=171.9Å, b=171.9Å, c=94.3Å $\alpha=\beta=90.0^{\circ}$, $\gamma=120^{\circ}$
V_M (two molecules per Asymmetric unit) ($Å^3$ Da^{-1})	4.93
Percentage solvent	74%
Total Measurements	382465 (51909)
Unique reflections	61659 (8886)
$\langle I/\sigma \rangle$	4.8 (8.0)
Redundancy	6.2 (5.8)
Data completeness (%)	99.3 (98.5)
R_{merge}† (%)	9.5 (11.3)

† $R_{merge} = \sum_i [I_i - (I_i)]/ \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

The solvent content of the crystals was determined using the Matthews equation. For one dimer in the asymmetric unit the V_M value was 4.93$Å^3$ Da^{-1} with a corresponding solvent content of 74%.

4.5. Structure determination of S. typhimurium *with dTDP-D-glucose bound*

Due to the new space group and cell dimensions, the structure of RmlB from *S. typhimurium* with dTDP-D-glucose had to be solved by molecular replacement using *AMoRe* (Navaza, 1994) as implemented in the CCP4 program suite. The phasing model used was the recently refined 2.47Å structure of *S. typhimurium* holoprotein (Allard *et al.*, 2001) (PDB ref: 1G1A). Using a dimer search model and data in the resolution range 10-3.3Å, one clear solution for one dimer in the asymmetric unit was obtained.

The cross rotation function gave ninety-nine solutions, one of which stood out: (i) α=15.76, β=88.55, γ=140.32 with a correlation coefficient of 10.8, equivalent to 11.6 standard deviations. Slight unease about the V_M 4.93Å^3 Da^{-1} meant that *AMoRe* was run a number of times with monomers and structural domains as the search models to see if any other solutions would be evident. However, despite numerous trials no other solutions were found.

Having found this one clear solution the translational search gave the following result (Table 4.3). At this point the space group was confirmed as $P6_1$.

Table 4.3. Solution from the translation function.

Solution	α	β	γ	x	y	z	c.c*	R_{factor}	#Height/σ
1	15.76	88.55	140.32	0.4569	0.1312	0.0000	48.6	43.6	38.2

**c.c refers to the correlation coefficient. # Refers to height of peak at this position divided by the r.m.s of the map.*

Following rigid body refinement in *AMoRe,* the final solution gave an R-factor and correlation coefficient of 41.8% and 55.1%, respectively (Table 4.4). The solution was applied to the dimer search model using MOLEMAN2 (Kleywegt, 1999) and the crystal packing checked using the program O (Jones *et al.*, 1991).

Table 4.4. Molecular replacement solution for one dimer following rigid body refinement.

Solution	α	β	γ	x	y	z	c.c*	R_{factor}
1	16.04	88.09	140.28	0.4575	0.1313	0.0000	55.1	41.8

**c.c refers to the correlation coefficient.*

4.6. Refinement of the 2.4Å structure of RmlB from S. typhimurium *with dTDP-D-glucose bound*

The refined 2.47Å data set of the *S typhimurium* holoprotein was used as the starting model against which to refine the data. The model was refined using *CNSsolve* (Brunger *et al.*, 1998) with a new random subset of data (5%) being omitted from all refinement calculations to provide an assessment of the refinement process (R_{free}). The model was initially put through the annealing component of *CNSsolve*, which reduced the R-factor to 34.6% and the R_{free} to 38.0%. SIGMAA (Read, 1986)

weighted $2F_o$-F_c and F_o-F_c electron density maps were generated and viewed using the program O (Jones *et al.*, 1991). The F_o-F_c map revealed density for both the NAD^+ and dTDP-D-glucose in the active site of each monomer (Figure 4.2).

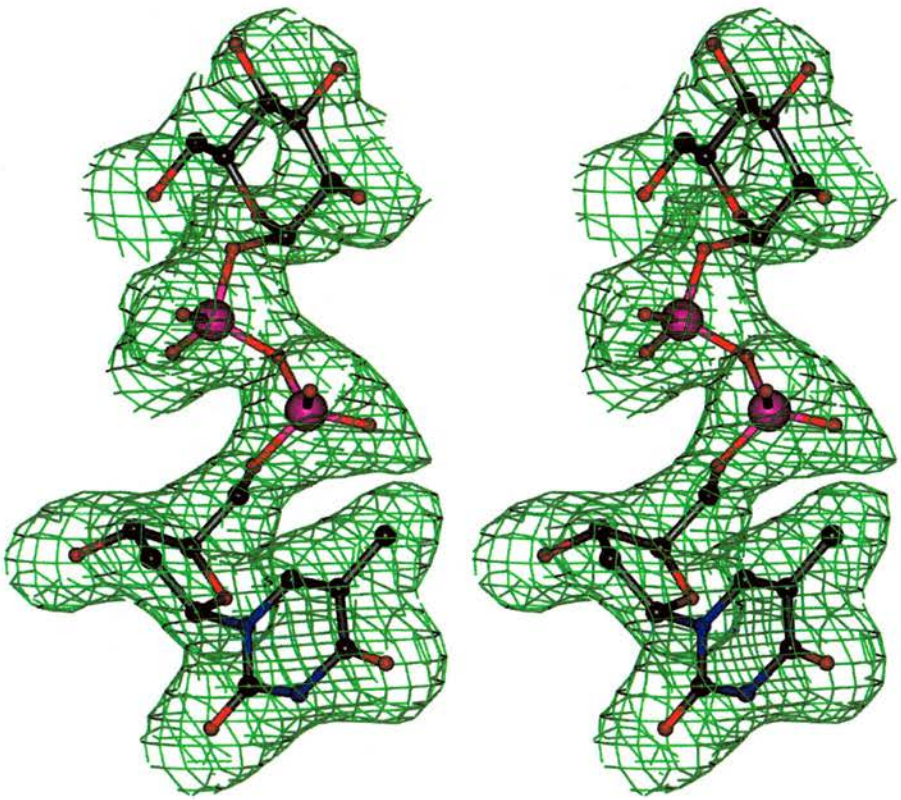


Figure 4.2. Stereo diagram showing the dTDP-D-glucose ligand bound in the active site of S. typhimurium *RmlB. The F_o-F_c map is contoured at 2.5σ.*

Originally the NAD^+ was added into the structure but with only dTDP in place of dTDP-D-glucose. This was done so as not to bias future electron density maps in favour of a pyranose ring. The model was further refined using alternate rounds of individual *B*-factor and restrained positional refinement. Positional non-crystallographic symmetry restraints were applied throughout the refinement process. This brought the R-factor down to 25.7%, and the R_{free} to 28.6%. At this point the $2F_o$-F_c and F_o-F_c electron density maps revealed strong density for residues 353-361 of the protein monomers (these residues had not been visible in the holoprotein structure), allowing all of the residues of RmlB to be built in confidently. In addition, clear movements in three areas of the substrate bound monomers were seen that needed to be rebuilt in O. The dTDP-D-glucose was now fully incorporated into the structure and waters were added by the water-pick procedure of *CNSsolve*. Further rounds of individual *B*-factor and restrained positional refinement followed including additional water picking. The model was continually checked with the program O and manual interventions made when required. Following the refinement procedure the model gave a final R-factor of 19.6% and an R_{free} of 21.9%. The final refinement statistics are summarised in Table 4.5.

Table 4.5. Refinement statistics for S. typhimurium *RmlB with dTDP-D-glucose bound.*

Resolution Range (Å)	34.175 – 2.40
Number of protein atoms in a.s.u	5758 (2 x 2879)
Number of water molecules	614
R-factor (working)[a] (%)	19.6
R_{free} (test)[b] (%)	21.9
R.m.s. deviations from ideal values:	
Bond lengths (Å)	0.008
Bond angles (deg)	1.45
Dihedrals (deg)	21.70
Improper dihedrals (deg)	1.08
Mean *B*-value (Å^2)	23
B-factor deviation bonds (Å^2)	1.20
B-factor deviation angles (Å^2)	2.00
Ramachandran:	
Residues in most favoured region (%)	92.0
Residues in allowed regions (%)	8.0

[a] R-factor = $\sum | F_{obs} - F_{calc} | / \sum | F_{obs} |$.
[b] 5% of reflection data.

4.7. Data collection of S. typhimurium *with dTDP bound*

Data were collected from the single frozen crystal at Daresbury SRS, station 14.2 using a wavelength of 0.978Å with an ADSC Quantum 4 CCD detector. As with all the *S. typhimurium* crystals to date, cryoprotection of the crystal was achieved by bringing the crystal up through increasing concentrations of glycerol (5, 10, 15 and 20%) with a 30s soak in each. The data were collected at 110K to a resolution of 1.8Å with 100% completeness. The crystal to detector distance was 140mm with data being recorded as 230 non-overlapping 15 s 0.3° rotation exposures to reduce spot overlaps. Data were processed with the programs MOSFLM (Leslie, 1992) and merged and scaled using SCALA from the CCP4 program suite. A total of 549500 measurements were recorded, with 140877 unique reflections (Table 4.6). The RmlB crystals showed clear hexagonal symmetry and were assigned to space group $P6_1$. A summary of the data collection statistics is given in Table 4.7.

Table 4.6. 1.8Å data statistics for S. typhimurium *RmlB with dTDP bound, Daresbury SRS, station 14.2.*

Resolution (Å)	No. Reflections	Completeness (%)	†R_{merge} (%)	Redundancy
43.4 – 5.69	4519	98.3	4.1	3.6
5.69 – 4.02	8239	100.0	3.9	3.9
4.02 – 3.29	10607	100.0	4.1	3.9
3.29 – 2.85	12531	100.0	4.6	4.0
2.85 – 2.55	14159	100.0	5.4	4.0
2.55 – 2.32	15667	100.0	6.3	4.0
2.32 – 2.15	16956	100.0	7.7	4.0
2.15 – 2.01	18256	100.0	9.7	4.0
2.01 – 1.90	19430	100.0	13.1	3.9
1.90 – 1.80	20513	100.0	18.5	3.7
43.4 – 1.80	140877	99.9	5.3	3.9

† $R_{merge} = \sum_i [I_i - (I_i)] / \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

Table 4.7. Summary of x-ray data collection statistics for RmlB from S. typhimurium *with dTDP bound (values in parentheses refer to the highest resolution shell).*

	Daresbury SRS, station 14.2
Wavelength (Å)	0.978
Space group	$P6_1$
Resolution (Å)	1.8
Cell parameters	a=169.9Å, b=169.9Å, c=92.8Å $\alpha=\beta=90.0^{\circ}$, $\gamma=120^{\circ}$
V_M (two molecules per Asymmetric unit) ($Å^3\ Da^{-1}$)	4.74
Percentage solvent	73%
Total Measurements	549500 (76465)
Unique reflections	140877 (20513)
$\langle I/\sigma \rangle$	8.7 (4.1)
Redundancy	3.9 (3.7)
Data completeness (%)	99.9 (100.0)
R_{merge}† (%)	5.3 (18.5)

† $R_{merge} = \sum_i [I_i - (I_i)]/ \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

4.8. Structure determination of S. typhimurium *with dTDP bound*

Despite the dTDP data having the same space group and almost the same cell dimensions as the dTDP-D-glucose structure of *S. typhimurium*, it proved necessary to solve the structure by molecular replacement using *AMoRe* (Navaza, 1994). The 2.4Å dTDP-D-glucose homodimer structure of *S. typhimurium* was used as the phasing model giving one solution for one dimer in the asymmetric unit.

The cross rotation function gave the solution: (i) α=44.65, β=91.86, γ=317.15 with a correlation coefficient of 12.9, equivalent to 18.5 standard deviations. When put into the translation search the following solution was obtained (Table 4.8).

Table 4.8. Solution from the translation function.

Solution	α	β	γ	x	y	z	c.c*	R_{factor}	#Height/σ
1	44.65	91.86	317.15	0.4602	0.3295	0.0000	63.1	36.4	36.7

**c.c refers to the correlation coefficient. # Refers to height of peak at this position divided by the r.m.s of the map.*

Following rigid body refinement in *AMoRe* the final solution gave an R-factor and correlation coefficient of 32.5% and 70.1%, respectively (Table 4.9). The solution was applied to the dimer search model using MOLEMAN2 (Kleywegt, 1999) and the crystal packing checked using the program O (Jones *et al.*, 1991).

Table 4.9. Molecular replacement solution following rigid body refinement.

Solution	α	β	γ	x	y	z	c.c*	R_{factor}
1	44.49	91.81	317.03	0.4583	0.3268	0.0000	70.1	32.5

**c.c refers to the correlation coefficient.*

4.9. Refinement of the 1.8Å structure of RmlB from S. typhimurium *with dTDP bound*

The refined 2.4Å data set of the *S typhimurium* dTDP-D-glucose structure was used as the starting model against which to refine the data. The same R_{free} flag was kept from the *S typhimurium* dTDP-D-glucose structure and extended to 1.8Å. As before, the model was refined using *CNSsolve* (Brunger *et al.*, 1998). The model was initially put through the annealing component of *CNSsolve*, which reduced the R-factor to 29.2%, and the R_{free} to 30.5%. SIGMAA (Read, 1986) weighted $2F_o$-F_c and F_o-F_c electron density maps were generated and viewed using the program O (Jones *et al.*, 1991). The F_o-F_c map revealed strong density for both the NAD^+ and dTDP at 3.0σ in the active site of each monomer (Figure 4.3). Both the NAD^+ and dTDP were added into the structure. The model was further refined using alternate rounds of individual *B*-factor and restrained positional refinement. Positional non-crystallographic symmetry restraints were applied throughout the refinement process. This brought the R-factor down to 26.5% and the R_{free} to 27.7%. As with the dTDP-D-glucose structure, the same structural movements as compared to the holoprotein were observed in O, indicating that it is the binding of the dTDP part of the substrate that induces the conformational movements of the enzyme monomers. Rounds of water-picking, individual *B*-factor and restrained positional refinement gave a final R-factor of 20.1% and an R_{free} of 21.6%. The final refinement statistics are summarised in Table 4.10.

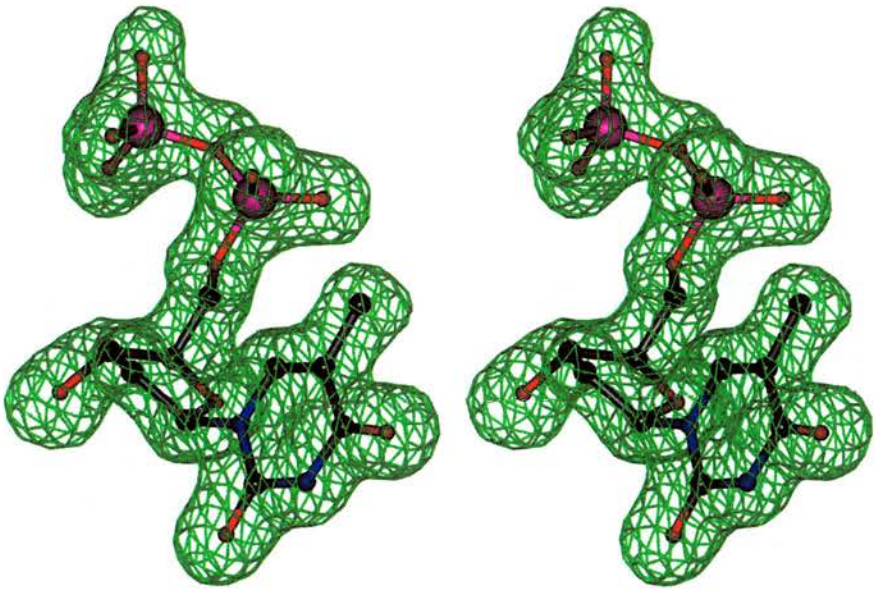


Figure 4.3. Stereo diagram showing the dTDP ligand bound in the active site of S. typhimurium *RmlB. The* F_o-F_c *map is contoured at* 3.0σ.

Table 4.10. Refinement statistics for S. typhimurium *RmlB with dTDP bound.*

Resolution Range (Å)	43.37 – 1.80
Number of protein atoms in a.s.u	5758 (2 x 2879)
Number of water molecules	702
R-factor (working)[a] (%)	20.1
R_{free} (test)[b] (%)	21.6
R.m.s. deviations from ideal values:	
Bond lengths (Å)	0.008
Bond angles (deg)	1.43
Dihedrals (deg)	22.20
Improper dihedrals (deg)	0.89
Mean *B*-value ($Å^2$)	24
B-factor deviation bonds ($Å^2$)	1.30
B-factor deviation angles ($Å^2$)	2.00
Ramachandran:	
Residues in most favoured region (%)	91.7
Residues in allowed regions (%)	8.3

[a] R-factor = $\sum | F_{obs} - F_{calc} | / \sum | F_{obs} |$.
[b] 5% of reflection data.

4.10. Analysis of the final S. typhimurium *models*

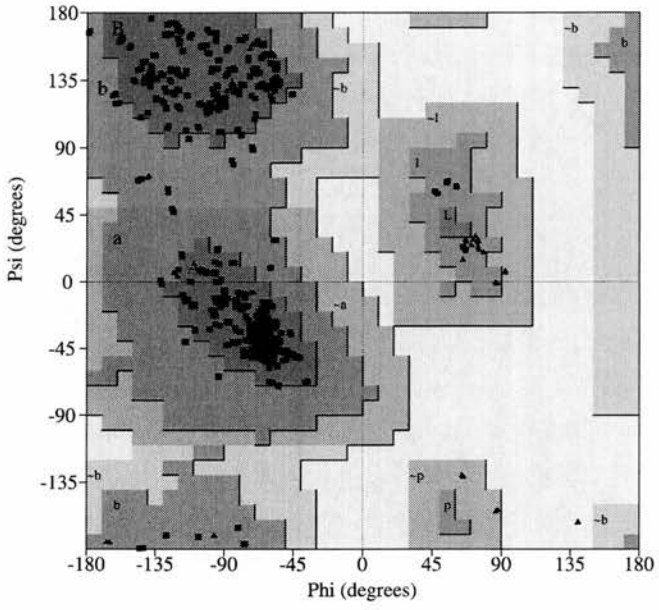
The final two structures both consisted of one homodimer in the asymmetric unit, each comprising two identical subunits. The dTDP-D-glucose structure contained 2 molecules of NAD^+, 2 molecules of dTDP-D-glucose and 614 solvent molecules. The dTDP structure contained 2 molecules of NAD^+, 2 molecules of dTDP, 3 molecules

of glycerol and 702 solvent molecules. Both structures had all 722 residues built in with 5758 non-hydrogen atoms. The final stereochemistry of the model was checked by PROCHECK (Laskowski *et al.*, 1993). WHATIF (Vriend, 1990) was used to check the best orientation of O* and N* atoms in residues such as Asn, Arg, Asp, Gln and His. VOIDOO (Kleywegt and Jones, 1994) and *CNSsolve* were used to calculate protein surface and volumes. Hydrogen bonds and salt bridges were analysed using *CNSSolve* and the Protein-Protein Interactions Server (www.biochem.ucl.ac.uk/bsm/PP/server).

The overall geometry of the models was good, with rms deviations from ideal geometry of 0.008Å and 1.449° for bond lengths and bond angles, respectively (Engh and Huber, 1991) for the dTDP-D-glucose model, and 0.008Å and 1.431° for bond lengths and bond angles for the dTDP model.

Ramachandran plots generated using PROCHECK showed that both structures had more than 90% of residues lying within the most favoured regions with no residues in disallowed regions (Ramakrishnan and Ramachandran, 1965), Figure 4.4. The average temperature factors of the refined models for all atoms were 23Å^2 and 24Å^2 for the dTDP-D-glucose and dTDP models, respectively. The mean errors in the atomic coordinates estimated from a Luzzati plot (Luzzati, 1952) were 0.26Å (working set) and 0.30Å (test set) for the dTDP-D-glucose model and 0.21Å (working set) and 0.23Å (test set) for the dTDP model.

(a)



(b)

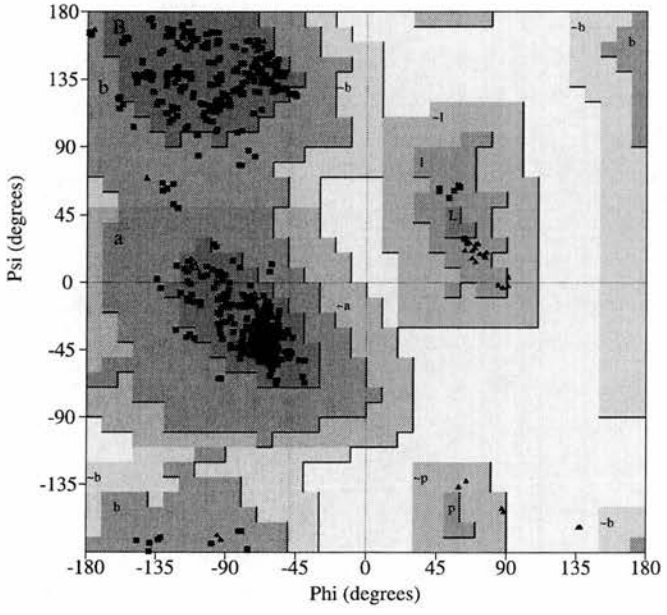


Figure 4.4. (a) Ramachandran plot for the 2.4Å structure of S. typhimurium *RmlB with dTDP-D-glucose. (b) Ramachandran plot for the 1.8Å structure of* S. typhimurium *RmlB with dTDP. Glycine residues (not restricted to any particular region of the plot) are represented by a ▲, non-glycine residues by a ■. The darker the shaded region, the more favourable the φ, ψ combination.*

4.11. The co-crystallisation of S. suis *with dTDP-D-glucose, dTDP and dTDP-xylose*

S. suis RmlB protein was incubated overnight at 277K with 2.5mM DTT, 4mM NAD^+ and either 10mM dTDP-D-glucose, 10mM dTDP or 7mM dTDP-α-D-pyrano-xylose (dTDP-xylose, kindly supplied by Dr A. Hegeman from The University of Wisconsin-Madison). I would like to again acknowledge the help of undergraduate student Konstantinos Beis in the laboratory for his work on *S. suis* RmlB. Initial crystallisation trials were carried out with the same conditions as used to obtain the holoprotein crystals; hanging drop method with protein at a concentration of 2.3mg ml^{-1} and a mother liquor containing: 0.3M ammonium acetate, 0.1M *tri*-sodium citrate dihydrate pH 5.4, 35% w/v PEG 4000 and the additive 3% w/v 1,6-hexanediol. Initial results gave promising small needles for the dTDP and dTDP-xylose co-complexes. After some optimisation, reasonable needles and some cubic crystals were obtained for the protein co-crystallised with dTDP or dTDP-xylose using the following conditions: 0.3M ammonium sulphate, 0.1M *tri*-sodium citrate dihydrate pH 5.4, 35% w/v PEG 4000, protein at 2.3mg ml^{-1} and an incubation temperature of 293.5K (Figure 4.5). In order to try and obtain crystals with dTDP-D-glucose bound, Hampton screens I and II were set up. Using protein at 2.3mg ml^{-1}, nice needle crystals were obtained in Hampton screen I condition 15: 0.2M ammonium sulphate, 0.1M sodium cacodylate pH 6.5, and 30% w/v PEG 8000 and an incubation temperature of 293.5K. Slight modification of this condition (pH 6.4 rather than 6.5) and using the hanging drop method for crystal growth (2μl protein mixed with 2μl mother liquor) gave reasonable diffracting needles.

Figure 4.5. Crystals of S. suis *RmlB co-crystallised with dTDP-xylose.*

4.12. Data collection of S. suis *with dTDP-D-glucose bound*

Data were collected from a single frozen crystal at Daresbury SRS, station 9.6 using a wavelength of 0.870Å with an ADSC Quantum 4 CCD detector (Table 4.11). Cryoprotection was achieved by soaking the crystal in mother-liquor containing 15% glycerol for 5 minutes. The data were collected at 110K to a resolution of 2.2Å with 99.8% completeness. The crystal to detector distance was 190mm with data being recorded as 200 90 s 0.5° rotation exposures. Data were processed with the programs MOSFLM (Leslie, 1992) and merged and scaled using SCALA from the CCP4 program suite. A total of 179806 measurements were recorded, with 55517 unique reflections. Table 4.12 gives a summary of the final data collection statistics.

Table 4.11. 2.2Å data statistics for S. suis *RmlB with dTDP-D-glucose bound, station 9.6, Daresbury SRS.*

Resolution (Å)	No. Reflections	Completeness (%)	†R_{merge} (%)	Redundancy
50.00 – 6.96	1912	98.8	6.8	2.7
6.96 – 4.92	3333	99.5	6.4	3.1
4.92 – 4.02	4216	99.7	6.2	3.2
4.02 – 3.48	4980	99.9	7.1	3.2
3.48 – 3.11	5586	99.9	9.1	3.3
3.11 – 2.84	6143	99.7	12.1	3.3
2.84 – 2.63	6640	99.6	15.4	3.3
2.63 – 2.46	7119	99.8	18.5	3.3
2.46 – 2.32	7588	99.8	23.5	3.3
2.32 – 2.20	8000	99.9	28.4	3.3
50.0 – 2.20	55517	99.8	11.2	3.2

† $R_{merge} = \sum_i [I_i - (I_i)] / \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

Table 4.12. Summary of x-ray data collection statistics for S. suis *RmlB with dTDP-D-glucose bound (values in parentheses refer to the highest resolution shell).*

	Daresbury SRS, station 9.6
Wavelength (Å)	0.870
Space group	$P2_12_12_1$
Resolution (Å)	2.20
Cell parameters	a=61.2Å, b=96.2Å, c=182.9Å $\alpha=\beta=\gamma=90^{o}$
V_M (four molecules per Asymmetric unit) ($Å^3$ Da^{-1})	3.50
Percentage solvent	63%
Total Measurements	179806 (26125)
Unique reflections	55517 (8000)
$\langle I/\sigma \rangle$	5.7 (2.6)
Redundancy	3.2 (3.3)
Data completeness (%)	99.8 (99.9)
R_{merge}† (%)	11.2 (28.4)

† $R_{merge} = \sum_i [I_i - (I_i)]/ \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

4.13. Data collection of S. suis *with dTDP bound*

Data were collected from a single frozen crystal at Daresbury SRS, station 9.6 using a wavelength of 0.870Å with an ADSC Quantum 4 CCD detector, as for the dTDP-D-glucose complex. The data were collected at 110K to a resolution of 1.8Å with 97.7% completeness (Table 4.13). The crystal to detector distance was 180mm with data being recorded as 180 90 s 0.5° rotation exposures. Data were processed with the programs MOSFLM (Leslie, 1992) and merged and scaled using SCALA from the CCP4 program suite. A total of 345981 measurements were recorded, with 104163 unique reflections. Table 4.14 gives a summary of the final data collection statistics.

Table 4.13. 1.8Å data statistics for S. suis *RmlB with dTDP bound, station 9.6, Daresbury SRS.*

Resolution (Å)	No. Reflections	Completeness (%)	†R_{merge} (%)	Redundancy
28.30 – 5.69	3652	99.6	5.6	3.5
5.69 – 4.02	6358	100.0	5.2	3.8
4.02 – 3.29	8086	100.0	5.3	3.9
3.29 – 2.85	9505	100.0	6.3	3.9
2.85 – 2.55	10749	100.0	7.6	3.7
2.55 – 2.32	11746	99.5	8.9	3.4
2.32 – 2.15	12554	98.4	10.3	3.2
2.15 – 2.01	13306	97.1	12.9	3.1
2.01 – 1.90	13967	95.8	17.7	3.0
1.90 – 1.80	14240	92.7	25.1	2.7
22.36 – 1.80	104163	97.7	7.3	3.3

† $R_{merge} = \sum_i [I_i - (I_i)] / \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

Table 4.14. Summary of x-ray data collection statistics for S. suis *RmlB with dTDP bound (values in parentheses refer to the highest resolution shell).*

	Daresbury, station 9.6
Wavelength (Å)	0.870
Space group	$P2_12_12_1$
Resolution (Å)	1.8
Cell parameters	a=62.3Å, b=99.0Å, c=184.6Å $\alpha=\beta=\gamma=90^o$
V_M (four molecules per Asymmetric unit) ($Å^3$ Da^{-1})	3.50
Percentage solvent	63%
Total Measurements	345981 (38348)
Unique reflections	104163 (14240)
$\langle I/\sigma \rangle$	6.9 (2.9)
Redundancy	3.3 (2.7)
Data completeness (%)	97.7 (92.7)
R_{merge}† (%)	7.3 (25.1)

† $R_{merge} = \sum_i [I_i - (I_i)]/ \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

4.14. Data collection of S. suis *with dTDP-xylose bound*

Data were collected from a single frozen crystal at Daresbury SRS, station 9.6 using a wavelength of 0.870Å with an ADSC Quantum 4 CCD detector, as for the dTDP-D-glucose complex. The data were collected at 110K to a resolution of 1.8Å with 97.0% completeness (Table 4.15). The crystal to detector distance was 180mm with data being recorded as 180 90 s 0.5° rotation exposures. Data were processed with the

programs MOSFLM (Leslie, 1992) and merged and scaled using SCALA from the CCP4 program suite. A total of 340418 measurements were recorded, with 102080 unique reflections. Table 4.16 gives a summary of the final data collection statistics.

Table 4.15. 1.8Å data statistics for S. suis *RmlB with dTDP-xylose bound, station 9.6, Daresbury SRS.*

Resolution (Å)	No. Reflections	Completeness (%)	†R_{merge} (%)	Redundancy
38.35 – 5.69	3617	99.8	5.6	3.3
5.69 – 4.02	6296	100.0	5.1	3.6
4.02 – 3.29	7984	100.0	5.4	3.7
3.29 – 2.85	9401	100.0	6.6	3.7
2.85 – 2.55	10621	100.0	8.4	3.7
2.55 – 2.32	11634	100.0	9.8	3.6
2.32 – 2.15	12601	99.6	11.8	3.4
2.15 – 2.01	13307	98.4	15.6	3.3
2.01 – 1.90	13797	95.9	21.6	3.1
1.90 – 1.80	12822	85.1	29.1	2.3
38.35 – 1.80	102080	97.0	7.9	3.3

† $R_{merge} = \sum_i [I_i - (I_i)] / \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

Table 4.16. Summary of x-ray data collection statistics for S. suis *RmlB with dTDP-xylose bound (values in parentheses refer to the highest resolution shell).*

	Daresbury, station 9.6
Wavelength (Å)	0.870
Space group	$P2_12_12_1$
Resolution (Å)	1.8
Cell parameters	a=62.4Å, b=97.2Å, c=185.5Å $\alpha=\beta=\gamma=90^o$
V_M (four molecules per Asymmetric unit) ($Å^3$ Da^{-1})	3.50
Percentage solvent	63%
Total Measurements	340418 (30104)
Unique reflections	102080 (12822)
$\langle I/\sigma \rangle$	6.9 (2.5)
Redundancy	3.3 (2.3)
Data completeness (%)	97.0 (85.1)
R_{merge}† (%)	7.9 (29.1)

† $R_{merge} = \sum_i [I_i - (I_i)]/ \sum_i (I_i)$, where I_i is the scaled intensity of the ith observation and (I) is the mean intensity for that reflection.

4.15. Refinement of the S. suis *RmlB complexes*

All the model complexes were isomorphous with the *S. suis* holoprotein structure. The first model to be refined was the 1.8Å data set of *S. suis* with dTDP. The structure of the 2.9Å *S. suis* holoprotein was used as the starting model against which to refine the data. The same R_{free} flag was kept from the *S. suis* holoprotein structure and extended to 1.8Å. The model was refined using *CNSsolve* (Brunger *et al.*, 1998).

The model was initially put through the annealing component of *CNSsolve*, which reduced the R-factor to 27.7%, and the R_{free} to 29.4% (the starting R-factor and R_{free} were 38.8% and 38.9%, respectively). SIGMAA (Read, 1986) weighted $2F_o$-F_c and F_o-F_c electron density maps were generated and viewed using the program O (Jones *et al.*, 1991). The F_o-F_c map revealed clear density for both the NAD^+ and dTDP in the active site of each monomer (Figure 4.6). Both the NAD^+ and dTDP were added into the model. Further refinement continued using alternate rounds of individual *B*-factor and restrained positional refinement and eventual addition of water molecules. Positional non-crystallographic symmetry restraints were applied throughout the refinement process. The final model has an R-factor of 19.7% and R_{free} of 21.7%; the refinement statistics are summarised in Table 4.17.

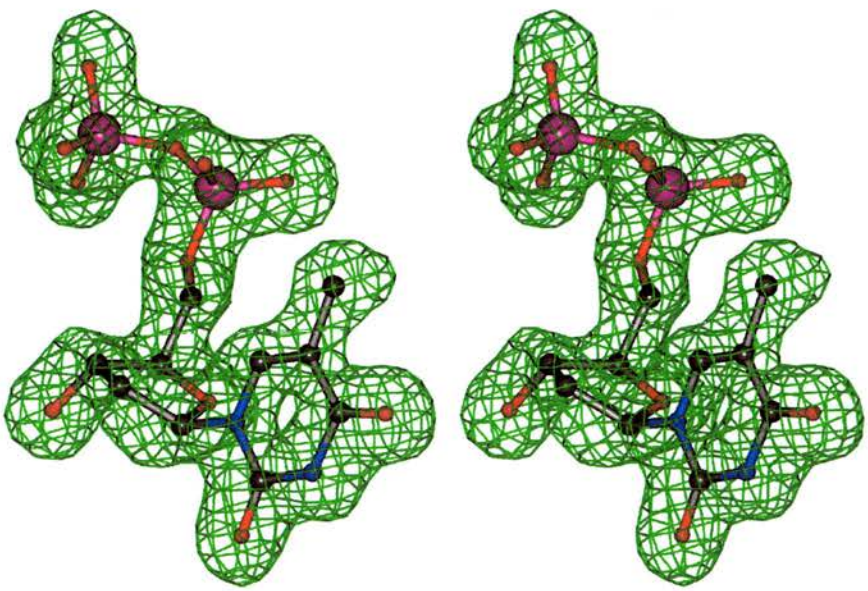


Figure 4.6. Stereo diagram showing the dTDP ligand bound in the active site of S. suis *RmlB. The F_o-F_c map is contoured at 3.0σ.*

Table 4.17. Refinement statistics for S. suis *RmlB with dTDP bound.*

Resolution Range (Å)	28.30 – 1.80
Number of protein atoms in a.s.u	5479 (2741+2738)
Number of waters	585
R-factor (working)[a] (%)	19.7
R_{free} (test)[b] (%)	21.7
R.m.s. deviations from ideal values:	
Bond lengths (Å)	0.008
Bond angles (deg)	1.25
Dihedrals (deg)	22.41
Improper dihedrals (deg)	0.822
Mean *B*-value ($Å^2$)	18
B-factor deviation bonds ($Å^2$)	1.00
B-factor deviation angles ($Å^2$)	1.45
Ramachandran:	
Residues in most favoured region (%)	89.8
Residues in allowed regions (%)	10.2

[a] R-factor = $\sum | F_{obs} - F_{calc} | / \sum | F_{obs} |$.
[b] 5% of reflection data.

The 2.2Å data set of *S. suis* with dTDP-D-glucose was refined using the dTDP 1.8Å *S. suis* structure as the starting model. *CNSsolve* (Brunger *et al.*, 1998) was used for refinement with the same R_{free} flag being kept from the *S. suis* dTDP structure. The model was put straight into the energy minimisation component of *CNSsolve*, which reduced the R-factor to 26.9%, and the R_{free} to 28.8% (the starting R-factor and R_{free} were 46.1% and 46.9%, respectively). SIGMAA (Read, 1986) weighted $2F_o$-F_c and F_o-F_c electron density maps were generated and viewed using the program O (Jones *et al.*, 1991). The F_o-F_c map revealed clear density for both the NAD^+ and dTDP-D-glucose in the active site of each monomer (Figure 4.7). Both the NAD^+ and dTDP-D-glucose were added into the model at full occupancy. Further refinement proceeded smoothly using alternate rounds of individual *B*-factor and restrained positional refinement and addition of water molecules. Positional non-crystallographic symmetry restraints were applied throughout the refinement process. The final model had an R-factor of 17.5%, and R_{free} of 20.6%; the refinement statistics are summarised in Table 4.18.

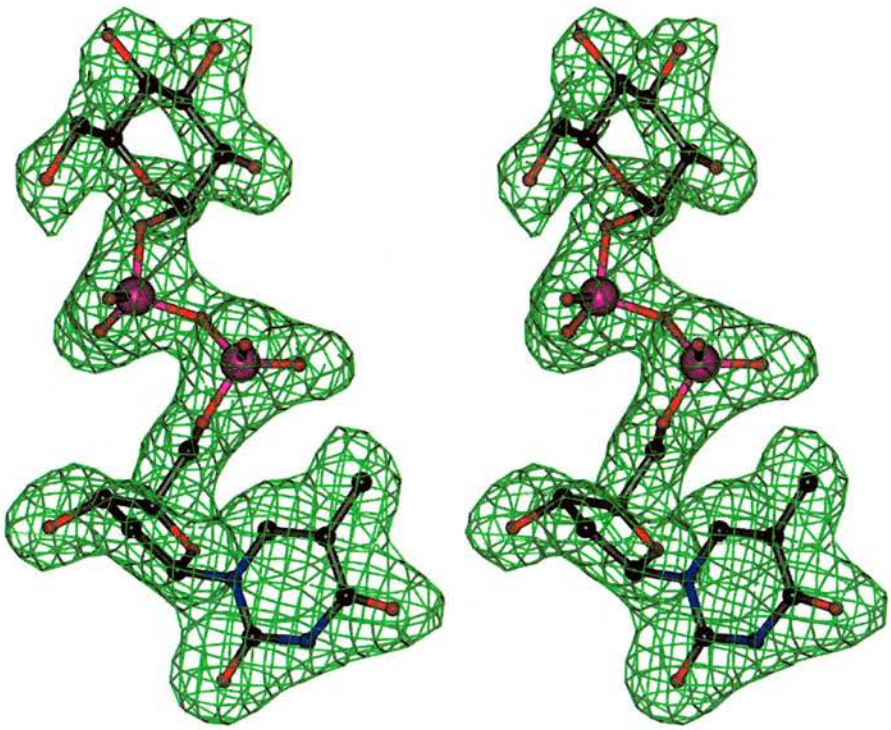


Figure 4.7. Stereo diagram showing the dTDP-D-glucose ligand bound in the active site of S. suis *RmlB. The F_o-F_c map is contoured at 3.0σ.*

Table 4.18. Refinement statistics for S. suis *RmlB with dTDP-D-glucose bound.*

Resolution Range (Å)	50.0 – 2.20
Number of protein atoms in a.s.u	5488 (2741+2747)
Number of waters	713
R-factor (working)[a] (%)	17.5
R_{free} (test)[b] (%)	20.6
R.m.s. deviations from ideal values:	
Bond lengths (Å)	0.005
Bond angles (deg)	1.22
Dihedrals (deg)	22.15
Improper dihedrals (deg)	0.784
Mean *B*-value ($Å^2$)	17
B-factor deviation bonds ($Å^2$)	1.00
B-factor deviation angles ($Å^2$)	1.45
Ramachandran:	
Residues in most favoured region (%)	90.2
Residues in allowed regions (%)	9.8

[a] R-factor = $\sum | F_{obs} - F_{calc} | / \sum | F_{obs} |$.
[b] 5% of reflection data.

The 1.8Å data set of *S. suis* with dTDP-xylose was refined using the dTDP-D-glucose 2.2Å *S. suis* structure as the starting model. The same R_{free} flag was kept from the *S. suis* dTDP structure, however the program REFMAC (Murshudov *et al.*, 1999) was used to refine the structure. The use of TLS (trans/liberation/screw-rotation) and restrained individual *B*-factor refinement in REFMAC quickly reduced the R-factors. The generated SIGMAA (Read, 1986) weighted $2F_o$-F_c and F_o-F_c electron density

maps showed clear density for both the dTDP-xylose and NADH in the active sites when viewed using the program O (Jones *et al.*, 1991); the ligands were added into the structure at full occupancy (Figure 4.8). Arp/warp was used to improve the model's quality further by adding water molecules (Perrakis *et al.*, 1999). The final model had an R-factor of 14.5% and R_{free} of 17.2%; the refinement statistics are summarised in Table 4.19.

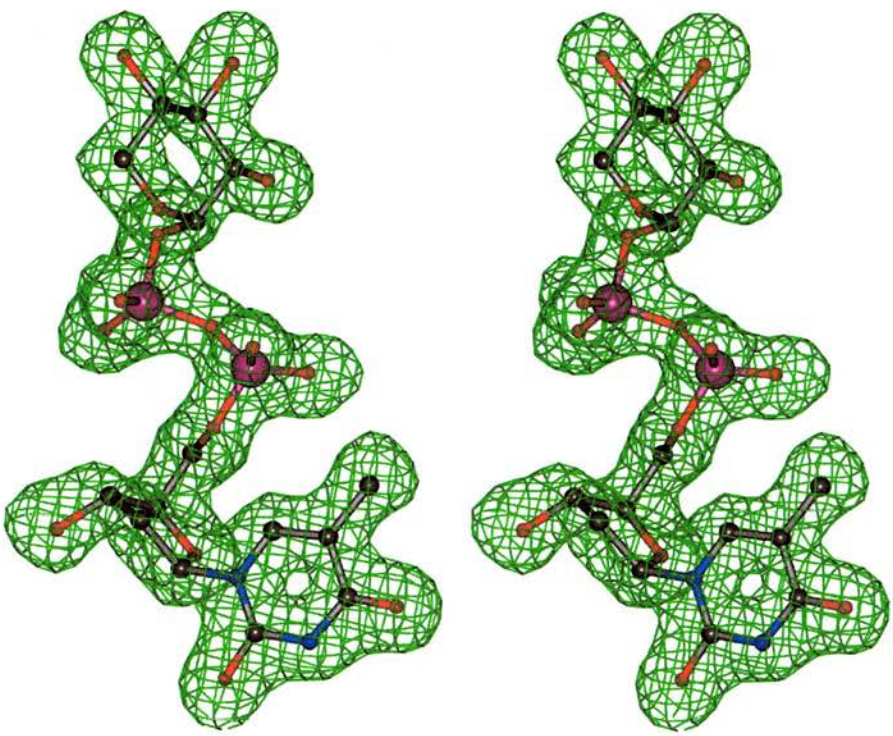


Figure 4.8. Stereo diagram showing the dTDP-xylose ligand bound in the active site of S. suis *RmlB. The F_o-F_c map is contoured at 3.0σ.*

Table 4.19. Refinement statistics for S. suis *RmlB with dTDP-xylose bound.*

Resolution Range (Å)	38.35 – 1.80
Number of protein atoms in a.s.u	5479 (2741+2738)
Number of waters	804
R-factor (working)[a] (%)	14.5
R_{free} (test)[b] (%)	17.2
R.m.s. deviations from ideal values:	
Bond lengths (Å)	0.007
Bond angles (deg)	1.28
Dihedrals (deg)	22.47
Improper dihedrals (deg)	0.840
Mean *B*-value (Å^2)	17
B-factor deviation bonds (Å^2)	0.90
B-factor deviation angles (Å^2)	1.30
Ramachandran:	
Residues in most favoured region (%)	91.3
Residues in allowed regions (%)	8.7

[a] R-factor = $\sum | F_{obs} - F_{calc} | / \sum | F_{obs} |$.
[b] 5% of reflection data.

4.16. Analysis of the final S. suis *models*

All the *S.suis* structures consist of one homodimer in the asymmetric unit, each comprising two identical subunits. The dTDP structure contains 2 molecules of NAD^+, 2 molecules of dTDP and 585 solvent molecules. The dTDP-D-glucose structure contains 2 molecules of NAD^+, 2 molecules of dTDP-D-glucose, 3 sulphate molecules and 713 solvent molecules. The dTDP-xylose structure contains 2 molecules of NAD^+, 2 molecules of dTDP-xylose, 3 sulphate molecules and 804 solvent molecules. The F_o-F_c density maps at 3σ for all the enzyme complex ligands were clearly interpretable, enabling all ligands to be built in confidently. All $2F_o$-F_c experimental density maps were of high quality at 1σ. However, for the *S. suis* dTDP and dTDP-xylose complexes, the first two residues in monomer A and the last and first residue of monomer B were unable to be built in due to diffuse electron density. The *S. suis* dTDP-D-glucose complex had all but the first two residues of monomer A and the first residue of monomer B built in.

The quality and stereochemistry of the models was checked using the same programs as used for the *S. typhimurium* complexes. The overall geometry of the models was good, with rms deviations of the models from ideal geometry of 0.008Å and 1.254° for bond lengths and bond angles, respectively (Engh and Huber, 1991) for the dTDP model, 0.005Å and 1.220° for the dTDP-D-glucose model and 0.007Å and 1.284° for the dTDP-xylose model.

Ramachandran plots generated using PROCHECK showed that all the structures had more than 90% of residues lying within the most favoured regions with no residues in

disallowed regions (Ramakrishnan and Ramachandran, 1965) (Figure 4.9). The mean errors in the atomic coordinates estimated from a Luzzati plot (Luzzati, 1952) were 0.20Å (working set) and 0.22Å (test set) for the dTDP model, 0.20Å (working set) and 0.24Å (test set) for the dTDP-D-glucose model and 0.19Å (working set) and 0.21Å (test set) for the dTDP-xylose model.

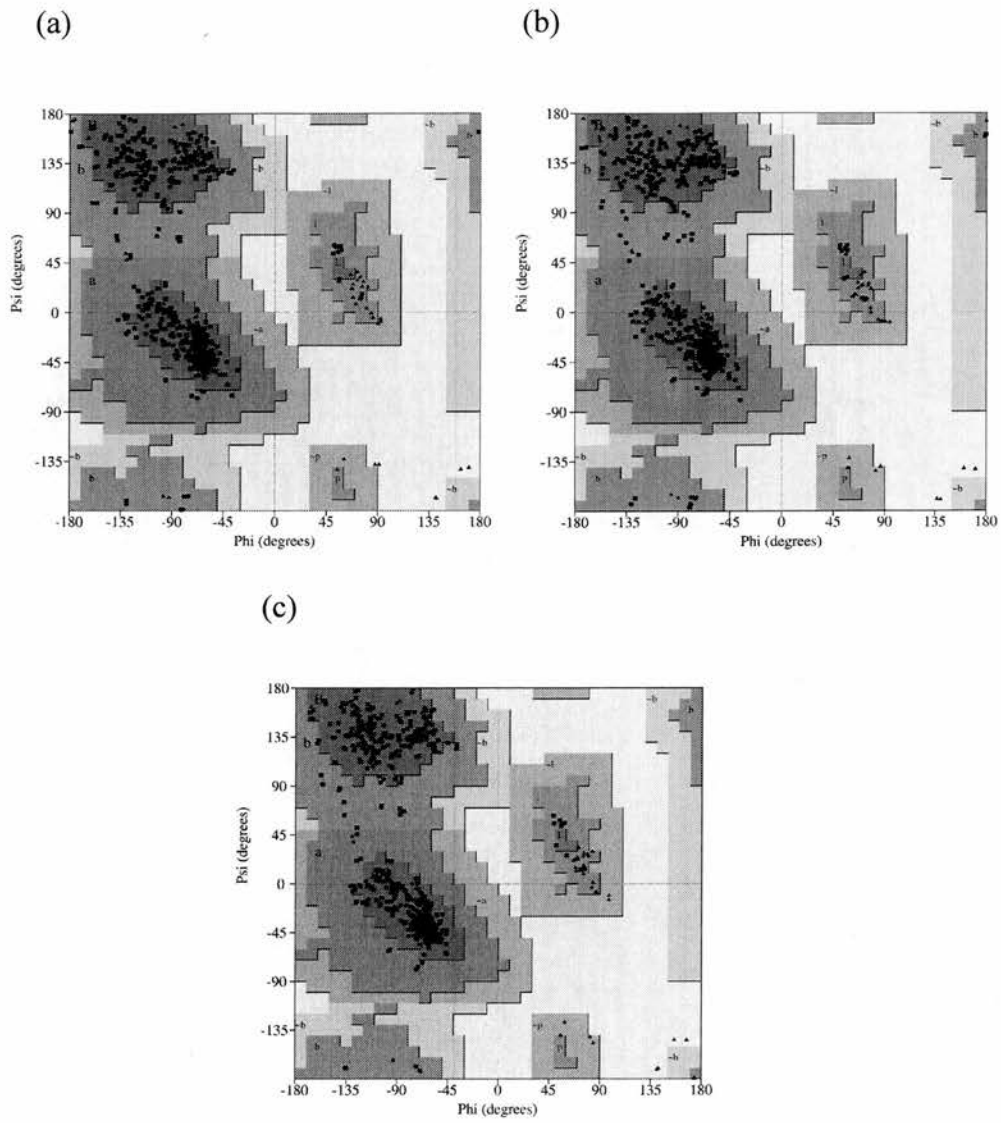


Figure 4.9. (a) Ramachandran plot for the 1.8Å structure of S. suis *RmlB with dTDP, (b) Ramachandran plot for the 2.2Å structure of* S. suis *RmlB with dTDP-D-glucose, and (c) Ramachandran plot for the 1.8Å structure of* S. suis *RmlB with dTDP-xylose. Glycine residues (not restricted to any particular region of the plot) are represented by a ▲, non-glycine residues by a ■. The darker the shaded region, the more favourable the φ, ψ combination.*

4.17. Results and discussion

Conformational aspects of the enzyme complexes

In chapter 3, the structure of the RmlB holoprotein was discussed. In essence, the RmlB monomer can be divided into two domains for functional purposes; an N-terminal NAD^+ binding domain containing the characteristic Rossmann fold and a C-terminal sugar-nucleotide binding domain (Figure 4.10) (the homodimer with substrate bound is shown in Figure 4.11) (Allard *et al.*, 2001). The fact that the substrate dTDP-D-glucose was originally modelled into the active site of *S. typhimurium* RmlB left some uncertainty as to what structural changes may occur when the enzyme actually binds substrate or substrate analogue in the active site.

The question of whether it is substrate or product bound in the dTDP-D-glucose complexes is interesting. The density is not totally clear as to whether the substrate has been turned over, and although the density for the NAD^+ is planar suggesting it is in the oxidised form, the possibility that the substrate has been turned over cannot be ruled out. However, having obtained good complex structures, it was immediately obvious that there were major rearrangements of secondary structural elements in RmlB on binding substrate/substrate analogue or just dTDP.

(a)

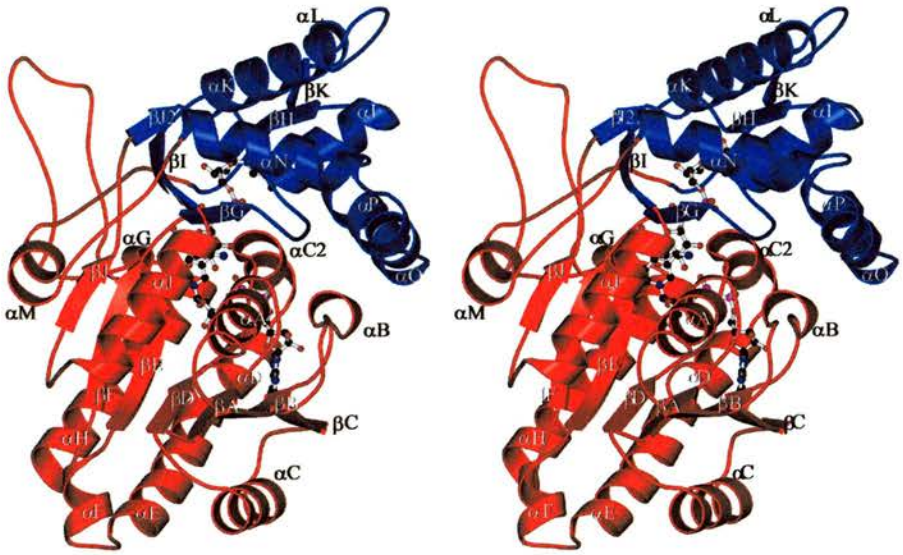


Figure 4.10. A stereo ribbon representation of the S. typhimurium *RmlB monomer with the coenzyme NAD^+ and substrate (dTDP-D-glucose) shown bound in ball and stick representation. The secondary structural elements are labelled as per the holoenzyme structure in chapter 3 proceeding from the N to C terminus. The N-terminal and C-terminal domains are colour coded in red and blue, respectively.*

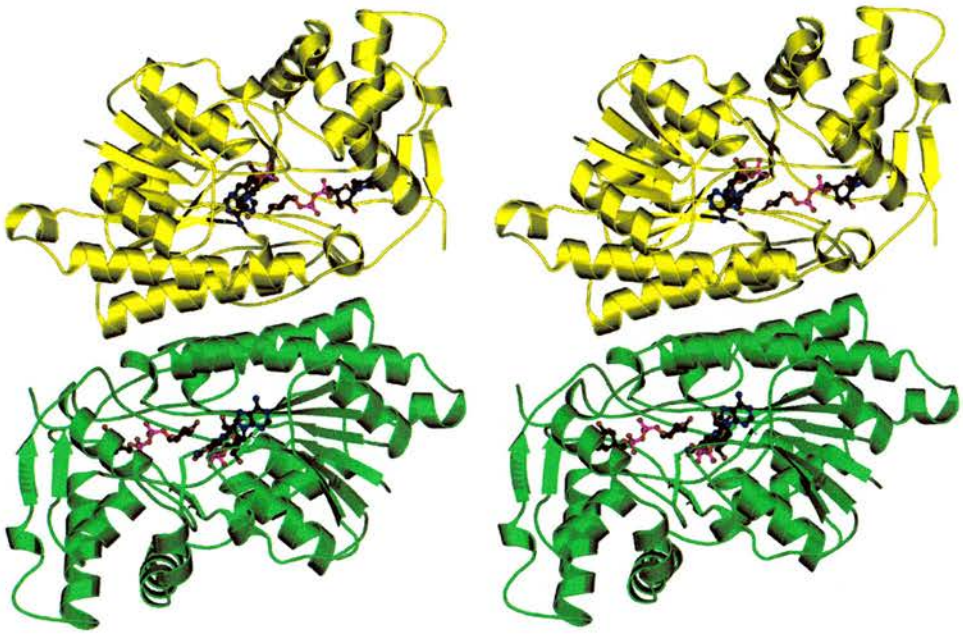


Figure 4.11. Stereo ribbon representation of the homodimer from S. typhimurium. *Each monomer is shown in a different colour, with the coenzyme* NAD^+ *and substrate (dTDP-D-glucose) shown bound in ball and stick representation.*

To understand what gross changes occur in the enzyme architecture on binding a ligand in the active site, it was useful to compare the complex structures with that of the holoprotein (only NAD^+ bound) structure of *S. typhimurium* RmlB. The co-crystallised structures show that it is the binding of the thymidine nucleotide, whether on its own as dTDP or as part of the sugar nucleotide dTDP-D-glucose, in the active site that is responsible for the conformational changes that take place. This is similar to GALE where the binding of uridine nucleotide in the epimerase active site induces a protein conformational change (Bertland and Kalckar, 1968; Frey, 1996; Kang *et al.*, 1975). These major areas of structural movement that take place in RmlB on binding dTDP are mainly confined to the C-terminal region (Figure 4.12) and occur in both monomers. On binding dTDP these structural changes result in an rmsd of 1.4Å between the resting and dTDP bound *S. typhimurium* enzyme (the same rmsd value is obtained between the dTDP-D-glucose bound and resting enzyme). However, the rmsd between the dTDP and dTDP-D-glucose bound *S. typhimurium* enzyme is only 0.25Å, which is the same value obtained between the dTDP and dTDP-D-glucose bound *S. suis* enzyme. The similarity between the dTDP and dTDP-xylose bound monomers of *S suis* is even closer giving an rmsd of only 0.12Å.

Importantly, the complexes from *S. suis* have the same spatial arrangement of their domains as those from *S. typhimurium*. However, as the *S. suis* structure with NAD^+ bound almost certainly contains dTDP bound in the active site (see Figure 3.9), the following discussion will focus principally on the enzyme from *S. typhimurium*.

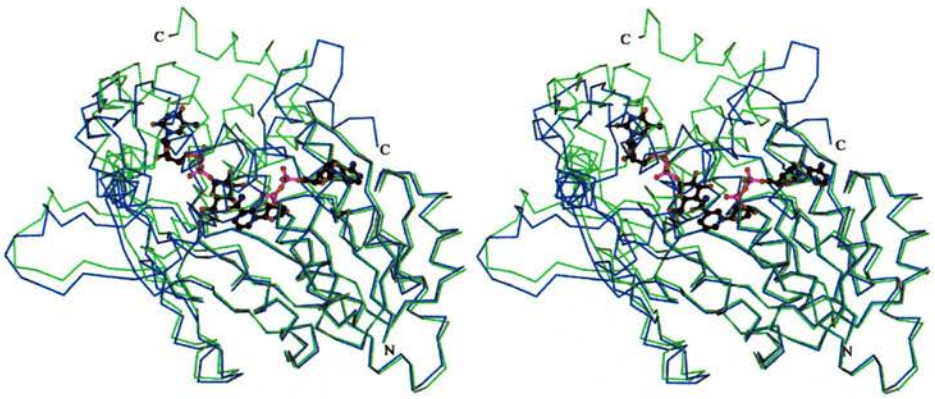


Figure 4.12. The stereo superposition of monomers of S. typhimurium *in the resting state (blue, NAD^+ bound) and with NAD^+ and substrate (dTDP-D-glucose) bound (green). The N and C-terminals are labelled.*

The dTDP structure shows that residues in the region His85 to Gly92 become ordered from an unstructured loop in the holoprotein into a new helix, αC2 (Figure 4.10). This helix becomes part of the active site pocket, with residue Ser86 hydrogen bonding to the O3 of the NAD^+ diphosphate group. The high Cα rmsd values for the individual residues in helix αC2 illustrate the large movement that has occurred, ranging from 4.7Å for His85 to 14.8Å for Ile90, the average rmsd being 9.6Å for the helix. The second main conformational movement occurs in the section constituting residues Glu290 to His300 where the RmlB enzyme consists of a small β-strand (βK, residues 291-294) and part of a loop. On binding dTDP the β-strand moves in towards the thymidine ring of the dTDP. This allows His300 (His287 in *S. suis*) to hydrogen bond to the 3'-hydroxyl group of the ribose sugar of dTDP, thereby controlling the selection for the deoxy-ribonucleotide sugar substrate in the binding site by using its imidazole side chain to sterically exclude oxy-ribonucleotide sugars (Figures 4.13 and 4.14). When modelling dTDP-D-glucose into the RmlB active site as discussed in chapter 3, it had originally been suggested that Asn266 might select

for the deoxy-ribonucleotide substrate. However, with these substrate complexes, it is now clear that this is not the case, although the side chain does hydrogen bond to the 3'-hydroxyl group of the ribose sugar.

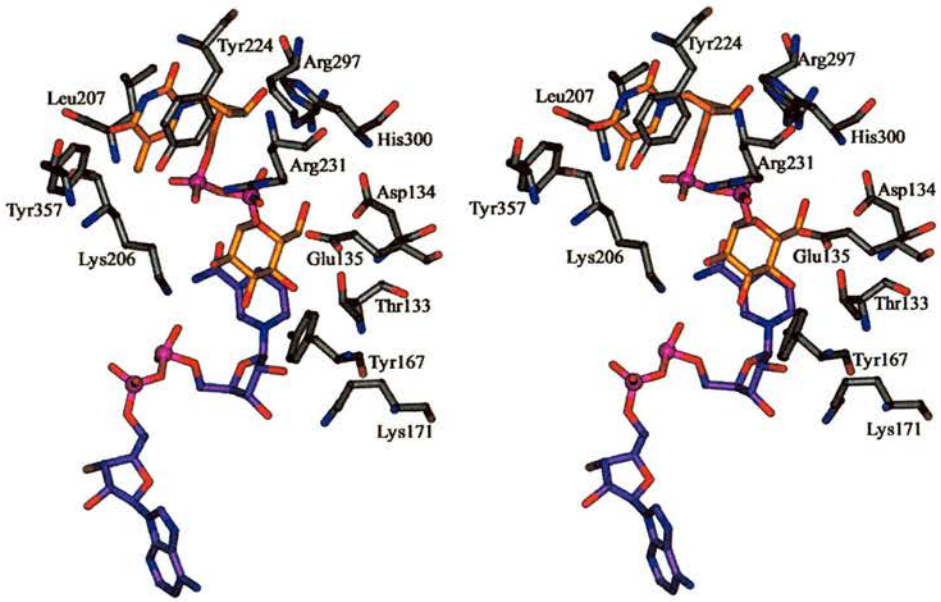


Figure 4.13. Stereo view of the S. typhimurium *active site showing important residues discussed in the text. The coenzyme NAD^+ is coloured orange and substrate dTDP-D-glucose purple. All bonds are coloured according to atom type.*

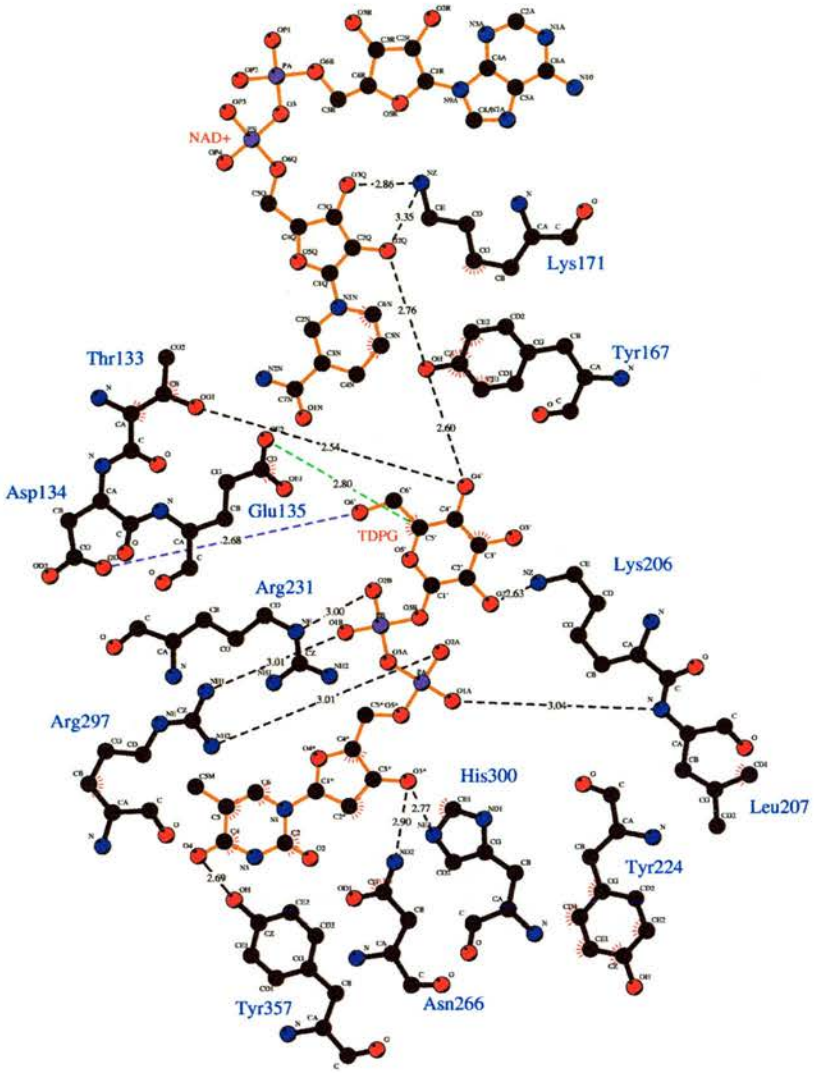


Figure 4.14. Schematic representation of some of the important residues involved in the binding of dTDP-D-glucose (TDPG) and NAD$^+$ in the active site of S. typhimurium *as generated by LIGPLOT (Wallace* et al., *1995). Hydrogen bonds are indicated as black broken lines and other important interactions are shown as green (Glu135 Oε2 to glucosyl-C5') and purple (Asp134 Oδ1 to glucosyl-C6' hydroxyl group) broken lines. Red shading around atoms indicates involvement in hydrophobic contacts. Atoms are colour coded as follows: black, carbon; blue, nitrogen; and red, oxygen. Ligand bonds are shown in orange.*

The importance of selecting for the deoxy-ribose sugar is demonstrated by incubating *E. coli* RmlB with the oxy-ribose sugar UDP-glucose as opposed to its normal substrate dTDP-D-glucose. This leads to a lowering of the second-order rate constant by 6 orders of magnitude and a 200-fold reduction of k_{cat} (Hegeman *et al.*, 2001). These results demonstrate the sensitivity of the reaction to the proper alignment of substrate, which would be perturbed by the ribosyl 2'-hydroxyl group.

The third gross change occurs in the C-terminal residues 347 to 357 of *S. typhimurium*, which forms an extended α-helix (αP) (*S. suis* residues 330 to 342, helix αQ); this helix caps the active site (Figure 4.10a). These residues are disordered and absent from the resting structure.

The active site of RmlB with dTDP-D-glucose bound

With the sugar substrate in the active site, a number of important hydrogen bonds are formed (Figure 4.13). In *S. typhimurium* the side chains of Thr133, Tyr167 and Lys206 hydrogen bond to the hexapyranose ring. The Oγ1 of Thr133 forms a short hydrogen bond to the glucosyl C4'-hydroxyl (2.5Å). The Tyr167 Oη also hydrogen bonds to the glucosyl C4'-hydroxyl group (2.6Å) and Lys206 Nζ hydrogen bonds to the glucosyl C2'-hydroxyl group (2.6Å). The side chains of residues Arg231 and Arg297 hydrogen bond to the phosphoryl oxygens along with a peptidic hydrogen bond supplied by Leu207. In both monomers of *S. typhimurium*, Tyr224 moves to form a stacking interaction with the thymidine ring of the substrate. This is also seen in *S. suis* RmlB, HGALE and SQD1 where the residues Tyr218, Phe226 and Tyr256 fulfil the same roles by stacking with the thymidine and uracil rings, respectively (Mulichak *et al.*, 1999; Thoden *et al.*, 2000). At the same time in RmlB, the active

site lysine (Lys171 in *S. typhimurium* and Lys165 in *S. suis*) hydrogen bonds to the 3'-hydroxyl group of the nicotinamide ribose moiety in the *S. suis* complexes, and to both the 2' and 3'-hydroxyl groups in the *S. typhimurium* complexes. The active site tyrosine (Tyr167 *S. typhimurium* and Tyr161 *S. suis*) also hydrogen bonds to the 2'-hydroxyl group of the nicotinamide ribose molecule and comes under the influence of the positive electrostatic field created by both the ε-ammonium ion of the lysine and the positively charged nicotinamide, which in turn lowers the pK_a of the tyrosine. Measurement of the pK_a of the corresponding tyrosine in *E. coli* RmlB (Tyr160) gives a value of 6.41 (Gerratana *et al.*, 2001). Importantly, the active site tyrosine in the complexes now forms hydrogen bonds to the respective glucosyl-C4' hydroxyl or xylosyl-C4' hydroxyl groups (2.6Å), which along with its lowered pK_a, enables it to act directly as the active site base.

S. suis *with dTDP-xylose bound*

Spectrophotometric analysis of the *S. suis* RmlB crystals with dTDP-xylose bound revealed the coenzyme to be in the reduced state NADH (Figure 4.15). This was suggested by the slightly buckled nicotinamide ring when refining the structure. However, the structure also suggests that it is dTDP-xylose in the active site and not dTDP-4-ketoxylose, indicating partial turnover of the dTDP-xylose. Additionally, a water molecule (S191) occupies the position where the glucosyl-C6' hydroxyl group would be if the substrate dTDP-D-glucose was bound in the active site (Figure 4.16). This demonstrates that the active site is designed around the catalysis of the glucose ring, of which more is discussed in the next chapter.

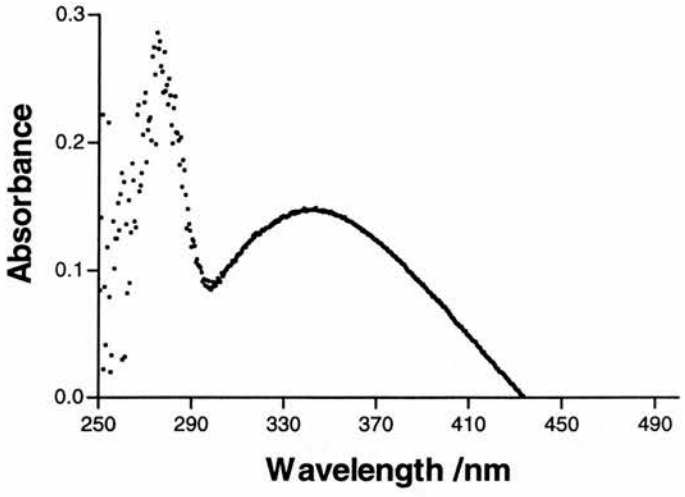


Figure 4.15. Spectrophotometric analysis of S. suis *RmlB crystals with dTDP-xylose bound. Crystals containing dTDP-xylose and NADH were mounted in capillaries and examined using a single crystal fibre optic photodiode array microspectrophotometer with a xenon arc light source at room temperature. The spectra clearly demonstrate peaks at 280nm and 340nm corresponding to the absorption maxima of the protein and NADH thus confirming the presence of NADH in the protein crystals. I would like to acknowledge Dr. R. Wilmouth at the University of Oxford for helping me with the spectrophotometric analysis of RmlB.*

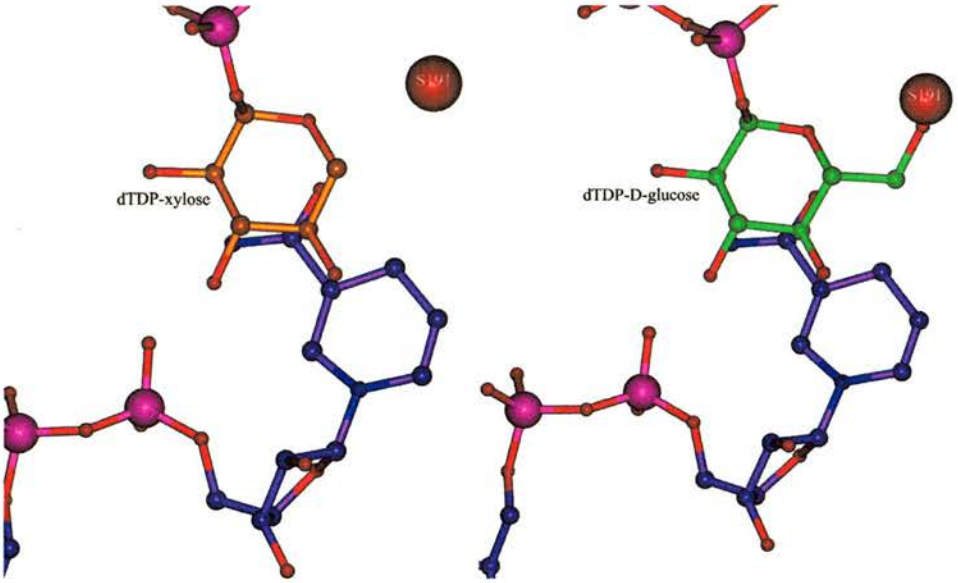


Figure 4.16. The binding of dTDP-xylose (left, orange) and dTDP-D-glucose (right, green) in the active site of S. suis *RmlB. The position of water S191 has been superpositioned into the dTDP-D-glucose active site (right). The coenzyme (NAD(H) is shown in purple.*

4.18. Conclusions

The co-crystallised complexes of RmlB reveal there to be a number of conformational changes when the enzyme binds a nucleotide such as dTDP in its active site. These structural movements result in the C-terminal of the enzyme becoming ordered and the protein having a much lower temperature factor. In addition, when the substrate is bound in the active site important active site residues, such as the catalytic tyrosine and an aspartic and glutamic acid pairing, move into positions that allow them to carry out the catalytic mechanism of the enzyme.

Chapter five

A molecular understanding of the mechanism of dehydratase enzymes

5.1. Summary

The conformational changes that occur in RmlB on binding sugar-nucleotide or just nucleotide in the active site have important implications as to how the enzyme carries out catalysis. The conserved tyrosine moves allowing it to act directly as the active site base. A highly conserved aspartic and glutamic acid pairing carries out the dehydration step of the mechanism. This pairing is unusual in that although the aspartic and glutamic acid do not contact each other directly, the structure shows that the aspartic acid residue must be protonated at pH 7.0. A comparison of these structures with other related structures has established that the trajectory for hydride transfer is much more specific than previously thought. It seems clear that the substrate is required to move within the active site to complete the catalytic cycle and that this movement is driven by the elimination of water. The structural information reveals the molecular basis of carbohydrate dehydration. The results have implications in the study of all members of the SDR superfamily.

5.2. Introduction

RmlB is an example of an enzyme that dehydrates a nucleotide-sugar, a necessary event in the synthesis of deoxysugars. The biogenesis of 6-deoxysugars requires the action of an NDP-hexose 4,6-dehydratase (dTDP-glucose 4,6-dehydratase, CDP-glucose 4,6-dehydratase, GDP-mannose 4,6-dehydratase or UDP-glucose 4,6-dehydratase; (Liu and Thorson, 1994)). The formation of NDP-6-deoxy-4-hexuloses such as dTDP-4-keto-6-deoxy-D-glucose (dTDP-6-deoxy-D-*xylo*-hexulose), have been the subject of intensive research for more than three decades (Kirschning *et al.*, 1997). The acknowledged mechanism (Figure 5.1) involves an initial oxidation at glucosyl-C4' by a tightly bound NAD^+ molecule. The 4-dehydrogenation of the substrate is a critical activating step for β-elimination of water between C5 and C6. Without this activation by the adjacent ketone functionality, deprotonation at C5 (pKa >40) would be impossible under any conceivable conditions in an enzymatic active site. The 4-keto functionality acidifies the C5 proton bringing its pK_a into the high teens (March, 1985) and also creates additional possibilities for the catalytic mechanism of dehydration. Finally, the hydride is transferred back from NADH to the C6 position of the pyranose ring, resulting in formation of the product dTDP-4-keto-6-deoxy-D-glucose and regenerating the NAD^+ coenzyme.

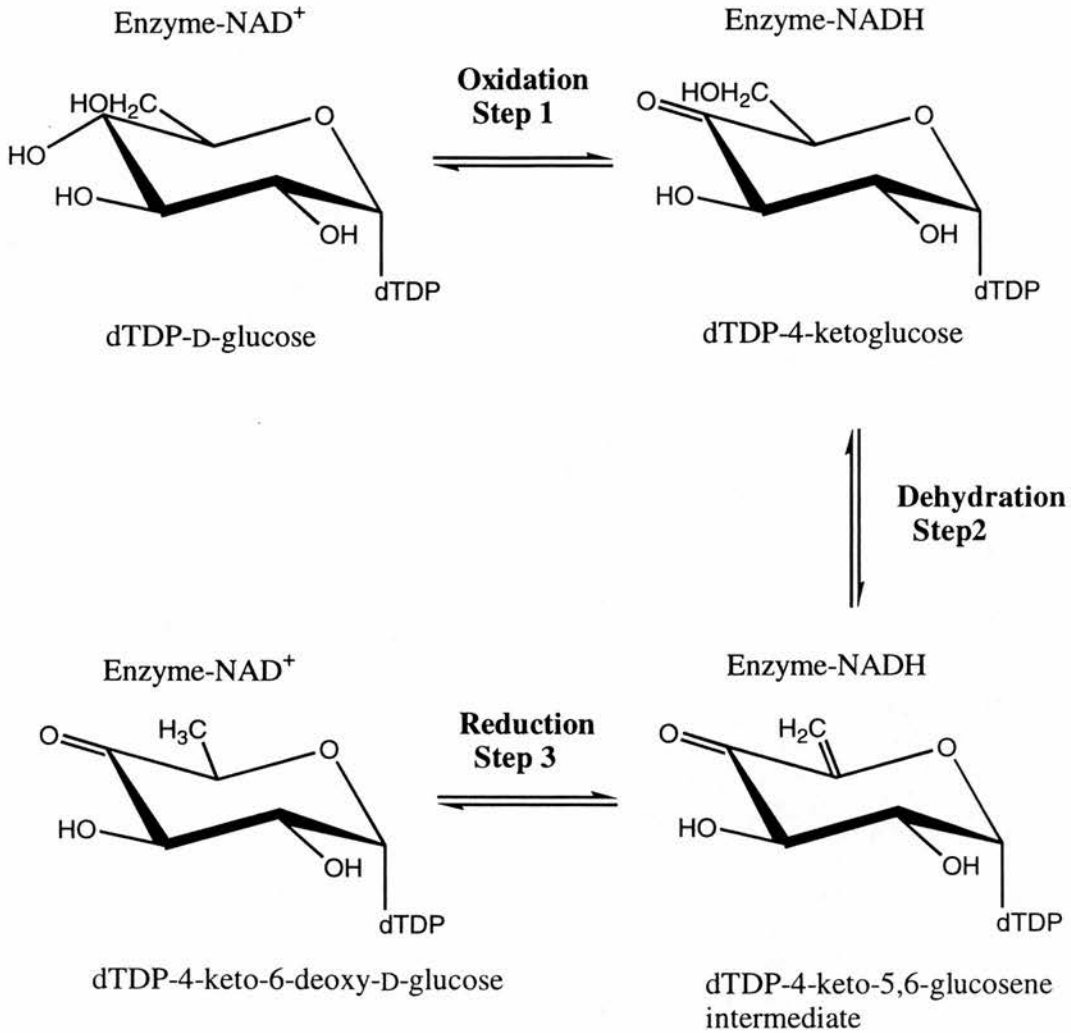
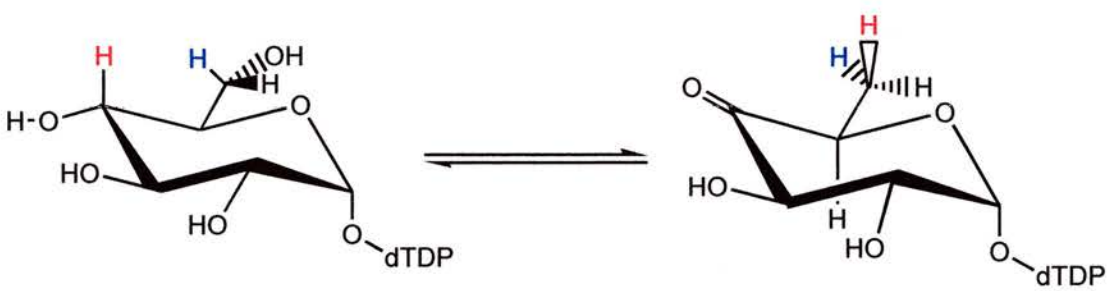


Figure 5.1. The three step reaction catalysed by RmlB.

Many labelling studies have supported this sequence of steps. Perhaps one of the most important being the demonstration that a deuterium or tritium label initially incorporated at the glucosyl-C4' position of the substrate was found in the C6' position of the product (Gabriel and Lindquist, 1968; Melo *et al.*, 1968). This confirms that a hydride is temporarily transferred to the NAD^+ coenzyme and redelivered to an alternate position in the product. Further labelling studies have also shown that the hydride transfer steps concerning the nicotinamide coenzyme involve the pro-S (B-side) position of NADH, and that the net transfer of hydride from C4 to C6 is exclusively intramolecular (Zarkowsky *et al.*, 1970). Finally, Snipes *et al.* (1977) demonstrated the stereochemical course of the reaction using dTDP-[4-^{2}H, 6-^{3}H]-glucose as the substrate. Configurational analysis of the chiral acetic acid, obtained by oxidation of the resulting 6-deoxysugar nucleotide, confirmed that the hydride shift from C4 to C6 occurs intramolecularly and showed that the migrating hydrogen species replaces the 6-hydroxyl group inversion of configuration (Figure 5.2).



dTDP-[4-^{2}H, 6-^{3}H]-glucose dTDP-4-keto-6-deoxy-D-glucose

Figure 5.2. Demonstration that a deuterium label at glucosyl-C4' (red) migrates intramolecularly to C6 of the glucosyl moiety and that replacement of the hydroxyl group at C6 by the hydrogen from C4 occurs with inversion of configuration at C6 (Snipes et al., *1977)*

5.3. The RmlB mechanism

The C6'-dehydration catalysed by RmlB is necessary for the biosynthesis of all 6-deoxyhexoses. The mechanism is a three-step process involving oxidation, dehydration and reduction and was originally proposed in 1963 (Glaser, 1963). In the RmlB-substrate complexes the substrate and coenzyme are aligned for catalysis as shown in Figure 5.3. The positions of the key catalytic residues are shown in Figure 5.4.

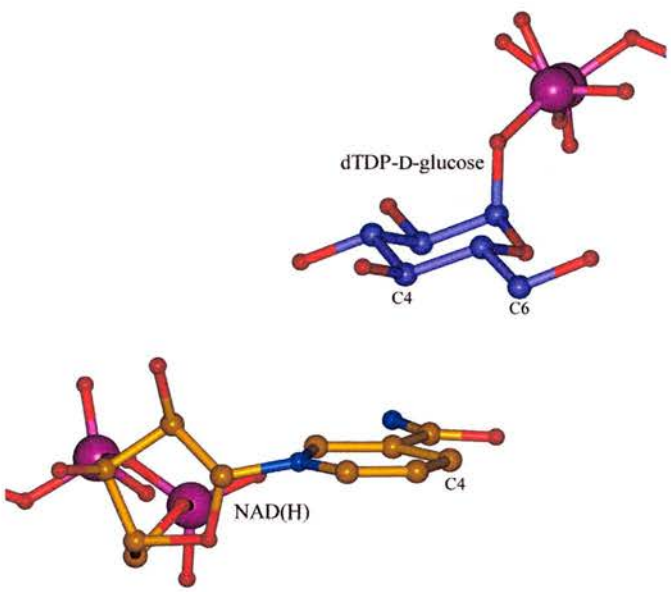


Figure 5.3. The alignment of the nicotinamide (orange) and pyranose (purple) rings of the coenzyme and substrate for hydride transfer in the S. typhimurium *dTDP-D-glucose complex.*

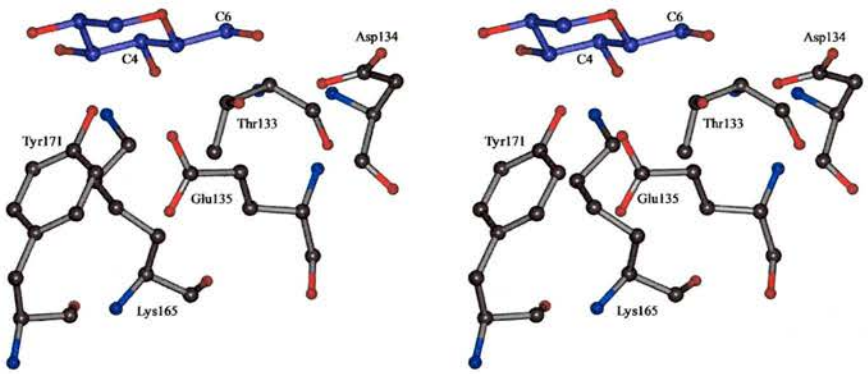


Figure 5.4. Stereo diagram showing the key residues from S. typhimurium *RmlB involved in catalysis. For clarity, just the glucose ring of the substrate (purple) is shown.*

Step one - oxidation

For the oxidation step to take place, an active site base is required to deprotonate the glucosyl-C4' hydroxyl and the NAD^+ must be in a position to accept a hydride from glucosyl-C4'.

As discussed in chapter four, movement of the RmlB active site on binding substrate brings the active site tyrosine into hydrogen bonding distance with the glucosyl-C4' hydroxyl (2.6Å), which along with its lowered pK_a enables it to act directly as the active site base (Figure 5.4). Without this movement at the active site the tyrosine would be too far from the C4'-hydroxyl group to act as a base; a conundrum first highlighted by Thoden *et al.* (1996a) with their studies on EGALE. To solve this problem, the active site serine was proposed to act as a proton shuttle. However, in 1999 Mulichak *et al.* observed that the active site tyrosine in SQD1 makes a close hydrogen bond (2.5Å) to the glucosyl C4'-hydroxyl group of both monomers of the enzyme homodimer. In addition, Thoden *et al.* (2000) showed that in HGALE the

active site tyrosine from one of the monomers was within hydrogen bonding distance of glucosyl-C4' hydroxyl (3.1Å) as compared to 4.1Å for the other monomer (Thoden *et al.*, 2000). This led both groups to propose that the active site tyrosine might act directly as the active site base. This work definitely establishes that the tyrosine acts as the base and also shows that nucleotide binding controls the conformational changes that enable it to do so.

The positioning of the nicotinamide ring for the oxidation step is also crucial. The enzyme is a class B dehydrogenase (Wang and Gabriel, 1970), in that the nicotinamide ring of the coenzyme adopts the *syn* conformation, presenting the *si* face for pro-S hydride transfer from glucosyl-C4' of the substrate to C4' of the dinucleotide cofactor. On binding substrate, the nicotinamide C4 atom of NAD^+ is positioned 3.5Å from the glucosyl-C4' in the *S. typhimurium* dTDP-D-glucose complex (Figure 5.3), 3.7Å from the glucosyl-C4' in the *S. suis* dTDP-D-glucose complex and 3.7Å from the xylosyl-C4' in the dTDP-xylose complex. Importantly, the angle that the glucosyl-C4' makes with the C4-N1 atoms of the planar nicotinamide ring is between 78° and 88° for the substrate and dTDP-xylose complexes, showing that the C4 of the nicotinamide ring is positioned almost directly above the C4 or C6 of the substrate for linear hydride transfer. This is very similar to the structures of the enzymes SQD1 (Mulichak *et al.*, 1999), HGALE (Thoden *et al.*, 2000) and AGME (Deacon *et al.*, 2000), all of which undergo the same initial oxidative step.

In each case the nicotinamide ring C4 atom of NAD^+ is positioned approximately 3.4Å to 3.7Å from the glucosyl-C4' (or C6' in the case of AGME) with a

corresponding angle of between 81° and 96°. As already discussed in chapter three, analysis of redox enzyme substrate complexes in the PDB (Allard *et al.*, 2001; Fraaije and Mattevi, 2000) (including flavin adenine dinucleotide and flavin mononucleotide-dependent proteins) reveals that the site that donates or accepts hydride to or from the co-factor is at a distance of between 3.5Å and 4.0Å from C4 of the nicotinamide or N5 of the flavin ring, values which agree well with these findings (it is essential to remember that all these distances are in non-productive complexes and that the likely hydride transfer distance in catalysis will be about 2.5Å). From first principles this makes sense as the hydride is transferred from glucosyl-C4' to the lowest unoccupied molecular orbital (LUMO) of the nicotinamide ring. This LUMO has *p*-orbital character and is therefore perpendicular to the plane of the ring. The LUMO of the NAD^+ C4 aligns with the C-H bond of glucosyl-C4' for linear hydride transfer as shown in Figure 5.5.

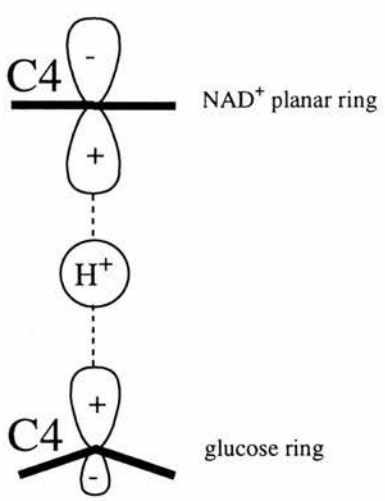


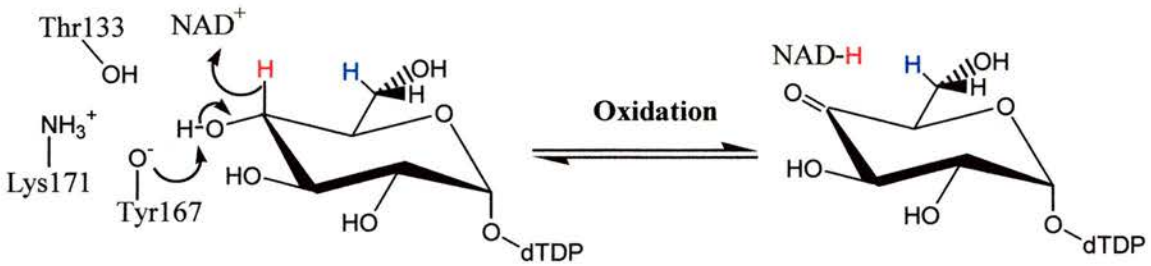
Figure 5.5. The positioning of the NAD^+ and glucose ring for linear hydride transfer from glucosyl-C4' to NAD^+-C4'.

With the active site tyrosine and nicotinamide ring suitably positioned the enzyme is able to perform the first step of catalysis, the removal of a proton from the glucosyl-C4' hydroxyl group by the active site base (the tyrosine) and concomitant hydride abstraction from glucosyl-C4' by NAD^+ (Figure 5.6). The creation of the 4'-ketosugar is important as it lowers the pK_a of the C5 proton of the substrate making the dehydration step possible.

With the tyrosine acting directly as the active site base there is no need to invoke the proton shuttle mechanism as proposed for EGALE. The question then remains as to the real role of the threonine in RmlB (and GMD that performs a very similar catalytic mechanism to RmlB but as yet no complex structures are available) or serine in HGALE.

Interestingly, despite threonine being found in all glucose 4,6-dehydratases, replacement of the threonine with a serine in *E. coli* RmlB has little effect on the activity of the enzyme (Gerratana *et al.*, 2001). Also, the position of the threonine in the RmlB substrate complexes reveals its Oγ1 to be 2.5Å from the glucosyl-C4' hydroxyl group, a situation also seen in HGALE and SQD1 where the corresponding distances for the serine and threonine are 2.5Å and 2.4Å from the glucosyl-C4' hydroxyl group, respectively. These short hydrogen bonds (often assigned "low barrier", but less frequently confirmed) have been proposed to be important in the catalytic mechanism of some enzymes by stabilising reaction intermediates or transition states (Cleland and Kreevoy, 1994; Cleland *et al.*, 1998). The observation that the threonine can be mutated to an alanine (Thr134Ala) in *E. coli* RmlB with a 200-fold lowering of k_{cat} (Gerratana *et al.*, 2001) seems to indicate that although the

threonine may not be essential for catalysis it might play a role in tuning the pK_a of the substrate.



dTDP-D-glucose dTDP-4-ketoglucose

Figure 5.6. The oxidation step carried out by RmlB. Residues are numbered as for S. typhimurium.

Step two - dehydration

The second step involves the dehydration of the dTDP-4-ketoglucose intermediate by the removal of water across glucosyl C5' and C6' to give the dTDP-4-ketoglucose-5,6-ene intermediate, the existence of which has recently been confirmed (Gross *et al.*, 2000). This step requires base catalysis for abstraction of the glucosyl-C5' proton and acid catalysis for protonation of the leaving glucosyl-C6' hydroxyl group.

One possibility suggested by Hegeman *et al.* (2001) was that the active site tyrosine could protonate the leaving glucosyl-C6' hydroxyl group, which would have the advantage of re-establishing negative charge adjacent to the NADH and thus drive the final chemical step to allow reformation of the tyrosinate-NAD^+ ion pair. However, these structures show that the tyrosine is more than 6Å from the glucosyl-C6' hydroxyl group. Instead, it is the positions of two other active site residues, Glu135 (*S. suis* Glu127) and Asp134 (*S. suis* Asp126), that hold the key to this dehydration step.

In both the *S. typhimurium* and *S. suis* dTDP-D-glucose complexes the glutamic acid is correctly aligned to deprotonate the glucosyl-C5'; the Oε2 of the glutamic acid being 2.8Å from glucosyl-C5' in both structures (Figure 5.4). The role of the glutamate in this reaction was originally inferred by site directed mutagenesis studies on GMD (Somoza *et al.*, 2000), and has been further substantiated by site directed mutagenesis studies on *E. coli* RmlB (Hegeman *et al.*, 2001). Indeed, when GMD is superimposed onto the *S. typhimurium* dTDP-D-glucose complex (rms deviation of 2.4Å over 294 Cα atoms), Glu133 of GMD is positioned almost exactly on top of

Glu135 of *S. typhimurium* RmlB (Figure 5.7). As GMD performs an almost identical dehydration step to RmlB (the conversion of GDP-D-mannose to the GDP-4-keto-5,6-manoseen intermediate), this is compelling evidence for the role of Glu133 in GMD being identical to that of Gu135 in RmlB. This glutamate is totally conserved in all RmlB and GMD enzymes looked at.

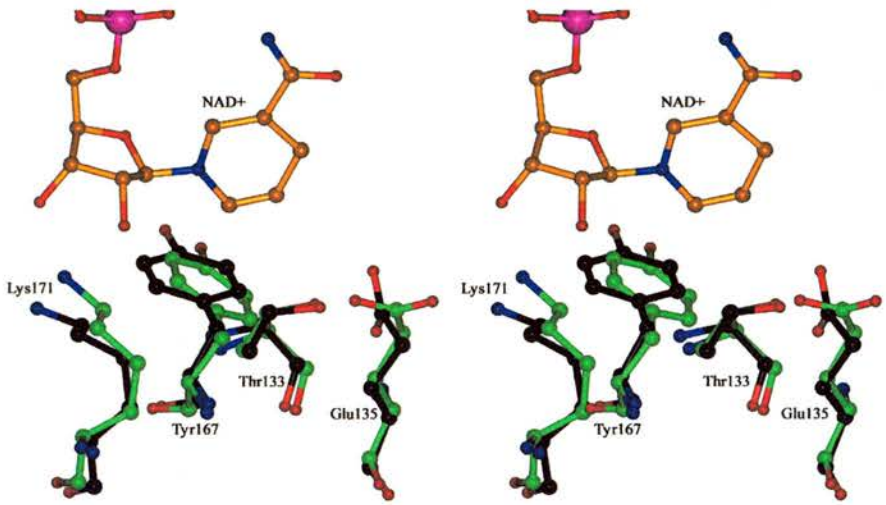


Figure 5.7. Stereo superposition of the active site residues GMD (green) onto the S. typhimurium *active site (black). Only four key active site residues are shown, numbering refers to the* S. typhimurium *RmlB. The* S. typhimurium *RmlB* NAD^+ *is shown in orange.*

The enzyme GFS catalyses the next reaction after GMD in the GDP-fucose pathway, the conversion of GDP-4-keto-6-deoxy-D-mannose to GDP-L-fucose. GFS is an interesting enzyme in that it catalyses a double epimerisation at positions C3 and C5 of GDP-4-keto-6-deoxy-D-mannose to give GDP-4-keto-6-deoxy-D-glucose and then a subsequent NADPH reduction at C4 to yield GDP-fucose (Somers *et al.*, 1998). The reaction catalysed by GFS encompasses the roles that both RmlC and RmlD fulfil in the dTDP-L-rhamnose pathway. GFS is similar to RmlB and GALE in that it is a

member of the SDR family and contains a Ser-Tyr-Lys (Ser107, Tyr136, Lys140 in *E. coli* GFS) catalytic triad. Two groups simultaneously published the holoenzyme structure of *E coli* GFS (Rizzi *et al.*, 1998; Somers *et al.*, 1998). Despite modelling substrate into the enzyme's active site, neither group was able to identify a residue that could act as a base in the epimerisation at C5' of the intermediate GDP-4-keto-6-deoxy-D-glucose. A structural superposition of GFS onto the RmlB holoenzyme from *S. typhimurium* reveals that the conserved Cys109 from GFS sits directly on RmlB Glu135 with the same orientation (Figure 5.8). Considering the highly conserved nature of this glutamate in RmlB and GMD, it seems highly probable that GFS Cys109 is the required catalytic base for the epimerisation at C5'of the GDP-4-keto-6-deoxy-D-glucose intermediate.

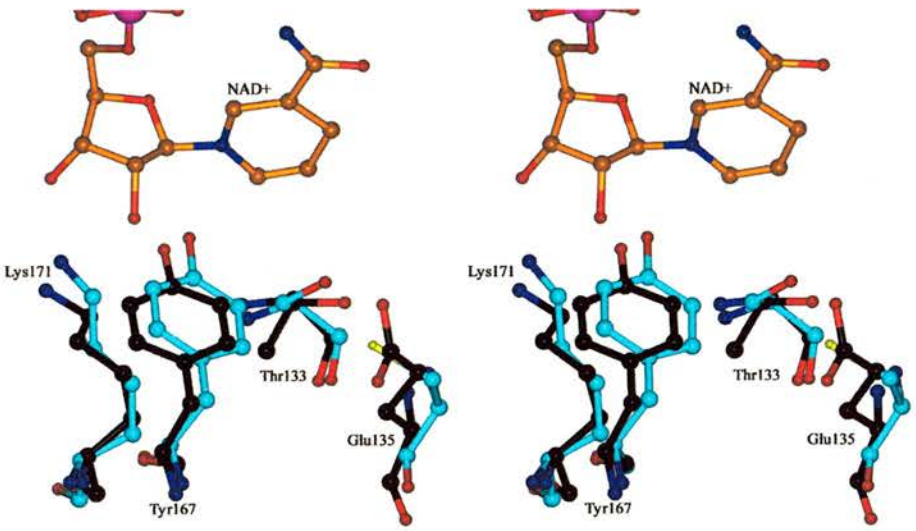


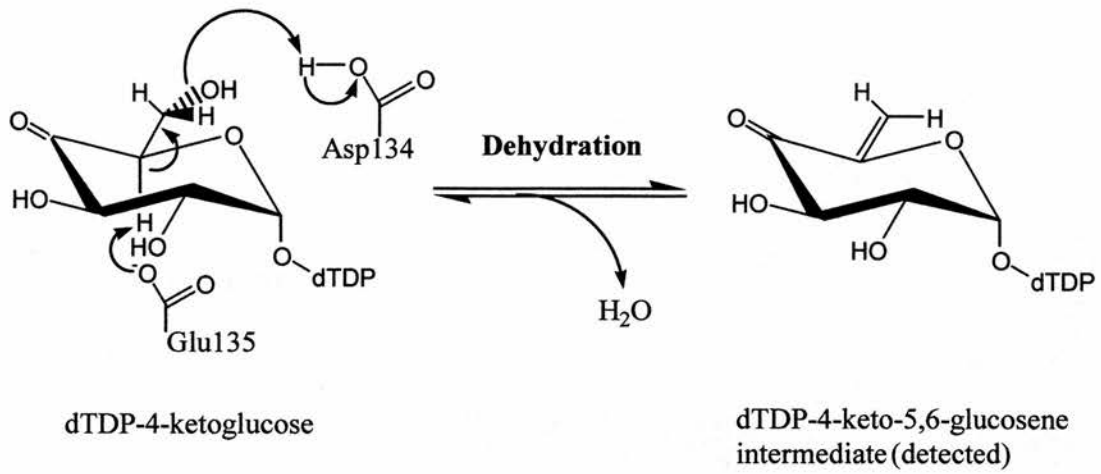
Figure 5.8. Stereo superposition of the active site residues GFS (cyan) onto the S. typhimurium *active site (black). Only four key active site residues are shown, numbering refers to the* S. typhimurium *RmlB. The* S. typhimurium *RmlB* NAD^+ *is shown in orange. The yellow atom (sulphur) belongs to Cys109 of GFS.*

Interestingly, in the *S. suis* dTDP-xylose complex the position of the active site glutamic acid (Glu127) is not aligned for the deprotonation of glucosyl-C5'. This structural alignment of Glu127 in the *S. suis* dTDP-xylose complex explains the much slower rate (120 fold) of solvent deuterium exchange with dTDP-xylose C5 (only $0.001s^{-1}$) than with product dTDP-4-keto-6-deoxy-D-glucose ($0.12\ s^{-1}$) (Gross *et al.*, 2001). As dTDP-xylose lacks the C6(OH) group of dTDP-D-glucose the enzyme reaction can only proceed to give a dTDP-4-ketoxylose species, hence the reaction does not progress to completion (Hegeman *et al.*, 2001).

For elimination of water to ensue the glucosyl-C6' hydroxyl group must be protonated by an acid as hydroxide is a poor leaving group. Inspection of the structure reveals that only Asp134 (*S. suis* Asp126) makes contact with this hydroxyl. The Oδ1 of Asp134 is 2.7Å from the C6'-hydroxyl group of the glucose. Importantly, the position of the aspartic acid is perfect for the *syn*-elimination of water. If the aspartic acid were charged it would result in the O6-hydroxyl proton being shared between the O6 and Oδ1, creating a partial negative charge on O6. This would make it all but impossible for the C6-O6 to cleave. The only logical conclusion is that the aspartic acid is protonated and that the proton is donated to O6 in a hydrogen bond. O6 then abstracts this proton and the C6-O6 bond cleaves, thus water is eliminated (Figure 5.9). Support for this structural hypothesis comes from the study of the turnover of the substrate analogue dTDP-6-fluoro-6-deoxyglucose by the native and Asp135Ala/Asp135Asn mutants of *E. coli* RmlB (Gross *et al.*, 2001). This substrate analogue eliminates the fluoride ion, which is a good leaving group and needs no proton. The pseudo substrate is turned over almost equally well by the Asp mutants and native enzyme. This is in marked contrast to the real substrate dTDP-D-glucose,

where turnover is approximately two orders of magnitude slower by the same aspartic acid mutants, confirming that the aspartic acid acts as an acid (Hegeman *et al.*, 2001). Interestingly, in the enzyme CDP-D-glucose 4,6-dehydratase from *S. typhimurium* an aspartic acid (Asp133) in a sequence based alignment is in a conserved position with RmlB Asp134. As this enzyme performs the same mechanistic reaction as RmlB (He *et al.*, 1996; Tanner, 2001) it seems likely that this conserved aspartic acid fulfils the same role in both organisms.

The dehydration step may occur by two possible mechanisms; enolisation of the transiently formed dTDP-4-ketoglucose intermediate followed by elimination (Gerlt and Gassman, 1992), or by the concerted 5,6 elimination of water from the intermediate (Gross *et al.*, 2001). It has now been shown using simultaneous kinetic analysis of glucosyl-C5(^{1}H/^{2}H) solvent hydrogen and C6(^{16}OH/^{18}OH) solvent oxygen exchange using MALDI-TOF mass spectrometry that dehydration clearly proceeds through the concerted elimination of water between C5 and C6 (Hegeman *et al.*, personal communication). Gerlt and Gassman first described concerted acid/base catalysis of enolisation as a means of explaining rapid catalytic turnover of β-elimination reactions despite large pK_a gaps between the substrate α-carbon acids and the active site bases (Gerlt and Gassman, 1992). They suggested that concerted acid/base catalysis of β-elimination reactions would resolve pK_a discrepancies in dehydration reactions without necessitating the formation of an enol intermediate. Concerted acid/base catalysis of either elimination or enolisation may be employed in enzyme-catalysed reactions to decrease energy barriers in cases where large pK_a differences exist between active site residues and substrates.



dTDP-4-ketoglucose

dTDP-4-keto-5,6-glucosene intermediate (detected)

Figure 5.9. The dehydration step carried out by RmlB. Residues are labelled as for S. typhimurium.

Step three - reduction

The final step involves hydride transfer back to glucosyl-C6' resulting in inversion of configuration at C6 of the product (Snipes *et al.*, 1977) and proton addition to C5 of glucose. The structures show that the nicotinamide ring sits above the plane of the hexapyranose ring (Figure 5.3), thus hydride abstraction and transfer occur on the same face of the molecule as predicted from labelling studies (Snipes *et al.*, 1977). Formation of the product dTDP-4-keto-6-deoxy-D-glucose is accompanied by regeneration of the coenzyme NAD^+.

With dTDP-D-glucose bound in the active site the NAD(H) C4 is positioned 3.7Å from the glucosyl-C6' in both the *S. typhimurium* and *S. suis* monomers, making an angle of between 108° and 118° with C4-N1 of the nicotinamide ring. Based on the EGALE UDP-glucose and UDP-galactose complex structures hydride transfer could be possible in this orientation, as the C4' atoms of the UDP-pyranose rings are in similar positions with respect to the nicotinamide ring when compared to the glucosyl-C6' in the RmlB complexes. If one accepts this, then in RmlB, Glu135 *S. typhimurium* (Glu127 *S. suis*) would then simply reprotonate glucosyl-C5' to complete the reaction. Water molecules would be required to reset the protonation states of Tyr167 and Asp134.

However, we have shown that the EGALE complexes are distorted; in the HGALE structure (where the tyrosine, Tyr157, may act directly as the active site base) the glucosyl-C4' of substrate is in an identical orientation with respect to the nicotinamide as the C4' of dTDP-D-glucose and dTDP-xylose in the RmlB complexes. Furthermore, when the sugar nucleotide complexes from RmlD

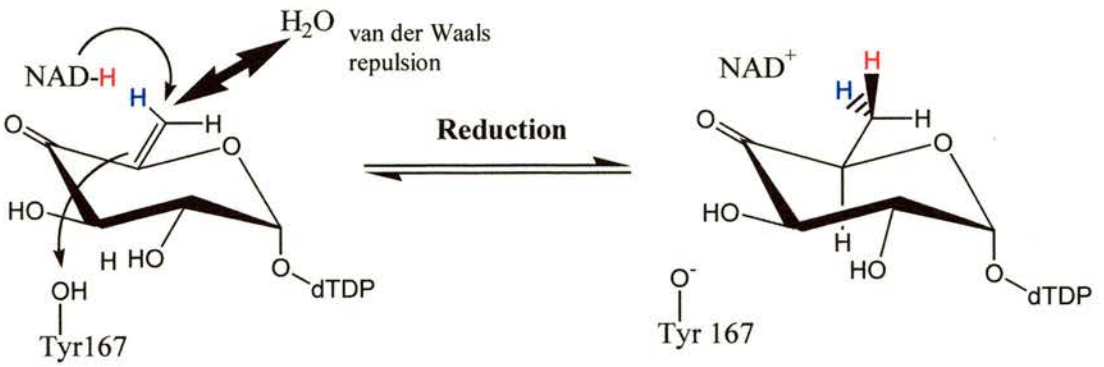
(reduction of rhamnosyl-C4') (Blankenfeldt & Naismith, personal communication), SQD1 (oxidation/reduction at glucosyl-C4') (Mulichak *et al.*, 1999) and AGME (significantly oxidation and reduction occurs at the C6 position) (Deacon *et al.*, 2000) are superimposed onto the RmlB substrate complexes, the carbon involved in hydride transfer is always located within a 1Å sphere of the glucosyl-C4' in the RmlB complexes. Taken together, this principle strongly implies that the trajectory of hydride transfer between the nicotinamide and a carbon atom is very specific. This suggests that the glucosyl-C6' position in our complexes is not in the correct position for hydride transfer. Thus the glucose ring of the RmlB substrate must undergo some movement after elimination of water to realign glucosyl-C6' for the optimum trajectory for hydride transfer from NADH to glucosyl-C6'. One possibility is that during the dehydration step there is a movement of glucosyl-C6' due to the shift of glucosyl-C5' from sp^3 to sp^2 which would change the angular nature of the surrounding bonds possibly resulting in glucosyl-C6' moving towards the NADH C4 atom (Burgi *et al.*, 1974).

The hexopyranosyl ring of dTDP-D-glucose from *S. typhimurium* can be rotated about the Pβ-O bond connecting dTDP with the anomeric oxygen atom of the 4-ketosugar so that glucosyl-C6' falls into the 1Å sphere centred on the position of glucosyl-C4' in the RmlB complexes allowing linear hydride transfer (Figure 5.10). In this new position there are no contacts less than 2.5Å between dTDP-D-glucose and the protein. The net result is that the glucosyl-C6' atom has moved by about 2.5Å. It has already been predicted that such a movement might be required for reduction (Allard *et al.*, 2001). This movement of the pyranose ring brings Tyr167 OH within 3.8Å of glucosyl-C5', a position that would enable it to protonate C5, thereby resetting itself

in the process ready for the next round of catalysis. The protonated Oε2 of Glu135 can also be brought to within 2.5Å of the Oδ1 of Asp134 enabling the proton to be transferred, thus resetting their protonation states. All that is required is for the $\chi 2$ torsion angle (CB-CD) to adopt an alternative energetic minima (180° to +60°). Such torsion flexibility in amino-acid side chains is well documented in proteins and occurs over very short timescales.

The rotation of the sugar ring requires a driving force. The driving force for the rotation of the 4-ketoglucose-5,6-ene ring is presumably the need to separate C6 from O6 (now H_2O). In dTDP-D-glucose, where the atoms are bonded this distance is 1.2Å, but as separate molecules (glucosene and water) this distance would be approximately 3.5Å (the normal van der Waals separation between a carbon and oxygen). As the O6 (water) is located inside a pocket in the protein and effectively capped by the glucose ring, the van der Waals repulsion cannot be relieved by expulsion of the newly created water from the binding site. The increase in separation to relieve the van der Waals strain is thus approximately 2.2Å, almost exactly the same as the movement required to position the glucosyl C6' for hydride transfer.

It is also possible that the attack of the hydride ion from NADH upon glucosyl-C6' at the double bond in the substrate may involve the formation of an enolate anion intermediate as has been invoked for the SDR enzyme enoyl reductase (Rafferty *et al.*,1995).



dTDP-4-keto-5,6-glucosene intermediate

dTDP-4-keto-6-deoxy-D-glucose

Figure 5.10. The reduction step carried out by RmlB. Residues are labelled as for S. typhimurium.

5.5. Final Conclusions

The reaction catalysed by RmlB is the first committed step in all 6-deoxysugar biosynthetic pathways, examples of which are found in all branches of life. This complex chemical reaction, which involves oxidation, water elimination and reduction, has long fascinated biologists and chemists. This work establishes in molecular detail the nature of each step of the reaction and the specific nature of hydride transfer. In the oxidation step, the substrate complexes show clearly that the active site tyrosine moves to enable it to act directly as the active site base. In the dehydration step, the role of a conserved Asp/Glu pairing has been shown to be central to the β-elimination of water across glucosyl-C5' and C6'. Finally in the reduction step, the strict parameters of hydride transfer indicate that the glucose ring must move so as to be able to accept a hydride from the NADH-C4' (Figure 5.11). This work not only shows how that movement can be brought about but also how all the residues can be reset to their correct protonation steps for the next round of catalysis. Crucially, many of these active site residues are conserved in other enzymes of the SDR family such as GMD, GALE, SDQ1 and GFS. The role of residues such as the active site glutamic and aspartic acid has important implications for the study of dehydratase enzymes as a whole.

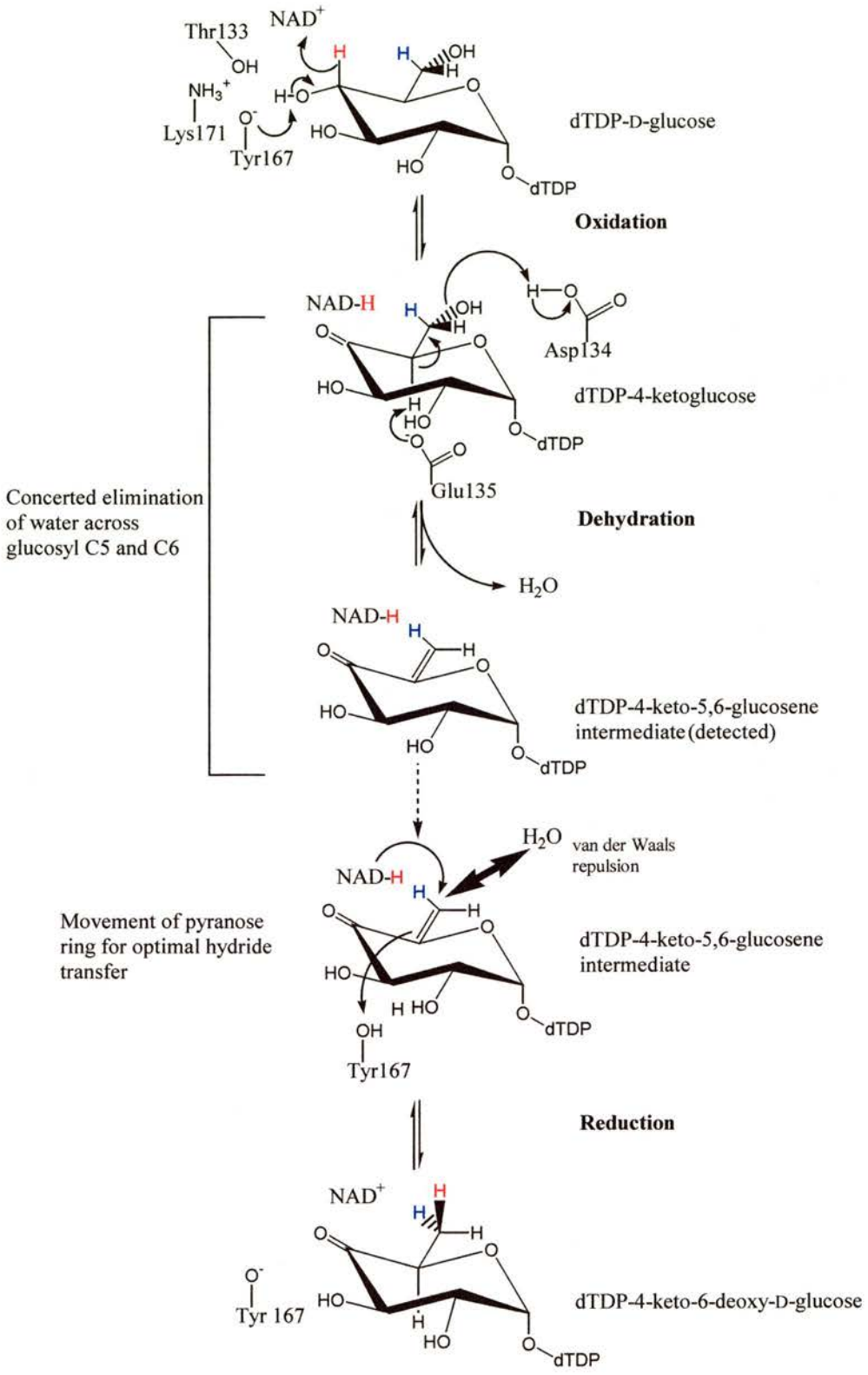


Figure 5.11. The mechanism of dTDP-D-glucose 4,6-dehydratase (RmlB). Residues are labelled as for S. typhimurium.

Appendices

Appendix I

Hampton Screen 1

1 30& MPD, 0.1M Na acetate pH 4.6, 0.02M Calcium Chloride
2. 0.4 M K, Na Tartrate
3. 0.4 M Ammonium Phosphate
4. 2.0 M Ammonium Sulphate, 0.1 M Tris HCl pH 8.5
5. 30% MPD, 0.1 M Na Hepes pH 7.5, 0.2 M Sodium Citrate
6. 30% PEG 4000, 0.1 M Tris HCl pH 8.5, 0.2 M Magnesium Chloride
7. 1.4 M Sodium Acetate, 0.1 M Na Cacodylate pH 6.5
8. 30% iso-Propanol, 0.1 M Na Cacodylate pH 6.5, 0.2 M Sodium Citrate
9. 30% PEG 4000, 0.1 M Na Citrate pH 5.6, 0.2 M Ammonium Acetate
10. 30% PEG 4000, 0.1 M Na Acetate pH 4.6, 0.2 M Ammonium Acetate
11. 1.0 M Ammonium Phosphate, 0.1 M Na Citrate pH 5.6
12. 30% iso-Propanol, 0.1 M Na Hepes pH 7.5, 0.2 M Magnesium Chloride
13. 30% PEG 400, 0.1 M Tris HCl pH 8.5, 0.2 M Sodium Citrate
14. 28% PEG 400, 0.1 M Na Hepes pH 7.5, 0.2 M Calcium Chloride
15. 30% PEG 8000, 0.1 M Na Cacodylate pH 6.5, 0.2 M Ammonium Sulphate
16. 1.5 M Lithium Sulphate, 0.1 M Na Hepes pH 7.5
17. 30% PEG 4000, 0.1 M Tris HCl pH 8.5, 0.2 M Lithium Sulphate
18. 20% PEG 8000, 0.1 M Na Cacodylate pH 6.5, 0.2 M Magnesium Acetate
19. 30% iso-Propanol, 0.1 M Tris HCl pH 8.5, 0.2 M Ammonium Acetate
20. 25% PEG 4000, 0.1 M Na Acetate pH 4.6, 0.2 M Ammonium Sulphate
21. 30% MPD, 0.1 M Na Cacodylate pH 6.5, 0.2 M Magnesium Acetate
22. 30% PEG 4000, 0.1 M Tris HCl pH 8.5, 0.2 M Sodium Acetate
23. 30% PEG 400, 0.1 M Na Hepes pH 7.5, 0.2 M Magnesium Chloride
24. 20% iso-Propanol, 0.1 M Na Acetate pH 4.6, 0.2 M Calcium Chloride
25. 1.0 M Sodium Acetate, 0.1 M Imidazole pH 6.5
26. 30% MPD, 0.1 M Na Citrate pH 5.6, 0.2 M Ammonium Acetate
27. 20% iso-Propanol, 0.1 M Na Hepes pH 7.5, 0.2 M Sodium Citrate
28. 30% PEG 8000, 0.1 M Na Cacodylate pH 6.5, 0.2 M Sodium Acetate
29. 0.8 M K, Na Tartrate, 0.1 M Na Hepes pH 7.5
30. 30% PEG 8000, 0.2 M Ammonium Sulphate
31. 30% PEG 4000, 0.2 M Ammonium Sulphate
32. 2.0 M Ammonium Sulphate
33. 4.0 M Sodium Formate
34. 2.0 M Sodium Formate, 0.1 M Na Acetate pH 4.6
35. 1.6 M Na, K Phosphate, 0.1 M Na Hepes pH 7.5
36. 8% PEG 8000, 0.1 M Tris HCl pH 8.5
37. 8% PEG 4000, 0.1 M Na Acetate pH 4.6
38. 1.4 M Sodium Citrate, 0.1 M Na Hepes pH 7.5
39. 2% PEG 400, 0.1 M Na Hepes pH 7.5, 2.0 M Ammonium Sulphate
40 20% iso-Propanol, 0.1 M Na Citrate pH 5.6, 20% PEG 4000
41. 10% iso-Propanol, 0.1 M Na Hepes pH 7.5, 20% PEG 4000
42. 20% PEG 8000, 0.05 M Potassium Phosphate
43. 30% PEG 1500
44. 0.2 M Magnesium Formate
45. 18% PEG 8000, 0.1 M Na Cacodylate pH 6.5, 0.2 M Zinc Acetate
46. 18% PEG 8000, 0.1 M Na Cacodylate pH 6.5, 0.2 M Calcium Acetate
47. 2.0 M Ammonium Sulphate, 0.1 M Sodium Acetate pH 4.6
48. 2.0 M Ammonium Phosphate, 0.1 M Tris HCl pH 8.5
49. 2% PEG 8000, 1.0 M Lithium Sulphate
50. 15% PEG 8000, 0.5 M Lithium Sulphate

Hampton Screen 2

1. 10% PEG 6000, 2.0 M Sodium Chloride
2. 0.5 M Sodium Chloride, 0.01 M CTAB, 0.01 M Magnesium Chloride
3. 25% Ethylene Glycol
4. 35% Dioxane
5. 5% iso-Propanol, 2.0 M Ammonium Sulphate
6. 1.0 M Imidazole pH 7.0
7. 10% PEG 1000, 10% PEG 8000
8. 10% Ethanol, 1.5 M Sodium Chloride
9. 2.0 M Sodium Chloride, 0.1 M Na Acetate pH 4.6
10. 30% MPD, 0.1 M Na Acetate pH 4.6, 0.2 M Sodium Chloride
11. 1.0 M 1,6 Hexanediol, 0.1 M Na Acetate pH 4.6, 0.01 M Cobalt Chloride
12. 30% PEG 400, 0.1 M Na Acetate pH 4.6, 0.1 M Cadmium Chloride
13. 30% PEG MME 2000, 0.1 M Na Acetate pH 4.6, 0.2 M Ammonium Sulphate
14. 2.0 M Ammonium Sulphate, 0.1 M Na Citrate pH 5.6, 0.2 M K/Na Tartrate
15. 1.0 M Lithium Sulphate, 0.1 M Na Citrate pH 5.6, 0.5 M Ammonium Sulphate
16. 2% Polyethyleneimine, 0.1 M Na Citrate pH 5.6, 0.5 M Sodium Chloride
17. 35% tert-Butanol, 0.1 M Na Citrate pH 5.6
18. 10% Jeffamine M-600, 0.1 M Na Citrate pH 5.6, 0.01 M Ferric Chloride
19. 2.5 M 1,6 Hexanediol, 0.1 M Na Citrate pH 5.6
20. 1.6 M Magnesium Sulphate, 0.1 M MES pH 6.5
21. 2.0 M Sodium Chloride, 0.1 M MES pH 6.5, 0.2 M Na/K Phosphate
22. 12% PEG 20,000, 0.1 M MES pH 6.5
23. 10% Dioxane, 0.1 M MES pH 6.5, 1.6 M Ammonium Sulphate
24. 30% Jeffamine M-600, 0.1 M MES pH 6.5, 0.05 M Cesium Chloride
25. 1.8 M Ammonium Sulphate, 0.1 M MES pH 6.5, 0.01 M Cobalt Chloride
26. 30% PEG MME 5000, 0.1 M MES pH 6.5, 0.2 M Ammonium Sulphate
27. 25% PEG MME 550, 0.1 M MES pH 6.5, 0.01 M Zinc Sulphate
28. 1.6 M Sodium Citrate pH 6.5
29. 30% MPD, 0.1 M Hepes pH 7.5, 0.5 M Ammonium Sulphate
30. 10% PEG 6000, 0.1 M Hepes pH 7.5, 5% MPD
31. 20% Jeffamine M-600, 0.1 M Hepes pH 7.5
32. 1.6 M Ammonium Sulphate, 0.1 M Hepes pH 7.5, 0.1 M Sodium Chloride
33. 2.0 M Ammonium Formate, 0.1 M Hepes pH 7.5
34. 1.0 M Sodium Acetate, 0.1 M Hepes pH 7.5, 0.05 M Cadmium Sulphate
35. 70% MPD, 0.1 M Hepes pH 7.5
36. 4.3 M Sodium Chloride, 0.1 M Hepes pH 7.5
37. 10% PEG 8000, 0.1 M Hepes pH 7.5, 8% Ethylene Glycol
38. 20% PEG 10,000, 0.1 M Hepes pH 7.5
39. 3.4 M 1,6 Hexanediol, 0.1 M Tris pH 8.5, 0.2 M Magnesium Chloride
40. 25% tert-Butanol, 0.1 M Tris pH 8.5
41. 1.0 M Lithium Sulphate, 0.1 M Tris pH 8.5, 0.01 M Nickel (II) Chloride
42. 12% Glycerol, 0.1 M Tris pH 8.5, 1.5 M Ammonium Sulphate
43. 50% MPD, 0.1 M Tris pH 8.5, 0.2 M Ammonium Phosphate
44. 20% Ethanol, 0.1 M Tris pH 8.5
45. 20% PEG MME 2000, 0.1 M Tris pH 8.5, 0.01 M Nickel (II) Chloride
46. 20% PEG MME 550, 0.1 M Bicine pH 9.0, 0.1 M Sodium Chloride
47. 2.0 M Magnesium Chloride, 0.1 M Bicine pH 9.0
48. 10% PEG 20,000, 0.1 M Bicine pH 9.0, 2% Dioxane

Appendix II

Stereochemistry of hydride transfer

The stereochemical parameters of hydride transfer were analysed using the following structures from the PDB data base: 1KVS (EGALE complexed with NAD^+ and UDP-glucose), 1UDC (EGALE complexed with NAD^+ and UDP-mannose), 1XEL (EGALE complexed with NADH and UDP-glucose; an abortive complex), 1A9Y (EGALE mutant S124A/Y149F complexed with UDP-Glucose), 1EK6 (HGALE complexed with NADH and UDP-Glucose), 1A9Z (EGALE mutant S124A/Y149F complexed With UDP-Galactose), 1DLI (UDP-Glucose dehydrogenase complexed with NAD^+ and UDP-xylopyranose), 2ACQ (Aldose reductase complexed with NADP and glucose-6-phosphate), 3LDH (Lactate dehydrogenase complexed with NAD^+ and pyruvate), 1LDN (Lactate dehydrogenase complexed with NADH, oxamate, and fructose-1,6-bisphosphate), 9LDT (Lactate dehydrogenase complexed with NADH and oxamate), 1BTO (Alcohol dehydrogenase complexed with NADH and (1s,3r)3-butylthiolane 1-oxide) and 1E13 (Human aldose reductase complexed with Idd384 inhibitor). The measurements made involved the distance between the C4 of the nicotinamide ring and the site of oxidative/reductive attack, and the angle that the N1 and C4 atoms of the planar nicotinamide make with the site of oxidative/reductive attack.

References

Adams, G. A., Quadling, C. and Perry, M. B. (1967) D-glycero-D-manno-heptose as a component of lipopolysaccharides from Gram-negative bacteria, *Can J Microbiol*, **13**, 1605-13.

Allard, S. T., Giraud, M. F., Whitfield, C., Messner, P. and Naismith, J. H. (2000) The purification, crystallization and structural elucidation of dTDP-D- glucose 4,6-dehydratase (RmlB), the second enzyme of the dTDP-L- rhamnose synthesis pathway from *Salmonella enterica* serovar typhimurium, *Acta Crystallogr D Biol Crystallogr*, **56**, 222-5.

Allard, S. T., Giraud, M., Whitfield, C., Graninger, M., Messner, P. and Naismith, J. H. (2001a) The Crystal Structure of dTDP-D-glucose 4,6-dehydratase (RmlB) from *Salmonella enterica* serovar Typhimurium, the second enzyme in the dTDP-L-Rhamnose Pathway, *J Mol Biol*, **307**, 283-295.

Allard, S. T. M., Giraud, M. F. and Naismith, J. H. (2001b) Epimerases: structure, function and mechanism, *Cell Mol Life Sci*, **58**, 1650-1665.

Amass, S. F., Wu, C. C. and Clark, L. K. (1996) Evaluation of antibiotics for the elimination of the tonsillar carrier state of *Streptococcus suis* in pigs, *J Vet Diagn Invest*, **8**, 64-7.

Andres, J., Moliner, V., Safont, V. S., Aullo, J. M., Diaz, W. and Tapia, O. (1996) Transition structures for hydride transfer reactions in vacuo and their role in enzyme catalysis, *J. Mol. Struct. (Theochem)*, **371**, 299-312.

Arends, J. P. and Zanen, H. C. (1988) Meningitis caused by *Streptococcus suis* in humans, *Rev Infect Dis*, **10**, 131-7.

Baron, E. J., Chang, R. S., Howard, D. H., Miller, J. N. and Turner, J. A. (1993) *Medical Microbiology: A short course*, John Wiley and Sons, New York.

Bauer, A. J., Rayment, I., Frey, P. A. and Holden, H. M. (1992) The molecular structure of UDP-galactose 4-epimerase from *Escherichia coli* determined at 2.5Å resolution, *Proteins*, **12**, 372-81.

Bellamacina, C. R. (1996) The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins, *Faseb J*, **10**, 1257-69.

Berger, E., Arabshahi, A., Wei, Y., Schilling, J. F. and Frey, P. A. (2001) Acid-base catalysis by udp-galactose 4-epimerase: correlations of kinetically measured acid dissociation constants with thermodynamic values for tyrosine 149, *Biochemistry*, **40**, 6699-705.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res*, **28**, 235-42.

Bertland, A. U. and Kalckar, H. M. (1968) Reversible changes of ordered polypeptide structures in oxidized and reduced epimerase, *Proc Natl Acad Sci U S A*, **61**, 629-35.

Bisno, A. L. and van de Rijn, I. (2000) Classification of Streptococci. In *Mandell, Douglas, and Bennett's principles and practice of infectious diseases.* 5th edit. (Mandell, G. L., Bennett, J. E. and Dolin, R., eds.), Vol. 2, pp. 2100-2105. Churchill Livingstone, Philadelphia.

Blankenfeldt, W., Asuncion, M., Lam, J. S. and Naismith, J. H. (2000) The structural basis of the catalytic mechanism and regulation of glucose-1-phosphate thymidylyltransferase (RmlA), *Embo J*, **19**, 6652-6663.

Brazeau, C., Gottschalk, M., Vincelette, S. and Martineau-Doize, B. (1996) In vitro phagocytosis and survival of *Streptococcus suis* capsular type 2 inside murine macrophages, *Microbiology*, **142**, 1231-7.

Brunger, 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 Biol Crystallogr*, **54**, 905-21.

Burgi, H., B., Dunitz, J. D., Lehn, J.-M. and Wipff, G. (1974) Stereochemistry of Reaction Paths at Carbonyl Centres, *Tetrahedron*, **30**, 1563.

Carnell, A. J. (1999) Stereoinversions using microbial redox-reactions, *Adv Biochem Eng Biotechnol*, **63**, 57-72.

Chanter, N., Jones, P. W. and Alexander, T. J. (1993) Meningitis in pigs caused by *Streptococcus suis*--a speculative review, *Vet Microbiol*, **36**, 39-55.

Chen, Z., Jiang, J. C., Lin, Z. G., Lee, W. R., Baker, M. E. and Chang, S. H. (1993) Site-specific mutagenesis of Drosophila alcohol dehydrogenase: evidence for involvement of tyrosine-152 and lysine-156 in catalysis, *Biochemistry*, **32**, 3342-6.

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**, 24608-12.

Cleland, W. W. and Kreevoy, M. M. (1994) Low-barrier hydrogen bonds and enzymic catalysis, *Science*, **264**, 1887-90.

Cleland, W. W., Frey, P. A. and Gerlt, J. A. (1998) The low barrier hydrogen bond in enzymatic catalysis, *J Biol Chem*, **273**, 25529-32.

Clifton-Hadley, F. A. and Enright, M. R. (1984) Factors affecting the survival of *Streptococcus suis* type 2, *Vet Rec*, **114**, 584-6.

Collaborative Computational Project, Number 4. (1994). The CCP4 Suite: Programs for Protein Crystallography, *Acta Crystallogr B*, **50**, 760-763.

Curd, H., Liu, D. and Reeves, P. R. (1998) Relationships among the O-antigen gene clusters of *Salmonella enterica* groups B, D1, D2, and D3, *J Bacteriol*, **180**, 1002-7.

de Moor, D. E. (1963) Septicaemic infection in pigs caused by haemolytic streptococci of new Lancefield groups R, S, and T., *Antonie Van Leeuwenhoek*, **29**, 272.

Deacon, A. M., Ni, Y. S., Coleman, W. G., Jr. and Ealick, S. E. (2000) The crystal structure of ADP-L-glycero-D-mannoheptose 6-epimerase: catalysis with a twist, *Structure Fold Des*, **8**, 453-62.

Donnenberg, M. S. (2000) Pathogenic strategies of enteric bacteria, *Nature*, **406**, 768-74.

Ducruix, A. and Giege, R. (1992) *Crystallisation of Nucleic Acids and Proteins, a practical approach*, IRL Press, Oxford.

Edman, P. and Begg, G. (1967) A protein sequenator, *Eur J Biochem*, **1**, 80-91.

Eisenstein, T. K., Huang, D. and Schwacha, M. G. (1996) Immunity to Salmonella infections. In *Enteric Infections and Immunity* (Friedman, H., Paradise, L. J., Bendinelli, M. and Brandsma, J., eds.), pp. 57-78. Plenum Press, New York.

Eklund, H., Horjales, E., Jornvall, H., Branden, C. I. and Jeffery, J. (1985) Molecular aspects of functional differences between alcohol and sorbitol dehydrogenases, *Biochemistry*, **24**, 8005-12.

Elliott, S. D. (1966) Streptococcal infection in young pigs. I. An immunochemical study of the causative agent (PM streptococcus), *J Hyg (Lond)*, **64**, 205-12.

Elliott, S. D. and Tai, J. Y. (1978) The type-specific polysaccharides of *Streptococcus suis*, *J Exp Med*, **148**, 1699-704.

Elliott, T., Hastings, M. and Desselberger, U. (1997) *Lecture notes on Medical Microbiology*. 3rd edit, Blackwell Science, Oxford.

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

Esnouf, R. M. (1999) Further additions to MolScript version 1.4, including reading and contouring of electron-density maps, *Acta Crystallogr D Biol Crystallogr*, **55**, 938-40.

Essigmann, B., Hespenheide, B. M., Kuhn, L. A. and Benning, C. (1999) Prediction of the active-site structure and NAD(+) binding in SQD1, a protein essential for sulfolipid biosynthesis in Arabidopsis, *Arch Biochem Biophys*, **369**, 30-41.

Etzioni, A., Frydman, M., Pollack, S., Avidor, I., Phillips, M. L., Paulson, J. C. and Gershoni-Baruch, R. (1992) Brief report: recurrent severe infections caused by a novel leukocyte adhesion deficiency, *N Engl J Med*, **327**, 1789-92.

Finlay, B. B. and Falkow, S. (1989) Salmonella as an intracellular parasite, *Mol Microbiol*, **3**, 1833-41.

Finlay, B. B., Heffron, F. and Falkow, S. (1989) Epithelial cell surfaces induce Salmonella proteins required for bacterial adherence and invasion, *Science*, **243**, 940-3.

Flowers, H. M. (1981) Chemistry and Biochemistry of D and L-fucose. In *Advances in Carbohydrate Chemistry and Biochemistry* (Stewart-Tipson, R. and Horton, D., eds.), Vol. 39, pp. 276-645. Academic Press.

Fraaije, M. W. and Mattevi, A. (2000) Flavoenzymes: diverse catalysts with recurrent features, *Trends Biochem Sci*, **25**, 126-32.

Frey, P. A. (1987) Complex pyridine nucleotide-dependent transformations. In *Pyridine Nucleotide Coenzymes: Chemical, Biochemical and Medical Aspects* (Dolphin, D., Poulson, R. and Avramovic, O., eds.), Vol. 2B, pp. 461-511. John Wiley & Sons, New York.

Frey, P. A. (1996) The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose, *Faseb J*, **10**, 461-70.

Gabriel, O. and Lindquist, L. C. (1968) Biological mechanisms involved in the formation of deoxy sugars. IV. Enzymatic conversion of thymidine diphosphoglucose-4T to thymidine diphospho-4-keto-6-deoxyglucose-6T, *J Biol Chem*, **243**, 1479-84.

Gerlt, J. A. and Gassman, P. G. (1992) Understanding enzyme-catalyzed proton abstraction from carbon acids: details of stepwise mechanisms from β-elimination reactions, *J. Am. Chem. Soc.*, **114**, 5928-5934.

Gerratana, B., Cleland, W. W. and Frey, P. A. (2001) Mechanistic Roles of Thr134, Tyr160, and Lys 164 in the Reaction Catalyzed by dTDP-Glucose 4,6-Dehydratase, *Biochemistry*, **40**, 9187-95.

Ghosh, D., Weeks, C. M., Grochulski, P., Duax, W. L., Erman, M., Rimsay, R. L. and Orr, J. C. (1991) Three-dimensional structure of holo 3a,20b-hydroxysteroid dehydrogenase: a member of a short-chain dehydrogenase family, *Proc Natl Acad Sci U S A*, **88**, 10064-8.

Giannella, R. A., Broitman, S. A. and Zamcheck, N. (1973a) Influence of gastric acidity on bacterial and parasitic enteric infections. A perspective, *Ann Intern Med*, **78**, 271-6.

Giannella, R. A., Formal, S. B., Dammin, G. J. and Collins, H. (1973b) Pathogenesis of salmonellosis. Studies of fluid secretion, mucosal invasion, and morphologic reaction in the rabbit ileum, *J Clin Invest*, **52**, 441-53.

Giannella, R. A., Gots, R. E., Charney, A. N., Greenough, W. B. and Formal, S. B. (1975) Pathogenesis of Salmonella-mediated intestinal fluid secretion. Activation of adenylate cyclase and inhibition by indomethacin, *Gastroenterology*, **69**, 1238-45.

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

Giraud, M. F., McMiken, H. J., Leonard, G. A., Messner, P., Whitfield, C. and Naismith, J. H. (1999) Overexpression, purification, crystallization and preliminary structural study of dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD), the fourth enzyme of the dTDP-L-rhamnose synthesis pathway, from *Salmonella enterica* serovar Typhimurium, *Acta Crystallogr D Biol Crystallogr*, **55**, 2043-6.

Giraud, M. F., Leonard, G. A., Field, R. A., Berlind, 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.

Gitzelmann, R., Steinmann, B., Mitchell, B. and Haigis, E. (1977) Uridine diphosphate galactose 4'-epimerase deficiency. IV. Report of eight cases in three families, *Helv Paediatr Acta*, **31**, 441-52.

Glaser, L. and Kornfeld, S. (1961) The enzymatic synthesis of thymidine-linked sugars, *J Biol Chem*, **236**, 1795-1799.

Glaser, L. (1963) Biosynthesis of deoxysugars, *Physiol Rev*, **43**, 215-242.

Gouet, P., Courcelle, E., Stuart, D. I. and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript, *Bioinformatics*, **15**, 305-8.

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**, 25069-77.

Griffith, F. (1928) The significance of Pneumococcal types, *J Hyg (Lond)*, **27**, 113-159.

Gross, J. W., Hegeman, A. D., Vestling, M. M. and Frey, P. A. (2000) Characterization of enzymatic processes by rapid mix-quench mass spectrometry: the case of dTDP-glucose 4,6-dehydratase, *Biochemistry*, **39**, 13633-40.

Gross, J. W., Hegeman, A. D., Gerratana, B. and Frey, P. A. (2001) Dehydration Is Catalyzed by Glutamate-136 and Aspartic Acid-135 Active Site Residues in *Escherichia coli* dTDP-Glucose 4,6-Dehydratase, *Biochemistry*, **40**, 12497-504.

Grundy, W. N., Bailey, T. L., Elkan, C. P. and Baker, M. E. (1997) Hidden Markov model analysis of motifs in steroid dehydrogenases and their homologs, *Biochem Biophys Res Commun*, **231**, 760-6.

Hallis, T. M. and Liu, H. (1999) Learning nature's strategies for making deoxy sugars: Pathways, mechanisms, and combinatorial applications, *Accounts. Chem. Res.*, **32**, 579-588.

Hancock, I. C. (1997) Bacterial cell surface carbohydrates: structure and assembly, *Biochem Soc Trans*, **25**, 183-7.

Hardie, J. M. and Whiley, R. A. (1997) Classification and overview of the genera Streptococcus and Enterococcus, *Soc Appl Bacteriol Symp Ser*, **26**, 1S-11S.

Hart, T. and Shears, P. (1996) *Diagnosis in Color: Medical Microbiology*, Mosby-Wolfe, London.

He, X., Thorson, J. S. and Liu, H. W. (1996) Probing the coenzyme and substrate binding events of CDP-D-glucose 4,6-dehydratase: mechanistic implications, *Biochemistry*, **35**, 4721-31.

He, X. M., Agnihotri, G. and Liu, H. W. (2000) Novel enzymatic mechanisms in carbohydrate metabolism, *Chem Rev*, **100**, 4615-4661.

Heard, T. (1984) *Streptococcus suis* type 2 in British pig herds, *In Pract*, **6**, 69-71.

Hegeman, A. D. (2001) The mechanism and catalysis of water elimination in the reaction of *Escherichia coli* dTDP-Glucose 4,6-Dehydratase RffG. Ph.D Thesis, University of Wisconsin-Madison.

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, *Biochemistry*, **40**, 6598-610.

Henderson, M. J., Holton, J. B. and MacFaul, R. (1983) Further observations in a case of uridine diphosphate galactose-4-epimerase deficiency with a severe clinical presentation, *J Inherit Metab Dis*, **6**, 17-20.

Hess, J. and Kaufmann, S. H. (1996) *Salmonella enterica* infection, *Res Immunol*, **147**, 581-6.

Higgins, R. and Gottschalk, M. (1995) Distribution of *Streptococcus suis* capsular types in 1994, *Can Vet J*, **36**, 320.

Hutchinson, E. G. and Thornton, J. M. (1996) PROMOTIF--a program to identify and analyze structural motifs in proteins, *Protein Sci*, **5**, 212-20.

Ingraham, J. L. and Ingraham, C. A. (1995) *Introduction to Microbiology*, Wadsworth Publishing Company, London.

Jiang, X. M., Neal, B., Santiago, F., Lee, S. J., Romana, L. K. and Reeves, P. R. (1991) Structure and sequence of the rfb (O antigen) gene cluster of *Salmonella enterica* serovar typhimurium (strain LT2), *Mol Microbiol*, **5**, 695-713.

Johnson, D. A. and Liu, H. (1998) Mechanisms and pathways from recent deoxysugar biosynthesis research, *Curr Opin Chem Biol*, **2**, 642-9.

Joiner, K. A. (1988) Complement evasion by bacteria and parasites, *Annu Rev Microbiol*, **42**, 201-30.

Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard. (1991) Improved methods for binding protein models in electron density maps and the location of errors in these models, *Acta Crystallogr A*, **47**, 110-9.

Jones, T. A. (1992) A, yaap, asap, @#*? A set of averaging programs. In *Molecular Replacement* (Dodson, E. J., Gover, S. and Wolf, W., eds.), pp. 91-105. SERC Daresbury Laboratory, Warrington, UK.

Jornvall, H., Persson, M. and Jeffery, J. (1981) Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type, *Proc Natl Acad Sci U S A*, **78**, 4226-30.

Jornvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J. and Ghosh, D. (1995) Short-chain dehydrogenases/reductases (SDR), *Biochemistry*, **34**, 6003-13.

Jornvall, H. (1999) Multiplicity and complexity of SDR and MDR enzymes, *Adv Exp Med Biol*, **463**, 359-64.

Jornvall, H., Hoog, J. O. and Persson, B. (1999) SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature, *FEBS Lett*, **445**, 261-4.

Kabsch, W. (1976) A solution for the best rotation to relate two sets of vectors., *Acta Crystallogr A*, **32**, 922-923.

Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers*, **22**, 2577-637.

Kang, U. G., Nolan, L. D. and Frey, P. A. (1975) Uridine diphosphate galactose-4-epimerase. Uridine monophosphate- dependent reduction by α- and β-D-glucose, *J Biol Chem*, **250**, 7099-105.

Karsan, A., Cornejo, C. J., Winn, R. K., Schwartz, B. R., Way, W., Lannir, N., Gershoni-Baruch, R., Etzioni, A., Ochs, H. D. and Harlan, J. M. (1998) Leukocyte Adhesion Deficiency Type II is a generalized defect of de novo GDP-fucose biosynthesis. Endothelial cell fucosylation is not required for neutrophil rolling on human nonlymphoid endothelium, *J Clin Invest*, **101**, 2438-45.

Kirschning, A., Bechthold, A. F. W. and Rohr, J. (1997) Chemical and biochemical aspects of deoxysugars and deoxysugar oligosaccharides, *Topics in Current Chemistry*, **188**, 1-84.

Klein, J. (1990) *Immunology*, Blackwell Science, Oxford.

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

Kleywegt, G. J. (1999) Experimental assessment of differences between related protein crystal structures, *Acta Crystallogr D Biol Crystallogr*, **55**, 1878-84.

Lancefield, R. C. (1933) A serological differentiation of human and other groups of hemolytic streptococci, *J Exp Med*, **57**, 571-595.

Langford, P., Williams, A. E. and Kroll, J. S. (1991) Superoxide dismutases of pathogenic and non-pathogenic *Streptococcus suis* type 2 isolates, *FEMS Microbiol Lett*, **61**, 347-50.

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

Leloir, L. F. (1951) The enzymatic transformation of uridine diphosphate glucose into a galactose derivative, *Arch Biochem Biophys*, **33**, 186-190.

Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. In *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography, Number 26.* SERC Daresbury Laboratory, Warrington, UK.

Liu, H. W. and Thorson, J. S. (1994) Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria, *Annu Rev Microbiol*, **48**, 223-56.

Liu, Y., Thoden, J. B., Kim, J., Berger, E., Gulick, A. M., Ruzicka, F. J., Holden, H. M. and Frey, P. A. (1997) Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from *Escherichia coli*, *Biochemistry*, **36**, 10675-84.

Lutticken, R., Temme, N., Hahn, G. and Bartelheimer, E. W. (1986) Meningitis caused by *Streptococcus suis*: case report and review of the literature, *Infection*, **14**, 181-5.

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

Ma, Y., Mills, J. A., Belisle, J. T., Vissa, V., Howell, M., Bowlin, K., Scherman, M. S. and McNeil, M. (1997) Determination of the pathway for rhamnose biosynthesis in mycobacteria: cloning, sequencing and expression of the *Mycobacterium tuberculosis* gene encoding α-D-glucose-1-phosphate thymidylyltransferase, *Microbiology*, **143**, 937-45.

Macpherson, D. F., Manning, P. A. and Morona, R. (1994) Characterization of the dTDP-rhamnose biosynthetic genes encoded in the rfb locus of *Shigella flexneri*, *Mol Microbiol*, **11**, 281-92.

March, J. (1985) Acids and bases. In *Advanced Organic Chemistry*, pp. 218-236. John Wiley and Sons, New York.

Maskell, D. and Allen, A. (1997) Molecular biology of lipopolysaccharide biosynthesis in Salmonella and Bordetella, *Biochem Soc Trans*, **25**, 850-6.

Matthews, B. W. (1968) Solvent content of protein crystals, *J Mol Biol*, **33**, 491-7.

McEver, R. P. (1997) Selectin-carbohydrate interactions during inflammation and metastasis, *Glycoconj J*, **14**, 585-91.

McGrath, B. C. and Osborn, M. J. (1991) Localization of the terminal steps of O-antigen synthesis in *Salmonella typhimurium*, *J Bacteriol*, **173**, 649-54.

McNeil, M., Daffe, M. and Brennan, P. J. (1990) Evidence for the nature of the link between the arabinogalactan and peptidoglycan of mycobacterial cell walls, *J Biol Chem*, **265**, 18200-6.

Melo, A., Elliott, W. H. and Glaser, L. (1968) The mechanism of 6-deoxyhexose synthesis. I. Intramolecular hydrogen transfer catalyzed by deoxythymidine diphosphate D-glucose oxidoreductase, *J Biol Chem*, **243**, 1467-74.

Mergaert, P., Van Montagu, M. and Holsters, M. (1997) The nodulation gene nolK of *Azorhizobium caulinodans* is involved in the formation of GDP-fucose from GDP-mannose, *FEBS Lett*, **409**, 312-6.

Michalek, S. M., Morisaki, R. L., Gregory, S., Kimura, C. C., Harmon, S., Hamada, S., Kotani, S. and McGhee, J. R. (1984) Oral adjuvants enhance salivary IgA responses to purified *Streptococcus mutans* antigens., *Protides Biol. Fluids Proc. Colloq*, **32**, 47-52.

Mulichak, A. M., Theisen, M. J., Essigmann, B., Benning, C. and Garavito, R. M. (1999) Crystal structure of SQD1, an enzyme involved in the biosynthesis of the plant sulfolipid headgroup donor UDP-sulfoquinovose, *Proc Natl Acad Sci U S A*, **96**, 13097-102.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. and Dodson, E. J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT, *Acta Crystallogr D Biol Crystallogr*, **55**, 247-55.

Navaza, J. (1994) *AMoRe*: an automated package for molecular replacement. Acta Cryst., *Acta Crystallogr A*, **50**, 157-163.

Nelsestuen, G. L. and Kirkwood, S. (1971) The mechanism of action of the enzyme uridine diphosphoglucose 4- epimerase. Proof of an oxidation-reduction mechanism with direct transfer of hydrogen between substrate and the B-position of the enzyme-bound pyridine nucleotide, *J Biol Chem*, **246**, 7533-43.

Novagen. (1999) *pET System manual*. 8 edit, Novagen, Madison.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymology* (Carter, C. W. and Sweet, R. M., eds.), Vol. 276, pp. 307-326. Academic Press.

Perrakis, A., Morris, R. and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat Struct Biol*, **6**, 458-63.

Persson, B., Krook, M. and Jornvall, H. (1995) Short-chain dehydrogenases/reductases, *Adv Exp Med Biol*, **372**, 383-95.

Persson, B., Nordling, E., Kallberg, Y., Lundh, D., Oppermann, U. C., Marschall, H. U. and Jornvall, H. (1999) Bioinformatics in studies of SDR and MDR enzymes, *Adv Exp Med Biol*, **463**, 373-7.

Petry, K. G. and Reichardt, J. K. (1998) The fundamental importance of human galactose metabolism: lessons from genetics and biochemistry, *Trends Genet*, **14**, 98-102.

Power, S. B. (1978) *Streptococcus suis* type 2 infection in pigs, *Vet Rec*, **102**, 215-6.

Pritchard, D. G., Gregory, R. L., Michalek, S. M. and McGhee, J. R. (1986) Biochemical aspects of serotype carbohydrate antigens of *Streptococcus mutans*. In *Molecular microbiology and immunology of Streptococcus mutans* (Hamada, S., Michalek, S. M., Kiyono, H., Menaker, L. and McGhee, J. R., eds.), pp. 39-49. Elsevier Science Publishers, Amsterdam.

Rafferty, J. B., Simon, J. W., Baldock, C., Artymiuk, P. J., Baker, P. J., Stuitje, A. R., Slabas, A. R. and Rice, D. W. (1995) Common themes in redox chemistry emerge from the X-ray structure of oilseed rape (*Brassica napus*) enoyl acyl carrier protein reductase, *Structure*, **3**, 927-38.

Ramakrishnan, C. and Ramachandran, G. N. (1965) Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units, *Biophys J*, **5**, 909-33.

Read, R. J. (1986) Improved Fourier Coefficients from Maps Using Phases from Partial Structures with Errors, *Acta Crystallogr A*, **42**, 140-149.

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.

Rizzi, M., Tonetti, M., Vigevani, P., Sturla, L., Bisso, A., Flora, A. D., Bordo, D. and Bolognesi, M. (1998) GDP-4-keto-6-deoxy-D-mannose epimerase/reductase from *Escherichia coli*, a key enzyme in the biosynthesis of GDP-L-fucose, displays the structural characteristics of the RED protein homology superfamily, *Structure*, **6**, 1453-65.

Romana, L. K., Santiago, F. S. and Reeves, P. R. (1991) High level expression and purification of dthymidine diphospho-D- glucose 4,6-dehydratase (rfbB) from *Salmonella* serovar typhimurium LT2, *Biochem Biophys Res Commun*, **174**, 846-52.

Rossmann, M. G., Liljas, A., Branden, C. I. and Banaszak, L. J. (1975) Evolutionary and structural relationships among dehydrogenases. In *The Enzymes* 3rd edit. (Boyer, P. D., ed.), Vol. 11, pp. 61-102. Academic Press, New York.

Sadovskaya, I., Brisson, J. R., Thibault, P., Richards, J. C., Lam, J. S. and Altman, E. (2000) Structural characterization of the outer core and the O-chain linkage region of lipopolysaccharide from *Pseudomonas aeruginosa* serotype O5, *Eur J Biochem*, **267**, 1640-50.

Sambrook, J. and Russell, D., W. (2001) *Molecular cloning: a laboratory manual.* Third edit, Cold Spring Harbor Laboratory Press, New York.

Sanders, D. A., Staines, A. G., McMahon, S. A., McNeil, M. R., Whitfield, C. and Naismith, J. H. (2001) UDP-galactopyranose mutase has a novel structure and mechanism, *Nat Struct Biol*, **8**, 858-63.

Sandlin, R. C., Goldberg, M. B. and Maurelli, A. T. (1996) Effect of O side-chain length and composition on the virulence of *Shigella flexneri* 2a, *Mol Microbiol*, **22**, 63-73.

Schein, C. H. and Noteborn, M. H. M. (1988) Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperature, *Bio/Technology*, **6**, 291-294.

Schwartz, M. F. and Jornvall, H. (1976) Structural analyses of mutant and wild-type alcohol dehydrogenases from *Drosophila melanogaster*, *Eur J Biochem*, **68**, 159-68.

Seyfarth, A. M., Wegener, H. C. and Frimodt-Moller, N. (1997) Antimicrobial resistance in *Salmonella enterica* subsp. enterica serovar typhimurium from humans and production animals, *J Antimicrob Chemother*, **40**, 67-75.

Sleigh, D. J. and Timbury, M. C. (1998) *Notes on Medical Microbiology*. 5th edit, Churchill Livingstone, Edinburgh.

Smith, H. E., Damman, M., van der Velde, J., Wagenaar, F., Wisselink, H. J., Stockhofe-Zurwieden, N. and Smits, M. A. (1999) Identification and characterization of the cps locus of *Streptococcus suis* serotype 2: the capsule protects against phagocytosis and is an important virulence factor, *Infect Immun*, **67**, 1750-6.

Snipes, C. E., Brillinger, G. U., Sellers, L., Mascaro, L. and Floss, H. G. (1977) Stereochemistry of the dTDP-glucose oxidoreductase reaction, *J Biol Chem*, **252**, 8113-7.

Sohng, J. K. and Yoo, J.-C. (1996) Cloning, Sequencing and Expression of dTDP-D-Glucose 4,6-dehydratase Gene from *Streptomyces antibioticus* Tu99, a Producer of Chlorothricin., *J. Biochem. Mol. Biol.*, **29**, 183-191.

Somers, W. S., Stahl, M. L. and Sullivan, F. X. (1998) GDP-fucose synthetase from *Escherichia coli*: structure of a unique member of the short-chain dehydrogenase/reductase family that catalyzes two distinct reactions at the same active site, *Structure*, **6**, 1601-12.

Somoza, J. R., Menon, S., Schmidt, H., Joseph-McCarthy, D., Dessen, A., Stahl, M. L., Somers, W. S. and Sullivan, F. X. (2000) Structural and kinetic analysis of *Escherichia coli* GDP-mannose 4,6 dehydratase provides insights into the enzyme's catalytic mechanism and regulation by GDP-fucose, *Structure Fold Des*, **8**, 123-35.

Staats, J. J., Feder, I., Okwumabua, O. and Chengappa, M. M. (1997) *Streptococcus suis*: past and present, *Vet Res Commun*, **21**, 381-407.

Stinson, M. W., Nisengard, R. J. and Bergey, E. J. (1980) Binding of streptococcal antigens to muscle tissue in vitro, *Infect Immun*, **27**, 604-13.

Stinson, M. W., Barua, P. K., Bergey, E. J., Nisengard, R. J., Neiders, M. E. and Albini, B. (1984) Binding of *Streptococcus mutans* antigens to heart and kidney basement membranes, *Infect Immun*, **46**, 145-51.

Sturla, L., Etzioni, A., Bisso, A., Zanardi, D., De Flora, G., Silengo, L., De Flora, A. and Tonetti, M. (1998) Defective intracellular activity of GDP-D-mannose-4,6-dehydratase in leukocyte adhesion deficiency type II syndrome, *FEBS Lett*, **429**, 274-8.

Swanson, B. A. and Frey, P. A. (1993) Identification of lysine 153 as a functionally important residue in UDP-galactose 4-epimerase from *Escherichia coli*, *Biochemistry*, **32**, 13231-6.

Tanner, M. E. (2001) Sugar Nucleotide-Modifying Enzymes, *Curr Org Chem*, **5**, 169-192.

Tayoro, J., Besnier, J. M., Laudat, P., Cattier, B. and Choutet, P. (1996) Infective endocarditis due to *Streptococcus suis* serotype 2 [letter], *Eur J Clin Microbiol Infect Dis*, **15**, 765-6.

Thatcher, D. R. and Sawyer, L. (1980) Secondary-structure prediction from the sequence of *Drosophila melanogaster* (fruitfly) alcohol dehydrogenase, *Biochem J*, **187**, 884-6.

Thoden, J. B., Frey, P. A. and Holden, H. M. (1996a) High-resolution X-ray structure of UDP-galactose 4-epimerase complexed with UDP-phenol, *Protein Sci*, **5**, 2149-61.

Thoden, J. B., Frey, P. A. and Holden, H. M. (1996b) Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from *Escherichia coli*: implications for the catalytic mechanism, *Biochemistry*, **35**, 5137-44.

Thoden, J. B., Frey, P. A. and Holden, H. M. (1996c) Crystal structures of the oxidized and reduced forms of UDP-galactose 4-epimerase isolated from *Escherichia coli*, *Biochemistry*, **35**, 2557-66.

Thoden, J. B., Gulick, A. M. and Holden, H. M. (1997a) Molecular structures of the S124A, S124T, and S124V site-directed mutants of UDP-galactose 4-epimerase from *Escherichia coli*, *Biochemistry*, **36**, 10685-95.

Thoden, J. B., Hegeman, A. D., Wesenberg, G., Chapeau, M. C., Frey, P. A. and Holden, H. M. (1997b) Structural analysis of UDP-sugar binding to UDP-galactose 4-epimerase from *Escherichia coli*, *Biochemistry*, **36**, 6294-304.

Thoden, J. B. and Holden, H. M. (1998) Dramatic differences in the binding of UDP-galactose and UDP-glucose to UDP-galactose 4-epimerase from *Escherichia coli*, *Biochemistry*, **37**, 11469-77.

Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L. and Holden, H. M. (2000) Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase, *Biochemistry*, **39**, 5691-701.

Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L. and Holden, H. M. (2001) Human UDP-Galactose 4-Epimerase: Accommodation of UDP-*N*-Acetylglucosamine within the active site, *J Biol Chem*, **276**, 15131-15136.

Thompson, M. W., Strohl, W. R. and Floss, H. G. (1992) Purification and characterization of TDP-D-glucose 4,6-dehydratase from anthracycline-producing streptomycetes, *J Gen Microbiol*, **138**, 779-86.

Threlfall, E. J., Rowe, B. and Ward, L. R. (1993) A comparison of multiple drug resistance in salmonellas from humans and food animals in England and Wales, 1981 and 1990, *Epidemiol Infect*, **111**, 189-97.

Tonetti, M., Sturla, L., Bisso, A., Zanardi, D., Benatti, U. and De Flora, A. (1998) The metabolism of 6-deoxyhexoses in bacterial and animal cells, *Biochimie*, **80**, 923-31.

Tsuda, H., Yamashita, Y., Toyoshima, K., Yamaguchi, N., Oho, T., Nakano, Y., Nagata, K. and Koga, T. (2000) Role of serotype-specific polysaccharide in the resistance of *Streptococcus mutans* to phagocytosis by human polymorphonuclear leukocytes, *Infect Immun*, **68**, 644-50.

Tsukioka, Y., Yamashita, Y., Oho, T., Nakano, Y. and Koga, T. (1997) Biological function of the dTDP-rhamnose synthesis pathway in *Streptococcus mutans*, *J Bacteriol*, **179**, 1126-34.

Turgeon, P. L., Higgins, R., Gottschalk, M. and Beaudoin, M. (1994) Antimicrobial susceptibility of *Streptococcus suis* isolates, *Br Vet J*, **150**, 263-9.

Van den Bosch, L., Manning, P. A. and Morona, R. (1997) Regulation of O-antigen chain length is required for *Shigella flexneri* virulence, *Mol Microbiol*, **23**, 765-75.

Vara, J. A. and Hutchinson, C. R. (1988) Purification of thymidine-diphospho-D-glucose 4,6-dehydratase from an erythromycin-producing strain of *Saccharopolyspora erythraea* by high resolution liquid chromatography, *J Biol Chem*, **263**, 14992-5.

Varughese, K. I., Xuong, N. H., Kiefer, P. M., Matthews, D. A. and Whiteley, J. M. (1994) Structural and mechanistic characteristics of dihydropteridine reductase: a member of the Tyr-(Xaa)3-Lys-containing family of reductases and dehydrogenases, *Proc Natl Acad Sci U S A*, **91**, 5582-6.

Vazquez-Torres, A., Jones-Carson, J., Baumler, A. J., Falkow, S., Valdivia, R., Brown, W., Le, M., Berggren, R., Parks, W. T. and Fang, F. C. (1999) Extraintestinal dissemination of Salmonella by CD18-expressing phagocytes, *Nature*, **401**, 804-8.

Vecht, U., Wisselink, H. J., Jellema, M. L. and Smith, H. E. (1991) Identification of two proteins associated with virulence of *Streptococcus suis* type 2, *Infect Immun*, **59**, 3156-62.

Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program, *J Mol Graph*, **8**, 52-6, 29.

Wallace, A. C., Laskowski, R. A. and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, *Protein Eng*, **8**, 127-34.

Wang, S. F. and Gabriel, O. (1969) Biological mechanisms involved in the formation of deoxy sugars. V. Isolation and crystallization of thymidine diphosphate-D-glucose oxidoreductase from *Escherichia coli* B, *J Biol Chem*, **244**, 3430-7.

Wang, S. F. and Gabriel, O. (1970) Biological mechanisms involved in the formation of deoxysugars. VI. Role and function of enzyme-bound nicotinamide adenine dinucleotide in thymidine diphosphate D-glucose oxidoreductase, *J Biol Chem*, **245**, 8-14.

Wee, T. G., Davis, J. and Frey, P. A. (1972) Studies on the mechanism of action of uridine diphosphate-galactose-4- epimerase. I. An ambiguity in the chemical trapping of a proposed keto- intermediate by NaB-3H4, *J Biol Chem*, **247**, 1339-42.

Wee, T. G. and Frey, P. A. (1973) Studies on the mechanism of action of uridine diphosphate galactose 4- epimerase. II. Substrate-dependent reduction by sodium borohydride, *J Biol Chem*, **248**, 33-40.

Whitfield, C. and Valvano, M. A. (1993) Biosynthesis and expression of cell-surface polysaccharides in gram- negative bacteria, *Adv Microb Physiol*, **35**, 135-246.

Whitfield, C. (1995) Biosynthesis of lipopolysaccharide O antigens, *Trends Microbiol*, **3**, 178-85.

Williams, A. E. (1990) Relationship between intracellular survival in macrophages and pathogenicity of *Streptococcus suis* type 2 isolates, *Microb Pathog*, **8**, 189-96.

Williams, N. and Wander, J. (1980) Deoxy and Branched-chain Sugars. In *The Carbohydrates: Chemistry and Biochemistry* (Pigman, W. and Horton, D., eds.), Vol. 1B, pp. 761-798. Academic Press, New York.

Windsor, R. S. and Elliott, S. D. (1975) Streptococcal infection in young pigs. IV. An outbreak of streptococcal meningitis in weaned pigs, *J Hyg (Lond)*, **75**, 69-78.

Wong, S. S. and Frey, P. A. (1977) Fluorescence and nucleotide binding properties of *Escherichia coli* uridine diphosphate galactose 4-epimerase: support for a model for nonstereospedific action, *Biochemistry*, **16**, 298-305.

Xiang, S. H., Haase, A. M. and Reeves, P. R. (1993) Variation of the *rfb* gene clusters in *Salmonella enterica*, *J Bacteriol*, **175**, 4877-84.

Zarkowsky, H., Lipkin, E. and Glaser, L. (1970) The mechanism of 6-deoxyhexose synthesis IV. The role of pyridine nucleotide in substrate release, *Biochem Biophys Res Commun*, **38**, 787-93.