Molecular recognition in proteins

A thesis submitted for the degree of Doctor of philosophy

Iain Kerr

June 2003

Declarations

I, Iain Kerr, hereby certify that this thesis, which is approximately 35,500 words in length, has been written by me, that it is a record of work carried out by me and that it has not been submitted in any previous application for a higher degree

Date 25 June 2003 Signature of Candidate.

I was admitted as a research student in October 1999 and as a candidate for the degree of Ph.D. in October 2000; the higher study for which this is a record was carried out at the University of St.Andrews between 1999 and 2003

Date 25 June 2003 Signature of Candidate..

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

Date 25 JUNE 03 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 25 June 2003 Signature of candidate

Abbreviations

Amino acids

Ala (A) - alanine

Arg (R) - arginine

Asn (N) - asparagine

Asp (D) - aspartic acid

Cys (C) - cysteine

Gln (Q) - glutamine

Glycine (G) - glycine

Glu (E) - glutamic acid

His (H) - histidine

Ile (I) - isoleucine

Leu (L) - leucine

Lys (K) - lysine

Met (M) - methionine

Phe (F) - phenylalanine

Pro (P) - proline

Ser (S) - serine

Thr (T) - threonine

Tyr (Y) - tyrosine

Trp (W) - tryptophan

Val (V) - valine

ADP - adenosine diphosphate

AIDS - acquired immune deficiency syndrome

ARI - AIDS related illness

ATP - adenosine triphosphate

C - cytosine

CCD - charged coupled device

CDP - cytidine diphosphate

CDP-ME - 4-diphosphocytidyl-2-C-methylerythritol

CDS - Canada's drug strategy

CTP - cytidine triphosphate

DBD - DNA binding domain

DMAP - dimethylallyl diphosphate

d_{min} - diffraction limit in high resolution shells

DNA - Deoxyribonucleic acid

DTT - dithiothreitol

EDTA - ethylenediaminetetra-aceticacid

ENR - enoyl-acyl carrier protein reductase

G - guanine

GDP - guanosine diphosphate

GlcNAc - N-acetyl glucosamine

HPLC - high pressure liquid chromotography

IPP - isopentenyl diphosphate

IPTG - isopropyl-β-D-thiogalactopyranoside

IR - isomorphous replacement

K_{cat} - the turnover number

KDO - 2-keto-3-deoxyoctonic acid

K_m - the Michaelis constant

LBHB - low barrier hydrogen bond

LPS - lipopolysaccharide

LUCA - last univeral cellular ancestor

MAC - membrane attack complex

MAD - multiwavelength anomalous diffraction/dispersion

MDR - multi-drug resistant

MEP - 2-C-methyl-D-erythtitol-4-phosphate

MRSA - methicillin resistant *Staphylococcus aureus*

mRNA - messenger ribonucleic acid

NAD^+ - nicotinamide adenine dinucleotide (oxidised form)

NADH - nicotinamide adenine dinucleotide (reduced form)

$NADP^+$ - nicotinamide adenine dinucleotide phosphate (oxidised form)

NADPH - nicotinamide adenine dinucleotide phosphate (reduced form)

NCS - non-crystallographic symmetry

NER - nucleotide excision repair

OB - oligonucleotide/oligopeptide/oligosaccharide

PBP - penicillin binding proteins

PCR - polymerase chain reaction

PDB - protein data bank

PMSF - phenylmethyl sulphonyl fluoride

PPi - pyrophosphate

RPA - replication protein A

rRNA - ribosomal RNA

SAD - single wavelength anoamlous diffraction/dispersion

SDR - short-chain dehydrogenase/reducatse

SDS-PAGE - sodium dodecyl sulphate polyacrylamide gel electrophoresis

SSB - single stranded DNA binding protein

ssDNA - single stranded DNA

*Sso*SSB - *Sulfolobus solfataricus* SSB

TB - tuberculosis

TBP - TATA-box binding protein

TFB - transcription factor B

TNF α - tumour necrosis factor alpha

tRNA - transfer RNA

TTP - thymidine triphosphate

TYE - tryptone yeast extract

UDP - undecaprenol

V_m - the Matthew's coefficient

V_{max} - maximum velocity of an enzymatic reaction following michaelis-menten kinetics

VRSA - vancomycin resistant *Staphylococcus aureus*

WHO - World Health Organisation

XP - xeroderma pigmentosum

Abstract

The elucidation of molecular recognition by proteins is critical in understanding enzyme mechanism, inhibitor design, and the functional significance of structural, non-enzymatic proteins. The work described herein deals with molecular recognition by the enzyme RmlD from the pathogenic *Salmonella typhimurium* and SSB from the crenarchaeote *Sulfolobus solfataricus*, a structural protein important in maintaining DNA integrity in the cell.

RmlD is last of four enzymes involved in rhamnose biosynthesis. The production of rhamnose is an appealing drug target. The sugar is exclusively synthesised by bacteria and is an essential cell wall component of many pathogenic species. RmlD and other enzymes in the rhamnose pathway are therefore promising templates against which novel antibacterials can be developed. The structure of the apo-enzyme, a complex with NADH and an abortive dTDP-L-rhamnose – NADPH complex are discussed and these were used to infer residues involved in catalysis and substrate binding. These residues were mutated and assayed for enzymatic activity compared to the wild-type protein.

SSBs are ubiquitous proteins that protect single stranded regions of the DNA duplex exposed during replication, transcription, recombination and DNA damage. The high resolution crystal structure of *Sulfolobus* SSB and its similarity to the DNA-binding domains of the human homologue, RPA, validate the archaeal information processing pathways as a good model in which to study the more complex eukaryotic systems. A model of *Sulfolobus* SSB in complex with ssDNA has been generated. Residues involved in ssDNA-binding have been

inferred from the model and their importance confirmed through biophysical means.

Acknowledgements

Firstly, I thank my supervisor Prof. Jim Naismith for all his help and guidance during my Ph.D and instilling in me his enthusiasm for crystallography. He also sprung for drinks and meals occasionally, which was pretty decent of him too.

I've had a lot of help and advice during my Ph.D.and I have a lot of people to thank for that, most importantly Drs Wulf Blankenfeldt, Marie-France Giraud, Huanting Lui, Miryam Ascuncion and ChangJiang Dong. I particularly thank Wulf for helping me initially get to grips with UNIX and crystallography software. Thanks also to Ross Wadsworth for his friendship and supplying me with all that pure SSB protein !.

My first year in St.Andrews was a time I'll never forget, partly because I lived in the worst university accomodation in Britain, but mostly because I made wonderful friends in both Andrew Hagen and John Rouvalis. In the latter stages of my Ph.D. I was very fortunate to meet Kostas Beis, Simon Newstead, Gareth Williams, Jo Parker and Alex Merkel. Thanks for all the great times and memories guys !. Simon and Kostas also provided critical comments and proofreading which I ignored, but thanks anyway guys !.

Lastly I'd like to thank my fiance Helen Walden for basically putting up with me. Hopefully we can get to know each other again and begin our lives together.

Contents

Chapter 1

Introduction

Summary

Antibiotic resistance, "the acquired ability of a micro-organism to grow in the presence of an antibiotic to which it is usually sensitive" (Madigan, Martinko et al. 1997), is a growing problem in human and veterinary medicine. The need for novel antibacterial chemotherapy means that research in the field of antibiotic resistance has shifted to unique targets. Many of these targets are not found in mammals or are synthesised by different pathways, allowing the development of drugs with selective toxicity to the invading pathogen. Structure-function relationships, elucidated through a combination of X-ray crystallography and biochemical techniques, are important in identifying and determining the validity of new targets. The information obtained on enzyme mechanism by probing active sites with substrates and their analogues is invaluable. Crystals containing protein-inhibitor complexes can tell us how the drug would bind in three-dimensional space, an aspect important in determining the interactions that allow inhibitors to bind preferentially over the natural substrate. This introductory chapter includes an overview of antibiotic resistance (sections 1.1-1.2) and the impact of crystallography on this growing problem (section 1.3-1.4), focusing on a biosynthetic pathway in *eubacteria* as a target for therapeutic intervention and an example of how crystallography has been able to determine the basis of enzyme inhibition by a drug.

The second part of this chapter is concerned with the close phylogenetic relationship between archaea and eukaryotes (section 1.5). In this section, discussion will be focused on the recognition of the archaea as a domain of life separate from *eubacteria* and the relationship between archaeal and eukaryotic systems. The remarkable similarities between eukaryotic proteins and their

archaeal homologues will also be discussed showing how the determination of archaeal protein structures has shed light on the more complex eukaryotic systems.

1.1 The threat of infectious disease

Over millions of years bacteria have adapted to exist in a variety of environmental and biological niches. Although relatively non-fastidious, bacteria like all organisms require the raw materials with which to generate energy and survive. With the appearance of higher, multi-cellular organisms and eukaryotic-prokaryotic symbiosis, pathogenic bacteria have evolved, elaborating an arsenal of extracellular factors that promote host colonisation and the release of essential nutrients either by the subversion of complex eukaryotic biochemistry or direct cellular damage.

Infectious disease spread by pathogenic bacteria (and other micro-organisms) is a leading cause of mortality and morbidity throughout the world. It was estimated that in the year 2000, infectious and parasitic diseases accounted for almost 19% of all deaths worldwide (WHO 2001). Developing countries are worst affected where there is insufficient infrastructure to facilitate treatment or prevention (*Fig.1.1.1*).

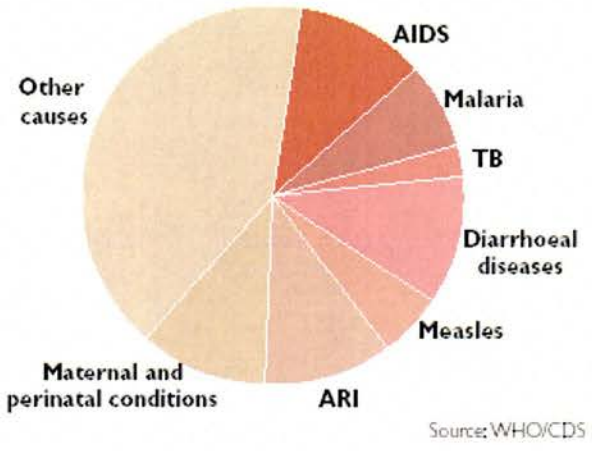

Fig.1.1.1 Causes of death among children and young adults in Africa and Southeast Asia. Reprinted from (WHO/CDS 2002)

The re-emergence of highly communicable diseases such as tuberculosis (TB) is a major cause for concern, prompting the establishment of a Global TB Monitoring and Surveillance Project by the World Health Organisation (WHO). It is projected that 1.7million people die of the disease each year with 8.8million developing active TB over the same time scale (WHO/CDS 2002). Although 2 billion worldwide carry the causative agent, *Mycobacterium tuberculosis* can be contained effectively by alveolar macrophages in the lung. However, the growing AIDS (Acquired Immune Deficiency Syndrome) pandemic makes TB infection more difficult to control. An immunosuppressive event, such as the onset of AIDS, activates the organism which quickly spreads and patients develop rapid degeneration of the lung tissue (caseous necrosis), resulting in eventual death. If the treatment and control of TB does not improve significantly, 35 million people will die of the disease within the next 20 years (WHO/CDS 2002).

The economic burden associated with the ever increasing incidence of infectious disease is also a key consideration. To those in developing countries,

the result of continual bouts of illness can be crippling, seriously affecting the provision of food and family income. The total cost of TB infection and related deaths to the poor is estimated to be in the region of $12 billion per year (WHO/CDS 2002).

1.2 Antibiotic resistance

Widespread dissemination of antibiotic resistance is the primary factor responsible for the growing number of deaths and severe illness attributable to bacterial infectious disease. The Golden Age of medicine, during which a wide range of antibiotics were discovered and derivatised, was initiated by the re-discovery of Penicillin by Alexander Fleming in 1928. Noticing lysis of bacterial colonies occurred next to a contaminant mould in an old petri dish, Fleming hypothesised and later proved that the mould produced a substance that perfused through the growth medium, killing the cells (Fleming 1929). This was to be the first of a number of "Magic Bullets" that selectively killed bacteria, leaving the host unharmed. However, Fleming noted that some species of bacteria were unaffected (Fleming 1929). By 1940, Chain and Abraham had shown that bacteria were capable of secreting a product that destroyed penicillin (Abraham and Chain 1940). In a New York Times article in 1945 Fleming warned of the misuse of penicillin saying "The greatest possibility of evil in self-medication is the use of too small doses so that instead of clearing up infection the microbes are educated to resist penicillin and a host of penicillin-fast organisms is bred out which can be passed to other individuals and from them to others until they reach someone who gets a septicaemia or pneumonia which penicillin cannot save."

It is now clear that antibiotic resistance disseminates through the accumulation of adaptive mutations that give resistant strains a selective advantage over sensitive ones (Morris, Kellner et al. 1998). The spread of resistance, mediated by the acquisition of resistance genes, is controlled by three mechanisms: Conjugation - the transfer of genetic material through close cell to cell contact, Transduction - the transfer of resistance gene carrying-bacteriophage from one bacterium to another, and Transformation - the uptake of DNA by competent bacteria (Morris, Kellner et al. 1998). Antibiotic resistance genes may be carried on circular DNA such as plasmids or parasitic stretches of DNA such as transposons and integrons. Both plasmids and transposons may spread resistance determinants by conjugation. Transposons are particularly efficient because they are self-transmissible, encoding factors that allow them to excise themselves and integrate into plasmids or chromosomal DNA. Integrons are involved in the generation of large resistance gene clusters.

As Fleming warned, misuse of antibiotics in human and veterinary medicine has generated selection pressures that favour the spread of resistant strains far quicker than would occur along normal evolution through spontaneous mutation (Morris, Kellner et al. 1998; Guillemot 1999; WHO 2001). Additionally, the prescription of antibiotics for viral illnesses, including the common cold, is still widely practiced even though these drugs specifically target bacteria, and are therefore completely ineffective against viruses. In 1999, it was reported that more than 20% of antibiotic prescription to adults in the developed world was for viral respiratory diseases (Guillemot 1999). In one study, greater than 70% of infants had used at least one antibiotic only 200 days after birth. Clearly this pattern has to change.

The principal mechanisms of antibiotic resistance can be grouped into four main categories: modification of the drug, efflux-mediated antibiotic resistance, mutation of the drug target and cell wall impermeability. The latter mechanism is dealt with later in the text (sections 2.3.2 and 2.3.3).

1.2.1 Chemical modification of antibacterials

Molecules that are able to modify antibacterials are represented by two distinct groups, β-lactamases and aminoglycoside-modifying enzymes. As the name suggests, β-lactamases confer resistance to β-lactam antibiotics including penicillin, cephalosporins and ampicillin (Massova and Mobashery 1998) and are used to treat many common/severe bacterial infections including those of the upper respiratory tract, heart (endocarditis), brain (meningitis) and urinary tract (Mims, Playfair et al. 1993). β-lactams bind to and inactivate Penicillin Binding Proteins (PBPs) in the cell walls of gram-negative and gram-positive bacteria, by forming a stable acyl -enzyme complex between the β-lactam ring and the nucleophilic, active site serine (Massova and Mobashery 1998). PBPs are implicated in a number of processes involved in cell wall biosynthesis, including the transpeptidation reaction during peptidoglycan formation. β-lactamases, which are related to PBPs (Kelly, Dideberg et al. 1986), catalyse the hydrolysis of the lactam ring (*Fig. 1.2.1.1*).

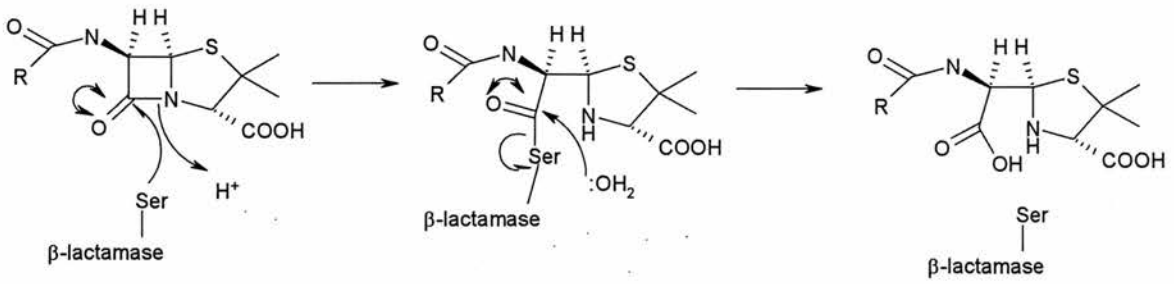

Fig.1.2.1.1 Inactivation of penicillin by a β-lactamase.

Formation of the acyl-enzyme complex in the β-lactamase active site (through a nucleophilic serine) is followed by de-acylation, cleaving the C-N bond in the β-lactam ring. By losing the conformation of the ring, the resulting compound is sterically hindered to undergo the initial nucleophilic attack at the acyl carbon by PBPs and is therefore a poor inhibitor of cell wall synthesis

Aminoglycosides are hydrophilic sugars that irreversibly bind prokaryotic ribosomal RNA (rRNA) interfering with translation (Kotra, Haddad et al. 2000). Kanamycin, gentamicin and neomycin bind the A-site, interfering with tRNA recognition by the ribosome. Streptomycin (used in TB treatment) causes misreading of mRNA codons. Aminoglycoside-modyfying enzymes catalyse the transfer of phosphate, nucleotide and acetyl moieties to the various sugar functional groups of this class of antibiotics, sterically hindering their interaction with rRNA (Kotra, Haddad et al. 2000).

1.2.2 Drug Efflux

This mechanism is employed by both gram-negative and gram-positive species. Efflux pumps are trans-membrane proteins that mediate the expulsion of antibacterials, detergents and solvents from the cell. Many transporters have the

ability to extrude several different types of chemically unrelated compounds (reviewed in (Borges-Walmsley and Walmsley 2001; Markham and Neyfakh 2001)).

TetL and Tet K, the tetracycline-specific transporters of *Bacillus subtilis* and *Staphylococcus aureus* are known to transport Na^+ and K^+ ions (Krulwich, Jin et al. 2001). Conversely, the MefA and MefE transporters expressed by *Streptococcus pneumoniae* are specific for macrolide antibiotics. The importance of these pumps is underscored by the fact that their presence in the cell wall leads to an almost 64-fold increase in macrolide resistance in *S. pneumoniae* (Sutcliffe, Tait-Kamradt et al. 1996)

In gram negatives, the major class of drug pumps are the proton-driven antiporters (Borges-Walmsley and Walmsley 2001). *E. coli* possesses over twenty such proteins. Drugs are expelled from the cytoplasm into the periplasmic space using proton-motive force generated during respiration (Borges-Walmsley and Walmsley 2001). Additional membrane proteins, such as TolC are thought to conclude expulsion of drugs from the cell.

1.2.3 Mutation of the drug target

Mutations in bacterial ribosomal proteins are responsible for resistance to a number of clinically important antibiotics. Examples include erythromycin (Chittum and Champney 1994), used in the treatment of atypical pneumonia (eg. *Legionella pneumophila*), and the aminoglycosides, spectinomycin (Bilgin, Richter et al. 1990) and streptomycin (Allen and Noller 1989) used in the treatment of various gram-positive infections. The effect of these antibiotics is to ultimately block bacterial protein synthesis (the aminoglycosides have been

discussed above). *E. coli* resistance to erythromycin, which blocks the release of tRNA from the ribosome, is due to the deletion of only three residues (MKR) in the ribosomal L22 protein (Chittum and Champney 1994). Recently this mutation was confirmed by mass spectrometry (Wilcox, Cavey et al. 2001). Similarly, mutations in ribosomal proteins from strains resistant to spectinomycin and streptomycin were identified. A change in the mass of ribosomal proteins from resistant strains was able to confirm that mutation of these proteins is important in antibiotic resistance (Wilcox, Cavey et al. 2001). Analysis of the mutated proteins following tryptic digestion was able to pinpoint the site of the mutations.

1.2.4 Multiple-drug resistance

The rise of multi-drug resistant (MDR) strains in a number of species is particularly worrying, with some only responsive to a few antibacterials. *Salmonella typhimurium* DT104, a common cause of mortality and morbidity worldwide, is one of the most common antibiotic resistant phage types (Briggs and Fratamico 1999). This species is increasingly common in immunocompromised/AIDS patients underscoring the need for effective therapeutics (Gallardo, Ruiz et al. 1999). Readily isolated from the rumens of cattle, strain DT104 has been documented to harbour large gene clusters conferring resistance to ampicillin, sulfonamides, tetracycline, streptomycin and chloramphenicol (Gallardo, Ruiz et al. 1999). Strains only resistant to one or two antibiotics have increased as much as 5-fold over a 7 year period (Gallardo, Ruiz et al. 1999). MDR TB is classically resistant to Isoniazid and rifampicin, the first line drugs used in TB combination-chemotherapy. The treatment of MDR-TB

costs hundreds of times more than that used for conventional TB (Dye. C 2002) and its spread is classified as more deadly than AIDS, an illness for which there is still no effective treatment.

The most acute danger posed by the growth and spread of antibacterial resistance is the appearance of bacterial species for which almost no effective chemotherapy exists. The first signs appeared only six years ago after the report of an isolate of *Staphylococcus aureus* resistant to vancomycin (VRSA) in Japan (Centres for Disease Control and prevention 1997). Prior to this, vancomycin was considered the last line of defence for patients infected with methicillin-resistant S. *aureus* (MRSA). Subsequent reports of VRSA in the US suggest that the focus of infection is not centered solely on Japan and is likely to spread further (Centres for Disease Control and prevention 1997).

1.2.4 The search for new antibacterials

The continued development of new therapeutics is essential to prevent the spread of infectious disease and emergence of resistance to current antibacterials. However, interest in the pharmaceutical industry has shifted to the development of "lifestyle" drugs (viagra, anti-obesity), which sell to a much broader market and with more frequency than antibacterials do (WHO 2001). The major problem lies with the lack of incentive given to industry to develop new drugs. There is no money to be made from poorer countries in the developing world and the call for the appropriate use of antibiotics is seen to have a negative impact on sales, prompting many companies to shut down drug development programmes altogether (WHO 2001).

Those that do continue research into developing antibacterials agree that it is time to focus on new targets. Oxazolidinones, the only new class of antibiotics to be recently developed still rely on an old target, protein synthesis (Allsop 1998). With the completion of many bacterial genomes in place, research is being driven away from analogues of old classes and towards the development of agents specific for completely novel targets (Allsop 1998). Of course, resistance will eventually develop against these drugs. This is fundamental bacterial genetics. However, if proper control is used in the prescription of these drugs, they should provide a window of effective treatment till a more suitable solution can be determined.

1.3 The non-mevalonate pathway as a target for therapeutic intervention

Isoprenes such as isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are the five-carbon precursors of isoprenoids, a diverse group of natural products. In eukaryotes the products of isoprenoid biosynthesis include steroids, sterols and chlorophylls (Spurgeon and Porter 1981). Bacterial isoprenoids include ubiquinones and undecaprenol (UDP), a lipid carrier involved in cell wall biosynthesis (Reusch 1984; Neidhardt, Curtis III et al. 1996). In vertebrates, isoprene biosynthesis is carried out by a mevalonate-dependant pathway starting from 2-acetyl-CoA (Goldstein and Brown 1990). Some bacterial species, such as *E. coli* and *Mycobacterium tuberculosis*, utilise a mevalonate-independent pathway (non-mevalonate) for isoprene biosynthesis using pyruvate and D-glyceraldehyde-3-phosphate as the initial substrates (*Fig*

1.3.1). The presence of such a divergent pathway in bacteria makes the enzymes in the mevalonate-independent pathway an obvious target for the development of novel antibacterials with selective toxicity.

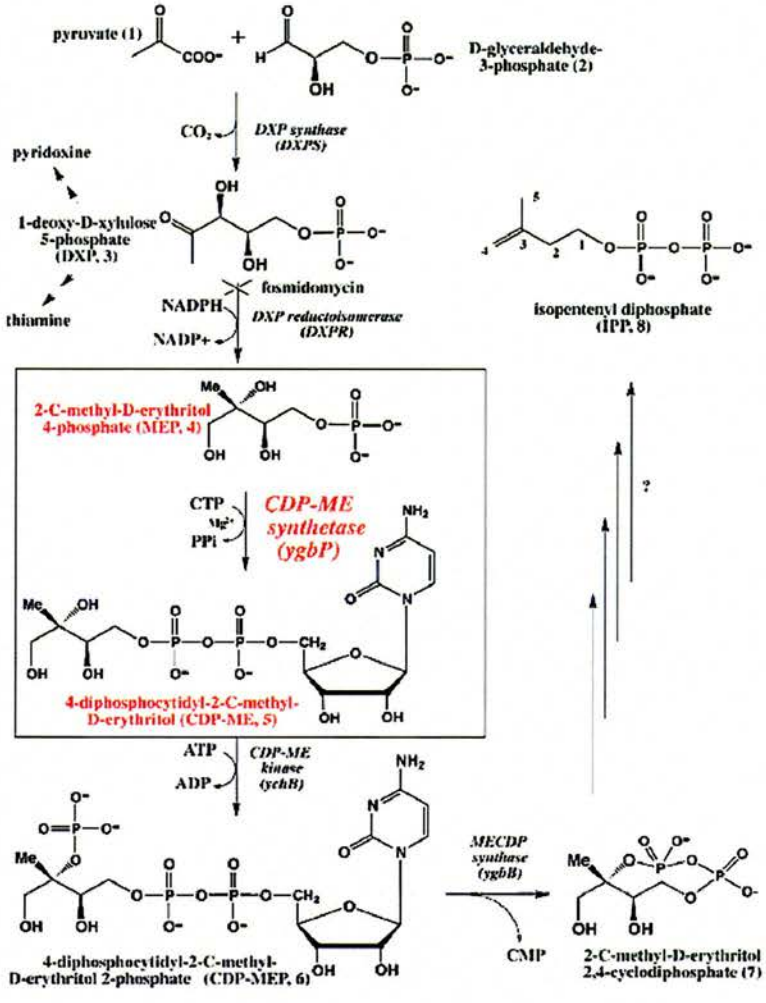

Fig. 1.3.1 The mevalonate-independent pathway. Reprinted from (Richard, Bowman et al. 2001). The last steps, annotated with arrows, remain to be elucidated but are suggested to involve the enzymes GcpE and LytB

The structures of three enzymes in pathway are known - DXP reductoisomerase (DXPR), CDP-ME synthetase (YgbP) and MECDP synthase (YgbB). The former two will be discussed.

1.3.1 DXP reductoisomerase (DXPR)

DXP reductoisomerase (1-deoxy-D-xylulose-5-phosphate reductoisomerase) catalyses the second step in the mevalonate independent pathway, an intramolecular rearrangement followed by an NADPH-mediated reduction at C4 as figure 1.3.1.1

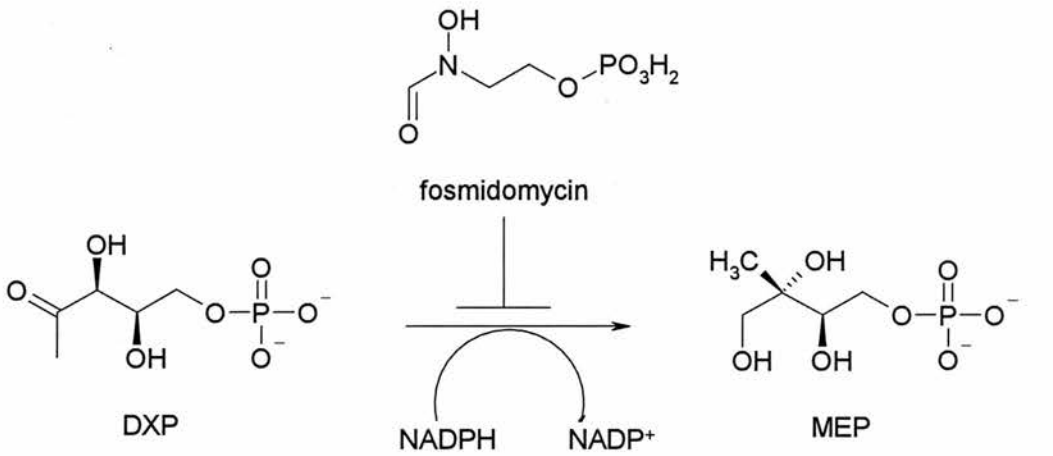

Fig.1.3.1.1 The DXPR reaction scheme. Reprinted from (Richard, Bowman et al. 2001)

The validity of DXPR as a drug target has long been known. The antibiotic fosmidomycin kills *Plasmodium falciparum*, the causal agent of cerebral malaria, and is able to cure mice infected with the rodent malaria parasite (Jomma, Weisner et al. 1999). The structure of DXPR from *E. coli* has been solved by two groups to resolutions of 2.2 and 2.5Å (Reuter, Sanderbrand et al. 2002; Yajima, Nonaka et al. 2002). *Yajima et al* fortuitously trapped a sulphate ion in the active site. Mutation of His209 to Gln, which forms a hydrogen bond with the anion, causes a 52,000 fold drop in k_{cat}/K_m (Kuzuyama, Takahashi et al. 2000). Therefore, the authors concluded that the position occupied by the sulphate may be where the phosphonate of the inhibitor/phosphate of the substrate lie in the

active site. Substrates lacking the phosphate function prevent NADPH oxidation (required for the C4 reduction) suggesting this group is important to anchor the substrate in the active site (Kuzuyama, Takahashi et al. 2000). Structural analysis reveals a nearby flexible loop (residues 209-212) containing Trp212 and His214 that may be important for recognition of the substrate carbon backbone.

Recently the coordinates of DXPR in complex with fosmidomycin were released on the protein data bank, although the primary citation remains to be published. Superimposition of the complex and apo-structures shows that the sulphate ion and phosphonate of the inhibitor occupy almost identical positions. However, a large conformational change in the loop containing His209 means that this residue points out of the active site instead of coordinating the phosphonate of the inhibitor. Trp212 undergoes torsional sidechain adjustments to make a hydrophobic interaction with the inhibitor backbone as proposed by *Yajima et al* (Yajima, Nonaka et al. 2002). Tight recognition of fosmidomycin is achieved through interaction with polar and hydrophilic side chains and two magnesium ions coordinated by a set of acidic residues.

1.3.2 CDP-Me synthetase (YgbP)

CDP-Me synthetase (4-diphosphocytidyl-2-C-methylerythritol synthetase) catalyses the third step in the pathway, the transfer of CTP (cytidine-tri phosphate) to MEP to form CDP-ME and pyrophosphate (*Fig. 1.3.2.1*).

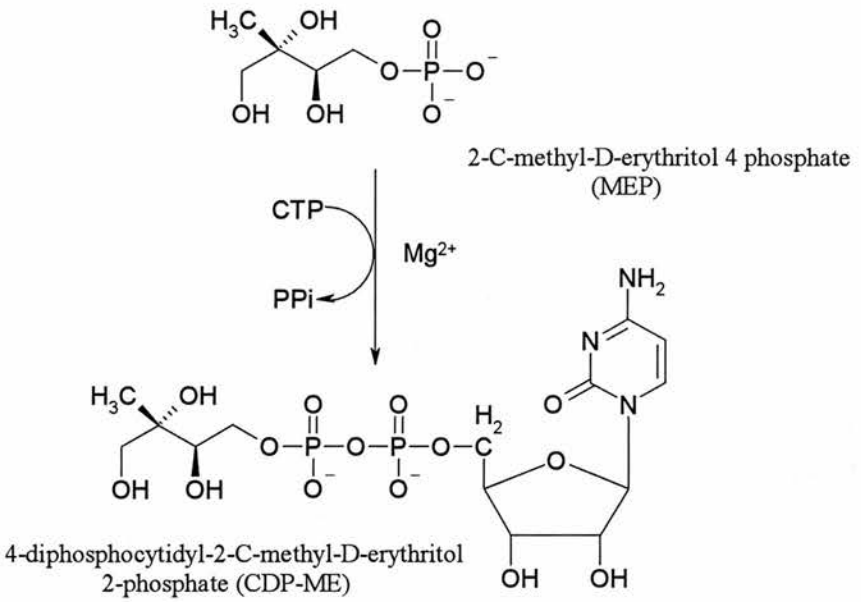

Fig.1.3.2.1 The reaction catalysed by CDP-Me synthetase

The high resolution structures from *E. coli* with substrate (CTP) and product (CDP-ME) give a detailed picture of the interactions required for recognition in the active site (Richard, Bowman et al. 2001) (*Fig.1.3.2.2*). The structure suggests that a bound magnesium ion, which is essential for activity, may be involved in stabilising a penta-coordinated transition state (*Fig.1.3.2.2*). Using this information two important residues (Lys27 and Lys213) were mutated to both alanine and serine (Richard, Bowman et al. 2001). All mutants were seriously compromised in their ability to turnover MEP proving that these are key residues in the active site.

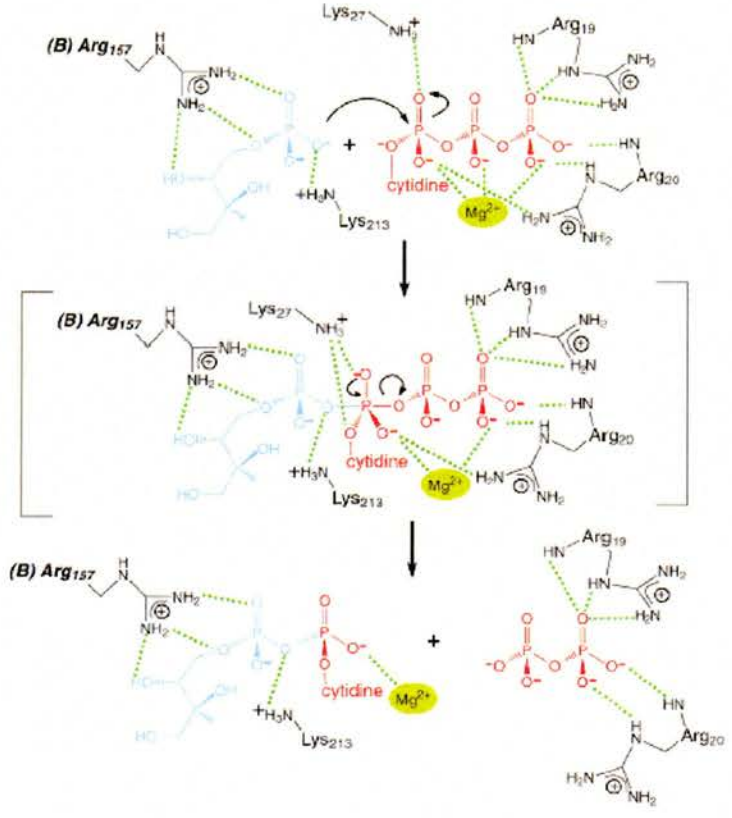

Fig.1.3.2.2 Schematic of a possible reaction mechanism for CDP-Me synthetase inferred from the structural data. Reprinted from (Richard, Bowman et al. 2001). The substrates MEP and CDP are shown in blue and red respectively.

Resolution of the role of these residues in catalysis would not have been possible without prior knowledge of how substrate and product bound in the active site. The structure of the enzyme-CTP complex determined by *Richard et al* shows that the α- phosphate of CTP is positioned for nucleophilic attack on Lys27, facilitating the formation of a covalent bond between CDP and MEP (*Fig.1.3.2.2*). The product complex with CDP-Me shows that Lys213 forms a hydrogen bond with the phosphate of MEP (Richard, Bowman et al. 2001). This may compensate for the negative charge so that the α-phosphate of CTP can approach the substrate to undergo nucleophilic attack.

1.4 Structural basis for the bactericidal properties of Triclosan

Triclosan is a broad-spectrum agent lethal to a number of medically important pathogens including *Staphylococcus aureus*, *Salmonella typhimurium* and *Shigella flexneri*. A large number of "over-the-counter" products contain triclosan including detergents, deodorants, cosmetics, lotions, creams, toothpastes and mouthwashes (http://www.lindachae.com/triclosan.htm). Until recent years, triclosan was believed to be a non-specific biocide (like detergents). *McMurry et al* showed through knockout complementation experiments that enoyl-acyl carrier protein reductase (ENR, FabI in *E. coli*) is the target for triclosan (McMurry, Oethinger et al. 1998). ENR catalyses the terminal step in bacterial (type II) fatty acid synthesis, the reduction of the C2-C3 carbon-carbon double bond in the enoyl moeity of the fatty acid precursor (*Fig.1.4.1*). The ENR substrate is linked to an acyl protein carrier (ACP).

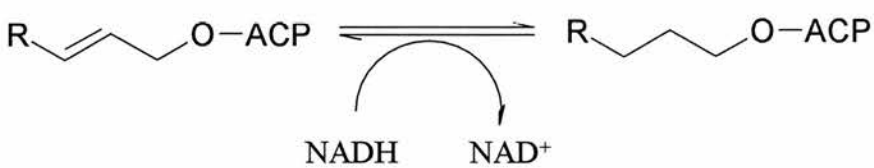

Fig.1.4.1 The ENR reaction

The structures of *E. coli* ENR in ternary complexes with triclosan and NADH (Stewart, Parikh et al. 1999) and triclosan and NAD^+ (Roujeinikova, Levy et al. 1999) have been determined. The two are identical with respect to co-factor and triclosan binding, with an rmsd of 0.4 over 257 α-carbons. The majority of the triclosan interactions with the active site and co-factor are shown in figure 1.4.2.

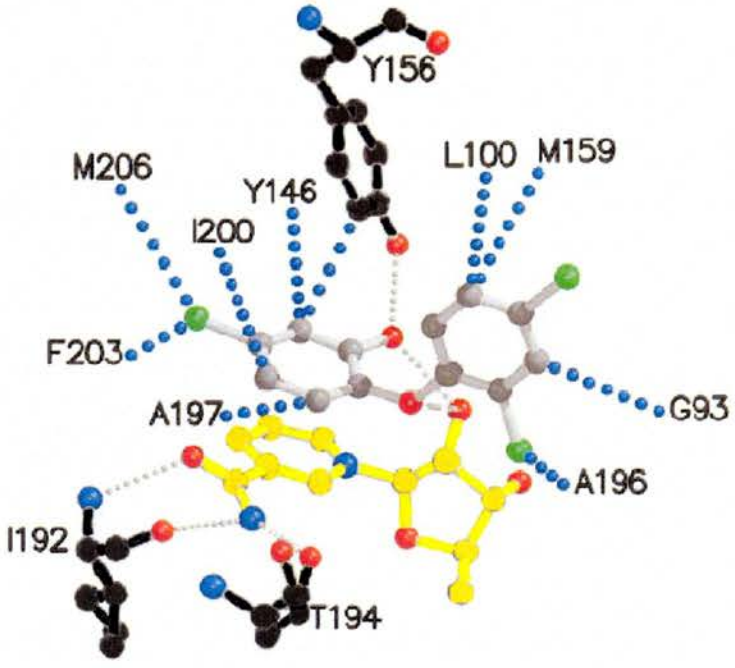

Fig. 1.4.2 Triclosan binding in the active site of ENR. Modified from (Stewart, Parikh et al. 1999). Thick dotted lines show hydrophobic interactions, hydrogen bonds are shown as thin, grey dotted lines.

The position of triclosan in the active site is in good agreement with the position of an enoyl moeity modelled into the active site of *E. coli* ENR by *Baldock et al* (*Fig. 1.4.3*).

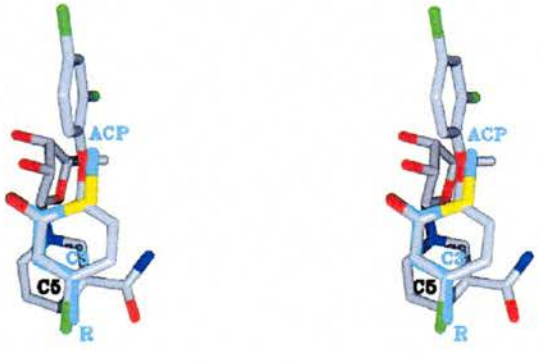

Fig. 1.4.3 Stereo picture showing the relative positions of bound triclosan and modelled substrate in the active site of *E. coli* ENR. Substrate carbons atoms are coloured cyan. Reprinted from (Baldock, Rafferty et al. 1998).

The modelled substrate superimposes almost exactly with C1-C4 on the 4-chloro phenol ring and the bridging oxygen of triclosan (Roujeinikova, Levy et al. 1999). The majority of interactions involved in triclosan recognition are hydrophobic, as expected for a molecule with two bulky non-polar rings (*Fig.1.4.2*). The phenol ring is held in place by π-stacking interactions with the nicotinamide ring of the co-factor. Interestingly, the C2-C3 double bond of the modelled enolate reduced during ENR catalysis, superimposes on the C5-C6 bond belonging to the phenolic ring of triclosan (Roujeinikova, Levy et al. 1999). Triclosan is thus positioned for hydride transfer from C4 of the nicotinamide ring on the co-factor, suggesting that the inhibitor effectively mimics the binding of the enolate anion present in the transition state of catalysis. Knowledge of the structural requirements for triclosan binding should make it possible to design new inhibitors that bind tightly to the active site of ENR. However the ability to select for triclosan resistant mutants in the lab (McMurry, Oethinger et al. 1998) suggests that triclosan-derivatives may be relatively ineffective, although this is not yet proven.

More recently the structure of ENR in complex with seven different diazaborines was solved (Levy, Baldock et al. 2001). These molecules inhibit gram-negative lipopolysaccharide synthesis (discussed in section 2.1.2). The finding that diazaborines are able to bind to the ENR active and the ubiquity of triclosan in the environment suggests that triclosan resistance may have a knock-on effect of creating broad spectrum resistance to other antimicrobials.

1.5 Organisation of the tree of life

Before the impact of science and evolutionary theory, the aborigines placed the division of life in the world between plants and animals. In 1866, Haeckel suggested a third branch for the tree of life, the protists (Haeckel 1866) a group of single-celled eukaryotes that we now know to include human parasites such as *Plasmodium spp.* (reviewed in (Hedges 2002)). Copeland and Whittaker later suggested that the tree of life harbour yet more branches represented by bacteria and fungi respectively (Copeland 1939; Whittaker 1959). Whittaker's suggestion of the five kingdoms of life led to the classical scheme defined by *Animalia*, *Plantae*, *Fungi*, *Protista* and *Monera* (bacteria) (*Fig 1.5.1*).

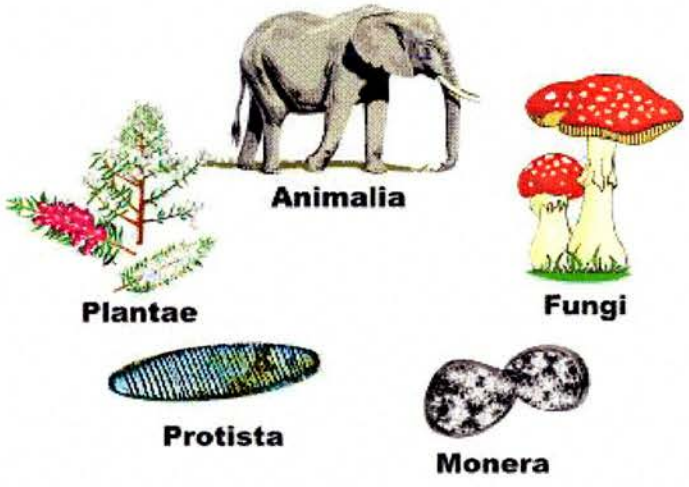

Fig 1.5.1 The Five kingdom tree of life (reproduced from http://www.palaeos.com/Systematics/Linnean/Kingdom.htm)

During the 1960s this model was further condensed, suggesting that all life could be partitioned into two domains: the *Eukaryotes* (single or multi-cellular, containing a membrane enclosed nucleus and other organelles) and the

Prokaryotes (single-celled, lacking a nucleus and other organelles) (Stanier and van Neil 1962; Murray 1968; Allsopp 1969).

1.5.1 The third domain of life

During the late 1970s, Carl Woese pioneered the field of molecular evolution. Woese was particularly interested in microbial evolution. In 1977, Woese and George Fox began studying methanogenic bacteria (methanogens) (Fox, Magrum et al. 1977). The methanogens were of interest because although morphologically diverse, they shared the common ability to grow anaerobically through the oxidation of hydrogen and reduction of carbon dioxide to form methane (Wolfe 1972; Zeikus 1977). To discern the relationship between members of the methanogens and other bacteria, Woese and Fox compiled a library of 16S ribosomal RNA (rRNA) (Fox, Magrum et al. 1977). This molecule was specifically used because of its ubiquitous nature and sequence specificity. Although sequence changes occur in rRNA, they do so slowly over time.

This comparison of 16S rRNA sequences showed that not only were the methanogens part of the same phylogenetic grouping but that they were divergent from other bacteria. Woese and Fox noted that the pattern of base modification was quite different from typical bacteria (Fox, Magrum et al. 1977). In parallel, co-author Magrum discovered that the methanogens lacked the tRNA TψCG sequence common to all *prokaryotes* and *eukaryotes*, instead having either a ψψCG or UψCG sequence. The divergence of the methanogens and bacteria was reinforced by biochemical studies - *Kandler et al* found that methanogenic cell walls did not contain peptidoglycan, a component essential to all bacteria (Kandler and Hippe 1977). Fox's co-author Balch (Fox, Magrum et

al. 1977), noted that upon studying a wide variety of tissues from many different organisms, certain coenzymes involved in methane production were only present in the methanogens (Fox, Magrum et al. 1977). In light of this evidence *Fox et al* (Fox, Magrum et al. 1977) suggested that the methanogens represented "the most ancient divergence yet encountered in the bacterial line of descent" and that they "existed at a time when an anaerobic atmosphere, rich in carbon dioxide and hydrogen, enveloped the planet and, if so, could have played a pivotal role in this planet's physical evolution".

This had obvious implications for the two-domain theory and the phylogenetic structure of the tree of life. Woese wondered about a common ancestor in both prokaryotic and eukaryotic domains from which these organisms arose (eventually leading to the notion of the Last Universal Common Ancestor - LUCA) (Woese and Fox 1977). Coining the term "*eubacteria*" ("eu" coming from the greek "ευ" meaning true), Woese, according to 16SrRNA sequence, grouped the blue-green bacteria, gram positive and gram negative bacteria into the same primary kingdom/domain (Woese and Fox 1977). Defined by 18S rRNA, Woese grouped the animals, plants, fungi and slime molds as the *urkaryotes* (eukaryotes phylogenetically, karyote/karya coming from the greek "καρυον" meaning "nut" or "kernel", corresponding to nucleus in the modern biological sense, thus eukarya/eukaryote "true nucleus").

A third domain was proposed by Woese, based on the unique phenotype and 16S rRNA sequences of methanogenic bacteria, saying "the apparent antiquity of the methanogenic phenotype plus the fact that it seems well suited to the type of environment presumed to exist on earth 3-4 billion years lead us tentatively to name this urkingdom the "*archaebacteria*" ("archaea" from

the greek adjective "αρχαιρξ" meaning "ancient" or "primitive") (Woese and Fox 1977). Woese also suggested the halophiles, which lacked peptidoglycan in their cell walls, be considered for the third domain.

In 1990, Woese proposed the abandonment of the term "*archaebacteria*" stating that it incorrectly suggested a specific relationship between these ancient organisms and the "*eubacteria*" (Woese, Kandler et al. 1990). Several lines of evidence suggested an intimate phylogenetic relationship between the *archaebacteria* and *eukaryotes* (discussed in the next section). Instead the term "*archaea*" was proposed and is to date the preferred name for the third domain of life. Woese further separated the *archaea* into two distinct lineages with the formal names *Euryarchaeota* (*euyarchaeotes*) and *Crenarchaeota* (*crenarchaeotes*) (Woese, Kandler et al. 1990). The *Euyarchaeotes* ("eury" from the greek "ευρυξ" meaning broad) comprise the methanogens, extreme halophiles, sulphate-reducing species and two thermophilic genera (*Thermoplasma* and *Thermococcus*). The "*Crenarchaeotes*", Woese proposed, comprise organisms whose niches are "entirely thermophilic" including those that live in extremely acidic and/or hot conditions (acidophiles and extreme thermophiles respectively) and sulphate-dependent archaea. Woese noted that the ability to live at high temperatures was a common phenotype of the archaea and suggested that this may have been the ancestral phenotype of the third domain (hence "crenarchaea", from the greek "κρηνη" meaning spring or fount, for the common phenotype of this kingdom and its resemblance to the presumed thermophilic ancestor).

1.5.2 Eukaryal-archaeal similarities

The discovery of a third domain led phylogeneticists to wonder about the origins of the *archaea* and their relationship to *prokaryotes* and *eukaryotes*. Shortly after Woese's work was published, Hiroshi Hori and Syozo Osawa (and other groups) began to construct phylogenetic trees to explain the evolutionary relationships between the three domains. Specifically, Hori and Osawa used 5S rRNA because of its conserved secondary structure and low substitution rate. Initial studies, using a relatively small library of molecules suggested a close link between eukaryotes and the archaea (Hori and Osawa 1979). From the rate of 5S rRNA nucleotide substitution and secondary structure, *Halobacterium cutirubrum* appeared more similar to *eukaryotes* than *prokaryotes* (Hori and Osawa 1979). This link was later confirmed in 1987 when Hori and Osaka employed a much larger library of 5S rRNA, encompassing 352 sequences with contributions for all three domains (Hori and Osawa 1987). By aligning the sequences to form a phylogenetic tree the two authors showed that the *archaea* form a distinct group that is phylogentically closer to *eukaryotes* than either are to the *eubacteria*. Hori and Osawa thus came to the conclusion that the *archaea* and *eukaryotes* were sistergroups (ie. they share a common ancestor) (Hori and Osawa 1987).

This link seemed to be confirmed in work by *Iwabe et al*, studying transcriptional elongation factors and ATPases (Iwabe, Kuma et al. 1989). Again, the construction of phylogenetic trees based on the primary structure of these molecules showed a positive correlation in sequence similarity and phylogenetic sisterhood between the two domains (Iwabe, Kuma et al. 1989). However this work became the subject of intense criticism; in one example, a subtype of ATPase previously only found in *archaea* and *eukaryotes* was detected in two

species of *eubacteria* (Kakinuma, Igarashi et al. 1991; Tsutsumi, Denda et al. 1991). More evidence of archaeal-eukaryotic sisterhood was needed.

1.5.3 Nucleic acid metabolism in the sistergroups

Archaeal and eubacterial genomes are incredibly similar - both package DNA into a single, circular chromosome often arranged with genes in the same pathway on operons (Edgell and Doolittle 1997). In comparing eukaryotes and the archaea one of the most surprising revelations is that DNA metabolism (replication, transcription, translation) in the archaea is much more similar to its eukaryal counterpart than the eubacterial machinery.

Archaeal transcription is the defining feature that links information processing in the two domains. Although, like *eubacteria*, *archaea* possess a single RNA polymerase, it is arranged in a manner more similar to the three eukaryotic polymerases. In general, eukaryotic and archaeal polymerases are made up of greater than ten homologous components (Zillig, Stetter et al. 1981; Huet, Sentenac et al. 1982; Zillig, Stetter et al. 1982). Additionally, like eukaryotic RNA polymerases, transcription by the archaeal homologues is stimulated by silybin (Schnabel, Sonnenbichler et al. 1982) and both are unaffected by antibiotics such as rifampicin that block transcription by eubacterial polymerases. *Huet et al* (Huet, Sentenac et al. 1982) demonstrated that antibodies raised against yeast RNA polymerases showed a more consistent pattern of cross reaction with archaeal polymerases than eubacterial ones (Schnabel, Sonnenbichler et al. 1982). In the case of RNA polymerase II, raised antibodies only cross-reacted with the archaeal homologues (Schnabel, Sonnenbichler et al. 1982). In later years, *Pulhler et al* showed that archaeal and

eukaryotic RNA polymerases were more similar in sequence to one another than either were to the eubacterial homologues studied (Puhler, Leffers et al. 1989). Puhler et al suggested, taken together with other evidence, that part of the eukaryotic genome may have ancient archaeal origins, although this is impossible to prove unequivocally.

In eukaryotes, transcription is directed and initiated by the binding of transcription factors to RNA polymerase II. Eubacterial genomes do not encode eukaryotic-like transcription factors and instead use sigma factors (σ factors), promoter-specific proteins that direct the transcription of specific groups of genes (eg. expression of eubacterial flagella) (Dale 1994). An essential protein in eukaryotic transcription is a factor called the TATA box-binding protein (TBP). TATAAA is a conserved DNA sequence in eukaryotic promoters that direct the transcription of coding genes (Stryer 1996). TBP binding to the TATA box is an essential initial step in the formation of a polymerase transcription complex, stimulating the binding of other transcription factors (Stryer 1996).

The revelation that archaea and eukaryotes harboured similar transcriptional components was further elaborated when in 1995, *Rowland et al* discovered an archaeal homologue of TBP in the hyperthermophile, *Pyrococcus woesei* (Rowlands, Baumann et al. 1994). This protein showed between 36 and 41% identity with known eukaryotic TBPs, remarkable considering the evolutionary differences between the two domains. In eukaryotes, the transcription factor TFB (known as TFIIB) stabilises the TBP-DNA complex. Functional homology was demonstrated by the formation of an archaeal TBP-TFB-DNA complex Eukaryotic TBPs are also known to bind transcriptional activators and repressors; *Rowland et al* showed that *P. woesei* TBP bound the

tumour suppressor protein p53 and the activation domain of the adenoviral E1A protein. A year later, Wettach et al published experimental evidence of the functional homology of Methanococcal and eukaryotic TBPs (Wettach, Gohl et al. 1995). Major findings included the discovery that yeast and human TBPs were able to replace the euryarchaeal TBP in cell-free transcription systems and that eukaryotic TBPs not only bind to the methanococcal promoter but specifically recognise the archaeal TATA box. The fact that TBP and other eukaryotic transcription factors are found in archaea, but not *eubacteria* suggests that these are ancient proteins derived from ancestral transcription factors.

Translational machinery also appears to be similar between archaea and eukaryotes. In 1995, Brown and Doolittle provided strong evidence to support the theory of archaea and eukaryotes being sistergroups by constructing phylogenetic trees from sequences of aminoacyl tRNA synthetases (RS) - molecules that catalyse the joining of an amino acid to its corresponding transfer RNA (Brown and Doolittle 1995). Trees constructed from IleRS proved to be more convincing proof of archaeal-eukaryotic sisterhood than those of *Iwabe et al* - the IleRS sequences are better models than ATPases or elongation factors because they are almost three times longer than those of the elongation factor genes used by *Iwabe et al.* Additionally, at the time, IleRS proteins could be isolated from a much broader range of organisms. This also seemed to be good evidence of eukaryotes and archaea using very similar translational machinery and that information processing pathways between the two domains are more similar than biochemical ones.

1.6 The crystal structure of the first eukaryotic/archaeal primase

DNA primases occur in all three domains of life and represent a special type of RNA polymerase. As DNA polymerases are unable to start the synthesis of new DNA strands, primases generate a small RNA primer that is used as a starting fragment by DNA polymerase to copy the ssDNA template (Dale 1994). In *eubacteria* and *eukaryotes*, primase catalytic activity is located on a single polypeptide, although eukaryotic primase have two subunits - the catalytic small subunit (~50kDa) and the large subunit (~60kDa). The small subunit differs completely from eubacterial and bacteriophage primases, but is very similar to the archaeal primase which has a molecular weight of roughly 40kDa (Kirk and Kuchta 1999).

Although the structures of T7 primase and several eubacterial primases are known, only recently was the structure of a eukaryotic like primase determined- *Pfu-prim* from the hyperthermophilic archeon *Pyrococcus furiosus* (Augustin, Huber et al. 2001). The overall structure and fold of the archaeal homologue is unrelated to that of the eubacterial and phage proteins. However, superposition of *Pfu-prim* with DNA polymerase β shows a similar arrangement of active site residues between the two. The fact that Kirk and Kutcha had already shown sequence homology between eukaryotic primases and this polymerase (Kirk and Kuchta 1999) suggests that the archaeal and eukaryotic primases are structural homolgues. More importantly the structure shows, for the first time, the active site architecture of a primase from the eukaryotic/archaeal family. A zinc-knuckle motif (consensus, C-X-[C, H]-X_{3-5}-C-X_2-C) is found in

the archaeal structure where sequence analysis suggests a classic zinc-finger motif in the eukaryotic primase. Three negatively charged residues in a conserved motif (E_{91}LVFDID_{97} in *Pfu-prim*) were previously shown to be essential for mouse primase activity (Copeland and Tan 1995). The archaeal structure suggests that Glu91 in this motif has a purely structural role because the side chain is deeply buried. Instead, Asp95 and 97 are proposed as catalytic residues, corresponding to two aspartic acid residues in DNA polymerase β that coordinate a catalytic cation in the active site (Steitz, Smerdon et al. 1994). According to the structure, Asp280 is also proposed to play an important catalytic role and this residue in mouse primase appears to be essential as mutation to alanine abolishes activity (Copeland and Tan 1995). To date, no eukaryotic primase structures are present in the database. Biochemical evidence from a eukaryotic primase, demonstrating the nature of proposed active site residues in the archaeal crystal structure, thus suggests *Pfu-prim* is a good model in which to study the less tractable eukaryotic homologue. As pointed out by the authors this structure is "an important piece of the puzzle of understanding the replication machinery and will facilitate biochemical and structural studies of the replisome".

1.7 A unique archaeal topoisomerase as a model for a meiotic recombination factor

DNA can exist in a number of topologically different forms including linear, circular and supercoiled. To package its vast length into comparatively small cells, DNA is traditionally supercoiled. Before processes such as replication and

transcription can occur supercoiled chromosomal DNA must be relaxed, a function assigned to topoisomerases (Stryer 1996). Type II topoisomerases are important in removing wound and knot-like structures during DNA metabolism, an ATP-driven process that involves generation of transient double-stranded breaks (DSBs).

The archaeal type II topoisomerase (Top6) was first discovered in *Sulfolobus shibatae* and has an A_2B_2 heterotetrameric composition (Bergerat, Gadelle et al. 1994). *S. shibatae* Top6 was notable in having no sequence homology to previously characterised type II topoisomerases and its homology with Spo11, a factor present during meiotic recombination (Bergerat, de Massy et al. 1997). Concurrently, *Keeney et al* were able to show that Spo11 was responsible for DSB activity during meiotic recombination (Keeney, Giroux et al. 1997) suggesting not only sequence but additionally functional homology with Top6.

Subsequently, determination of the Top6-A subunit (Top6A) crystal structure from the euryarchaeal *Methanococcus jannaschii* has allowed this topoisomerase to be used as a functional model for SpoII activity (Keeney, Giroux et al. 1997; Nichols, DeAngelis et al. 1999). The protein crystallises as a U-shaped dimer in the asymmetric unit (Nichols, DeAngelis et al. 1999). The proposed active site tyrosine is found in the N-terminal domain of Top6A (Nichols, DeAngelis et al. 1999) and this residue is conserved in Spo11 and members of the extended Top6A/Spo11 family; additional confirmation of functional homology comes from the fact that the analogous tyrosine in Spo11 is essential for DSB activity in yeast (Diaz, Alcid et al. 2002). Catalysis is proposed to involve the formation of a phospho-tyrosine linkage in the active site (Keeney

and Kleckner 1995; Liu, Wu et al. 1995). In the Top6A structure, an octa-coordinated Mg^{2+} ion is held in a Rossmann fold-like nucleotide-binding domain by a glutamic acid, two aspartic acid residues and a number of water molecules. As these acidic residues are conserved in Spo11 and homologues (including primases), it is probable that this enzyme also binds monovalent ions in a nucleotide binding pocket. This suggests that the binding of Mg^{2+} is catalytically important and mutations in this acidic pocket do indeed impair or abolish SpoII activity (Diaz, Alcid et al. 2002). Modelling of DNA into Top6A suggests a plausible model for SpoII-DNA interactions (*Fig 1.7.1*).

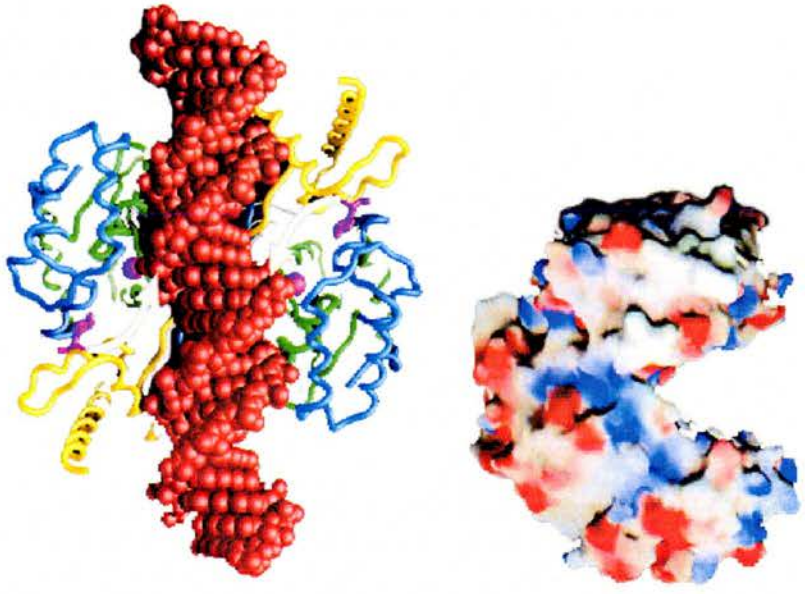

Fig 1.7.1 Left -Top6A dimer-DNA model. DNA is represented as a spacefilling model in red. Right - electrostatic surface of the Top6A dimer. Reprinted from (Nichols, DeAngelis et al. 1999)

A deep groove between the two Top6A monomers would appear to accommodate duplex DNA (Nichols, DeAngelis et al. 1999), allowing the molecule to remain significantly exposed on the surface. This would be the most logical place for DNA to bind as it constitutes the most electrostatically positive

part of Top6A. As stated by the authors, if the structures of SpoII and Top6A are indeed similar (implied by the sequence and functional similarity) then this would allow other recombination factors to interact with the bound DNA or regulate Spo11 activity. At least nine other proteins are required for the formation of double strand breaks (Haber 1997). The modelling also suggests a mechanism for DNA release by Spo11 as previously proposed, either by hydrolysis of the phospho-tyrosine linkage by other proteins active during recombination or direct cleavage by endonucleases (Keeney, Giroux et al. 1997). Thus, the structure of an archaeal protein demonstrates another example of the use of *archaea* as a model for eukaryotic systems.

Chapter 2

RmlD, a unique SDR involved in rhamnose biosynthesis

Summary

Enzymes in the Short-chain Dehydrogenase/Reductase (SDR) family are ubiquitous in nature owing to the essential function they perform in substrate catabolism. SDRs are capable of processing a number of substrates including alcohols (alcohol dehydrogenase, (Chen, Jiang et al. 1993; Benach, Atrian et al. 1998), dihydropteridines (dihydropteridine reductase, (Varughese, Xuong et al. 1994) and steroids (3α 20β-hydroxysteroid dehydrogenase (Ghosh, Weeks et al. 1991). Nucleotide co-factors are bound by an N-terminal Rossman fold (Rossmann, Moras et al. 1974) and are key to all SDR catalysed reactions which include oxidations, epimerisations, reductions and dehydrations. The co-factor is involved in hydride transfer, specifically either donating a hydride ion (NADP[H] > NAD[P]) or abstracting one ($NAD[P]^+$ > NAD[P]H) from the substrate in the transition state, facilitating subsequent steps during catalysis.

2.1 Rhamnose biosynthesis in pathogenic bacteria

2.1.1 The dTDP-L-rhamnose pathway

The production of rhamnose in pathogenic bacteria is an appealing therapeutic target. The sugar is not found or utilised in humans, presenting the opportunity to develop drugs with selective toxicity. A growing body of evidence suggests that rhamnose biosynthesis is essential to a large number of medically important pathogens. Initial studies by *Kornfeld and Glaser* in 1961 led to the elucidation of a biosynthetic pathway catalysed by four enzymes (Glaser and Kornfeld 1961; Kornfeld and Glaser 1961)(*Fig.2.1.1*), RmlA - α-D-glucose-1-phosphate thimidylyltransferase, RmlB - dTDP-D-glucose 4,6 dehydratase, RmlC - dTDP-

6-deoxy-D-xylo-4-hexulose 3,5 epimerase and RmlD - dTDP-6-deoxy-L-lyxo-4 hexulose reductase.

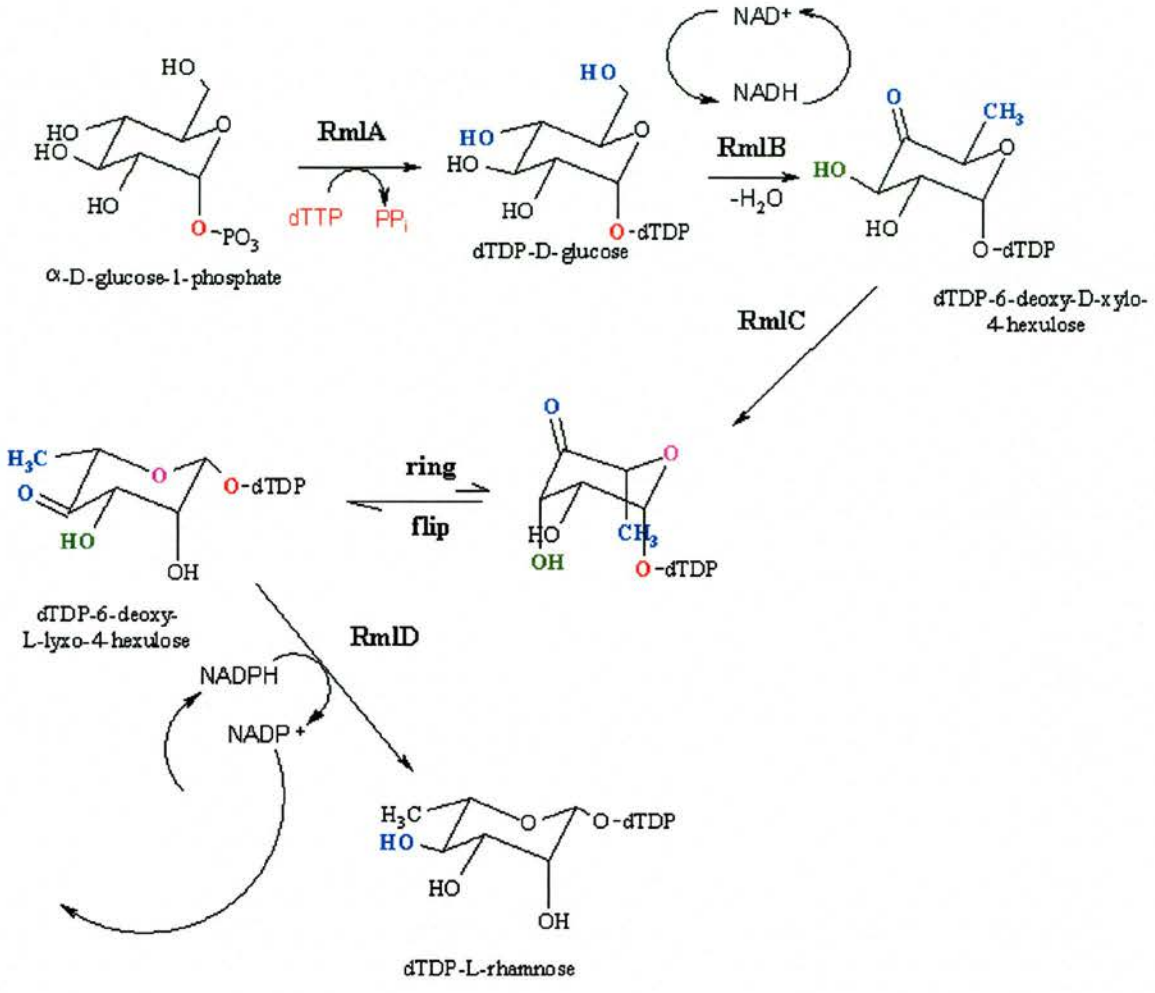

Fig.2.1.1 The dTDP-L-rhamnose biosynthetic pathway

Thymidine tri-phosphate (TTP) and α-D-glucose-1-phosphate are the initial substrates for the reaction. RmlA transfers a dTDP-moeity to C1 of the carbohydrate ring. This acts as a specific anchor to the substrate-binding domains of subsequent enzymes in the pathway. RmlA initiates the first step in almost all 6-deoxy sugar biosynthesis and is allosterically regulated by dTDP-L-rhamnose, generating a negative feedback loop that controls levels of the final product of the pathway (Melo G. 1965).

RmlB carries out the dehydration of dTDP-D-glucose to dTDP-4-keto-6-deoxy-D-glucose activating the sugar at positions 3 and 5 for epimerisation by RmlC which generates the 'natural' L-isomer (Giraud and Naismith 2000). The resulting configuration is thermodynamically strained with functional groups at positions 1, 3, 4 and 5 lying in the axial position. Spontaneous flipping of the ring is thought to occur at this stage. This generates the β-conformation required for the active site chemistry of RmlD which carries out the terminal step in the pathway - the NAD(P)H dependent reduction at the 4-keto position of dTDP-6-deoxy-L-lyxo-4-hexulose to give the product dTDP-L-rhamnose.

The four enzymes involved in rhamnose biosynthesis are encoded by genes on the rml (formerly rfb) cluster that are highly conserved throughout several bacterial species (Jiang, Neal et al. 1991; Reeves 1993; Schnaitman C. A. 1993; Szabo, Bronner et al. 1995; Mitchison M. 1997). As first shown by Jiang et al, the genes commonly lie in the order rmlBCAD. Sequence analysis of the rml cluster in *Salmonella typhimurium* LT2 implies a varied *G+C* content suggesting that the genes may have been acquired from an organism with a low *G+C* content (Jiang, Neal et al. 1991). This would explain the ubiquitous nature of these genes throughout the eubacterial domain.

2.1.2 Rhamnose in gram-negative bacteria

L-rhamnose is essential in the construction of surface glycoconjugates and polysaccharides in a variety of pathogenic bacteria and is commonly found as a component of gram-negative lipopolysaccharide (LPS, *Fig.2.1.2.1*), a molecule unique to this class of bacteria.

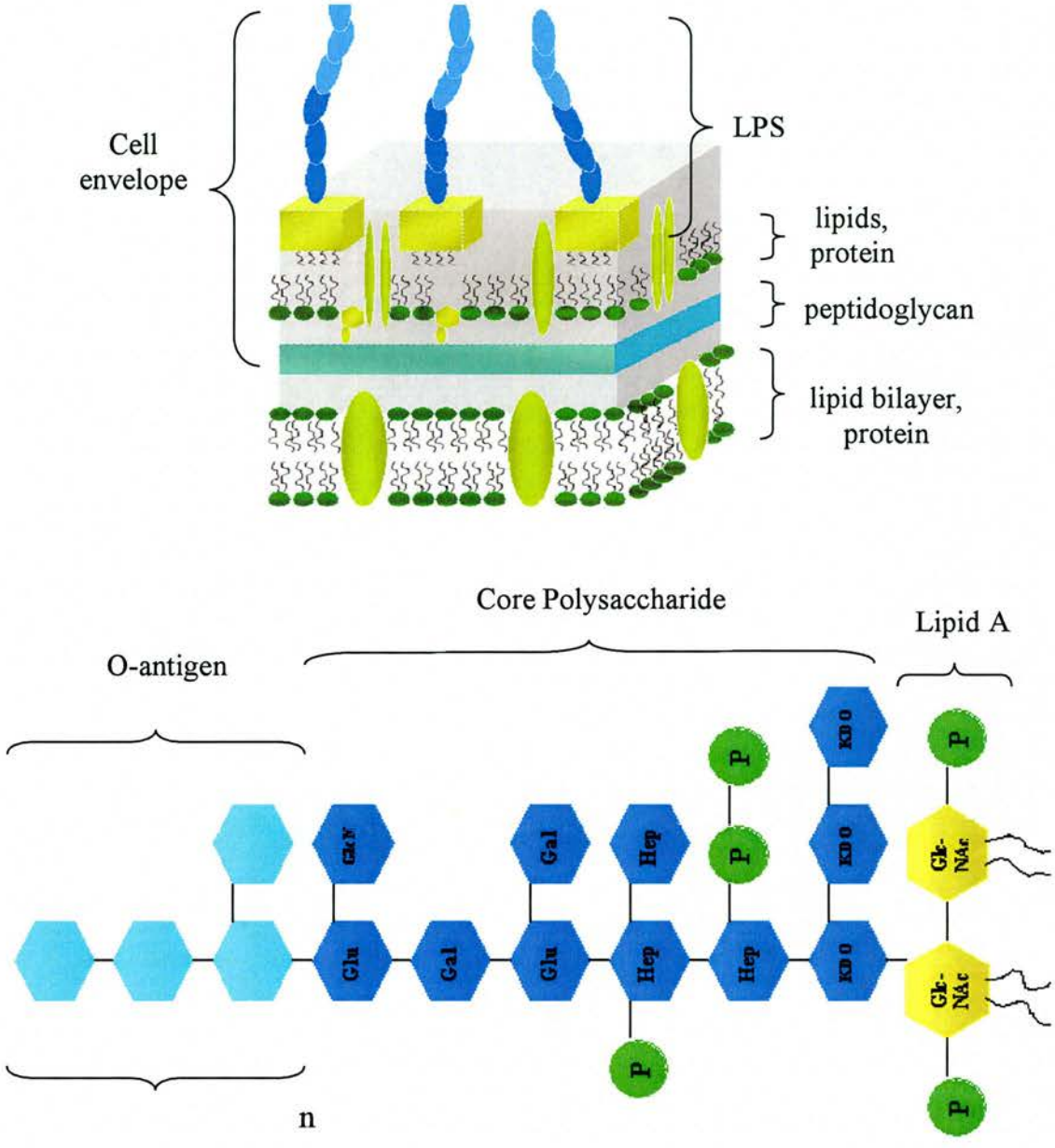

Fig.2.1.2.1 The Gram negative cell wall (top) and LPS (bottom)

LPS is one of the most toxic/reactogenic molecules known. It is a powerful activator of the inflammatory immune response which can be damaging to the host, as well as being important in the induction of a specific immunity. The systemic release of LPS, as a result of bacterial lysis, causes large amounts of TNF α to be released into the bloodstream leading to septic shock which, if untreated, terminates in massive organ failure and death (Janeway, Travers et al.

1999). Found in the outer leaflet of the cell envelope, LPS has three major components. Lipid A, composed of fatty acids ester linked to N-acetyl-glucosamine (GlcNAc), anchors LPS in the hydrophobic portion of the outer membrane (Madigan, Martinko et al. 1997). The Core polysaccharide has a net negative charge and contains a unique 8-carbon sugar - 2-keto-3-deoxyoctonic acid (KDO). Association with neighbouring LPS molecules is through Mg^{2+} ions that interact with the phosphate groups of the core sugars. L-rhamnose plays an important structural role in the third component of gram-negative LPS - the O-antigen, a homo/hetero-polymer of 5-6 carbon sugars. Found in the hydrophilic portion of the molecule, the O-antigen presents an important interaction surface that is impermeable to many hydrophobic antibiotics, detergents and fatty acids.

Rough LPS strains (with no O-antigen) are sensitive to the complement pathway found in serum, an enzymatic cascade involving, amongst other factors, inflammatory mediators and opsonins that help initiate a specific immune response. Smooth LPS strains, having a bimodal distribution of O-antigen subunits, are serum resistant (Taylor 1983). Complement killing is associated with insertion into the bacterial outer membrane of the terminal proteins of the complement cascade, the Membrane Attack Complex (MAC), through two pathways. The classical complement pathway is activated by the formation of an antibody-antigen complex. The alternative pathway is triggered in the absence of such a complex. The expression of smooth LPS may lower the fluidity of the outer membrane thus slowing the rate of insertion of MACs. It has also been suggested that long O-antigen sidechains in smooth LPS activate the complement cascade far from the cell surface, with the result that complement components

degrade before they are able to form functional MACs and assemble on the outer membrane (Morrison 1977; Taylor 1983).

Several lines of evidence link the deletion of rhamnose with weakening of the O-antigen and concomitant loss of serum resistance. *Burns et al*, have shown that a rhamnose negative strain of uropathogenic *E. coli*, generated by interruption of the rmlD gene, is serum sensitive (Burns and Hull 1998). Analysis by SDS-PAGE showed that the rhamnose$^-$ strain had a rough LPS phenotype. In 80% serum, the mutant was cleared within 2 hours, while the parental, wild-type strain continued to grow after 5 hours incubation (Burns and Hull 1998). *Chiang et al*, demonstrated that mutation of rmlB in *Vibrio cholerae*, the causative agent of cholera, inhibited intestinal colonisation (Chiang and Mekalanos 1999). LPS isolated from the mutant and analysed by SDS-PAGE was shown to lack the O-antigen. The rmlB mutant was 10,000 fold more sensitive to serum than the wild-type strain capable of synthesising the O-antigen. Additionally the mutant displayed dramatically reduced colonisation versus the wild-type strain in an infant mouse model. From these reports it is sensible to suggest that in the absence of rhamnose, the O-antigens of these organisms are truncated as further sugar subunits are unable to assemble on the surface. This results in a loss of bimodal O-antigen distribution and serum-sensitive phenotype that seriously attenuates these strains.

Pseudomonas aeruginosa, a common cause of opportunistic infection in immunocompromised patients, is promiscuous in its use of rhamnose and is able to simultaneously produce two distinct types of LPS: A-band and B-band (Rocchetta, Burrows et al. 1999). A-band LPS has a homopolymeric O-antigen consisting of D-rhamnose. *P. aeruginosa* is unique in utilising the D-enantiomer,

which is synthesized from GDP-D-mannose through the action of GDP-D-mannose dehydratase and GDP-4-keto-6-deoxy-D-mannose reductase (Rocchetta, Burrows et al. 1999). B-band LPS has a heteropolymeric O-antigen which, in some serotypes, contains L-rhamnose. The L-sugar is also a key residue in the core polysaccharide and part of a biosurfactant rhamnolipid secreted by the organism (Jarvis and Johnson 1949). *Rahim et al* have shown that *P. aeruginosa* PAO1 rmlC mutants synthesise truncated core polysaccharides unable to serve as acceptors for O-antigen, suggesting that L-rhamnose is an essential linking residue between the core and O-polysaccharides (Rahim, Burrows et al. 2000).

2.1.3 Rhamnose in Gram-positive bacteria

This class of bacteria also express polysaccharides on their cell surface, containing L-rhamnose. The Capsular polysaccharide (K-antigen, capsule) synthesised by some gram-positive (and gram negative) species provides a barrier function protecting the cell from external stresses and the immune system. Capsular polysaccharides often have anti- phagocytic properties, allowing pathogens to evade the action of activated macrophages. Like LPS, the K-antigen is an important virulence factor, capable of stimulating the release of inflammatory mediators.

Cariogenic strains of *Streptococcus mutans*, a major cause of dental caries, are known to possess capsules containing L-rhamnose (Tsukioka, Yamashita et al. 1997). If untreated, *S mutans* is able to spread to the heart and kidneys. The polysaccharide antigen is an important factor in host colonisation and the spread of the organism to major organs (Stinson, Nisengard et al. 1980;

Michalek, Morisaki et al. 1984). The backbone of the capsule is made up of 1,2 and 1,3 linked rhamnosyl residues synthesised via the dTDP-L-rhamnose biosynthetic pathway described earlier (Tsukioka, Yamashita et al. 1997; Tsukioka, Yamashita et al. 1997). The crucial role of L-rhamnose is underscored by the fact that inactivation of any of the rml genes inhibits capsular polysaccharide synthesis (Tsukioka, Yamashita et al. 1997; Tsukioka, Yamashita et al. 1997). Furthermore, rml mutants are only able to survive in sucrose supplemented media upon fusion of gtfB and C, abolishing the ability of these enzymes to synthesise the extracellular polysaccharide Glucan (Yamashita, Tomihisa et al. 1999).

Possibly the most novel use of rhamnose in bacteria is to be found in the mycobacterial cell wall, the outer surface of which has a highly unique structure. The low permeability and high hydrophobicity of the cell wall makes it highly resistant to penetration by several antibiotics. The development of new anti-mycobacterial drugs thus presents a significant challenge. The cell surface conveys an outer layer of mycolic acids that are connected to peptidoglycan, the main structural component of gram positive cell walls via an arabinogalactan polysaccharide (reviewed in(Chatterjee 1997))(*Fig.2.1.3.1*).

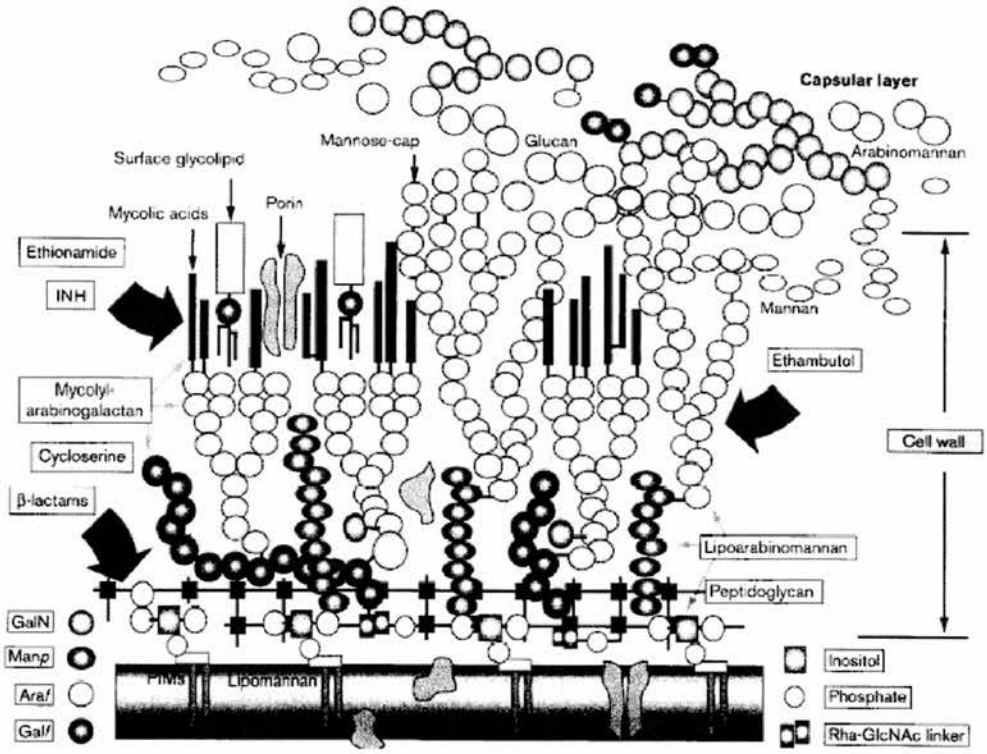

Fig.2.1.3.1 The mycobacterial cell wall, reprinted from (Chatterjee 1997)

This connection is essential for cell viability and growth is inhibited by ethambutol which prevents the attachment of the mycolic acids to the inner layer of peptidoglycan (Chatterjee 1997). An α-L-rhamnopyranosyl residue has been shown to be present in the linker that connects arabinogalactan and peptidoglycan. The linker is found in all mycobacteria and the rhamnose residue is arranged identically in each example (McNeil, Daffe et al. 1990). Recent studies by *Ma et al* have shown that interruption of dTDP-L-rhamnose biosynthesis inhibits growth of *Mycobacterium smegmatis* (Ma, Pan et al. 2002). An rmlD knockout strain is only able to grow in the presence of a rescue plasmid that carries the rmlD gene from *Mycobacterium tuberculosis*. The loss of the temperature-sensitive plasmid at 42°C prevents growth of the organism (Ma, Pan et al. 2002). These experiments suggest that inhibitors of the rhamnose pathway

may be good candidates for lead optimisation in the discovery of novel anti-mycobacterial drugs.

2.2 RmlD - the last enzyme in the rhamnose pathway

RmlD (dTDP-6-deoxy-L-lyxo-4-hexulose reductase) catalyses the terminal reaction in dTDP-L-rhamnose biosynthesis (Graninger, Nidetzky et al. 1999), the C_4 keto-reduction of dTDP-6-deoxy-L-lyxo-4-hexulose generating dTDP-L-rhamnose, the activated precursor for L-rhamnose *(Fig.2.2.1)*. NAD(P)H is required for the reductive biosynthesis.

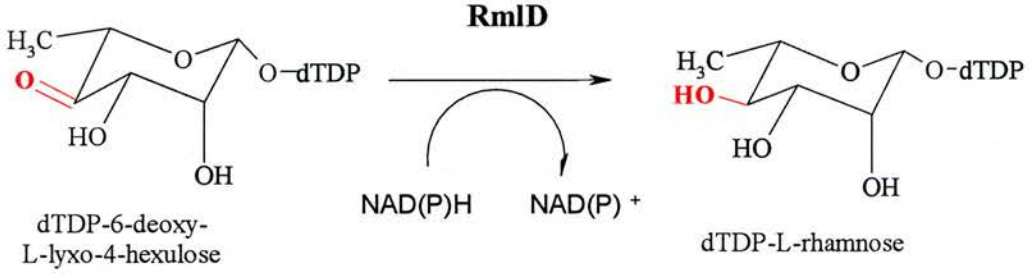

Fig.2.2.1 The reduction of the keto-sugar by RmlD

Kinetic analysis shows that the forward reaction is favoured (Graninger, Nidetzky et al. 1999). The substrate for RmlD is unstable and the reverse reaction (oxidation) has less than 1% of the activity of the forward reaction (Graninger, Nidetzky et al. 1999). RmlD has dual co-factor specificity and has a high affinity for both NADPH and NADH. Only a slight preference is shown for NADH (Graninger, Nidetzky et al. 1999).

Extensive biochemical studies by *Graninger et al (Graninger, Nidetzky et al. 1999)* on RmlD from *Salmonella typhimurium*, a 299 amino acid protein, has

provided a robust understanding of the enzyme. Overexpression in *E. coli* yielded a protein of 34.8 kDa upon purification and SDS-PAGE analysis (the molecular weight from the predicted amino acid sequence is 32.5kDa). The molecular weight from gel filtration was calculated to be 41.5kDa and does not agree with the predicted weight for the monomer or the dimer (65.2kDa). This anomalous result may be due to interaction with the gel filtration column. In an RmlB, C, D-coupled enzyme assay the absolute specificity of RmlD, and other enzymes in the pathway, were demonstrated by using alternative substrates including UDP-D-glucose, CDP-D-glucose and ADP-D-glucose. The apparent activity of the coupled RmlB-D system upon reaction with these analogues was less than 1% of the native activity with dTDP-D-glucose.

Only recently has the reducing activity of RmlD been separated from the 3, 5 epimerising activity of RmlC, which catalyses the preceeding step in the pathway. It was originally believed that these two proteins formed a dimeric, rhamnose synthase. However, *Jiang et al* have shown that two distinct gene products are required for the conversion of α-D-glucose-1-phosphate to dTDP-L-rhamnose (Jiang, Neal et al. 1991). Several experiments by *Graninger et al* prove this hypothesis. Using the coupled B, C, D assay, the consumption of NAD(P)H was only possible in the presence of excess RmlD (Graninger, Nidetzky et al. 1999). In addition, analysis of mixtures of RmlC and D by non-denaturing PAGE showed no evidence of complex formation. Addition of Ficoll, which enhances interactions in macromolecular complexes, to a coupled RmlC-D assay showed no influence on enzyme activity.

Sequence alignment data show that RmlD has three characteristic motifs– a Y-X_3-K motif containing the proposed catalytic tyrosine and lysine, a Wierenga

motif (here, G-X_2-G-X_2-G) (Wierenga and Hol 1983; Wierenga, De Maeyer et al. 1985) within a Rossmann fold (Rossmann, Moras et al. 1974) and a unique STDYVF moiety (Graninger, Nidetzky et al. 1999). The Wierenga motif plays a crucial role in co-factor binding, while the STDYVF contains a serine and threonine proposed to be important in catalysis. The Y-X_3-K motif unequivocally identifies RmlD as a member of the Reductase/Epimerase/Dehydrogenase (RED) superfamily (Labesse, Vidal-Cros et al. 1994). In particular, RmlD has similarities with the Short chain Dehydrogenase/Reductase (SDR) branch (Jornvall, Persson et al. 1995; Persson, Krook et al. 1995)(discussed below). RmlD may be the most appealing drug target in the rhamnose pathway; it acts on the most rhamnose like sugar and therefore has the most specific binding-pocket for the activated precursor.

2.3 The short chain dehydrogenase/reductase family

The family of enyzmes known as the short chain dehydrogenases/reductases (SDRs) have been known for over twenty years and are defined by a number of conserved sequences and motifs. *Jornvall et al* first described the family in 1981 based on the finding that dehydrogenases binding alcohols and those binding polyols could be subdivided as having polypeptides with short and long subunits respectively (Jornvall, Persson et al. 1981). Two notable findings in the poorly defined SDRs were the presence of a highly conserved Y-X_3-K motif (Jornvall, Persson et al. 1981) and, from secondary structure predictions, a co-factor/co-enzyme binding site in the N-terminus (Thatcher and Sawyer 1980). Widespread interest in the SDRs was only activated upon the finding of mammalian sequence homologues, notably prostaglandin (Krook, Marekov et al. 1990) and steroid

dehydrogenases (Peltoketo, Isomaa et al. 1988). The result was that the previous peak in SDR numbers (roughly 15 in 1988) doubled to almost 30 known sequences (reviewed in (Persson, Krook et al. 1995)). The most recent review cites 3000 known SDR sequences, with 30 three-dimensional structures determined (Oppermann, Filling et al. 2003).

Sequence identity in the SDRs ranges from 15-30% reflecting the widespread diversity of the family (Jornvall, Persson et al. 1995). Nevertheless, two highly conserved sequence motifs and strictly conserved amino acids suggest a common origin and justify the grouping of these enzymes into the one family. The current SDR model has 250-350 amino acids, an α/β folding pattern with a Rossmann fold and two highly conserved sequence motifs (Oppermann, Filling et al. 2003) - a variable G-X_3-G-X-G motif (G-X_2-G-X_2-G in RmlD) identified by Rik Wierenga (Wierenga and Hol 1983; Wierenga, De Maeyer et al. 1985) and a catalytic triad of Tyr-Lys-Ser/Thr with the tyrosine and lysine found in a Y-X_3-K motif. The Y-X_3-K motif contains the proposed catalytic tyrosine and lysine common to many SDRs (SDR catalysis is described in detail in section 2.8.2). The variable G-X_3-G-X-G moiety is part of an N-terminal Rossmann fold involved in cofactor binding. This motif is almost identical in the majority of SDRs studied (Jornvall, Persson et al. 1995), suggesting the spatial importance of this region in the molecule. In general Rossmann folds have a β_1-αA-β_2 secondary structure (Rossmann, Moras et al. 1974), with the Wierenga motif being crucial in the positioning of the cofactor in relation to the protein backbone (Wierenga and Hol 1983; Wierenga, De Maeyer et al. 1985). The Wierenga motif generally constitutes the first few residues of α-helix A and forms a loop between β_1 and the helix. The driving force for co-factor binding comes from the

interaction of the negative charge on the pyrophosphate with the helix dipole (Hol, van Duijnen et al. 1978; Wierenga, De Maeyer et al. 1985) Roles for the three conserved glycine residues were first suggested by Wierenga et al (Wierenga, De Maeyer et al. 1985). The first glycine forces a tight turn in the main chain between αA and β_1. A side-chain at this position would sterically hinder binding of the ribose moeity (Wierenga and Hol 1983). The presence of the second glycine prevents steric hinderance with the pyrophosphate of the co-factor, which would occur with a larger side chain at this position. The third glycine allows αA and β_1 to exist in close proximity, compacting the Rossmann fold. The importance of this motif is underscored in work by *Chen et al,* which showed a loss of activity in recombinant proteins without this pattern (Chen, Jiang et al. 1993).

The recent structural determination of NmrA, a transcriptional regulator, shows the protein to have a striking structural similarity to the SDR fold (Stammers, Ren et al. 2001). However the active site tyrosine is replaced by a methionine, making it unlikely that this is a catalytically active dehydrogenase or reductase and instead it appears that the SDR fold is adapted to perform another function, most likely protein-protein interaction. This finding suggests that the SDR fold is more widespread than was previously thought, suggesting that other functions remained to be elucidated.

Materials and Methods/Results

2.4 Isolation and purification of the pET-rmlD construct

2.4.1 The expression construct

The pET-rmlD vector was constructed by our collaborators in the lab of Professor Chris Whitfield at the University of Guelph, Canada (Graninger, Nidetzky et al. 1999). The rmlD gene from *Salmonella typhimurium LT2* was introduced into the multiple cloning site of the pET 30a(+) vector (Novagen) between the restriction sites XbaI and SacI (*Fig.2.4.1.1*). This placed the rmlD gene under the control of the IPTG-inducible T7 promoter for overexpression of the protein. The construct also contained a kanamycin resistance gene to allow selection of resistant colonies.

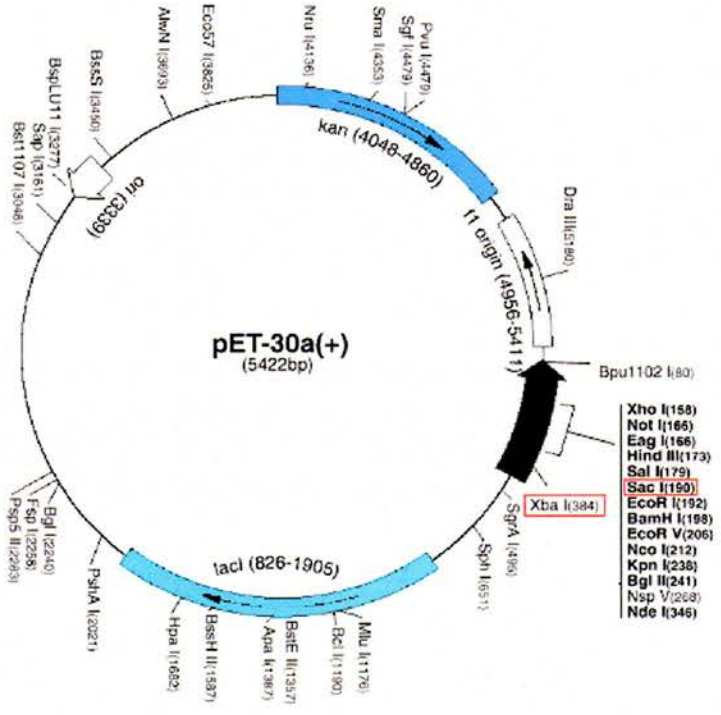

Fig.2.4.1.1 The pET 30a(+) vector

2.4.2 Transformation of competent *E. coli*

Chemically-competent *E. coli* JM109 cells (Stratagene) were transformed with the construct provided using a Heat Shock protocol (Kushner 1978; Hanahan 1983) – 0.1μg of plasmid DNA were added to 100μl of cells and incubated on ice for 20 minutes. The chilled cells were then incubated at 42°C in an unstirred water bath for 2 minutes then transferred to ice for 2 minutes. The transformed cells were added to 1ml of Luria-Bertani medium (LB) (Sambrook, Fritsch et al. 1989). After an hour's incubation at 37°C, with constant shaking, 100μl of cultured cells were spread on an L-agar plate containing 60μg/ml Kanamycin (Sigma) and incubated at 37 °C overnight.

2.4.3 Isolation of the pET-rmlD construct from *E.* coli JM109

A single colony was picked from the plate and used to inoculate a 10ml starter culture of LB containing 60μg/ml Kanamycin. The culture was incubated overnight at 37°C, with constant shaking. From the overnight culture, 1ml was used to inoculate 2 x 100ml of LB each. The cultures were incubated overnight at 37°C, with constant shaking.

DNA was isolated from overnight cultures by the alkaline lysis method (Birnboim and Doly 1979; Sambrook, Fritsch et al. 1989), using the QIAGEN maxi prep kit. The isolated DNA pellet was resuspended in 500μl of TE buffer, pH 8.0 (see Appendix). The purity and concentration of the DNA was assessed by running 5μl on a 1% agarose gel (*Fig 2.4.3.1*). To verify the presence of the rmlD gene in the construct, 5μl of each prep. was digested with the restriction enzymes XbaI and SacI. After an hour's incubation at 37°C, the enzymes were

inactivated by heating at 65°C for 10 minutes and 5µl of each digest was run on a 1% agarose gel (*Fig 2.4.3.2*).

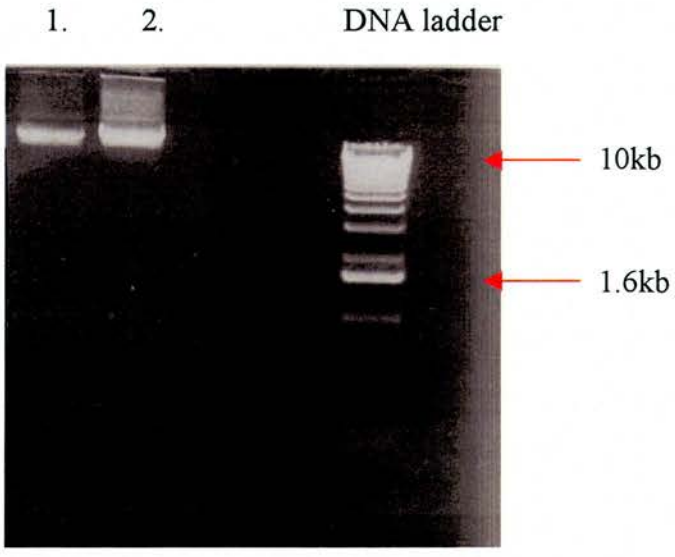

Fig 2.4.3.1 Two preparations of the pET-rmlD construct purified from E. coli JM109 Cells

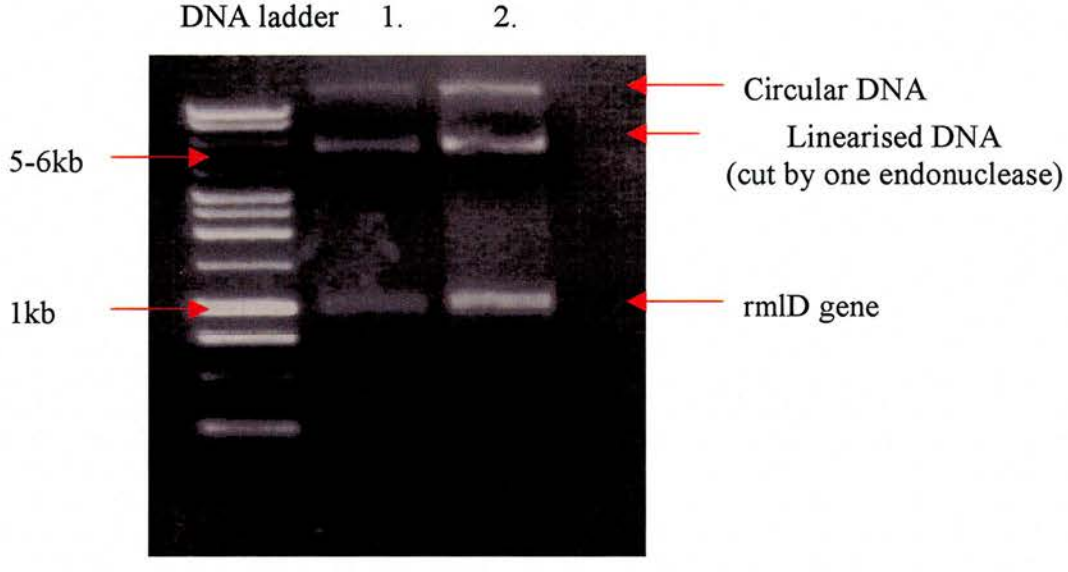

Fig 2.4.3.2 Restriction enzyme digests of the purified, plasmid DNA proving the presence of the RmlD gene in the double-stranded DNA construct.

Using the 1.6kb marker as a guide (10ng/µl) the purified DNA was quantitated and determined to be at a concentration of 28ng/µl (Fig 2.4.3.1, lane 1.) and

42ng/μl (Fig 2.4.3.1, lane 2.). Figure 2.4.3.2 shows that the construct contained the rmlD gene between the restricition sites XbaI and SacI. The double digest gave rise to a band at 1kb, corresponding to the rmlD gene which is 897bp in length. The single band at 5-6kb represents the linear DNA (only cut by one enzyme) expected to run at 6.3kb.

2.5 Site-directed mutagenesis/characterisation of mutants

The crystal structure of RmlD from *Salmonella typhimurium* was originally solved by Wulf Blankenfeldt in our lab. However the initial data were poor and some regions

of the model exhibited flexibility. To probe the enzyme mechanism, mutations were made at Ser103 and Thr104 to alanine and Tyr128 to phenylalanine. These are three of the proposed catalytic residues. Further mutations to alanine were made in the rhamnose-binding pocket at Val67, Asp68, Asp105 and Trp153 to establish the role (if any) of these residues in sugar binding.

2.5.1 Mutagenesis

The QuikChange Site-Directed Mutagenesis protocol (Stratagene) was used to introduce point mutations in the rmlD gene from *S. typhimurium* LT2 (*Fig.2.5.1.1*). This method allowed the use of the pET-rmlD construct as the DNA template and generates large quantities of mutated DNA using the Polymerase Chain Reaction (PCR). As non-mutated, parental DNA is methylated by *E. coli* it can be digested by addition of the endonuclease DpnI to the plasmid

pool (McClelland and Nelson 1992). The mutated DNA is not methylated by the PCR reaction and therefore not targeted by the enzyme for degradation. The following reaction mix was used for the PCR: 20ng pET-rmlD construct, 5µl of 10 x Pfu DNA polymerase buffer, 6µM dNTPs (nucleotide substrate for polymerase), 125ng Sense primer (5'-3'), 125ng Anti-sense primer (3'-5'). The mixture was made up to 49µl with water and 1µl of Pfu DNA polymerase was added. The PCR reaction was carried out in a GeneAmp PCR thermocycler (Applied Biosystems) over 12 cycles using the method detailed in Fig 2.5.1.1.

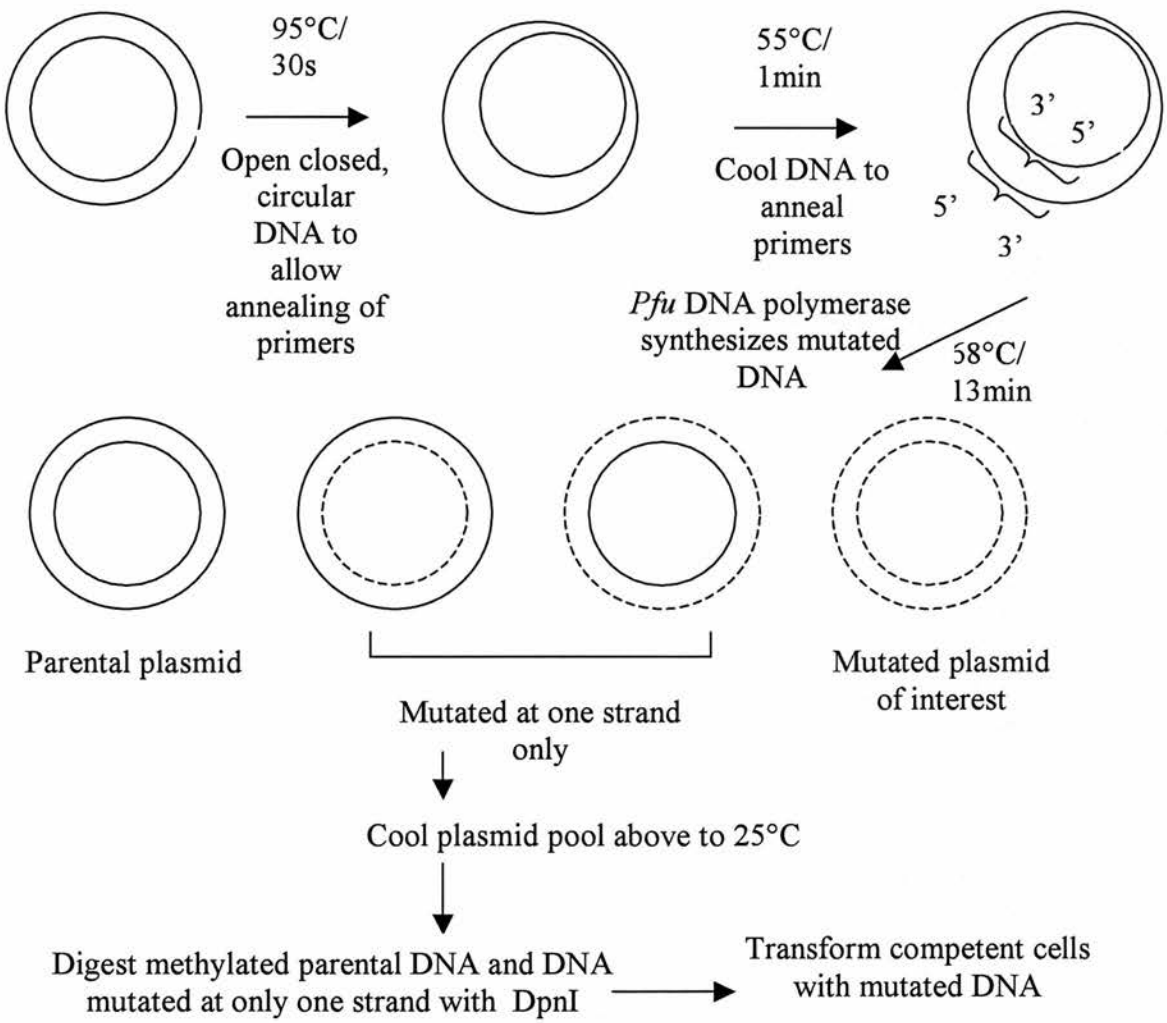

Fig.2.5.1.1 PCR mutagenesis

After 12 cycles of PCR, the success of the PCR reaction was assessed by running 5µl of the reaction mix on a 1% agarose gel - the remaining parental and partially mutated DNA was removed from the plasmid pool by addition of DpnI to the reaction mix. After digestion, 5µl of the PCR mix was checked on an agarose gel (*Fig 2.5.1.2, 2.5.1.3*).

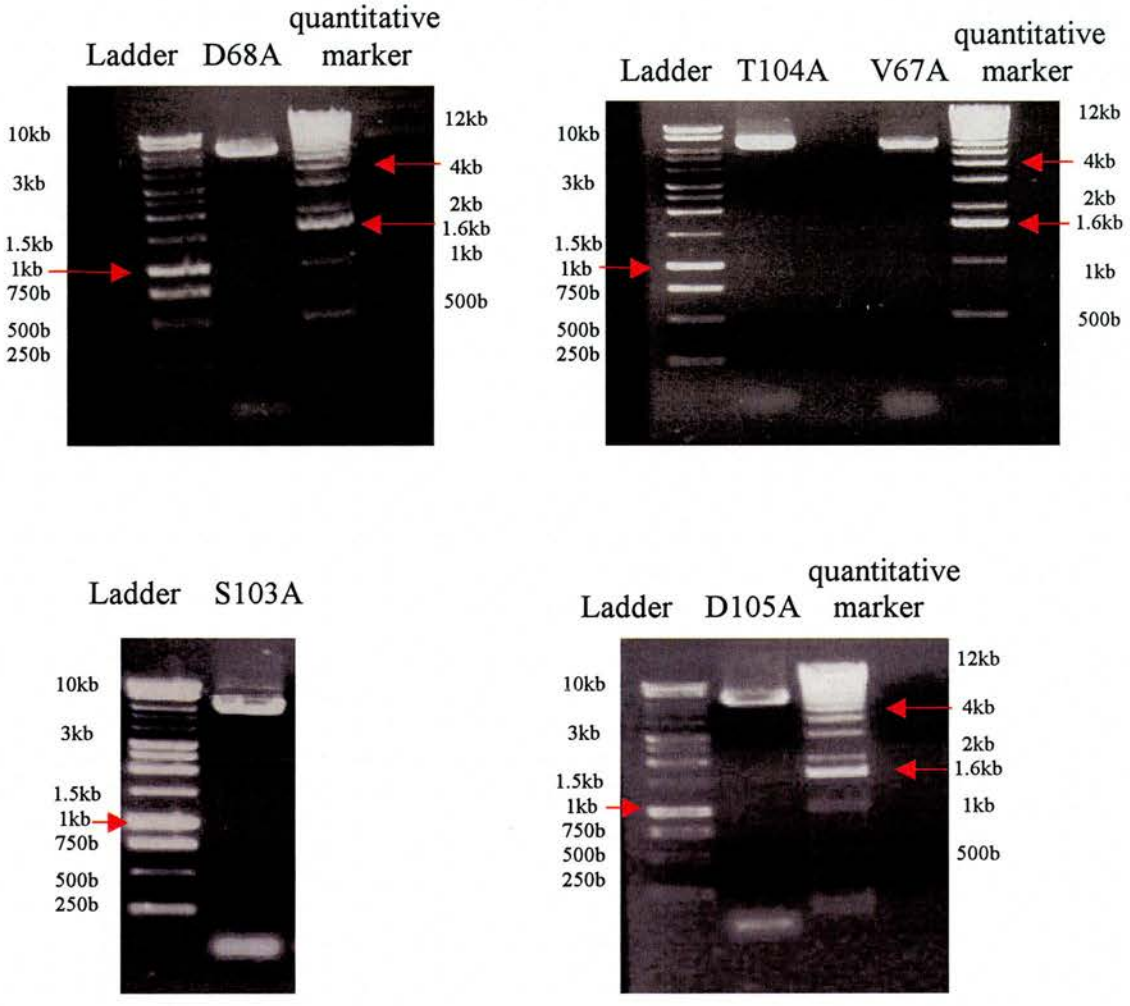

Fig.2.5.1.2 Gel analysis of the PCR products for RmlD mutants D68A, T104A, V67A, S103A and D105A. The band corresponding to 10kb on the DNA ladder represents the circular plasmid product.

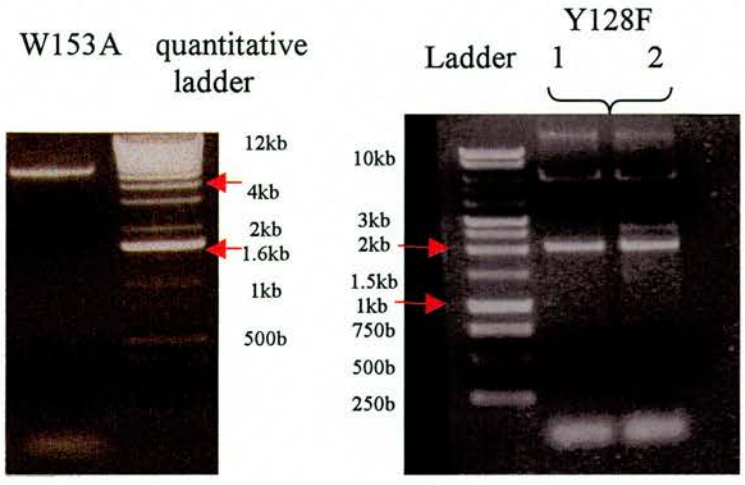

Fig.2.5.1.3 Gel analysis of the PCR products for RmlD mutants W153A and Y128F. Gels are annotated as the previous figure.

All of the PCR reactions were successful. In performing this experiment two bands are expected for each PCR mutant - one at 10kb for the amplified plasmid DNA (top in the gels) and another near the end of the gel representing the PCR primers. Figures 2.5.1.2 and 2.5.1.3 show the results of the PCR reaction after incubation of the plasmid pool with DpnI. The electrophoretic properties of the plasmid pool before DpnI digestion are exactly the same. Two bands are observed as above and only a slight decrease is noticeable in the amount of plasmid DNA on the gel.

The only anomalous result was that obtained for the Y128F mutant. As can be seen in Figure 2.5.1.3, there is an additional band at 2kb. Analysis of the Y128F forward primer and the pET 30a(+) sequence in DNAMAN suggested that this primer annealed elsewhere on the vector (*Fig.2.5.1.4*).

Y128F : pET 30a(+) identity= 67%

```
2       GTCGCCGCTGAATGTCTTTGGCAAAACCAAACTGGCG Y128Fprimer
        ||||  ||||| || | ||| ||   |||  |||||
1005    GTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCC pET 30a(+)
```

Fig.2.5.1.4 A region of complementarity between the Y128F forward primer and pET 30a(+) sequence outside the rmlD gene

The pET 30a(+) sequence in Figure 2.5.1.4 corresponds to a region on the lacI gene from bases 1005 to 1041. Therefore it is possible that a small linear stretch of DNA was amplified from some of the plasmid pool by non-specific binding of the primer to a region outwith the intended annealing site of the rmlD gene. This would explain the extra band on the gel. This linear DNA did not affect transformation of the mutated vector or subsequent expression of the Y128F mutant. Sequencing proved the incorporation of the correct mutation in the rmlD gene.

2μl of each reaction was used to transform 1ml of competent *E. coli* Novablue cells (Novagen) using the Heat Shock protocol described in section 2.4.2. After an hour's incubation at 37°C, 100μl of cells were spread on an L-agar plate containing 30μg/ml Kanamycin and incubated at 37°C overnight. The next day 9ml of LB + 30μg/ml kanamycin were inoculated with a colony from the plate and incubated overnight at 37°C.

2.5.2 Isolation and characterisation of mutated plasmid DNA

Plasmid DNA was isolated from overnight cultures (500ml) using the methods described previously for purification of the native construct. DNA was eluted in 250-400µl of water.

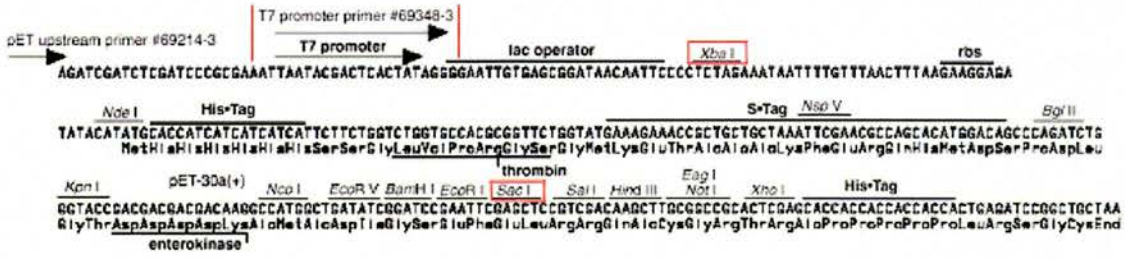

Fig 2.5.2.1 pET 30a(+) cloning/ expression region

5µl of each maxiprep was used for gel analysis of the plasmid DNA (*Fig 2.5.2.2*). The purified plasmid DNA for the following mutants: D105A, T104A and W153A, was highly viscous (suggesting that the DNA was highly concentrated) and therefore diluted for agarose gel analysis.

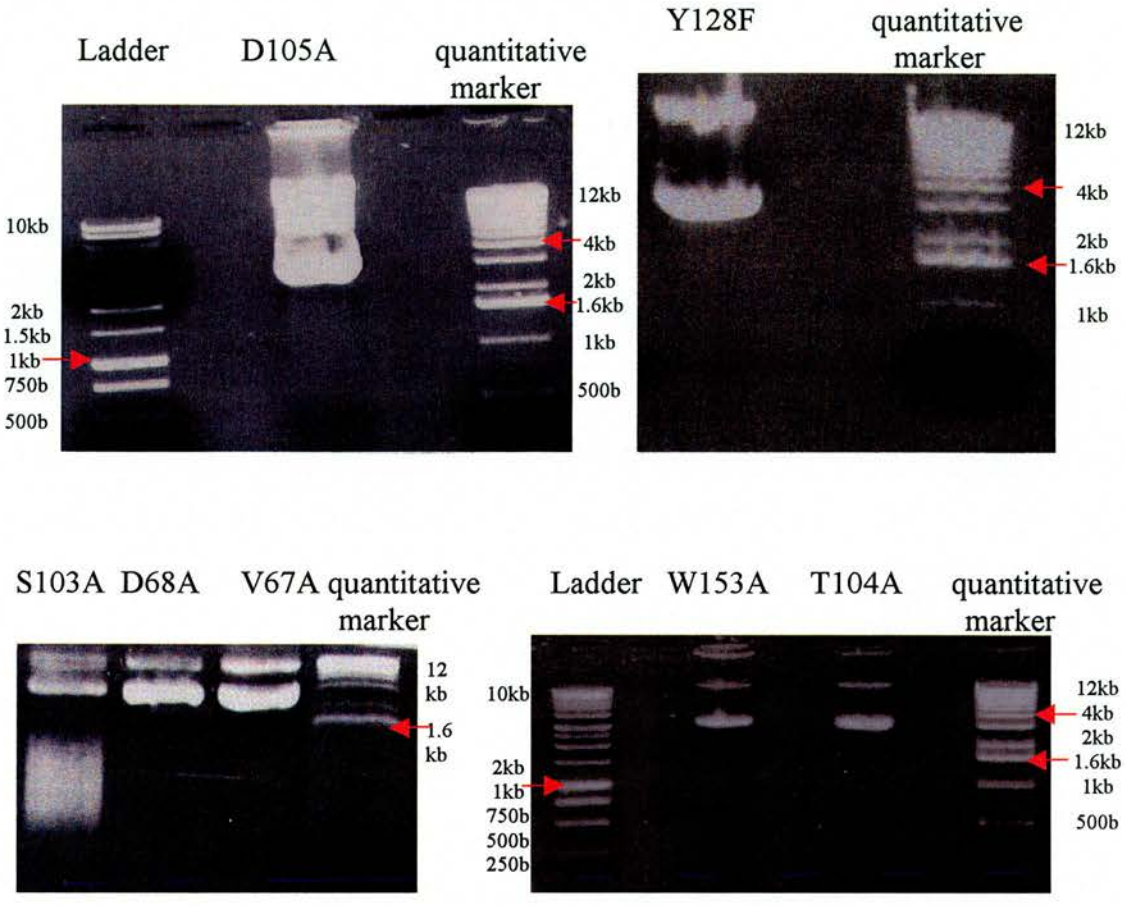

Fig.2.5.2.2 Purified plasmid DNA for the RmlD mutants. Each sample shows circular DNA at 10-12kb and supercoiled DNA at 2-4kb.

Figure 2.5.2.2 shows the purified rmlD mutant constructs. For each mutant the top band represents the circular plasmid DNA and the band below represents the supercoiled plasmid DNA. Both are observed in the successful purification of plasmid DNA.

The purified DNA was sequenced to check for the incorporation of the relevant point mutation using an ABI PRISM 377 DNA Sequencer. The reaction mix. contained: 300-500ng plasmid DNA and 5pmol T7 forward primer in 12μl of water. The T7 primer reads from the T7 promoter, upstream of the rmlD gene (*Fig 2.5.2.1*). To read the 3' end of the gene, sequencing was carried out as above with the T7 reverse primer which reads from the opposite end of the sequence.

Analysis of the sequences was carried out using the programs CHROMAS and DNAMAN.

2.5.3 Expression and purification of RmlD mutants

Competent *E. coli* BL21(DE3) pLysS cells (Novagen) were transformed with the mutated pET-rmlD construct using the protocol described in 2.4.2. The concentrated DNA was difficult to quantify, but the gel analysis suggested that 2-4µl of maxiprep DNA was sufficient for each transformation. *E. coli* BL21(DE3) pLysS is resistant to chloramphenicol, via the chloramphenicol acetyl transferase gene on the pLysS vector. This allows selection of pLysS which encodes T7 lysozyme, a natural inhibitor of T7 RNA polymerase, preventing background expression of proteins in uninduced cells (Studier, Rosenberg et al. 1990)(*Fig 2.5.3.1*).

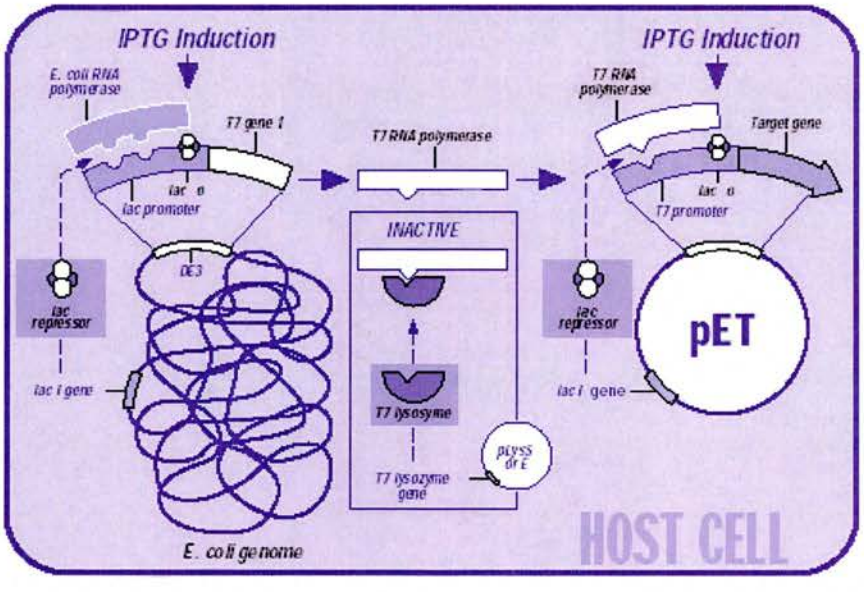

Fig.2.5.3.1 Protein expression in the pET-pLysS system (reprinted from Calbiochem protein expression manual)

Expression of mutant RmlD proteins was achieved using the method outlined by *Giraud et al* for the wild-type enzyme (Giraud, McMiken et al. 1999). A 10ml starter culture, containing 30μg/ml kanamycin (Sigma) and 34μg/ml chloramphenicol (Sigma), was inoculated with a single, transformed colony. The culture was incubated overnight at 37°C with constant shaking. 5ml of the overnight culture was used to inoculate a final culture volume of 500ml, containing 495ml of 2xTYE broth, 30μg/ml kanamycin and 34μg/ml chloramphenicol.

Cells were induced by addition of 1mM IPTG at an OD_{600} of 0.6-0.8 and grown for 3 hours, with constant shaking at 37°C. The cells were then pelleted by centrifugation for 10 minutes at 6000g, 4°C and lysed in a mixture of 150mM NaCl, 2mM dithiothreitol (DTT), 5mM phenylmethyl sulphonyl-fluoride (PMSF), 20μM lysozyme and 20mM Tris-HCL pH 9.0. The mixture was incubated at 4°C for 30 minutes, and lysis completed by sonication (4 x 30second cycles, each interrupted by 1 minute on ice). The lysate was centrifuged for 30 minutes at 30,000g, 4°C and the resulting supernatant containing the contents of the lysed cells was brought to 20% ammonium sulphate saturation, with constant, light stirring and incubated for at least 1 hour at 4°C. The saturated mixture was centrifuged for 20 minutes at 20,000g, 4°C and the supernatant dialysed against 2 X 2L of 20mM Tris-HCl pH 9.0 (the second dialysis was carried out overnight).

The dialysed supernatant was filtered through a 2μm syringe filter (Amicon) and applied to an HQ-anion exchange column (POROS) using a BiocadSprint HPLC. Proteins were eluted with an increasing gradient of NaCl. A 20μl sample of each fraction was taken and prepared for SDS-PAGE by addition

of 7µl of loading dye and 3µl of antioxidant (a reducing agent, which helps to unfold the protein through breakdown of disulphide bonds). The samples were heated at 70°C for 10 minutes and ran on a pre-cast gel (4-12% Bis-Tris) using the NuPAGE system (Invitrogen). Fractions corresponding to the molecular weight of RmlD (32.5kDa) and a single peak on the HPLC trace were pooled and concentrated. A second step, involving hydrophobic chromatography was omitted as the protein could not be recovered from the column. The purity of the concentrated protein was confirmed by SDS-PAGE as above (*Fig.2.5.3.2, 2.5.3.3*). Overall, the typical yield of pure protein was 12-15mg per litre of culture. The concentrated protein was aliquoted and stored at -70°C in 50% glycerol for further use.

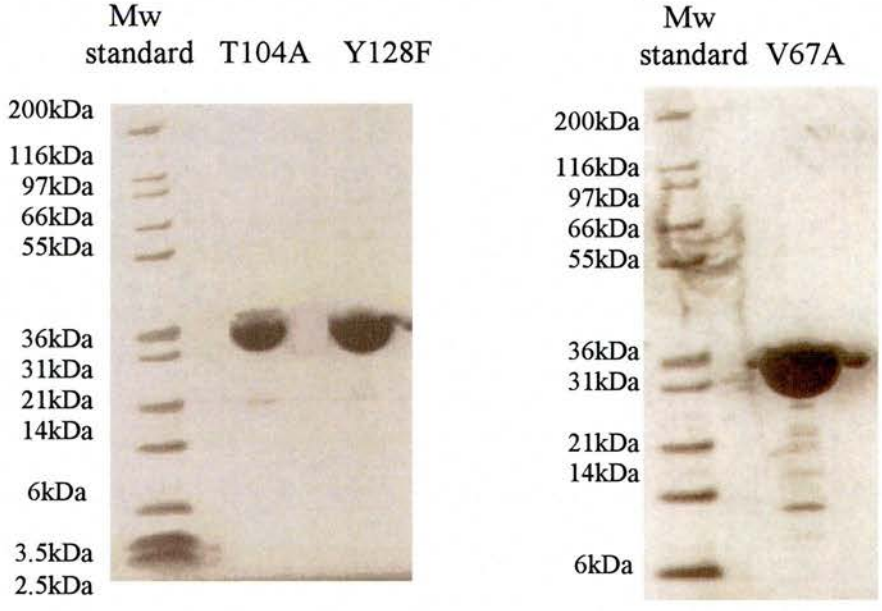

Fig.2.5.3.2 SDS-PAGE gels showing the purified, concentrated RmlD mutant proteins

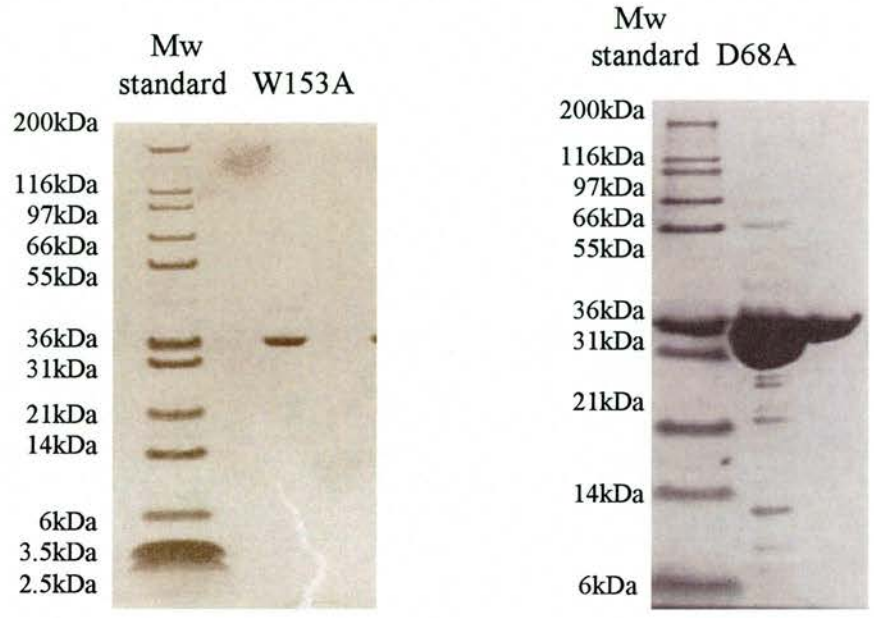

Fig.2.5.3.3 SDS-PAGE gels showing the purified, concentrated RmlD mutant proteins

The T104A, Y128F and W153A mutants were purified to homogeneity as confirmed by SDS-PAGE analysis. In each case, a single Coomassie-stained band was observed between 31 and 36kDa. This agreed with the molecular weight of RmlD predicted from the protein sequence to be approximately 32.6kDa for the monomer. The V67A and D68A mutants were purified as far as was possible. However the band corresponding to RmlD is by far the major species.

2.5.4 Kinetic bioassay of RmlD mutants

To measure the effect of the active site and substrate binding pocket mutations, a coupled RmlB, C, D assay was used to describe the kinetic properties of RmlD mutants under the Michaelis-Menten model (Blankenfeldt, Kerr et al. 2002). The assay is similar to that used by *Graninger et al* (Graninger, Nidetzky et al. 1999)

and is based on the hydrolysis of NADPH by RmlD. The reaction mixture contained 20 mM Tris-HCl pH 6.5, 10 mM $MgCl_2$, 0.072 mM NADPH, and 1.47 pM RmlD. To ensure the initial rate was dependent on the concentration of RmlD, both RmlB and RmlC were use in 24-fold molar excess. The assay was determined to obey steady state kinetics on upon the observation that the rate using wild-type and mutant proteins was halved when the concentration of RmlD was halved and doubled when the concentration of RmlD was doubled. Background hydrolysis of NADPH was measured by omitting RmlD from the reaction and was found to be negligible.

To determine the kinetic parameters of the wild-type enzyme and each mutant protein, dTDP-D-glucose (the substrate for RmlB) was added to the above reaction at concentrations ranging from 0.025mM to 1mM and the consumption of NADPH was measured at 340nm using a Helios Scanning Spectrophotometer (UNICAM). The results were analysed using VISION software (UNICAM) to determine the initial rate of each reaction. All reactions were performed in triplicate and the mean initial rate was used for calculation of Michaelis-Menten parameters. To obtain measurable rates, the V67A and D68A RmlD mutants were used at 6 and 2-fold higher concentration relative to other RmlD mutants. K_m and V_{max} were calculated by non-linear regression using Graphpad Prism (Graphpad software)(*Fig 2.5.4.1* and *2.5.4.2*).

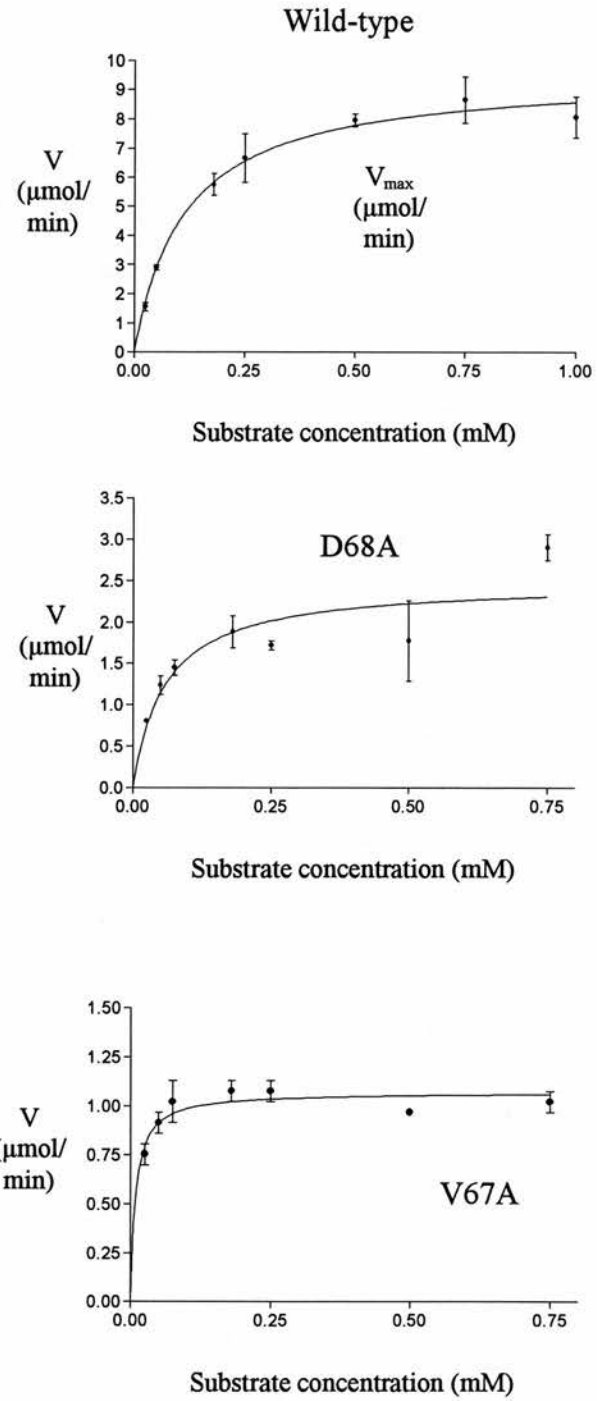

Fig.2.5.4.1 Michaelis-Menten curves for Wild-type RmlD , the D68A mutant and the V67A mutant.

Because the assay is coupled to three enzymes only apparent kinetic parameters could be derived. The substrate for RmlD is unstable and difficult to synthesize (Graninger, Nidetzky et al. 1999), complicating the design of a more direct assay for activity. In this experiment it was important to determine the activity of the RmlD mutants relative to wild-type, not the absolute kinetic parameters.

Enzyme	V_{max} (μmol/min/ml)	Apparent K_m (mM)	Relative specific activity
Wild type	9.55 ± 0.50	0.11 ± 0.02	1 ± 0.05
D68A	2.48 ± 0.23	0.06 ± 0.02	0.1 ± 0.01
V67A	1.07 ± 0.04	0.008 ± 0.003	0.02 ± 0.001
T104A		activity not detected	
Y128F		activity not detected	
W153A		activity not detected	

Table.2.5.4.2 Michaelis-Menten parameters for wild-type and mutant RmlD proteins

Figure 2.5.4.1 and Table 2.5.4.2 show the experimental data for those mutants that displayed measurable enzyme activity. In the context of these results, specific activity is defined as:

$$\frac{V_{max}}{\text{Total enzyme concentration}}$$

Where V_{max} is the maximum velocity of the reaction in μmol/min/ml and the total enzyme concentration is expressed in mg/ml.

For clarity, these values were then expressed in relative terms, expressing wild-type RmlD specific activity as 1 (100% active) and V67A and D68A as a

fraction of wild-type activity It was not possible to obtain measurable rates for T104A, Y128F and W153A even when used in >100 fold molar-excess relative to wild-type RmlD.

Discussion

2.6 Overall structure

RmlD, is a member of the SDR family, performing the final step in dTDP-L-rhamnose biosynthesis, a pathway for therapeutic intervention. The crystal structure of RmlD from *Salmonella typhimurium*, solved in our lab, confirms the presence of a two-domain protein with an N-terminal Rossman fold and a C-terminal substrate binding cleft (Blankenfeldt, Kerr et al. 2002)(*Fig.2.6.1*).

A six-stranded parallel β-sheet interspersed with α-helices forms the characteristic β-α-β repeating unit of the SDR Rossman fold (the larger of the domains) and contains the residues required for co-factor binding and those implicated in the catalytic mechanism - Ser103/Thr104, Tyr128 and Lys132. This domain is largely similar to those found in other SDRs.

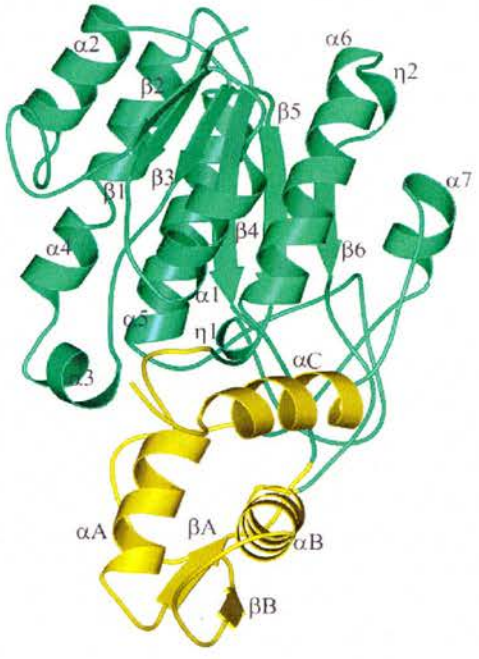

Fig.2.6.1 The two-domain structure of the RmlD monomer annotated according to secondary structure.The N-terminal co-factor binding domain is coloured green and the C-terminal substrate binding domain coloured yellow

As mentioned in section 2.3, the conserved Wierenga motif is important in correct positioning of the pyrophosphate backbone. Thr65 from the opposite side of the nucleotide binding pocket makes a hydrogen bond with the pyrophosphate, with the effect that the protein clamps down on the co-factor (*Fig.2.6.2*).

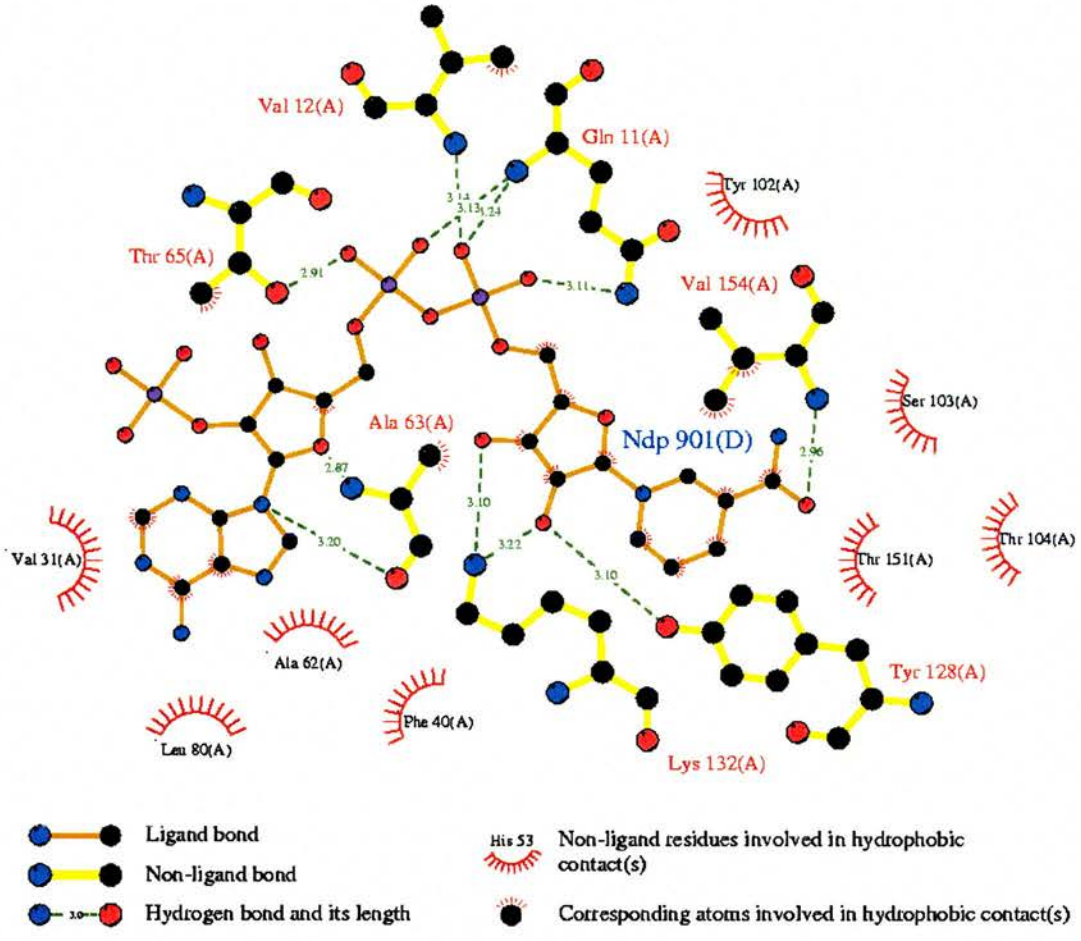

Fig.2.6.2 Ligplot representation of interactions between RmlD and NADPH (Ndp 901[D])

The ribose moeity connected to the adenine ring is sandwiched between Gly7 and 10 on one side and Ala62 and 63 on the other. The adenine ring itself is located in a hydrophobic pocket consisting of Val31, Phe40, Ala62, Ala65 and Leu 80. Tight recognition is conferred by hydrogen bonds with the backbone

nitrogen of Phe40 and the carboxylate of Asp39. Recognition of the nicotinamide ring and corresponding ribose moeity comes mainly from the active site residues and substrate. This is discussed later.

RmlD is uniquely able able to bind and use both NADH and NADPH nearly equally well as co-factors in the reductive biosynthesis of dTDP-L-rhamnose (Graninger, Nidetzky et al. 1999). The additional phosphate of NADPH points out towards the solvent and therefore no selective switch exists to prevent the entry of this group. Other SDRs such as RmlB have a bulge in an extended loop (between the equivalent of β2 and α2 in RmlD) that closes over this solvent exposed region and therefore favours binding of NAD(H) over NADP(H) (*Fig.2.6.3*).(Allard, Giraud et al. 2001). The RmlD-NADH complex shows that the ribose unit near the adenine ring is tightly bound by the terminal carboxyl motif of Asp30 (*Fig.2.6.4*). This residue is highly conserved in SDRs and governs co-factor specificity (Jornvall, Persson et al. 1995).

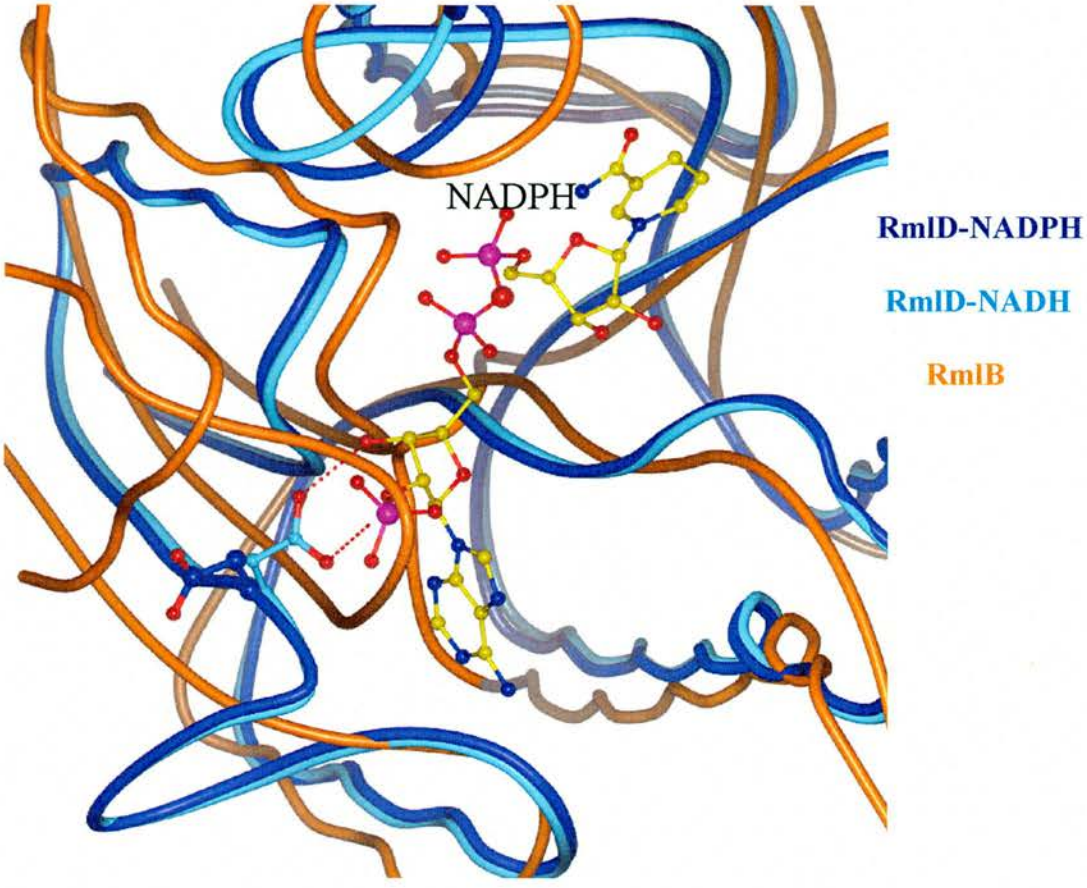

Fig.2.6.3 NAD(P)H binding by RmlD occluded by an elongated loop in RmlB (orange). RmlD-NADH structure is coloured cyan, while the RmlD-NADPH structure is coloured royal blue. NADPH and the conserved Asp30 in RmlD are depicted in ball and stick.

Through side chain torsional adjustments, Asp30 points out of the co-factor binding site upon NADPH binding to avoid an accumulation of negative charge with the additional phosphate group of the dinucleotide (*Fig.2.6.3*). A weaker interaction with water molecules replaces the displaced Asp30 and may explain the slight preference that RmlD has for NADH.

The substrate binding domain is a solvent exposed cleft generated by a parallel β-sheet packed adjacent three α-helices. The sheet pairs two strands that

lie in a partially disordered region comprising residues 174-179 and 239-258 (*Fig.2.6.4*).

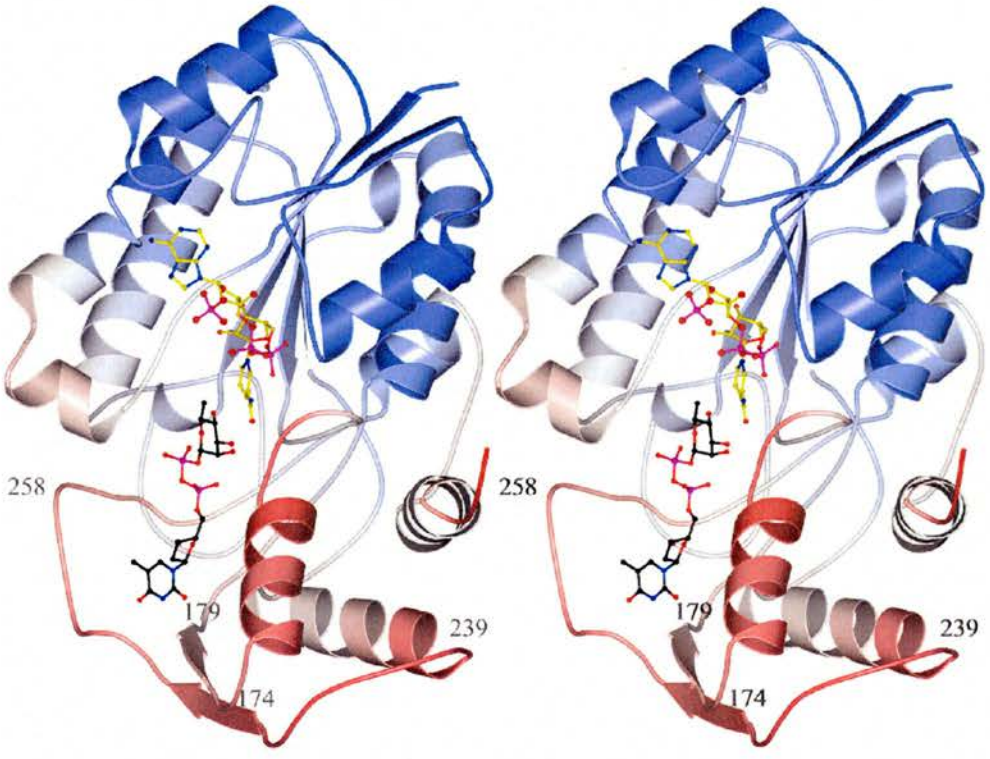

Fig.2.6.4 Stereo view of the RmlD monomer with NADPH (yellow backbone) and dTDP-L-rhamnose (black backbone) bound. Secondary structure is coloured by isotropic thermal-factors, indicating the flexibility of the substrate-binding domain (red) and the comparatively well-ordered co-factor binding domain (blue)

The open conformation of the active site allows the co-crystallised product (dTDP-L-rhamnose) to extend across the length of the substrate binding domain, parallel to helix αA. The TDP moeity points out of the cleft and is therefore disordered, while the carbohydrate ring sits deep inside and is comparatively tightly bound.

2.7 Dimerisation

Although only one RmlD monomer is located per asymmetric unit, application of crystallographic symmetry operators shows that a neighbouring monomer in an adjacent asymmetric unit is oriented to generate a possible dimer interaction. The crystal structure shows that three glutamic acid residues (Glu15, Glu190 and Glu292) from each monomer bind a single magnesium ion in the centre of the dimer interface via water mediated contacts (*Fig.2.7.1*)

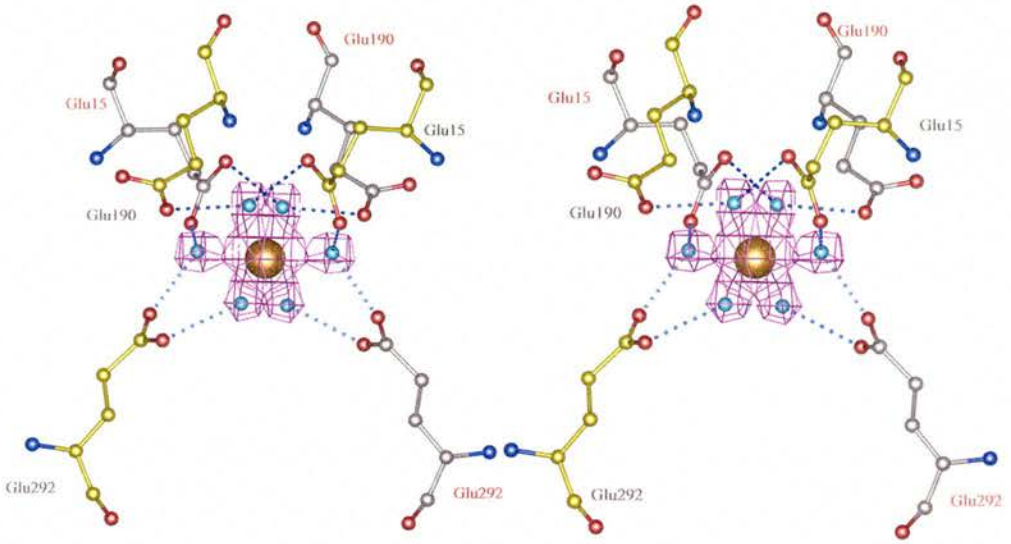

Fig.2.7.1 Stereo view of the RmlD dimer interface stabilised by a hexa-coordinated Mg^{2+} ion. Residues from each monomer are coloured yellow and grey respectively.

Glu15 and Glu190 are found in the Rossman fold on helices α1 and α 6 respectively, while Glu292 is found in the substrate-binding domain on helix αC. The free negative charge present at the interface is stabilised by positive charges from Arg18 and Arg194. This method of dimerisation has not been described for any other SDR. Those that exist as dimers (Varughese, Xuong et al. 1994) or

tetramers (Ghosh, Wawrzak et al. 1994) generally present a four helix bundle at the interface, two from the Rossman fold of each monomer (α4 and α5 in RmlD, see *Fig.2.6.1*), held together by mainly hydrophobic interactions. The interface is thus generated at the opposite face of the molecule when compared with the dimer interface in RmlD. In each monomer, helix αF holds the highly conserved and catalytically important tyrosine and lysine. The arrangement of a four -helix bundle in other SDRs serves two purposes, i) to prevent interaction between the two active sites by positioning the active site residues from each monomer away from one another and ii) to position the substrate binding domains distal to one another. No similar interaction is observed in RmlD crystals.

The fact that relatively high concentrations of $MgCl_2$ (100mM) are required for kinetic assay of RmlD in this study points to the essential nature of Mg^{2+} in enzyme activity. Our collaborators in the lab of Chris Whitfield have shown that RmlD can be reversibly inactivated by addition of excess EDTA, presumably sequestering the available magnesium ions required for dimerisation (Graninger, Nidetzky et al. 1999). Full activity is restored upon addition of excess $MgCl_2$ (Graninger, Nidetzky et al. 1999).

Nevertheless, it is not immediately apparent why RmlD would function as a dimer. All the residues required for ligand binding and catalysis are located in the monomer and dimerisation would appear to place the two active sites dangerously close to one another when compared with other SDRs. It is possible that magnesium binding is a way of compensating for the negative charge carried by the three glutamic acid residues at the dimer interface. Additionally, the structure highlights a number of hydrophilic interactions between the α-helices involved in dimerisation, suggesting that magnesium binding may be required for

correct positioning of the substrate and co-factor binding domains in each monomer. The only structural characterisations of an SDR functional as a monomer are Porcine Carbonyl reductase and Human Biliverdin Ixβ reductase (Ghosh, Sawicki et al. 2001; Pereira, Macedo-Ribeiro et al. 2001). Porcine Carbonyl reductase has an additional three helical "insertion domain" in the Rossmann fold, not present in RmlD (Ghosh, Sawicki et al. 2001). Two of the helices partially shield αE and αF preventing the formation of the four-helical bundle described above. Human Biliverdin Ixβ reductase is a single domain protein consisting of a single Rossmann fold comprising a seven-stranded parallel β-sheet flanked by helices, a short antiparallel strand and short parallel β-sheet (Pereira, Macedo-Ribeiro et al. 2001). The variable substrate binding domain common to most SDRs is missing making the overall tertiary structure of this enzyme quite differrent from the conventional SDR model.

2.8 Enzyme mechanism and substrate recognition

2.8.1 Mutational/structural analysis of the substrate-binding domain

This study involved the identification of important substrate-binding residues in RmlD. Such residues were inferred from initial structural data, and the resulting model was highly disordered. Only latterly was it possible to determine the correct topology of the substrate-binding domain, although some regions remained flexible. Nevertheless, those residues inferred from the initial model and mutated yielded some interesting results most of which can be rationalised by analysis of the final, refined model. All mutated residues are highly conserved in RmlD.

Fig.2.8.1.1 Primary sequence alignment of RmlD proteins from a selection of medically important pathogens showing conservation of residues in the substrate-binding pocket. Residues marked with a star were mutated to alanine in this study and monitored for enzymatic activity.

Asp68 found on α3, and on the opposite side of the substrate-binding cleft from αA, was mutated to alanine to eliminate the hydrophilic side chain. In the apo-enzyme and abortive complex this residue forms a salt bridge with the highly conserved Arg260 coordinating it to donate a proton from its guanidium head to O3 on the phosphate backbone of the co-crystallised product (*Fig.2.8.1.2*). In fact no other contacts occur with the phosphate backbone and negligible electron density is observed around this part of the ligand. Therefore, in the coupled enzyme assay, the D68A mutant shows a 10 fold drop in specific activity presumably because coordination of Arg260 is lost. No other residue appears to interact with Arg260 apart from perhaps Tyr106 which may make a hydrophobic interaction with the side-chain of Arg260.

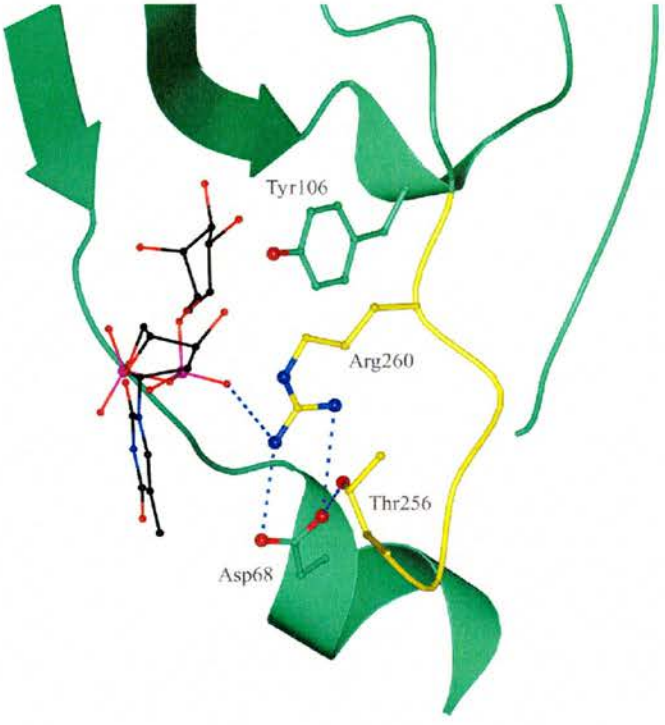

Fig.2.8.1.2 Possible important interactions of Asp68 with residues in the substrate-binding pocket.

Interestingly, the apparent Km of the D68A mutant is almost half that of the wild-type enzyme, however if the error in Km is taken into account then it does not differ much from the wild-type Km.

The V67A mutant shows a 50-fold drop in specific activity compared to wild-type. This residue forms a hydrophobic pocket around the methyl group on C6 of the co-crystallised product with Thr65, Tyr106 and Tyr128 (*Fig.2.8.1.3*). Mutation of Val67 to alanine may weaken this interaction. Additionally the V67A mutant displays a 14-fold decrease in apparent Km compared to the wild-type enzyme and it is difficult to rationalise this purely by analysis of the available structural data.

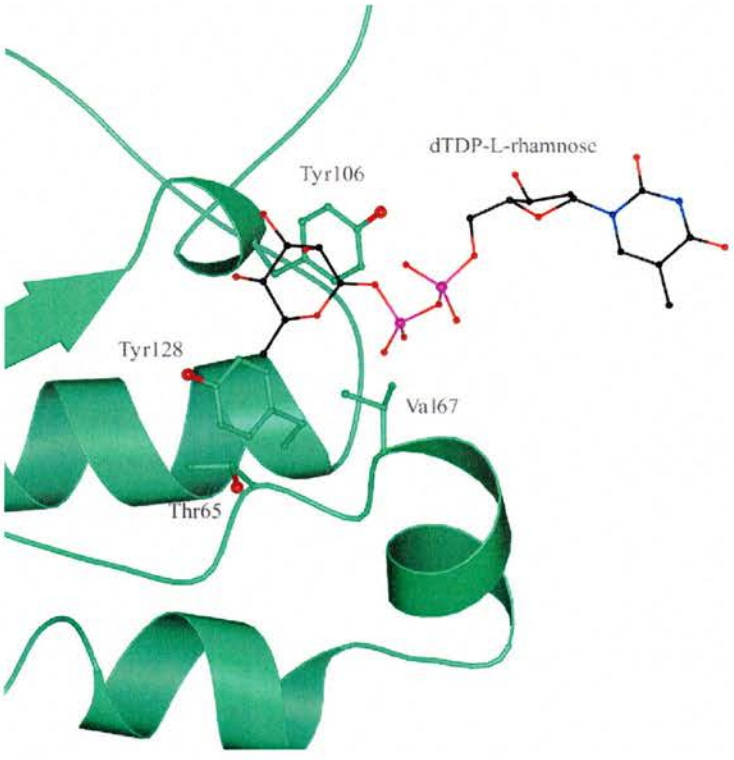

Fig.2.8.1.3 Val67 is proposed to be an important residue in a hydrophobic pocket that interacts with the carbohydrate ring on dTDP-L-rhamnose.

Trp153, found on the short loop immediately preceeding αA, does not seem to play a direct role in substrate binding, but is rather involved in the orientation of some important residues in the carbohydrate-binding pocket. Mutation to alanine in this study effectively inactivated RmlD and no measurable activity was detected in the kinetic assay. This residue is part of a hydrophobic pocket including Phe162, Trp223 and Tyr226 (*Fig.2.8.1.4*)

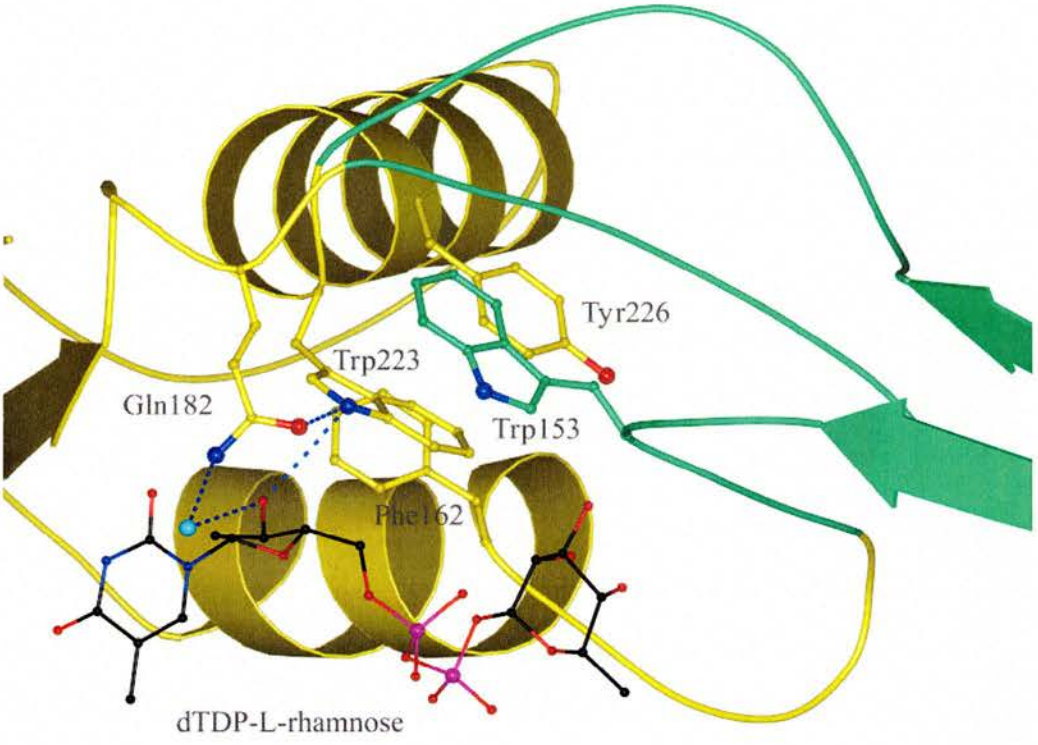

Fig.2.8.1.4 Interaction of Trp153 with residues lining the substrate-binding pocket. The figure is coloured as fig 2.6.1.

Trp153 stacks against Trp223 which donates a proton from the nitrogen on its indole ring to O3 on the ribose ring of dTDP-L-rhamnose (*Fig.2.8.1.4*). Additionally Trp223 makes a hydrogen bond with Oε1 on the amide group of Gln182 which proceeds through a water molecule to make an additional hydrogen bond interaction with the ribose ring on the TDP-moeity of the product (*Fig.2.8.1.4*). The combined elimination of these interactions in the W153A mutant may be sufficient to destablise the structure of the substrate binding domain (by disturbing a nearby hydrophobic pocket) and compromise the affinity of RmlD for an already weakly-bound substrate.

2.8.2 Mutational/structural analysis of the active site

The mechanism of action and exact residues involved in SDR catalysis have been the subject of much debate. It is widely accepted that, at least in oxidations and reductions, a tyrosinate ion is required for acid-base catalysis, involving the transfer of a proton in the transition state. An absolutely conserved tyrosine (Tyr128 in RmlD) which lies four residues upstream from a highly conserved lysine (RmlD Lys132) in SDRs supports this theory.

Mutagenesis studies confirm the requirement for an active site residue able to donate or abstract a proton from the substrate. *Drosophila melonogaster* alcohol dehydrogenase (DADH) which catalyses the oxidation of alcohol, the reverse reaction of RmlD, is inactivated following mutation of the catalytic tyrosine (Tyr152) to phenylalanine, an action that eliminates the phenolic group of tyrosine from proton transfer in the active site (Chen, Jiang et al. 1993). The same mutation in the SDR UDP-galactose-4-epimerase (GalE) has a significant impact on catalysis with a 10, 000 fold decrease in k_{cat} compared to the wild-type enzyme (Liu, Thoden et al. 1997). Mutation of the proposed catalytic tyrosine in rat 11β-hydroxysteroid dehydrogenase to phenylalanine or serine completely inactivates the enzyme (Obeid and White 1992). Interestingly a DADH Y152C mutant retains trace activity, suggesting the need for a deprotonatable residue. Cysteine perhaps is too reactive or has a pKa outwith the required range to stably perform this function. Mutation of the downstream Lys156 in DADH to isoleucine, also inactivates the enzyme but mutation to arginine allowed the measurement of some residual activity suggesting that a basic residue is important at this position. Again similar trends are observed upon mutation of this residue in homologous enzymes (Obeid and White 1992). The analogous

tyrosine and lysine are absolutely conserved in RmlD proteins over several bacterial genera (*Fig.2.8.2.1*).

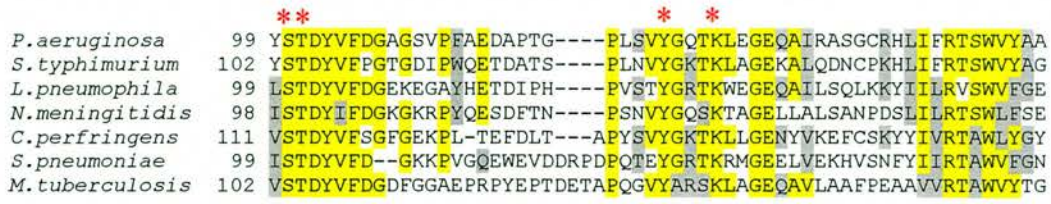

Fig.2.8.2.1 Primary sequence alignment of RmlD proteins from a selection of medically important pathogens showing conservation of residues in the active site. Residues marked with a star were mutated in this study and monitored for enzymatic activity

To obtain a more detailed model of the RmlD catalytic mechanism, mutagenesis studies were undertaken to determine the importance of some of the proposed catalytic residues. Due to the relatively basic reaction performed by the enzyme, the reduction of a single keto group to a hydroxyl, RmlD is a good model to study the mechanism of SDR acid-base catalysis. In this study the RmlD Y128F mutant is, like its SDR counterparts, rendered catalytically inactive (Blankenfeldt, Kerr et al. 2002). It was not possible to measure any detectable NADPH oxidation using the coupled enzyme assay described earlier. This correlates with structural evidence, inferring the essential nature of this residue, suggesting that the substrate is capable of approaching the catalytic tyrosine to accept a proton in the transition state (*Fig.2.8.2.2*).

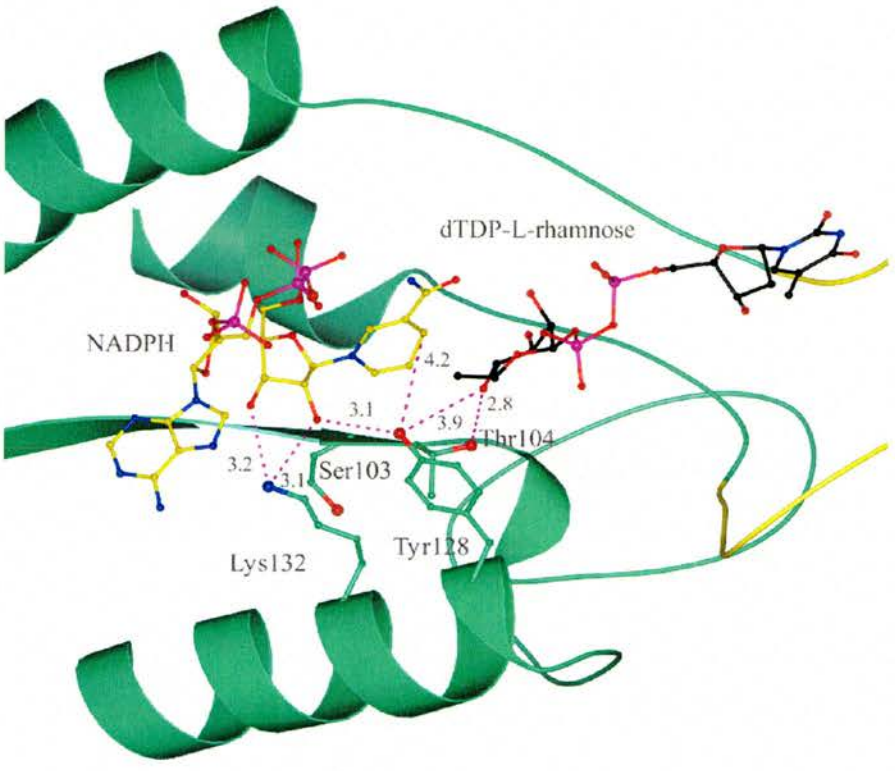

Fig.2.8.2.2 The active site of RmlD from *Salmonella typhimurium* showing the bound NADPH co-factor and product of the reaction, dTDP-L-rhamnose. Possible hydrogen-bonding contacts are labelled as distances between participating non-hydrogen atoms in angstroms.

Although not mutated in this study, the role of the conserved Lys132 is clear upon analysis of the RmlD structure (and structures of other SDRs) in conjunction with mutagenesis studies performed by other groups. For the catalytic tyrosine to participate in acid-base catalysis it must be ionised (to tyrosinate) in the transition state. After substrate turnover, the tyrosinate is reset to the fully protonated tyrosine and a new molecule of NAD(P)H is bound (thus the dinucleotide is a co-factor and not a co-enzyme). Under normal conditions the pKa of tyrosine is 10 and therefore Oη would remain protonated during

catalysis. The downstream, highly conserved lysine is proposed to be responsible for lowering the pKa of the catalytic tyrosine.

Several structural studies on SDRs have shown that Nζ on the lysine is too distant from Oη to directly interact with the phenol ring of the catalytic tyrosine (Ghosh, Wawrzak et al. 1994; Varughese, Xuong et al. 1994; Benach, Atrian et al. 1998; Allard, Giraud et al. 2001). This is also the case in RmlD, where these functional groups are separated by a distance of 4.5Å in the apo-enzyme and 5.3Å in the ternary, abortive complex with NADPH and dTDP-L-rhamnose (Blankenfeldt, Kerr et al. 2002). In comparison with GalE, it would seem that the role of Lys132 is more passive and that a combination of the positive charge on Nζ and the emerging positive charge on N1 of the co-factor nicotinamide ring (after hydride transfer to the substrate) creates an electrostatic field that stabilises the tyrosinate, allowing Tyr128 to act as the active site general acid in catalysis (Liu, Thoden et al. 1997; Blankenfeldt, Kerr et al. 2002).

The abortive RmlD complex shows that Tyr128 is too distant from C4 on the sugar ring (3.9Å) to directly transfer a proton. The ternary complex of GalE with NADH and UDP-glucose shows that a slight movement is required in the substrate binding domain to allow the catalytic tyrosine to abstract a proton from the C4 hydroxyl (the opposite reaction to RmlD) (Thoden, Wohlers et al. 2000). Similar conformational adjustments are observed in RmlB from *Streptococcus suis* and *Salmonella typhimurium* (Allard, Beis et al. 2002). Slight adjustments may be required in the active site of RmlD, however this is not observed possibly because crystal contacts prevent the movement of the C-terminal domain. It is probable that in the transition state, which would represent the true interaction between the substrate (intermediate) and Tyr128, the substrate and active site

residues are able to approach each other more closely. If this is the case, then it is possible that Oη and the C4 keto group share a proton equally through a low barrier hydrogen bond (LBHB, *Fig.2.8.2.3*). It has been shown in non-abortive complexes of *Streptococcus suis* RmlB that the catalytic tyrosine is able to form a short hydrogen bond of 2.5Å with the substrate analogue, dTDP-xylose (Allard, Beis et al. 2002). In comparison with the more recent work on RmlD (Blankenfeldt, Kerr et al. 2002) this short hydrogen bond can be thought of as a LBHB.

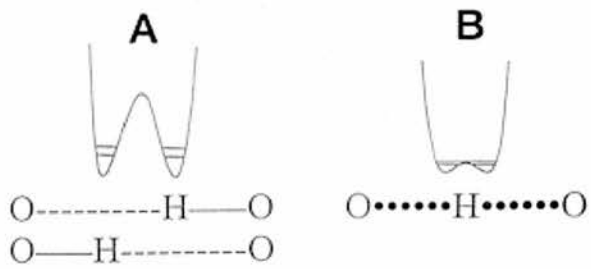

Fig.2.8.2.3 Energy diagrams for A- a traditional hydrogen bond of length 2.80Å and B - a low barrier hydrogen bond of length 2.55Å. Reprinted from (Cleland, Frey et al. 1998). The upper and lower horizontal lines are the zero energy levels for hydrogen and deuterium respectively. Curves denote the energy barrier that must be overcome to transfer a proton from one oxygen to the other

In enzymes, the 'traditional' hydrogen bond is primarily electrostatic in nature, formed between two groups of opposite charge. In such an interaction, the proton is more strongly bound to the donating atom, and therefore an energy barrier has to be overcome for proton transfer to occur. The energy released upon LBHB formation essentially lowers the activation energy of proton transfer and thus drives the otherwise unfavourable reaction. The proton is thus equally shared

between the two participating groups (more like a covalent bond) and there is virtually zero energy required to facilitate its transfer (Cleland, Frey et al. 1998; Cleland 2000). This type of interaction would be well suited to general acid-base catalysis in SDRs and it is sensible to assume that a family of enzymes that catalyse a large number of essential reactions would possess such an efficient mechanism.

Other attempts to rationalise the large distance between the catalytic acid/base and substrate in SDRs include the proposal of a "proton shuttle" between the substrate and catalytic tyrosine. In RmlD, this residue was thought to be the highly conserved Thr104. However, in all the structurally determined SDRs, the equivalent residue has been shown to be too distant from the catalytic tyrosine to fulfil this role. In RmlD, the distance between Oγ of Thr104 and Oη of Tyr128 is 4.4Å. However, in the absence of an active, ternary complex it cannot be categorically said whether or not this residue moves closer to the active site tyrosine during catalysis. Nevertheless, there is no movement observed in this region in the ternary complex of GalE with NADH and UDP-glucose. At 4.8Å, the distance between the "proton shuttle" and catalytic tyrosine is still too great (Thoden, Wohlers et al. 2000). Another proposed proton shuttle is the highly conserved Ser103 in RmlD, however this points away from the active site. Mutation of Thr104 in this study completely inactivates the enzyme. Mutation of the equivalent Ser132 in GalE leads to almost a 3000-fold decrease in k_{cat} (Liu, Thoden et al. 1997). In conjunction with the high conservation of a serine or threonine at this position and the close proximity of this residue to the substrate C4 hydroxyl/carbonyl in SDR ternary complexes (Breton, Housset et al. 1996; Thoden, Wohlers et al. 2000; Blankenfeldt, Kerr et al. 2002), this

biochemical evidence suggests an essential role in catalysis for this residue. As proposed for the equivalent residue in GalE, Oγ on Thr104 O4 is close enough to O4 on the substrate to form a hydrogen bond. The distance involved in the abortive RmlD complex is 2.8Å. This interaction is proposed to fine-tune the pKa of the substrate (pKa between 12 and -6 depending on the chemical state) to match that of the ionised Tyr128 (pKa ~6). It is essential that the pKa of the two participating groups are matched otherwise a LBHB is not formed (Cleland, Frey et al. 1998). Therefore mutation of threonine to alanine in this study would disturb this fine-tuning and prevent proton transfer. This may explain why the T104A mutant is catalytically inactive and it would be interesting to determine the effect of mutating this residue to serine.

Future work

This work provides a detailed insight into the nature of important catalytic and substrate-binding residues in RmlD. However some questions still remain. The unequivocal role of the proposed catalytic lysine (Lys132) remains unproven. Although the essential role of this residue is clear from the crystal structures of RmlD (and other SDRs) and mutational analysis in homologous proteins, enzymatic assay of a K132A mutant should provide robust biochemical evidence. The need for a basic residue at this position to generate the electrostatic field required to lower the pKa of the catalytic tyrosine could be proven by generating and assaying a K132R mutant. The role of Ser103, if any, in catalysis is still unclear. A proton shuttling mechanism involving this residue appears impossible as this residue points out of the active site. The enzymatic assay of a S103A mutant will determine if this residue does indeed play any role. Although the mutated DNA for a S103A mutant was generated, it was not possible to express this mutant from the vector. The expression clearly must be optimised. Similarly expression of the D105A mutant must be optimised. This residue plays an important role in substrate binding accepting a proton from the equatorial O3' on the rhamnosyl sugar ring.

The crystal structures of RmlD (apo and complex) still leave some questions about the catalytic mechanism unanswered. How is catalysis achieved when Tyr128 is outwith hydrogen bonding distance of C4 on the product (dTDP-L-rhamnose) and the reactive site on the co-factor ?. The formation and crystallographic determination of a ternary complex of RmlD-NAD(P)H-substrate may show that the active site residues are able to approach co-factor and substrate more closely. However, such ternary complexes tend to turnover in

the crystal. Additionally the substrate for RmlD is highly unstable. The formation of a low-barrier hydrogen bond in catalysis still needs to be proven. Neutron scattering would allow the positions of hydrogen atoms in RmlD to be determined although this requires crystals that diffract to higher resolution than at present. Theoretical calculations may shed light on whether a low barrier hydrogen bond is energetically possible between the catalytic tyrosine and substrate.

Knowledge of the RmlD crystal structure and complexes have allowed our collaborators (Richard Lee, University of Tennesee/Mike McNeil, Colorado State University) to begin designing specific inhibitors for this enzyme, validated by a combination of virtual screening and enzymatic assay. Virtual screening shows that many of the hits generated are co-factor mimics and this may be one of the general problems in designing inhibitors against enzymes with nucleotide-binding folds. Nevertheless the crystal structures of RmlD described herein show that the substrate binding pocket is highly specific for 6-deoxysugars containing a methyl group on the sugar ring making this enzyme a promising target for specific inhibition.

Chapter 3

Insights into ssDNA recognition by the OB fold

Summary

DNA is constantly under assault by a wide variety of cellular and environmental factors causing disruption of the DNA duplex. The double-helix can be distorted or damaged by exposure to ultraviolet radiation (reviewed in (Lindahl and Wood 1999)) chemical attack (reviewed in (Lindahl 1993)) and oxidative stress (reviewed in (Lindahl 1993; Finkel and Holbrook 2000)). In addition, single stranded regions are exposed during DNA replication, transcription (Wold 1997) and repair (Sandigursky, Mendez et al. 1996; Wold 1997; Genschel, Curth et al. 2000). To combat this, single stranded DNA-binding proteins (SSBs) have evolved to protect and sequester single-stranded DNA (ssDNA) until the duplex can be reformed. ssDNA binding by SSBs is non-sequence specific although different proteins have affinities for differing lengths of ssDNA.

The essential role of SSBs is underscored by their ubiquity in nature with examples from bacteria, eukaryotes, the archaea and even viruses. Although it is difficult to assign a universal role, SSBs do appear to be important factors during DNA replication. Experiments by *Gutierrez et al* have demonstrated that in-vitro replication of *Bacillus subtilis* phage ϕ29 is stimulated by *E. coli* SSB, Human replication protein A (hsRPA), T4 gp32 and adenovirus DNA-binding protein (AdSSB/Ad DBP) as well as the ϕ29-encoded p5 (Gutierrez, Martin et al. 1991). The diversity and common features present in this family are demonstrated by a number of well-studied examples discussed later (Bochkarev, Pfuetzner et al. 1997; Raghunathan, Ricard et al. 1997; Yang, Curth et al. 1997; Raghunathan, Kozlov et al. 2000; Bochkareva, Belegu et al. 2001).

3.1 SSB topology and domain organisation

Single stranded DNA binding proteins are identified by the presence of an 'OB-fold' (Oligonucleotide/Oligosaccharide/Oligopeptide binding fold) (Murzin 1993)(*Fig.3.1.1*), the principal DNA-binding unit. A striking feature of the fold is the poor conservation of any extended sequence motifs excepting a few conserved amino acids. This may explain the non-sequence specific recognition of ssDNA by SSBs and the broad range of substrates utilised by the fold. Typically the OB-fold comprises five β-strands in a Greek key arrangement that coil to form a closed β-barrel. This common core is capped by an α-helix between strands β3 and β4 (*Fig 3.1.1*).

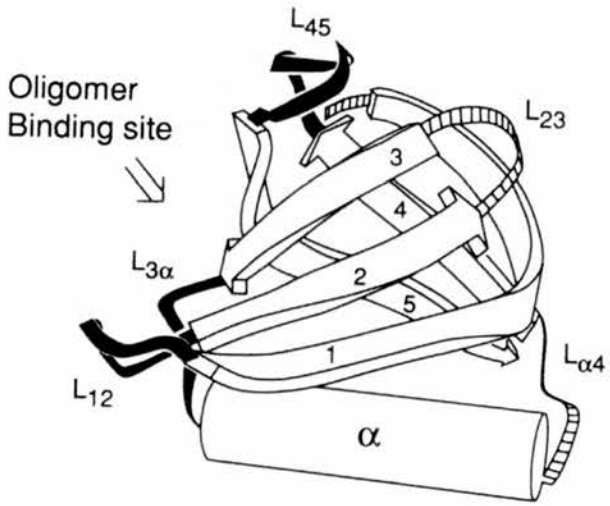

Fig.3.1.1 Illustration of the typical OB-fold topology as originally suggested by Murzin. Reproduced from (Murzin 1993).

SSBs from different domains of life use varying numbers of OB-folds. Bacterial SSBs possess a single OB-fold per polypeptide chain and form homotetramers. A BLAST search (Altschul, Madden et al. 1997) shows that mesophilic bacteria only encode one SSB. Some slight sequence variation is noted in *E. coli* SSB

from different strains. SSBs characterised from thermophilic bacteria possess two OB-folds in a single polypeptide (Dabrowski, Olszewski et al. 2002), and form homodimers. This may be a thermostable adaption, brought about by compaction of two ssDNA-binding units in the one monomer, in contrast to dimer assembly from two separate monomers in the mesophilic counterpart.

Eukaryotic SSBs differ widely in the number of OB folds they use. Replication protein A (RPA) is heterotrimeric in structure with each subunit encoded by a separate gene (Umbricht, Erdile et al. 1993); in total six OB folds have been identified. The Human mitochondrial SSB (HmtSSB) encodes a single OB fold per polypeptide similar in sequence to *E. coli* SSB (Yang, Curth et al. 1997). HsmtSSB monomers associate to form a homotetramer similar to the *E. coli* homologue (Yang, Curth et al. 1997). Recently, collaborators in the lab of Malcolm White at the University of St.Andrews have identified a human SSB distinct from RPA and the mitochondrial SSB. Although the open reading frame encodes only a single OB fold, gel filtration experiments suggest the formation of a dimer. The archaea are split into two major subdivisions - the *euryarchaea* and *crenarchaea*. The former have been found to express either monomeric (Kelly, Simancek et al. 1998) or heterotrimeric (resembling RPA) (Kelman, Pietrokovski et al. 1999) SSBs in solution.

3.2 *Escherichia coli* SSB

The crystal structures of *E. coli* SSB confirm the presence of four OB-folds that are brought together to form the functional homotetramer (Raghunathan, Ricard et al. 1997; Raghunathan, Kozlov et al. 2000). Each monomer is approximately 19kDa. ssDNA wraps around the tetramer in an elegant manner (*Fig.3.2.1*).

Dimers (as discussed here) are formed between chains A and B and chains C and D respectively (*Fig.3.2.1*). Evidence for a biological tetramer comes from the naturally occurring ssb-1 mutant (H55Y) (Glassberg, Meyer et al. 1979; Vales, Chase et al. 1980) This mutation causes a temperature sensitive defect in DNA replication that arises from SSB tetramer destabilisation (Williams, Murphy et al. 1984). This mutant is also defective in recombination (Glassberg, Meyer et al. 1979). The apo-SSB crystal structure (Raghunathan, Ricard et al. 1997) suggests that replacement of His55 with the bulkier tyrosine would cause either steric clashes in this region or a change in the contacts between monomers. The molecular weight of *E. coli* SSB from non-denaturing polyacrylamide is 76kDa. This native form is retained upon heating at 100°C for 2 minutes (Weiner, Bertsch et al. 1975) with only a 10% reduction in DNA-binding activity (Weiner, Bertsch et al. 1975).

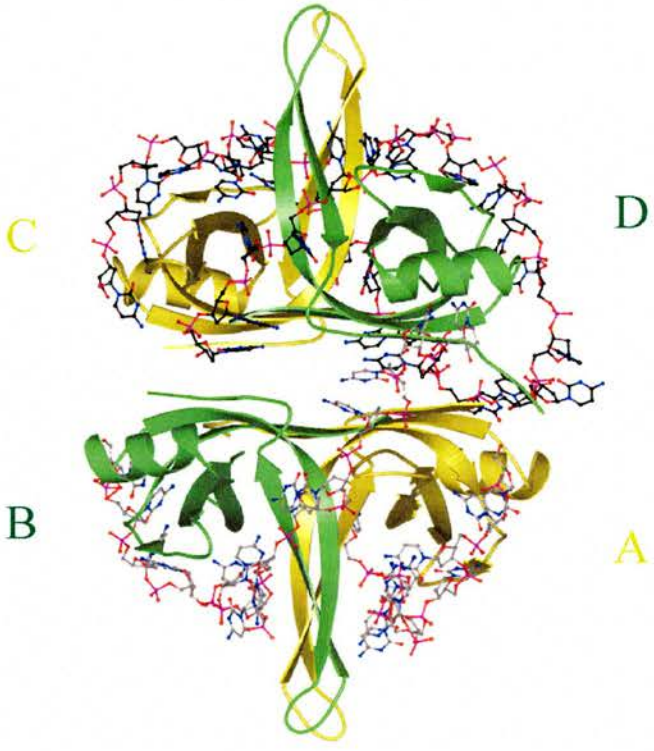

Fig.3.2.1 The E. coli SSB tetramer bound to two 35mer oligodeoxycytidylates. Monomers are labelled A-D

At low salt concentrations (<10mM NaCl), *E. coli* SSB saturates ssDNA with 'unlimited' cooperativity, forming long clusters along the strand (Griffith, Harris et al. 1984; Lohman and Overman 1985). Known as the $(SSB)_{35}$ binding mode, this state is consistent with the occlusion of 35nt per tetramer and ssDNA interaction with only two of the four subunits of the tetramer (Lohman and Overman 1985; Bujalowski and Lohman 1986). In the $(SSB)_{65}$ binding mode, at high salt concentrations (>0.2M NaCl), all four subunits interact with ssDNA (*Fig.3.2.1*) to occlude 65nt per tetramer and 'limited' inter-tetramer cooperativity is observed, leading to the formation of beaded structures along the single DNA strand (Chrysogelos and Griffith 1982). A mixture of these two species is observed in the range between these concentrations of NaCl (Lohman and Overman 1985). An intermediate mode has been identified with an apparent binding site size of 56 ± 3 nt per tetramer, $(SSB)_{56}$,(Bujalowski and Lohman 1986) although this is only apparent using $MgCl_2$ in the range 4-50mM. A fourth mode, binding 40nt, is also apparent although only at 37°C (Bujalowski and Lohman 1986). *Bujalowski et al* state that all three binding-mode transitions are within the range of salt concentrations found in *E. coli* and therefore may form in-vivo (Bujalowski and Lohman 1986).

The C-terminus of *E. coli* SSB forms a long tail (Sancar, Williams et al. 1981) consisting of a flexible Gly/Pro spacer and an acidic region, presumed to have helical secondary structure. This region has never been fully resolved in crystallographic studies. The tail becomes more sensitive to proteolysis upon ssDNA binding (Williams, Spicer et al. 1983) and deletion is correlated with severe repair deficient phenotypes, with no concomitant deleterious effect on DNA-binding affinity, suggesting a role in protein-protein interactions. The ssb-

113 mutant (P176S) located in this region has pleiotropic effects including loss of the ability to amplify RecA in response to DNA damage (Baluch, Chase et al. 1980). Additionally this mutant is unable to bind the repair protein exonuclease I, an interaction that has been mapped to the last ten residues of the C-terminal tail {Genschel, 2000 #185

3.3 Replication protein A (RPA)

Structural and biochemical analysis of human RPA has revealed a very different domain organisation and mechanism of ssDNA-binding in comparison with *E. coli* SSB. The heterotrimer has a total binding site size of 30nt {Wold, 1997 #4} and is composed of subunits of approximately 70, 32 and 14kDa (RPA70, RPA32 and RPA14 respectively) (Wold 1997) (*Fig.3.3.1*). Stable assembly of the apo-heterotrimer is thought to be achieved by formation of a helical bundle, each helix donated by a different domain (Bochkareva, Korolev et al. 2002). Three of these domains (DBD-C, DBD-D and RPA14) are implicated in forming a trimerisation core (*Fig.3.3.1*) during ssDNA binding (Bochkareva, Korolev et al. 2002). A helix from DBD-B and a transient helix from the N-terminal domain of RPA70 (RPA70N) may contribute to the helical bundle to stabilise the globular form of the apo-heterotrimer (Bochkareva, Korolev et al. 2002).

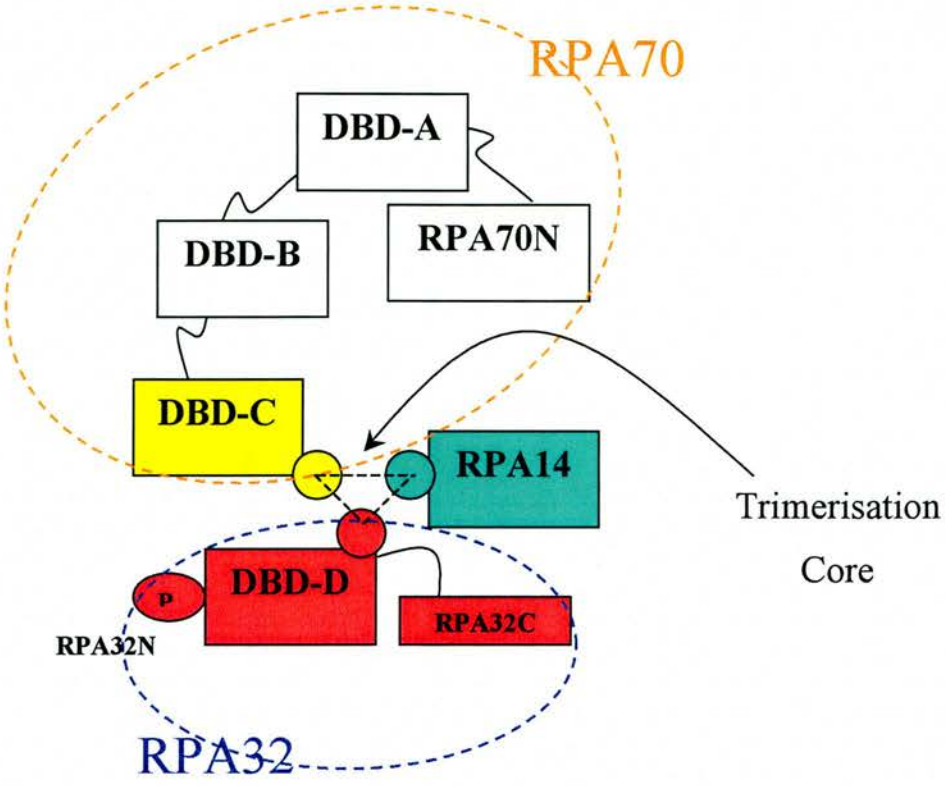

Fig.3.3.1 Domain organisation of the RPA heterotrimer modified from (Bochkareva, Korolev et al. 2002). Coloured circles represent helices contributing towards the trimerisation core.

Human RPA exhibits low cooperativity with respect to ssDNA binding (Kim and Wold 1995). Soluble expression of the individual subunits of the human protein in *E. coli* is not possible (Henricksen, Umbricht et al. 1994). Simultaneous expression of all three leads to a soluble complex with similar properties to native RPA. The N-terminus of RPA32 is phosphorylated in a cell-cycle dependent manner upon DNA-binding (Brush, Anderson et al. 1994; Brush, Morrow et al. 1996) (*Fig.3.3.1*), suggesting a mechanism for RPA regulation. Experiments by Treuner et al suggest that phosphorylation causes separation of RPA70 and RPA32 (Treuner, Findeisen et al. 1999). Domain deletion proves that the major ssDNA-binding activity is held in RPA70 (Gomes and Wold 1996).

RPA can localise in the nucleus and in replication foci (reviewed in (Wold 1997)). It has been shown to interact with viral replication initiator proteins including Epstein-Barr EBNA1 and the bovine papilloma virus E2 protein (reviewed in (Wold 1997)). A role in elongation during DNA replication is also implicated as antibodies raised against RPA decrease the elongation of replication forks (Walther, Bjerke et al. 1999). Like *E. coli* SSB, RPA plays a major role in DNA-repair, notably its interaction with proteins that function during nucleotide excision repair (NER). NER is the major pathway responsible for removal of bulky adducts from the DNA duplex that accumulate on exposure to ultraviolet radiation (reviewed in (Sancar 1996)). Defects in NER, are associated with Xeroderma Pigentosum (XP) (Cleaver 1968). XP patients display extreme hypersensitivity to ultraviolet radiation. Seven genes identified in XP patients encode proteins that function during NER (*XP* A-G) (reviewed in (Friedberg 2001)). XP is attributed to mutations in these genes as studied in-vivo using a mouse model (reviewed in (Friedberg 2001)). *Patrick et al* have shown that an RPA-XPA complex shows a higher affinity for damaged DNA than either RPA or XPA alone (Patrick and Turchi 2002). The RPA-XPA module is believed to be involved in damage recognition (He, Henricksen et al. 1995). Binding of the nucleases XPF and XPG by RPA (He, Henricksen et al. 1995; Matsunaga, Park et al. 1996) suggest that this module is required for assembly of an 'excision complex'. RPA also functions in gap-filling during NER, possibly binding to and stabilising the exposed ssDNA until the duplex is resolved (Guzder, Habraken et al. 1995; Mu, Park et al. 1995).

In total, RPA is thought to be composed of six OB folds. RPA70 has four: three ssDNA binding domains - DBD-A, DBD-B, DBD-C and RPA70N.

DBD-A and B are the structurally homologous principal DNA binding domains of RPA70 and form the core of this subunit (Gomes and Wold 1996; Philipova, Mullen et al. 1996) (*Fig 3.3.2*). The domains are joined by a flexible linker. DBD-B is attached to the third OB fold, DBD-C, by a similar linker which becomes resistant to proteolysis on ssDNA binding (Bochkareva, Korolev et al. 2002).

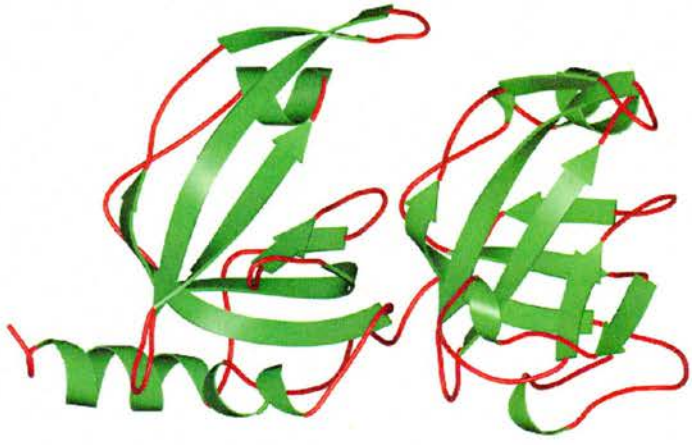

Fig.3.3.2 The crystal structures of DBD-B (left) and DBD-A (right).

DBD-C (*Fig.3.3.3*) is located in the C-terminal domain of RPA70 (RPA70-CTD) (Bochkareva, Korolev et al. 2002). This domain contains a conserved zinc-ribbon motif ($CX_4CX_{14}CX_2C$), outwith the core OB-fold, formed by three antiparallel β-strands. Zinc binding is proposed to stabilise the domain tertiary structure and modulate RPA70-CTD ssDNA binding (Bochkareva, Korolev et al. 2000; Bochkareva, Korolev et al. 2002).

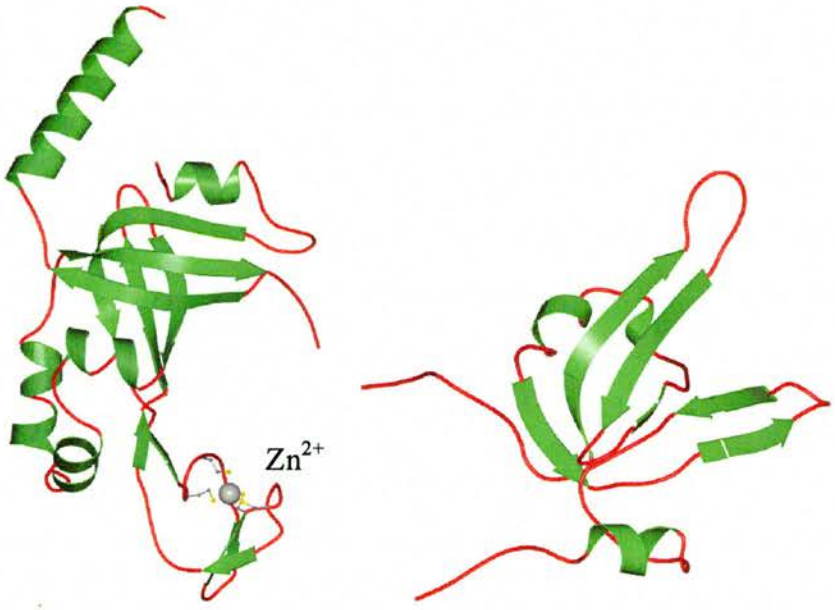

Fig.3.3.3 Crystal structures of DBD-C (left) and RPA70N (right)

The fourth OB fold is found in RPA70N. The topology of this domain in the average, minimized NMR structure (Jacobs, Lipton et al. 1999) (*Fig.3.3.3*) would appear to be slightly different from the other RPA subunits. However it is possible that this simply represents increased flexibility in solution. The crystal structure of this domain may show a more structured topology stabilised by crystal contacts. RPA70N is implicated as the protein-protein interaction domain of RPA70. This subunit has been shown to interact with the SV40 T antigen (Braun K.A., He Z., Ingles C.J., Wold M.S. unpublished observations). Additionally, a monoclonal antibody that binds the N-terminal domain has been shown to inhibit RPA stimulation of DNA polymerase α (Kenny, Schlegel et al. 1990).

DBD-D (*Fig.3.3.4*) is the central OB fold of RPA32, a three-domain subunit. A complex of DBD-D and RPA14 show weak ssDNA binding activity (Bochkareva, Frappier et al. 1998). This activity is masked (Bochkareva, Frappier et al. 1998) by the phosphorylated N-terminal domain (RPA32N)

(Brush, Anderson et al. 1994; Brush, Morrow et al. 1996) and the C-terminal protein-protein interaction domain (RPA32C) of the intact RPA32 (Li, Lu et al. 1995; Nagelhus, Haug et al. 1997).

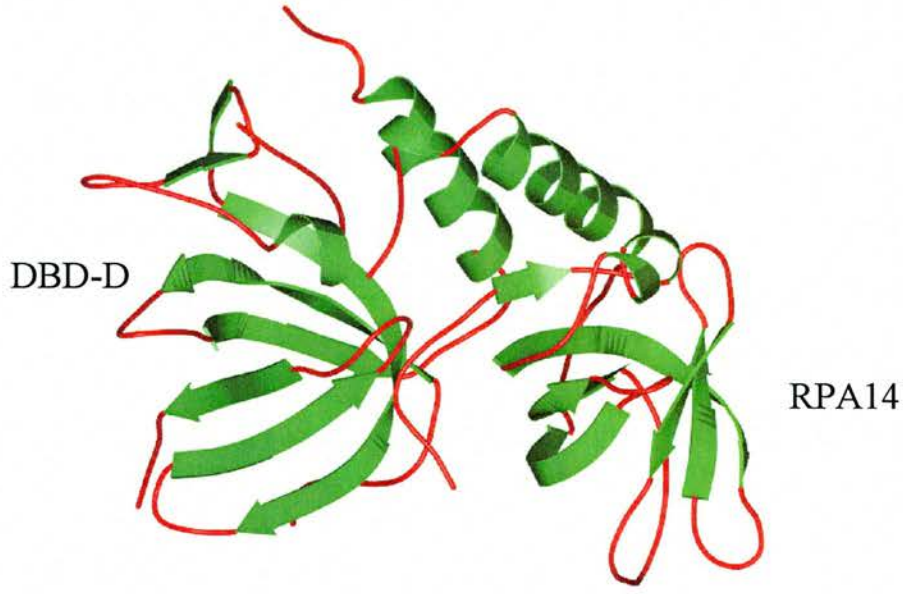

Fig.3.3.4 Crystal structure of a DBD-D -RPA14 complex

The small subunit, RPA14, contains the last OB fold (*Fig.3.3.4*) (Bochkarev, Bochkareva et al. 1999). Limited proteolysis suggests RPA14 is folded into a single structural domain (Bochkareva, Frappier et al. 1998), however only the C-terminal half is required for viability in yeast (Philipova, Mullen et al. 1996). The role of RPA14 is still unclear, although it is able to partially solubilise RPA70 in solution (Henricksen, Umbricht et al. 1994). Therefore the role of RPA14 may be purely structural.

Based on detailed structural and biochemical studies, a model for domain interaction and ssDNA binding in RPA is emerging (*Fig.3.3.1*). Electron microscopy techniques have been used to visualise an 8-10nt complex (the 8nt-binding mode) and a 30nt complex with ssDNA (Blackwell, Borowiec et al. 1996). The 8nt-binding mode is most likely a precursor to the 30nt complex and

is probably mediated by RPA70 as first indicated by glutaraldehyde cross-linking experiments (Blackwell and Borowiec 1994). In agreement with this, structural analysis confirms that DBD-A and B assemble in tandem to occlude 3nt each with a further 2nt bridging the space between the two OB-folds (Bochkarev, Pfuetzner et al. 1997) (*Fig.3.3.5*). DBD-A contacts the 5' end of DNA, while DBD-B assembles immediately downstream (Bochkarev, Pfuetzner et al. 1997).

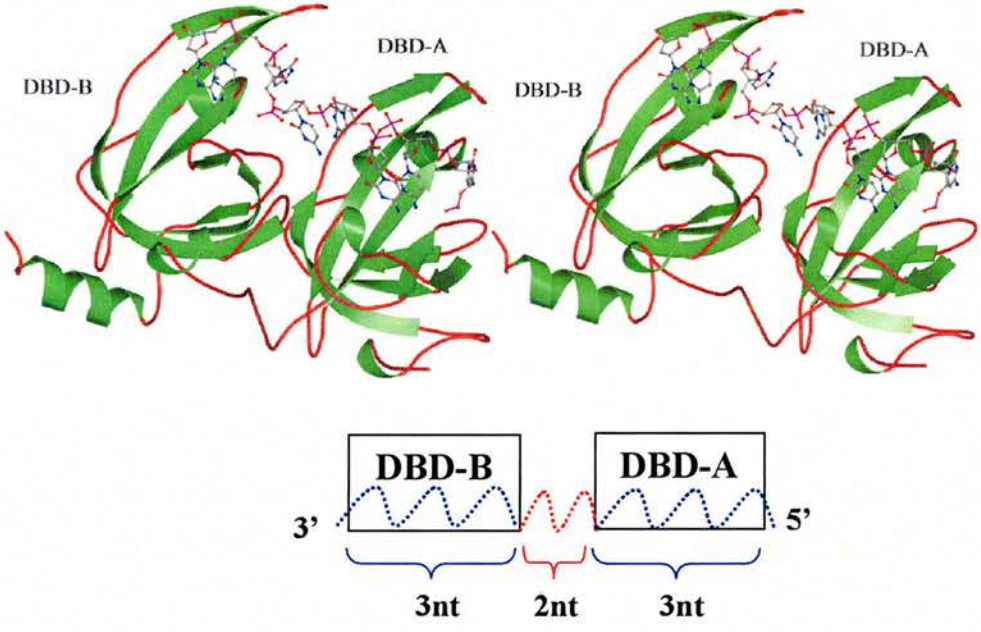

Fig.3.3.5 The DBD-A - DBD-B complex with ssDNA (top) and a schematic of the 8nt binding mode (bottom). In the schematic, DNA is depicted by a dashed line; nucleotides bound by the OB folds are coloured blue, while ssDNA between domains is coloured red.

DBD-C is next assembled on ssDNA in tandem with the first two OB folds, allowing RPA70 to contact between 13-14 and 22nt in total (Bastin-Shanower and Brill 2001).

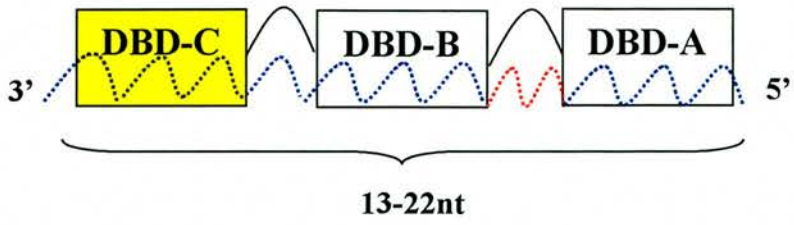

Fig.3.3.6 Model representing the assembly of DBD-C on ssDNA

As implicated by the crystal structure of the three-domain complex, DBD-C, DBD-D of RPA32 and RPA14 form the trimerisation core of RPA (*Fig.3.3.1*) (*Fig.3.3.7*).

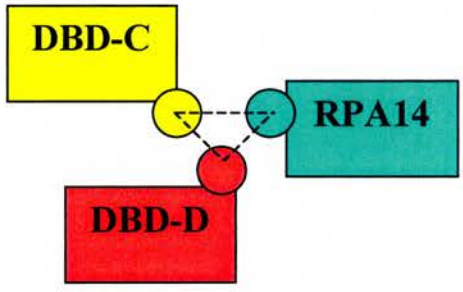

Fig.3.3.7 Schematic of the RPA trimerisation core. Coloured as fig. 3.3.1

The fact that RPA70 does not express solubly on its own and that RPA32 and RPA14 form a soluble complex (Henricksen, Umbricht et al. 1994) suggest that formation of the trimerisation core is essential for correct RPA70 folding and formation of a functional RPA heterotrimer. Additionally, an DBD-C-RPA32-RPA14 complex leads to a 3-5 fold increase in ssDNA binding affinity in comparison with an RPA32-RPA14 complex (Bochkareva, Frappier et al. 1998). Each domain in the trimerisation core is proposed to donate one helix (Bochkareva, Korolev et al. 2002) (*Fig.3.3.7*, *Fig.3.3.8*).

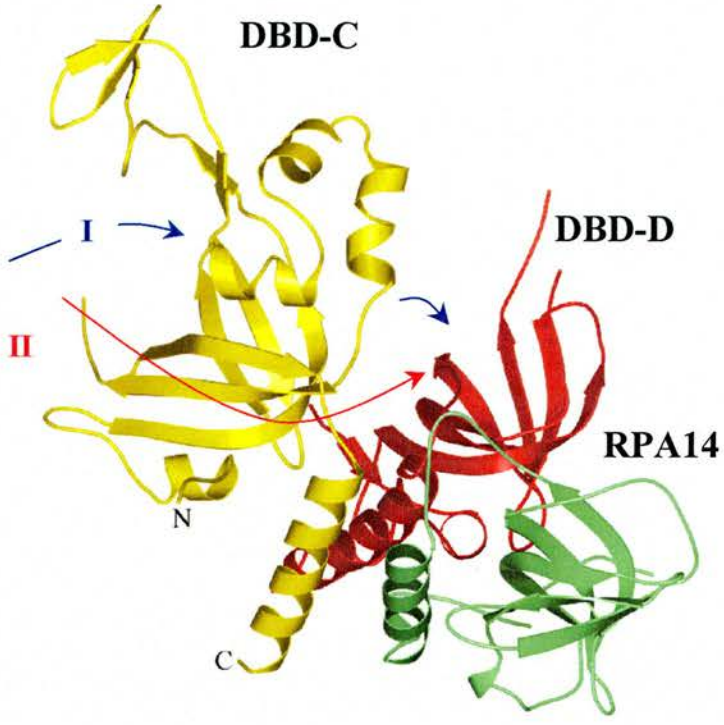

Fig.3.3.8 Crystal structure of the RPA trimerisation core. Two possible paths for DNA to contact DBD-D by wrapping round the core are shown as arrows.

Single stranded DNA is suggested to wrap around the trimerisation core to contact the last DNA-binding OB-fold, DBD-D (coloured red above), allowing RPA to bind 30nt in total (Bochkareva, Korolev et al. 2002). However the authors do not state how the DNA would wrap around the core (Bochkareva, Korolev et al. 2002). From the crystal structure of the trimerisation core, it would appear that two paths are possible (*Fig.3.3.8*). Path I allows the ssDNA to tunnel through the basic DNA binding cleft of DBD-C as one would expect. Only slight kinking is required to reach the analogous cleft in DBD-D (*Fig.3.3.9*). The basic cleft is bypassed in Path II and instead ssDNA wraps round the opposite face which is highly negatively charged. It seems unlikely that ssDNA would follow Path II, due to repulsion between the negatively charged phosphates of ssDNA and negatively charged amino acids of the protein.

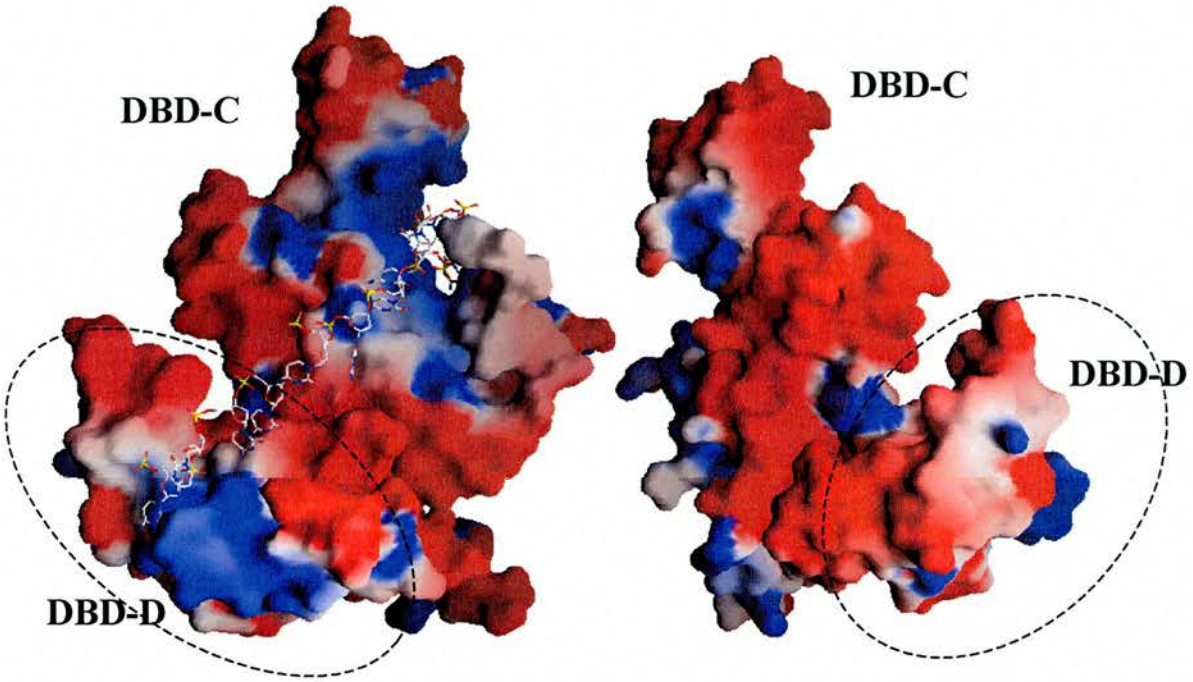

Fig.3.3.9 Electrostatic surfaces of DBD-C - DBD-D complex in the trimerisation core. DBD-D is highlighted by a dashed line in each figure Left - modelled ssDNA following a path through the basic DNA-binding clefts with slight kinking to fold round to DBD-D. Right - the opposite face of DBD-C is highly electronegative. The electrostatic surface is contoured from -3kT/e to +3kT/e.

This last step involving wrapping of ssDNA around RPA may explain the conformational change observed upon shifting from the 8 to 30nt binding mode (Blackwell, Borowiec et al. 1996) and the large number of nucleotides required to contact all four DBDs when in a sequential model of ssDNA binding only 18-20nt would be required.

3.4 *Sulfolobus solfataricus* SSB

The focus of this study is the single stranded DNA-binding protein from the crenarcheote, *Sulfolobus solfataricus* (*Sso*SSB). Recently, this protein has been

the subject of biochemical characterisation (Wadsworth and White 2001; Haseltine and Kowalczykoski 2002). The protein is encoded by a single open reading frame yielding a single polypeptide of 16kDa (Wadsworth and White 2001). The abundance of this protein in Sulfolobus cells (approximately 2-5% of the total protein) (Wadsworth and White 2001) suggests an important role in DNA metabolism. Retardation of ssDNA electrophoretic mobility by *Sso*SSB confirmed DNA-binding activity and this data in conjunction with tryptophan fluorescence experiments showed a binding site size of approximately 4-6 nucleotides of ssDNA (Wadsworth and White 2001). *Sso*SSB possesses a C-terminal tail similar to the mesophilic bacterial SSBs. This tail is rich in glycine and proline residues, conferring a high degree of flexibility. The tail is sensitive to proteolytic digestion, cleaving the last 29 amino acids from the C-terminus. The truncated, recombinant protein retains the ability to bind ssDNA (Wadsworth and White 2001).

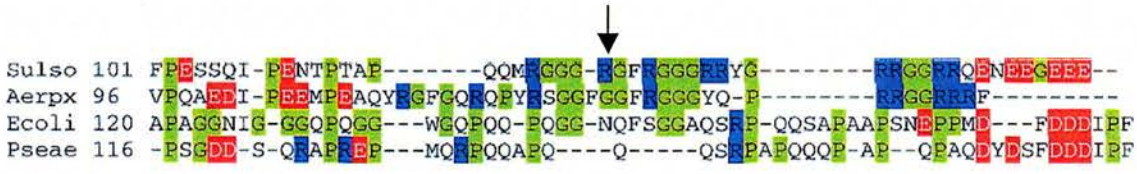

Fig.3.4.1 Sequence alignment showing the C-terminal diversity present in crenarchaeal and bacterial SSBs. Sulso - *Sulfolobus solfactaricus*, Aerpx - *Aeropyrum pernix*, Ecoli - *Escherichia coli*, Pseae - *Pseudomonas aeruginosa*. The cleavage site of *Sso*SSB is shown with an arrow.

The biologically significant oligomeric state of *Sso*SSB has been up till now undetermined. Glutaraldehyde crosslinking experiments suggest the dominant species to be monomeric, both in the presence and absence of DNA (Wadsworth

and White, unpublished observations). This would agree with gel filtration data, showing a single peak corresponding to the monomer. *Haseltine et al* have shown dimer and tetramer formation by analytical gel filtration (Haseltine and Kowalczykoski 2002). Higher oligomeric states were only seen upon heating the protein. This may be consistent with the high temperatures at which this thermophile exists.

By determining the crystal structure of *Sso*SSB, one of the main aims of this study was to determine the core fold and minimal biological unit of a crenarchaeal SSB relative to the homotetrameric bacterial SSBs and eukaryotic heterotrimers. It has already been suggested that crenarchaeal SSBs would resemble the bacterial homologues in overall folding and subunit composition, due to the presence of a common C-terminal tail in their sequences (Haseltine and Kowalczykoski 2002). However in contrast to *E. coli SSB*, the biochemical data do not overwhelmingly support the formation of a higher oligomeric state by *Sso*SSB. It was therefore hypothesised that *Sso*SSB would be more similar in structure to one of the RPA OB folds. As discussed in chapter one, a stronger similarity between crenarchaeal and eukaryotic SSBs would also be in agreement with the closer phylogenetic relationship between the archaea and eukaryotes already discussed.

Materials and Methods/Results

3.5 Crystallisation of native and selenomethionyl *Sso*SSB

3.5.1 Purification and crystallisation of native protein

Purified, recombinant *Sulfolobus solfataricus* SSB was provided by Ross Wadsworth from the lab of Malcolm White (all protein purification, modification, and assay prior to crystallisation was carried out by Ross Wadsworth). The recombinant protein is identical in size and sequence to that obtained from the native organism and retains the ability to bind ssDNA (Wadsworth and White 2001).

Preliminary crystallisation trials were performed using the full length (148 aa), 16kDa recombinant protein. Trials were undertaken using 1-60 mg/ml of protein and a drop size of 4µl under the sitting-drop vapour diffusion method (McPherson 1982) under Hampton crystal screens one and two (Jancarik and Kim 1991; Cudney, Patel et al. 1994). Crystals were not obtained under these conditions. Even at low protein concentrations, the full-length protein readily precipitated heavily in the majority of the two screens. Various Hampton grid screens were pursued (including the Ammonium sulphate screen), varying the protein concentration as above and using the hanging drop vapour diffusion method (McPherson 1982) with drop sizes of 4-6 µl. These attempts were also unsuccessful.

Analysis of the amino acid sequence had already shown the protein to possess a flexible, C-terminal extension susceptible to trypsin digestion (Wadsworth and White 2001). The recombinant protein was incubated with

trypsin, cleaving the last 29 residues. In all cases the truncated protein consists of 119 amino acids, as determined by mass spectrometry analysis carried out by Wadsworth, and retains the ability to bind ssDNA.

Crystals of the truncated protein were obtained in the Hampton Ammonium Sulphate grid screen after 3 days incubation at 20°C. Hexagonal bypyramids were obtained in 100mM Citric Acid pH 5.0 with 2.4M ammonium sulphate as the precipitant. Hexagonal rods were obtained in 100mM Hepes pH 7.0 and 2.4M ammonium sulphate. Crystals were more robust than expected probably due to the high salt from which they were grown and the relatively low solvent content, (calculated after collection of the initial diffraction data). Izit crystal dye (Hampton research) was added to each crystal drop and after a period of 30mins to 1 hour the crystals had been stained blue (*Fig.3.5.1.1*).

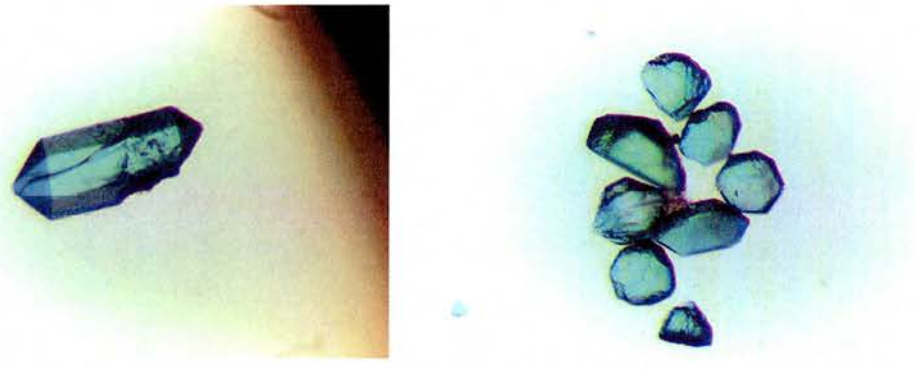

Fig.3.5.1.1 SSB crystals stained with Izit dye. Right - a fragment of a hexagonal rod, Left - some degraded hexagonal bipyramids

The dye enters the solvent channels in protein crystals and binds to the macromolecule. The absence of such channels in salt crystals means that they are unable to take up the dye. An optimised grid was devised around a Linbro 24-well tissue culture plate using the hanging drop vapour diffusion method.

Precipitant concentration ranged from 1.4-2.4M ammonium sulphate horizontally along the plate, with pH ranging from 3.5 to 5.0 vertically down the plate, using 100mM citric acid as the buffer. Bipyramidal and rod-shaped crystals (*Fig.3.5.1.2*) were obtained in a variety of conditions, with growth beginning after only a few days. Morphologically the "best crystals", with the sharpest edges, were obtained using a protein concentration of 7mg/ml. Under these conditions the largest rods had dimensions of 0.3mmx0.3mmx1mm. Crystals were judged to be of a suitable size for in-house data collection after 5-7 days.

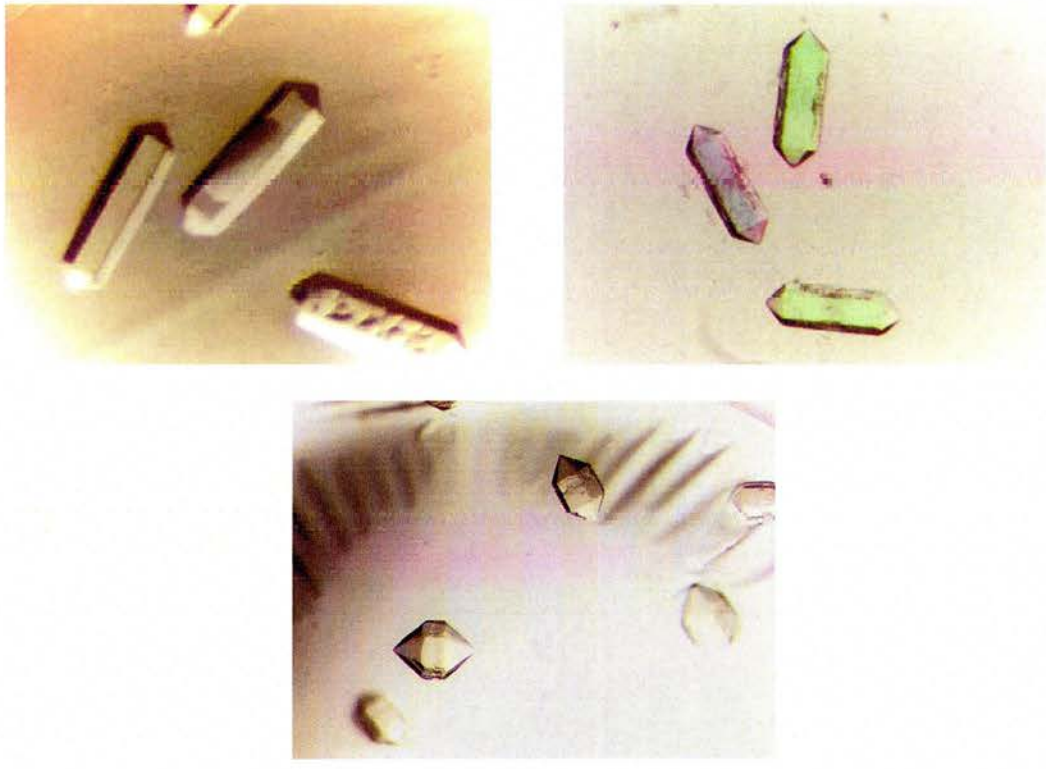

Fig.3.5.1.2 Hexagonal rods and hexagonal bipyramids obtained in the optimised ammonium sulphate-citric acid screen. Top right - crystals display optical birefringence under polarised light. Bottom - bipyramidal crystal grown in the optimised screen.

3.5.2 Purification and crystallisation Selenomethionyl protein

Selenomethionyl protein was prepared by Wadsworth, expressing the recombinant protein from *E. coli* Rosetta cells (Novagen) and following the protocol described by Van Duyne (Van Duyne, Standaert et al. 1993). The heavy atom derivative was purified as described for native, recombinant SSB except that all purification buffers contained 4mM DTT to prevent selenomethionine oxidation. Protein containing selenomethionine crystallised identically to the native (*Fig.3.5.2.1*), although the best crystals were obtained using a protein concentration of 10mg/ml.

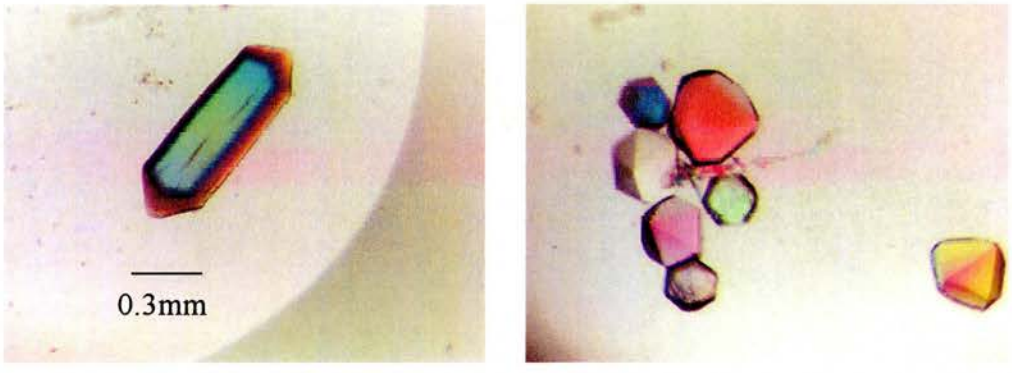

Fig.3.5.2.1 Selenomethionyl crystals grown in Left - 1.6M ammonium sulphate and 100mM Citric acid pH 3.5 and Right - 2.4M ammonium sulphate and 100mM Citric acid pH 5.0.

3.6 Data collection

3.6.1 Data collection from native crystals

Data were collected to 1.98Å from a single, frozen hexagonal rod (Table 3.6.1.1, Native-1)(Kerr, Wadsworth et al. 2001). Due to the high salt in the mother liquor, SSB crystals were difficult to cryoprotect. Initial attempts were made to

freeze crystals in mother liquor and glycerol (5-20%). Unfortunately, glycerol was not sufficient to protect the drop from the air and the ammonium sulphate rapidly salted out. Cryoprotection was achieved by transferring the crystal quickly to a fresh drop of mother liquor (typically 5-10μl) and adding 2-3μl of paraffin oil. "Oil-rings" in the diffraction pattern were usually eliminated by pulling the crystal through a fresh drop of mother liquor to dispose of excess oil on the looped crystal. Data were collected as 10min, 1° oscillations using an in-house source (operating at 50kV, 90mA) with a Nonius DIP2000 image plate/rotating anode and Osmic mirrors. The crystal to detector distance was 94mm. The data were integrated in DENZO and merged using SCALEPACK (Otwinowski and Minor 1996).

This crystal form had a primitive hexagonal reciprocal lattice, with unit cell parameters a=b=75.3Å, c=69.0Å $\alpha=\beta=90°$ $\gamma=120°$ and the systematic absences identified the crystallographic space group to be either $P6_1$ or $P6_5$. Assuming the asymmetric unit contained two SSB monomers, the Matthews coefficient (V_M) calculated from the cell dimensions was 2.18 $Å^3$ Da^{-1} (Matthews 1968), giving a solvent content of approximately 43%. Assuming a larger or smaller number of SSB monomers in the asymmetric unit gave values far outside those first suggested by Matthews.

	Native-1 (In-house)	Native-2 (Daresbury PX9.6)
Wavelength (Å)	1.54	0.87
Resolution (Å)	20-1.98 (1.98-1.95)	48-1.2 (1.23-1.20)
Total observations	307696	479333
Unique observations	16159	70162
Space group	$P6_{1/5}$	$P6_{1/5}$
Temperature (K)	130	130
Detector	DIP2000	ADSC - CCD
Cell constants	a=b=75.40Å, c=69.06Å $\alpha=\beta=90°$ $\gamma=120°$	a=b= 75.80Å c= 70.12Å $\alpha=\beta=90°$ $\gamma=120°$
V_m† (Å^3 Da^{-1})	2.18	2.25
I/σ	32 (6.6)	4.4 (1.3)
Average multiplicity	10.7 (5.9)	7.0 (3.5)
Data completeness (%)	96.3 (99.9)	99.9 (99.9)
R_{merge}‡(%)	7.1 (26)	8.0 (57)

‡$R_{merge} = \sum \sum I$ <h>i – (I <h>) / $\sum \sum I$ <h>i, where I<h>i is the measured diffraction intensity and the summation includes all observations. † assuming two molecules in the asymmetric unit.

Table 3.6.1.1 Data collected from native SSB crystals.

A 1.2Å dataset was collected from a single frozen bipyramidal crystal at Daresbury station PX 9.6,using a CCD detector (Table 3.6.1.1, Native-2)

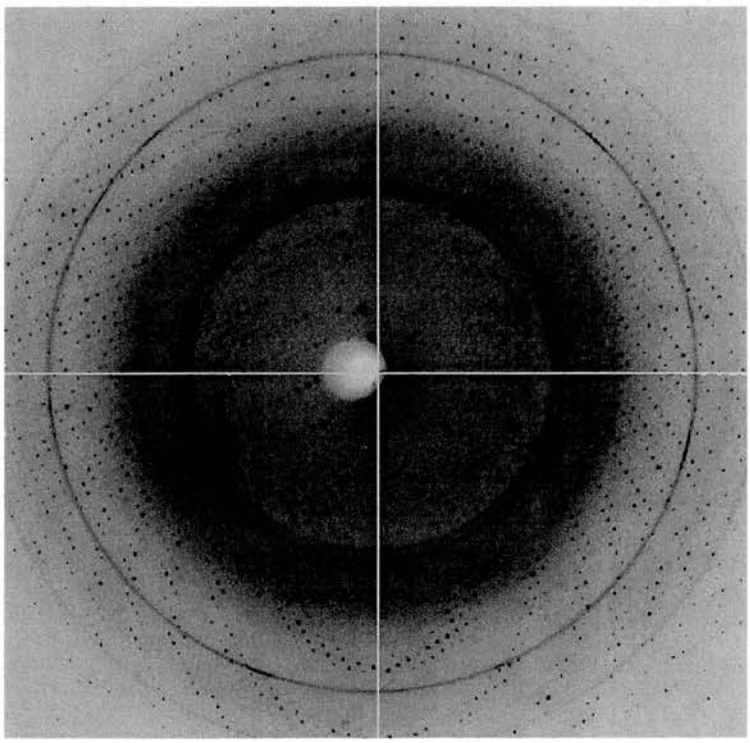

Fig.3.6.1.1 Diffraction of SSB crystals to 1.2Å

Data were recorded over four passes (47.9-1.9Å, 2 x 26-1.7Å, and 2.5-1.2Å) as 1° oscillations per frame to account for the presence of 'overloaded' reflections and 'oil rings' in the lower resolution shells and ensure high redundancy and completeness for all the data (Table 3.6.1.2).

The data from each pass were integrated separately in MOSFLM (Leslie 1992) and scaled together and merged in SCALA (Evans 1997). This crystal was determined to have roughly the same unit cell parameters as the rod form. The same crystallographic symmetry was identified from the systematic absences of the merged data showing the space group to be either $P6_1$ or $P6_5$.

	Low	Med 1	Med 2	High
Wavelength (Å)	0.87	0.87	0.87	0.87
Resolution (Å)	47.7-1.9	25.7-1.7	25.7-1.7	2.5-1.2
High Shell (Å)	1.95-1.90	1.74-1.70	1.74-1.70	1.23-1.20
Total observations	71855	99917	98447	226515
Unique observations	17835	24882	24697	61343
Space group	$P6_{1/5}$	$P6_{1/5}$	$P6_{1/5}$	$P6_{1/5}$
Temperature (K)	130	130	130	130
Cell constants (Å)	a=75.80 b=75.80 c= 70.12 α=β=90° γ=120°			
V_m† (Å^3 Da^{-1})	2.25			
I/σ	5.7 (4.4)	5.3 (2.5)	7.6 (5.1)	5.1 (1.3)
Average multiplicity	4.0(4.0)	4.0 (4.0)	3.9 (4.0)	3.6 (3.5)
Data completeness (%)	99.5 (99.5)	99.7 (99.7)	99.4 (99.4)	88.6 (88.6)
R_{merge}‡(%)	7.4 (15.7)	7.9 (29.3)	5.7 (14.4)	8.5 (57.2)

‡$R_{merge} = \sum \sum I <h>i - (I <h>) / \sum \sum I<h>i$, where $I<h>i$ is the measured diffraction intensity and the summation includes all observations. † assuming two molecules in the asymmetric unit.

Table 3.6.1.2 The four separate passes of diffraction data collected from crystal Native-2

3.6.2 Merohedral twinning in *Sso*SSB crystals

In assigning the correct space group to SSB crystals it was observed that the 1.2Å native diffraction data behaved anomalously. When the data were reduced to the asymmetric unit in Laue group 6/m the overall merging R-factor was 8.0%

and the systematic absences showed the true space group to be either $P6_1$ or $P6_5$. In the higher symmetry Laue group, 6/mmm, the data merged in P6222 and $P6_122$ with merging R-factors of 20.6%, considerably lower than would be expected when the wrong space group is assigned. This phenomenon is indicative of merohedral twinning where the rotational symmetry of the crystal lattice (the holohedry) is greater than the underlying Laue symmetry of the crystal (here, 6/m) (Yeates 1997). Twinning by merohedry is common in crystals that have a primitive hexagonal lattice, due to the equivalence of the a and b axes. There is no correlation between the extent/presence of twinning in SSB crystals and morphology or crystal growth conditions. At least one crystal from each condition in the optimised ammonium sulphate/citric acid screen has been tested for twinning. A typical strategy was to collect 30° of diffraction data (half a dataset) and use the Yeates crystal twinning server to determine the extent of twinning (see below) in the crystal (Yeates 1997). Morphologically, merohedral twinning is impossible to detect as the two twin domains align in three dimensions (Yeates and Fam 1999). As a result the reciprocal lattices are superimposable and twinning is therefore not apparent in the resulting diffraction pattern (*Fig.3.6.2.1*).

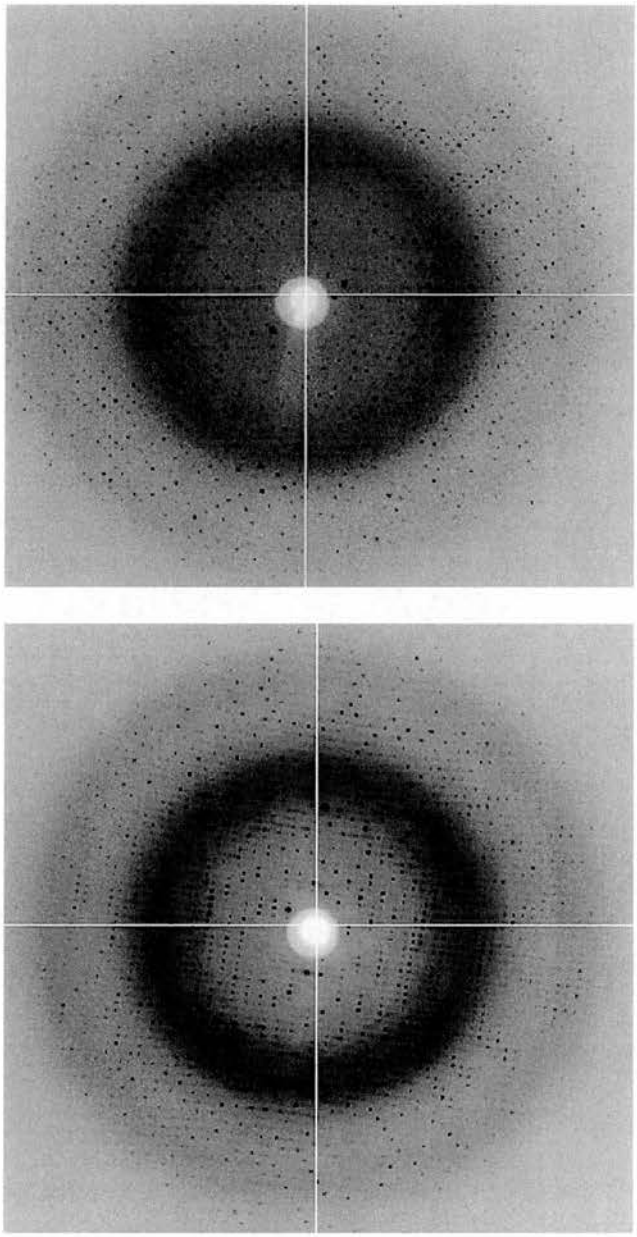

Fig.3.6.2.1 Diffraction pattern from *Sso*SSB crystals, d_{min}=1.7Å. Top - the reciprocal lattice is devoid of twinning. Bottom - the reciprocal lattice is merohedrally twinned although the character of the spots is indistinguishable in comparison with the top image.

The intensity of each measured reflection is the result of two non-equivalent overlapping reflections (the scattered X-rays that give rise to twin related reflections from the twin domains do not interfere). The degree of twinning is defined by the twin fraction, α, which describes the volume of the twin taken up

by the smaller of the two contributing crystals (Rees 1982). A perfect twin has a twin fraction of, or approaching, 0.5. The perfect superimposition of twin-related reflections allows the data to obey and be processed to erroneously high symmetry (Yeates and Fam 1999). Partial twins have a twin fraction not equal to 0.5 and may be successfully detwinned with knowledge of α and the twinning operator (the twin law) that aligns the two reciprocal lattices. The twin operation can be any symmetry operation in the holohedry but not in the true space group of the crystal. Detwinning the data inflates the errors in the measured intensities by a factor of $1/(1-2\alpha)$ (Yeates and Fam 1999); the errors become larger as α approaches 0.5 (Fisher and Sweet 1980). The 1.2Å diffraction data were submitted to the Yeates crystal twinning server and a test for partial merohedral twinning gave a twin fraction of approximately 0.30 (Yeates 1997). Presumably the inference of higher symmetry in *Sso*SSB crystals is a consequence of this high twin fraction. Data from a perfect twin would presumably fit higher crystal symmetry with a merging R-factor of much less than 20.6%.

A second more quantitative indicator of merohedral twinning is the extraction of unusual intensity distributions from the diffraction data (Rees 1982; Chandra, Acharya et al. 1999; Yang, Dauter et al. 2000). Wilson showed that diffraction data from a single crystal generally follow a certain intensity distribution for centric and acentric reflections (Wilson 1949). A plot of $N(z,\alpha)$ vs z where z represents the intensity relative to the mean intensity and $N(z,\alpha)$ is the fraction of reflections having intensity less than z demonstrates this (Rees 1980). From the plot, generated by the program TRUNCATE (CCP4), it is apparent that there are a smaller number of weaker reflections measured from twinned crystals (because the intensities of the overlapping reflections are

cumulative). Non-twinned data show that the theoretical $N(z,\alpha)$ curve is far closer to the experimental distribution, especially for acentric reflections (*Fig.3.6.2.3*).

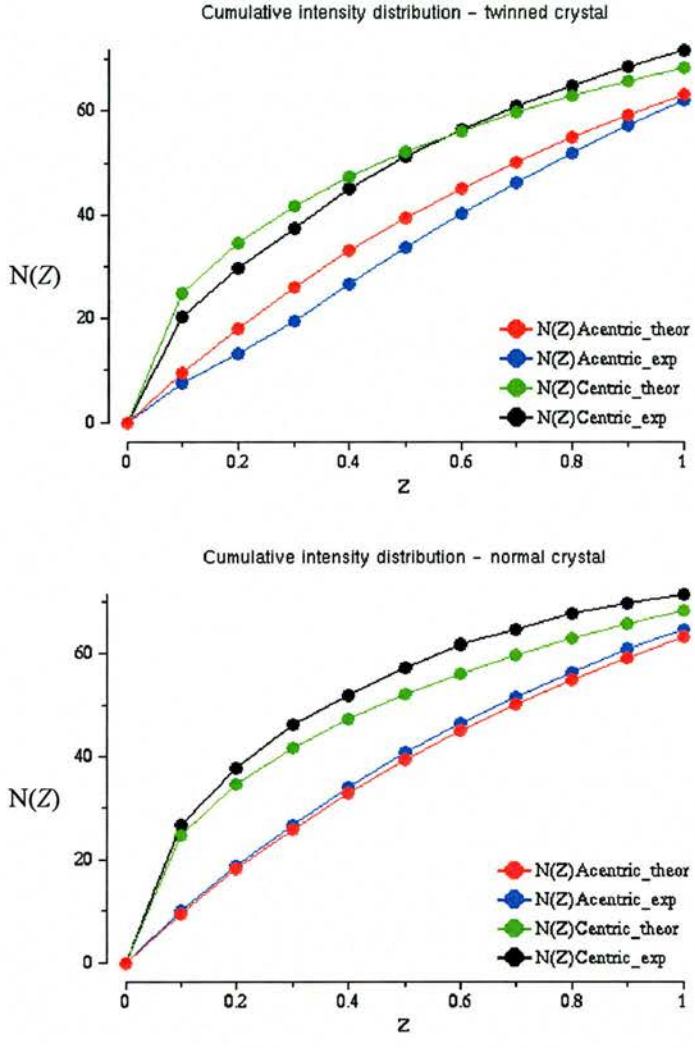

Fig.3.6.2.3 Plot of the Cumulative intensity distributions from top - a twinned SeMet SSB crystal and Bottom the SeMet SSB crystal devoid of twinning

The diffraction data collected from crystal Native-2 (1.2Å) were subjected to the CCP4 program DETWIN in an attempt to deconvolute the two overlapping lattices (Taylor and Leslie 1998). Analysis of the *S(H)* vs *H* plot generated by DETWIN (*Fig.3.6.2.4*)(Rees 1982; Yeates and Fam 1999; Yang, Dauter et al.

2000) suggested that the data were between 20% and 30% twinned based on the expected *S(H)* curves for various tested twinning fractions.

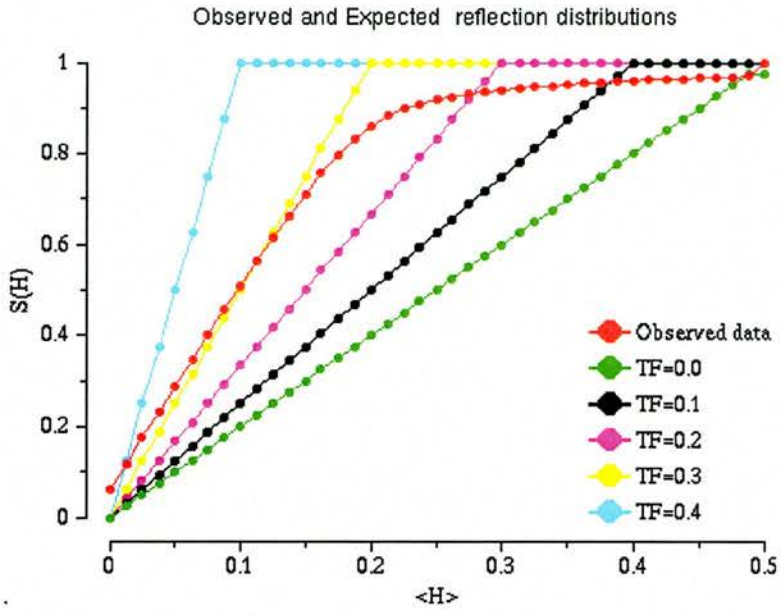

Fig.3.6.2.4 Plot of S(H) vs H showing the twin test performed by DETWIN. TF is the twin fraction. H is the fractional difference between a pair of twin-related intensities, S(H) is the cumulative distribution of H. The red curve represents the merged data collected from crystal Native-2

The data were successfully detwinned, assuming a twin fraction of 0.30 during the deconvolution process, as judged by the Yeates crystal twinning server and by analysis of the intensity distribution in TRUNCATE.

3.6.3 Data collection from selenomethionyl crystals

Crystals were screened at Daresbury station PX 14.2 for the presence of twinning. After 30° of data collection, using an oscillation angle of 1° and exposure time of 5 seconds/frame, the data were indexed in MOSFLM (Leslie

1992) and merged in SCALA (Evans 1997). The data were submitted to the Yeates twinning server for analysis and twinning by merohedry was confirmed in conjunction with analysis of the intensity statistics from TRUNCATE. All hexagonal rods screened were between 8 and 30% twinned. Only two full, three-wavelength MAD datasets were collected as detailed below.

Data were first collected from a single frozen hexagonal rod (Table 3.6.3.1, Se-twin). Using the Yeates crystal twinning server, this crystal was determined to be approximately 30% twinned. Oil rings were eliminated from the diffraction pattern by using 50% PEG 600 as the cryoprotectant instead of paraffin oil. The crystal was looped from the mother liquor and pulled through the PEG solution prior to freezing. A fluorescence scan of the crystal was obtained to determine the optimal wavelengths for MAD data collection. Three wavelengths were used; one at the peak of the scan where the anomalous difference (f '') is greatest, one at the inflection point where f ' (the real component of anomalous scattering) is lowest and one at a remote, high energy wavelength to optimise the dispersive difference between inflection and remote (Table 3.6.3.1). To ensure accurate measurement of the relatively small differences arising from anomalous scattering, 270 images were collected at each wavelength as 5 second, 1° non-overlapping oscillations (257 images at the peak wavelength). The crystal to detector distance was 140mm.. The data from each wavelength were integrated separately in MOSFLM (Leslie 1992) and merged in SCALA (Evans 1997).

Se-twin	λ1	λ2	λ3
Wavelength (Å)	0.9780	0.9785	0.9600
Resolution (Å)	37-1.69 (1.73-1.69)	37-1.69 (1.73-1.69)	37-1.66 (1.70-1.66)
Total observations	261665	412335	434900
Unique observations	24637	24479	25887
Space group	$P6_{1/5}$	$P6_{1/5}$	$P6_{1/5}$
Temperature (K)	130	130	130
Detector	ADSC - CCD	ADSC - CCD	ADSC - CCD
Cell constants (Å)	a=74.92 b=74.92 c= 69.29α=β=90° γ=120°		
V_m† ($Å^3$ Da^{-1})	2.15		
I/σ	4.0 (2.5)	6.8 (2.1)	5.8 (1.9)
Average multiplicity	10.6 (10.5)	16.8 (16.7)	16.8 (16.6)
Data completeness (%)	99.3 (99.3)	98.9 (98.9)	98.6 (97.9)
R_{merge}‡(%)	9.0 (29.4)	6.0 (35.6)	6.9 (38.6)
Anom. completeness*	98.8 (97.8)	98.6 (97.7)	98.6 (97.9)
Δano/σΔano	0.94	1.10	0.95

$R_{merge} = \sum\sum I <h>i - (I <h>) / \sum\sum I <h>i$, where $I<h>i$ is the measured diffraction intensity and the summation includes all observations. † assuming two molecules in the asymmetric unit. * Anomalous completeness corresponds to the fraction of possible acentric reflections for which an anomalous difference has been measured.

Table 3.6.3.1. Three wavelength MAD data collected from a single twinned selenomethionyl crystal

From the peak wavelength data, 68 batches with either $I/\sigma < 1$ or R_{merge} >13% were excluded. This led to improvements in R_{merge} (from 12.2% to 9.0%) and I/σ (3.2 to 4.0).

Data were also collected from a single, frozen bipyramidal crystal (Table 3.6.3.2, Se-norm). Initial analysis of the data proved that the crystal was devoid of twinning. An identical data collection strategy to that employed for the twinned crystal was used. 270 images were collected at each wavelength and the data from each wavelength were integrated separately in MOSFLM (Leslie 1992) and merged in SCALA (Evans 1997)(Table 3.6.3.2).

Se-norm	λ1	λ2	λ3
Wavelength (Å)	0.9780	0.9786	0.9600
Resolution (Å)	37-1.69 (1.73-1.69)	37-1.69 (1.73-1.69)	37-1.66 (1.70-1.66)
Total observations	408408	407283	432531
Unique observations	24609	24521	25989
Space group	$P6_{1/5}$	$P6_{1/5}$	$P6_{1/5}$
Temperature (K)	130	130	130
Detector	ADSC - CCD	ADSC - CCD	ADSC - CCD
Cell constants (Å)	a=b=74.69 c=69.06 α=β=90° γ=120°		
V_m† ($Å^3$ Da^{-1})	2.14		
I/σ	8.1 (1.6)	8.3 (1.4)	8.1 (1.4)
Average multiplicity	16.6 (16.3)	16.6 (16.3)	16.6 (16.3)
Data completeness (%)	100 (100)	100 (100)	100 (100)
R_{merge}‡(%)	5.4 (45.6)	5.6 (54.6)	5.5 (53.5)
Anom. completeness*	100	100	100
f'/f'' (refined)	-5.71/4.05	-4.69/4.64	-1.75/3.53
Δano/σΔano	1.13	1.16	1.03

$R_{merge} = \sum \sum I <h>i - (I <h>) / \sum \sum I<h>i$, where $I<h>i$ is the measured diffraction intensity and the summation includes all observations. † assuming two molecules in the asymmetric unit. * Anomalous completeness corresponds to the fraction of possible acentric reflections for which an anomalous difference has been measured.

Table 3.6.3.2. Three wavelength MAD data collected from a single selenomethionyl crystal devoid of twinning.

3.7 Structure solution and Refinement

3.7.1 Se-norm (normal crystal)

The three-wavelength, pre-merged MAD data were used to 1.7Å, to determine the positions of the anomalous scatterers in SOLVE (Terwilliger and Berendzen 1999). The program was instructed to refine anomalous scattering factors and search for a total of six anomalous scatterers, corresponding to an expected three methionines/monomer from the amino acid sequence. SOLVE initially failed to scale the separate wavelengths together. The three datasets were merged using CAD (CCP4) and submitted to SCALEIT for a thorough statistical analysis (Howell and Smith 1992).

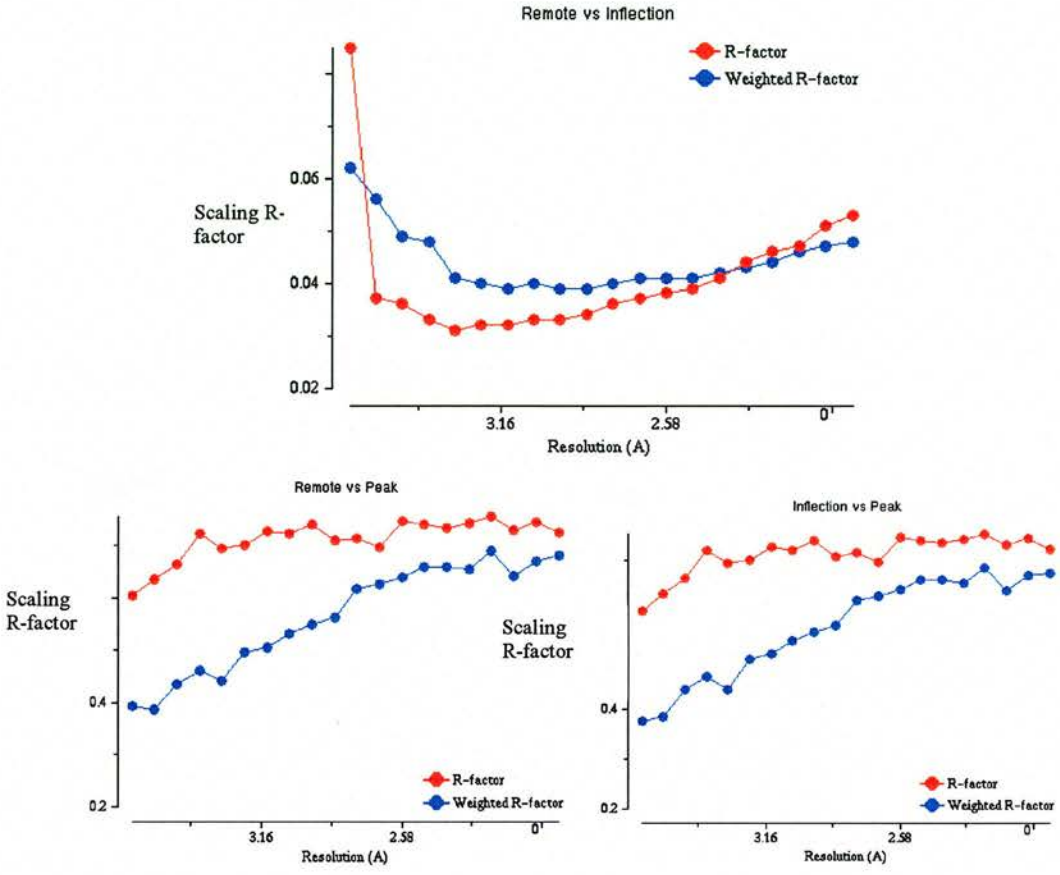

Fig.3.7.1.1 Amplitudes from the three wavelength data scaled against one another.

A cross comparison showed the peak wavelength to be the outlier as indicated by high R-factors (*Fig.3.7.1.1*) and standard deviations from the scaling of peak versus the inflection and remote wavelengths. As $P6_1$ and $P6_5$ are polar space groups, it is possible that the peak data had been indexed with the 6-fold screw axis oriented in the opposite direction when compared with the inflection and remote data. Using the CCP4 program REINDEX, the data were reduced to the asymmetric unit with the correct polarity applying the matrix k,h,-l to all reflections. The three wavelength data, merged in $P6_1$ (including the re-indexed peak), were submitted to SOLVE, as previously described and the top solution output four sites with a figure of merit of phasing (FOM) of 0.51 and a Z-score of 11.46 (Table 3.7.1.1). When the data were used to search for heavy atom sites

with $P6_5$ symmetry operators, the top solution gave the same figure of merit and a slightly lower Z-score (10.24).

X	Y	Z	Occupancy	B-factor	Peak Height (σ)
0.958	0.663	0.000	0.585	18.0	32.3 red
0.653	0.932	0.086	0.594	24.5	30.8 blue
0.798	0.646	0.150	0.262	37.0	11.4 green
0.349	0.944	0.007	0.161	25.9	10.5 purple

Table 3.7.1.1 Sites and statistics for the top solution found by SOLVE in $P6_1$.

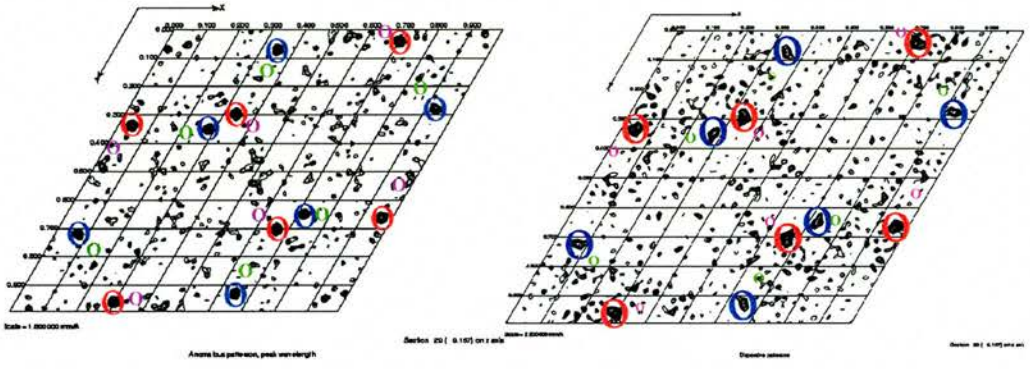

Fig.3.7.1.2 Patterson maps of harker section w=1/6, calculated from (left) the anomalous differences in the peak wavelength data and (right) the dispersive differences between the inflection and remote wavelengths. Peaks in blue and red denote cross-vectors between the equivalent positions of two strong, readily interpretable sites. Yellow and purple boxes denote the positions of cross-vectors between sites found by SOLVE, but not sufficiently strong to identify in the Patterson maps. Equivalent positions were calculated with VECTORS (CCP4).

Phases calculated from the four positions found by SOLVE were used to calculate electron density maps. At this stage, heavy atoms sites phased in both hands showed the true space group to be $P6_1$. The $P6_5$ density was

uninterpretable, while density derived from the sites found in $P6_1$ showed some evidence of secondary structure and a partial protein-solvent boundary (*Fig.3.7.1.3*). The SOLVE phases calculated in $P6_1$ were used as input to RESOLVE for solvent flattening of the resulting electron density (Terwilliger 2000) using a statistical density modification algorithm as opposed to traditional phase recombination methods. Electron density maps calculated from the adjusted $P6_1$ phases were of excellent quality, showing a clear protein-solvent boundary (*Fig.3.7.1.3*) and regions of extended secondary structure with well-defined side-chains (*Fig.3.7.1.4*).

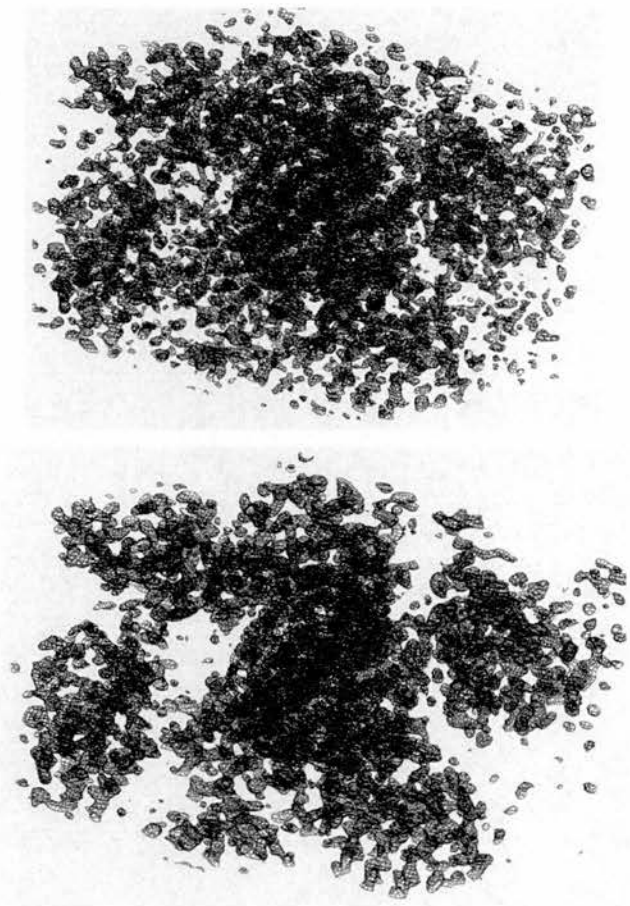

Fig.3.7.1.3 Electron density calculated in $P6_1$ from top- the raw, experimental SOLVE phases and bottom - the RESOLVE phases. The density for one SSB monomer is located in the centre of the map. Density is contoured at 1.5σ.

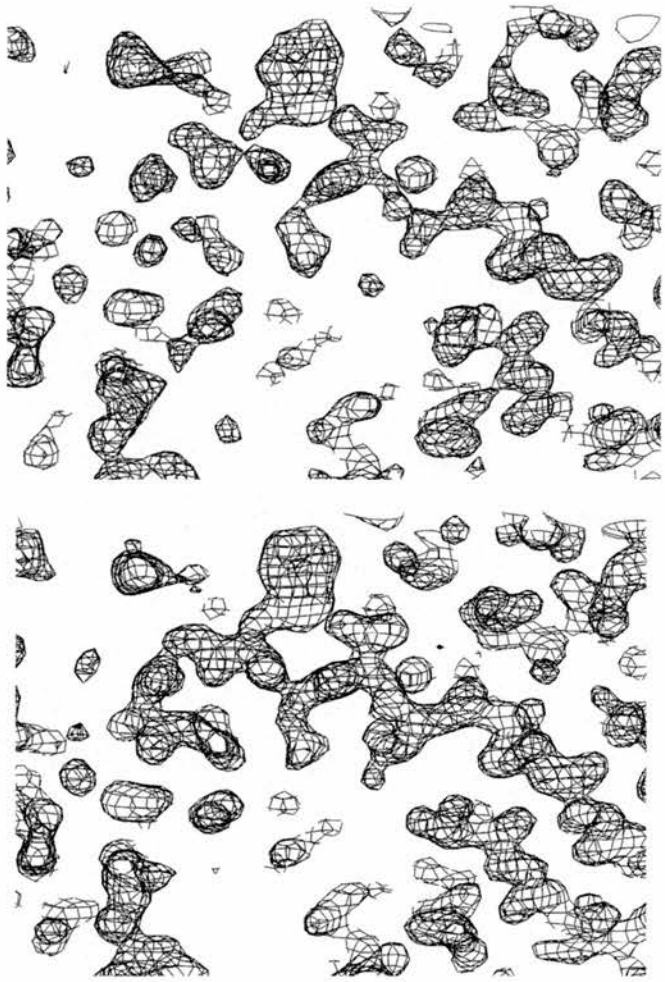

Fig.3.7.1.4 Side-chain density, contoured at 1.5σ, markedly improved when the raw, experimental phases (Top) were input to RESOLVE (Bottom).

3.7.2 Se-twin (twinned crystal)

Structure determination from twinned diffraction data poses a great challenge to the crystallographer. Merohedral twinning is often not detected until the refinement stage. At this point, the R-factor and R_{free} do not converge satisfactorily. Molecular replacement methods are generally applicable in the presence of twinning (Ito, Komiyama et al. 1995; Luecke, Richter et al. 1998; Ban, Nissen et al. 1999), although the errors in measured intensities will still

cause problems during structure refinement (Yeates and Fam 1999). In such cases successful detwinning of the data (as described earlier) has facilitated smooth refinement of atomic models. The problem of structure solution from twinned data is more complicated when phase information is obtained from isomorphous replacement (IR) methods and anomalous scattering. Nevertheless successful attempts have resulted in structures phased by IR methods (Reynolds, Remington et al. 1985; Yeates and Rees 1987). In one example high quality initial electron density maps were obtained by screening for heavy atoms derivatives with the lowest possible twin fractions (Terwisscha van Scheltinga, Valegard et al. 2001). Systematic errors incorporated in the intensity measurements by detwinning the observed structure factors (Grainger 1969) obviously affect the quality of the patterson map and subsequent phasing.

Problems also arise if the native and derivative datasets used in IR have different twin fractions as this may contribute to the intensity changes that are observed between the native and derivative diffraction patterns (Rees 1982). Features in the Patterson map may be due to such variances in the twin fractions and not the intensity differences due to heavy atom incorporation (Terwisscha van Scheltinga, Valegard et al. 2001). As pointed out by Rees, an extreme case would be where a native crystal appears to be a heavy atom-substituted derivative due to the anomalous intensity distribution from the twin. Rees et al were able to circumvent this problem using a $K_2Pt(NO_2)_4$ derivative with twin fraction 0.42 and native set with twin fraction 0.10 (Rees and Lipscomb 1980). By perfectly twinning the native dataset, the authors were able to markedly improve the quality of the difference Patterson.

The use of anomalous diffraction methods to generate phase information from merohedrally twinned data is particularly difficult. Anomalous scattering by heavy atoms results from a breakdown of Friedel's Law, which states that the reflections (h,k,l) and the symmetry related (-h,-k,-l) reflected from the opposite side of the same set of planes in the crystal are exactly equal in intensity and opposite in phase (Friedel 1913). Under anomalous scattering these Friedel (Bijvoet) pairs are no longer equal in intensity (Coster, Knol et al. 1933) or exactly opposite in phase. Anomalous scattering by an atom adds an imaginary component to the atomic scattering factor, the anomalous difference (f'') (Hendrickson 1991) that describes the difference in intensity between Friedel pairs of reflections. The anomalous differences arising from the anomalous scatterers can be used to locate the heavy atoms by patterson methods (single wavelength anomalous diffraction, SAD). This information can be used to obtain initial phase estimates for the heavy atom substructure. The location of the heavy atoms is improved by measuring dispersive (wavelength-dependent) differences (f') between reflections with the same miller indices at different wavelengths (Drenth 1999). Using this Multiwavelength Anomalous Diffraction method (MAD), the combination of the anomalous and dispersive difference information in the heavy atom search reduces the noise in the resulting Patterson map allowing the calculation of more accurate phase angles (Drenth 1999).

The intensity differences measured under anomalous scattering are small and therefore the anomalous signal needs to be accurately measured. The error in intensity measurements incorporated by merohedral twinning can therefore be drastic. At the time of writing there were only two reports of structure determination by MAD with twinned data (Yang, Dauter et al. 2000; Yang,

Forrer et al. 2000; Rudolph, Kelker et al. 2003). Yang et al describe the phasing of gpD, the capsid stabilising protein of bacteriophage λ, from pseudomerohedrally twinned crystals ($P2_1$) with a twinning fraction of 35-36% (Yang, Dauter et al. 2000). Interpretable electron density maps were obtained from both the twinned and detwinned data, although the maps calculated from the detwinned data gave better results. A cross-comparison of phasing programs using the twinned data showed that SOLVE (Terwilliger and Berendzen 1999) performed the best. In each case, density modification was performed by DM (Cowtan 1994). The resulting model was refined against the twinned data using *SHELXL* (Sheldrick and Schneider 1997), fitting the sum of the calculated structure factors for the individual twin domains, to the observed structure factors. This is preferable to refinement against detwinned data, due to the incorporation of systematic errors in the measured intensities by such a process (Jameson 1982).

Similarly, Rudolph et al were able to solve and refine the structure of an Interleukin-1β mutant using selenomethionine and native crystals with twinning fractions of approximately 0.35 and 0.40 respectively. Crystals belonged to space group $P4_3$. Selenium sites were located with both SOLVE and *SHELXD* (Schneider and Sheldrick 2002) using the twinned data. It is notable that attempts by both Yang et al and Rudolph et al included the use of high resolution data for structure phasing.

The three wavelength, twinned *Sso*SSB selenomethionine data were submitted to SOLVE (Terwilliger and Berendzen 1999) to identify heavy atom sites. The data were merged before submission and used to the full resolution

(1.7Å). Searching for 6 sites in $P6_1$, the top solution had a Z-score of 11.44 and a figure of merit of phasing of 0.38 as table 3.7.2.1.

X	Y	Z	Occupancy	B-factor	Peak Height (σ)
0.340	0.043	0.000	0.450	20.2	24.0
0.722	0.069	0.078	0.401	24.5	21.7
0.153	0.350	0.015	0.072	15.0	8.8

Table 3.7.2.1 Sites and statistics for the top solution found by SOLVE in $P6_1$

After phase improvement in RESOLVE (Terwilliger 2000), the figure of merit was 0.58. However, electron density maps calculated from both the SOLVE and RESOLVE phases were uninterpretable. The twinned crystal form differed (hexagonal rod) from the 'normal' crystal form (bipyramidal) used to solve the structure. Therefore the possibility still remained that this crystal actually belonged to the spacegroup $P6_5$. The solution in $P6_5$ (table 3.7.2.2) had the same figure of merit, but a lower Z-score (8.55).

X	Y	Z	Occupancy	B-factor	Peak Height (σ)
0.704	0.660	0.000	0.452	20.2	23.0
0.847	0.650	0.152	0.075	15.0	9.0
0.278	0.931	0.089	0.400	24.4	22.7

Table 3.7.2.2 Sites and statistics for the top solution found by SOLVE in $P6_5$

RESOLVE gave an improved figure of merit of 0.57, however the maps in this hand were also uninterpretable. Initially, the use of DETWIN to correct the observed intensities completely removed the anomalous signal, even when the twin fraction was grossly underestimated (α=0.10). Presumably the error lies in

improper deconvolution of the Friedel pair intensities (CCP4 2000). However, an anomalous difference Patterson calculated from the peak, twinned data identified the same distribution of peaks as is seen in figure 3.7.1.2 for the non-twinned data.

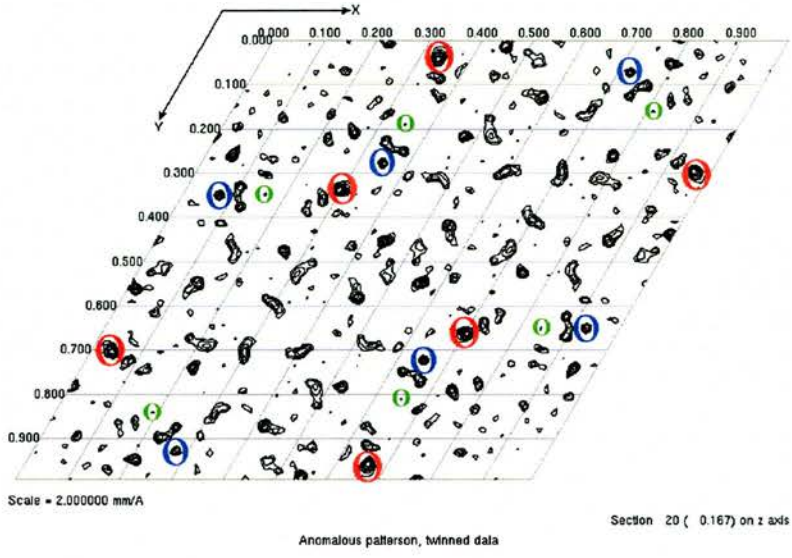

Fig.3.7.2.1 Anomalous difference patterson calculated from the peak, twinned data. Equivalent positions were determined using VECTORS (CCP4) and the sites in table 3.7.2.3 (next page). Peaks are coloured as figure 3.7.1.2 to highlight that these data were indexed in the opposite polarity

Therefore, the raw data files for the integrated three wavelength data were re-submitted to SCALA (Evans 1997) and the intensities were scaled but not merged using the UNMERGED ORIGINAL keywords to keep the friedel pairs separate. The three scaled, unmerged wavelengths were again used as input to SOLVE (Terwilliger and Berendzen 1999) (P61 setting) as above. The top solution was as Table 3.7.2.3 with a figure of merit of phasing of 0.35 and a Z-score of 12.45. When the equivalent positions were calculated in VECTORS, it

was clear that the twinned data were indexed in the opposite polarity when compared to the non-twinned data (figure 3.7.1.2, table 3.7.1.1) as the U and V coordinates of the patterson vectors were inverted.

X	Y	Z	Occupancy	B-factor	Peak Height (σ)
0.355	0.199	0.166	0.163	21.1	13.4 green
0.348	0.279	0.062	0.409	23.4	26.4 blue
0.044	0.704	0.152	0.441	19.8	28.8 red

Table 3.7.2.3 Sites and statistics for the top solution found by SOLVE in $P6_1$, using the unmerged data to 1.7Å.

The first site maps to a small feature coloured green in figure 3.7.2.1. The second and third sites map to stronger peaks representing harker vectors between two unique sites. However, analysis of the dispersive and anomalous differences as a function of resolution as output by SOLVE indicated that a significant amount of noise was included in the phasing procedure by using all the data to 1.7Å. SOLVE was therefore run again at 2.0Å and the top solution (Table 3.7.2.4) had greatly improved figure of merit (0.47, Z-score 11.59)

X	Y	Z	Occupancy	B-factor	Peak Height (σ)
0.042	0.703	0.000	0.459	19.5	21.8
0.069	0.347	0.077	0.423	23.0	18.0
0.060	0.652	0.156	0.097	12.3	7.6
0.348	0.195	0.013	0.049	1.0	7.2

Table 3.7.2.4 Sites and statistics for the top solution found by SOLVE in $P6_1$, using the unmerged data to 2.0Å.

Both the 1.7Å and 2.0Å phases were submitted to RESOLVE giving improved figures of merit of 0.55 and 0.72.

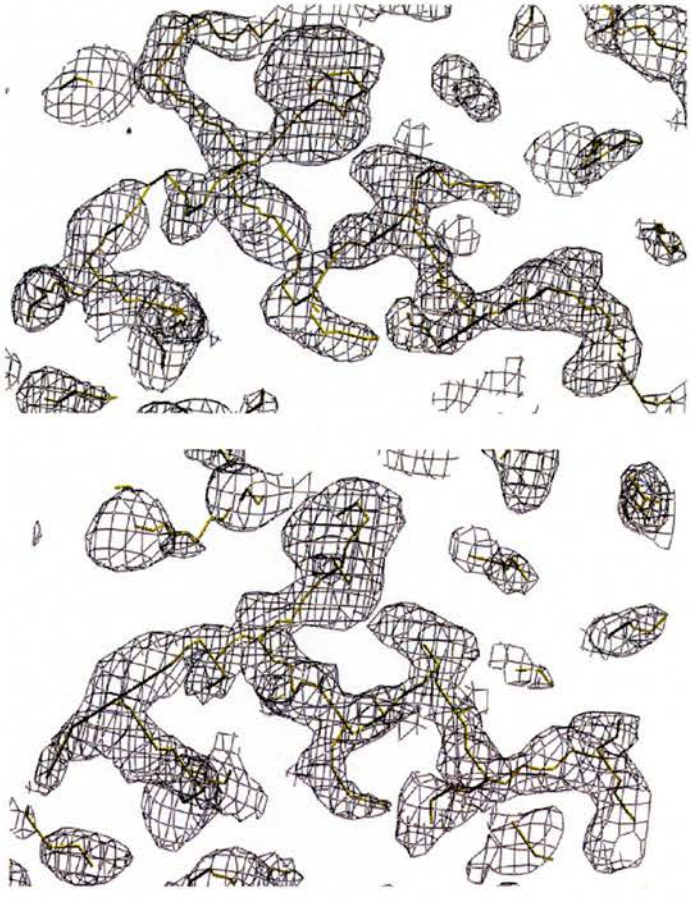

Fig.3.7.2.2 Electron density calculated from the RESOLVE phases at (top) 1.7Å and 2.0Å (bottom). The figures are centred on Trp56. Density is contoured at 1.5σ.

At this stage electron density calculated from both sets of phases was interpretable with clear density for the backbone and some sidechains. Skeletonisation of the maps using MAPMAN (Kleywegt and Jones 1996) showed good connectivity with 8 (to 1.7Å) and 14 (to 2.0Å) fragments consisting of 100 or more points. At 1.7Å, ARP/wARP builds a maximum of

187 residues (of a possible 238) as 7 chains with a connectivity index of 0.93 (*Fig.3.7.2.3*).

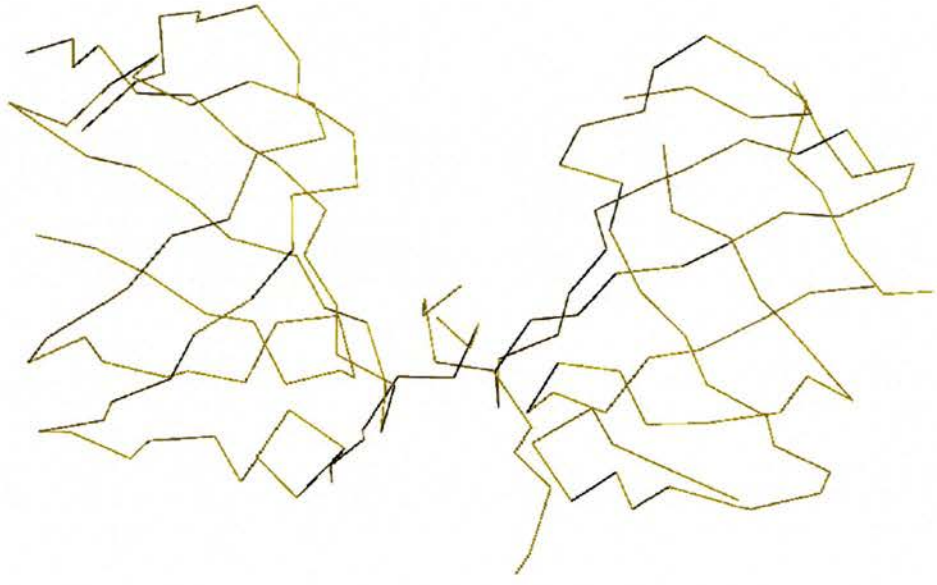

Fig.3.7.2.3 Part of the two molecules in the asymmetric unit traced by ARP/wARP.

Further density improvement by 2-fold non-crystallographic averaging will be pursued at a later date to improve the building, followed by atomic refinement using *SHELXL*.

3.7.3 Model building and refinement

The phases and structure factor amplitudes from RESOLVE (solution from crystal Se-norm in section 3.7.1) were used to automatically trace the contents of the asymmetric unit using ARP/wARP (Perrakis, Morris et al. 1999). Using all the data to 1.7Å, an initial Cα trace was obtained using the "*warpNtrace*" mode, running iterative cycles of automated model building in real space with

reciprocal space refinement performed by REFMAC5 (Murshudov, Vagin et al. 1997). The position of side-chain atoms are automatically filled with waters. This cyclic procedure ran for 99 cycles, yielding a Cα model with an R-factor of 17.4%. Electron density maps calculated from the *warpNtrace* phases were of excellent quality (*Fig.3.7.3.1*).

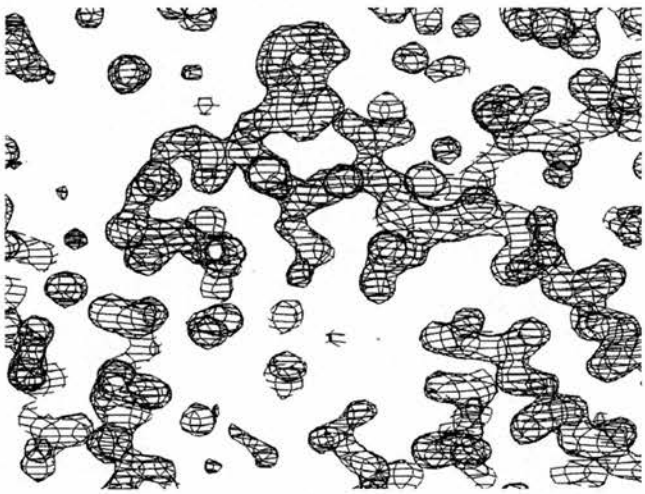

Fig.3.7.3.1 Side-chain density, contoured at 1.5σ , calculated from the *warpNtrace* phases. Same view as *Fig.3.7.1.4*

The extent of building was deemed to be excellent, with a final connectivity index of 0.95. Cα positions not automatically traced were included manually using 'O' (Jones, Zou et al. 1991). After removing all waters automatically added to the model during Cα-tracing, amino acid side-chains were built using the "*side dock*" option in ARP/wARP. This generated a complete crystallographic dimer, monomer A consisting of 115 residues and monomer B consisting of 114 residues.

The model was refined in REFMAC5 against the detwinned, native 1.2Å data (Native-2) in a gradual stepwise manner, starting at d_{min}=2.0Å (*Fig.3.7.3.2*).

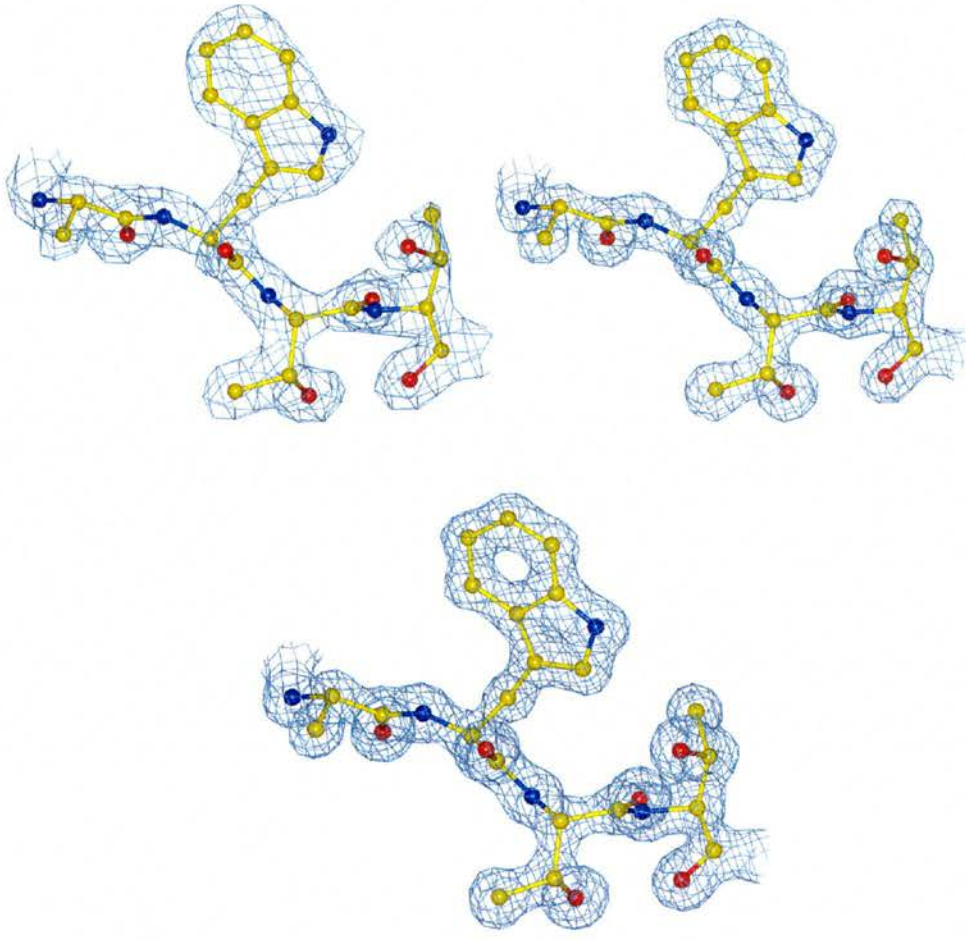

Fig.3.7.3.2 Electron density and model during the course of refinement. At each stage, refinement has reached convergence. Density contoured at 2.0σ using all the data to 2.0Å (top left), 1.5Å (top right) and 1.2Å (bottom)

The refinement approached convergence when the data were extended to 1.8Å and at this stage waters were added to the model using an automated script running REFMAC5 (refinement) and arp_waters (add waters module of ARP/wARP). The procedure searched for waters in density above 2.5σ x r.m.s density and removed atoms in density below 1σ x r.m.s density. Once the refinement had extended to 1.5Å, thermal factors were refined anisotropically. Hydrogen atoms were included in their riding positions, but not in the final model as they could not be located in the electron density.

The use of non-crystallographic (NCS) restraints and constraints was explored, however R_{free} (Brunger 1992) only decreased when they were removed; the observation to parameter ratio, calculated as below (*1.1.*), was approximately 7.0 (3.5 if refining anisotropic thermal factors) suggesting the need for NCS restraints was unwarranted.

$$\textit{Observation:Parameter} - \frac{\textit{Number of unique observations}}{\textit{Number of atoms in the asymmetric unit x } P}$$

where P is the number of parameters per atom, 5 for isotropic, restrained refinement (x, y, z, occupancy, 1 thermal parameter) and 10 for anisotropic restrained refinement (x, y, z, occupancy and 6 thermal parameters)

(*1.1.*)

Resolution	65-1.2
Space group	$P6_1$
R_{free}†	21.9 (35.6)
R-factor‡	19.1 (34.3)
Rmsd bonds (Å)/angles(°)	0.015/1.6
B factor deviation bonds (Å^2)/angles	
Main chain	1.5/2.2
Side Chains	2.8/4.1
Residues in Ramachandran core (%)*	96.4%
Protein atoms	1731
Water atoms	273
Ligand atoms	15
Average B factor (Å^2)	11

† $R_{free} = \Sigma_{hkl} \| F_{obs}(hkl) - | k | F_{calc}(hkl) \| / \Sigma_{hkl} | F_{obs}(hkl) |$ where F_{obs} and F_{calc} are the observed and calculated structure factors respectively. ‡ R-factor $= \Sigma_{(hkl)} \varepsilon T \| F_{obs}(hkl) - | k | F_{calc}(hkl) \| / \Sigma_{(hkl)} \varepsilon T | F_{obs}(hkl) |$ as implemented by Brunger (Brunger 1992), where $\Sigma_{(hkl)} \varepsilon T$ is 5% of the reflections omitted during refinement. * Ramachandran core refers to the most favoured region in the Φ/Ψ Ramachandran plot as defined by Laskowski et al. (Laskowski, MacArthur et al. 1993).

Table 3.7.3.1 Refinement statistics for the final model

The final model (*Fig.3.7.3.3*) was refined to an R_{free} of 21.9% and an R-factor of 19.1%.

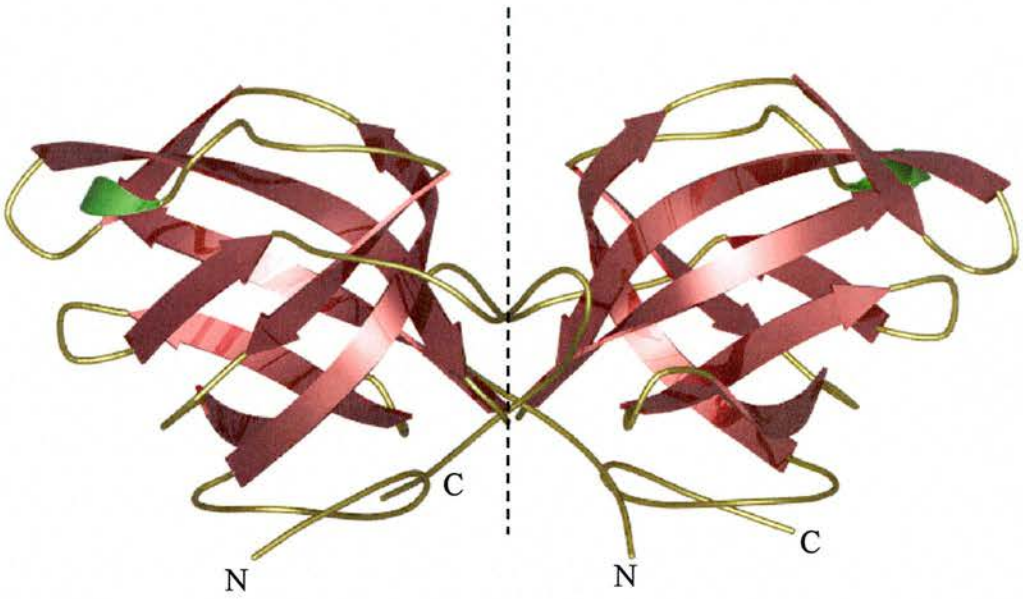

Fig.3.7.3.3 The two SSB molecules built in the asymmetric unit, viewed down an axis perpendicular to the 2-fold non-crystallographic axis (dashed line)

Model geometry and stereochemistry were judged to be excellent as determined by PROCHECK (*Fig.3.7.3.4*) (Laskowski, MacArthur et al. 1993) and WHATIF (Vriend 1990).

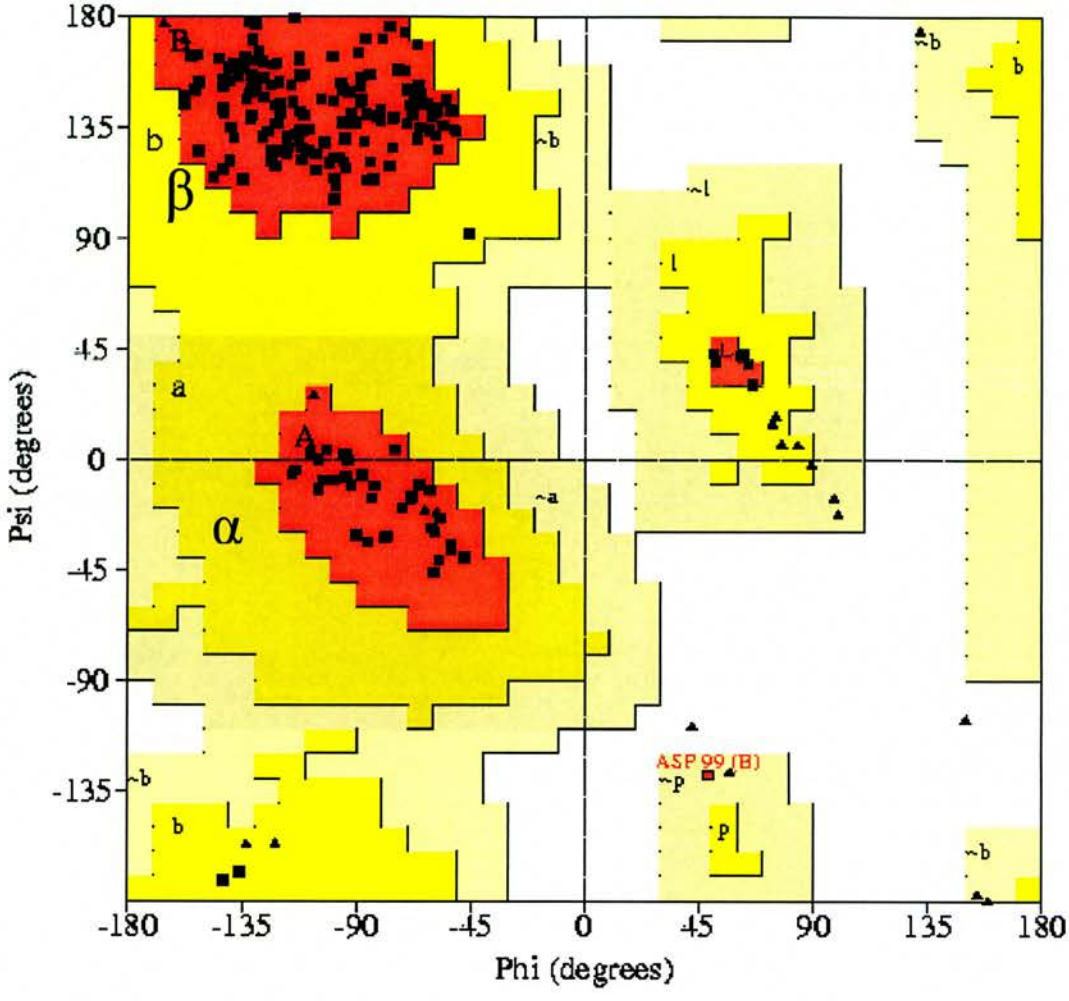

Fig.3.7.3.4 Ramachandran plot of the refined SSB model. Regions are defined as: **A**- core alpha, **a**- allowed alpha, ~**a**-generous alpha, **B**-core beta, **b**- allowed beta, ~**b**- generous beta, **L**- core left-handed alpha, **l**- allowed left handed alpha, ~**l**- generous left-handed alpha, **p**- allowed epsilon, ~**p**- generous epsilon.

The slightly higher R-factors in the high resolution shells were presumed to be due to the inevitable errors incorporated by the detwinning process (Jameson 1982). The coordinates have been submitted to the PDB (ID 1O7I) along with observed structure factor amplitudes for the three wavelength of MAD data and the 1.2Å native data (twinned and detwinned).

Discussion

3.8 Overall structure

The crystal structure of the single stranded DNA-binding protein from *Sulfolobus solfataricus* has been determined to 1.2Å, shedding light on ssDNA binding and protection in the archaea (*Fig.3.8.1*).

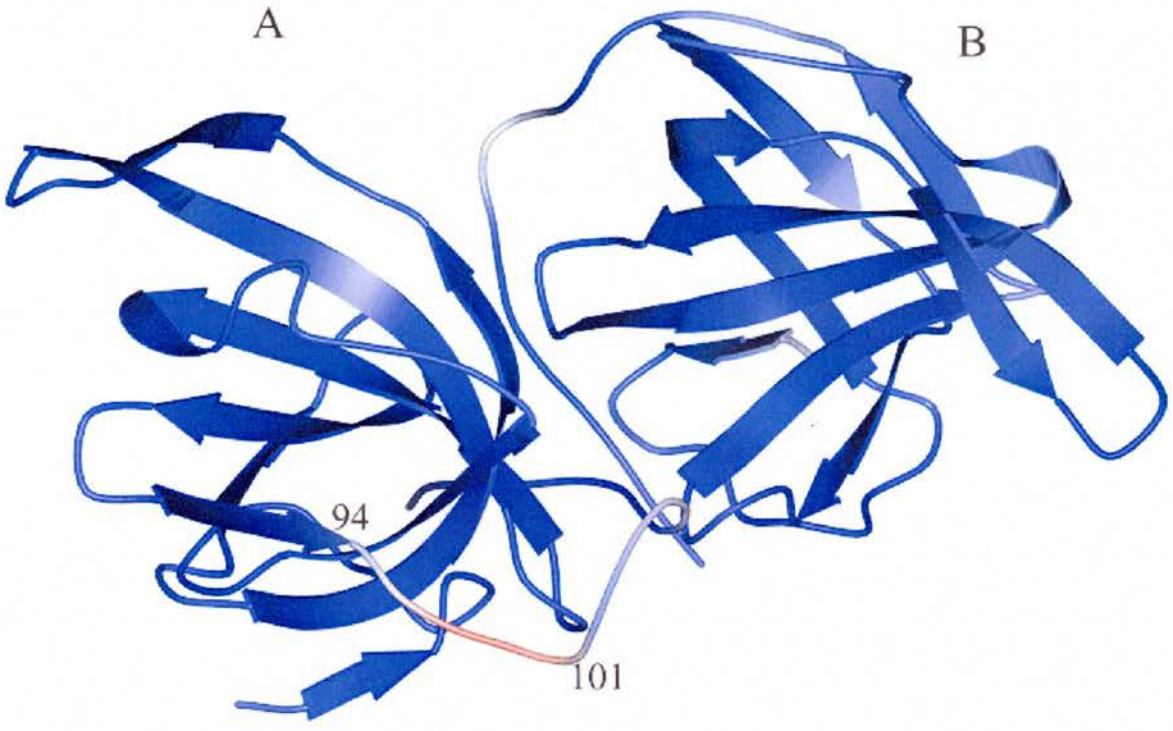

Fig.3.8.1 The two *Sso*SSB monomers traced in the asymmetric unit. A region of disorder at the C-terminus of monomer A is labelled. Blue areas have low B-factors, while white/red denotes areas with progressively high B-factors.

To date, this represents the highest resolution structure of a single-stranded DNA binding protein. The crystallised protein contains residues 1-119 - the C-terminal 29 residues were cleaved by trypsin digestion. This region is proposed to be involved in protein-protein interactions (Wadsworth and White 2001; Haseltine

and Kowalczykoski 2002). *Sso*SSB crystallises with two monomers in the asymmetric unit, related by two-fold non crystallographic symmetry - monomer A (residues 1-115) and monomer B (residues 1-114) (*Fig.3.8.1*). The remaining C-terminal residues are presumably disordered and could not be located in the electron density. Figure 3.8.1 shows the structure to be well ordered as implied by low thermal factors.

Brute-force superimposition of the two monomers (all Cα atoms from 1-114 inclusive) gives a root mean square deviation (rmsd) of 0.82Å. This seemingly larger difference can be attributed to the N-terminal three residues, which adopt a slightly different conformation in the two monomers, and two flexible regions comprising residues 31-36 and 94-100 (*Fig.3.8.1*). Excluding these three regions using the 'EXPLICIT' command in LSQMAN (Kleywegt 1996) gives an rmsd of 0.43Å, matching 98 α-carbons. Residues 94-100 are found on the surface of the protein and are involved in mainly water-mediated crystal contacts. According to thermal factors and poor electron density, this area is more flexible than the rest of the structure. Residues 31-36 are also found on the surface and form a crucial crystal-packing interface. The presence of a dimer in the asymmetric unit is an artefact of crystal packing and does not represent a biologically significant oligomer. The total buried surface area between the two monomers in the asymmetric unit is approximately 800Å^2, and only a few direct protein-protein interactions are observed.

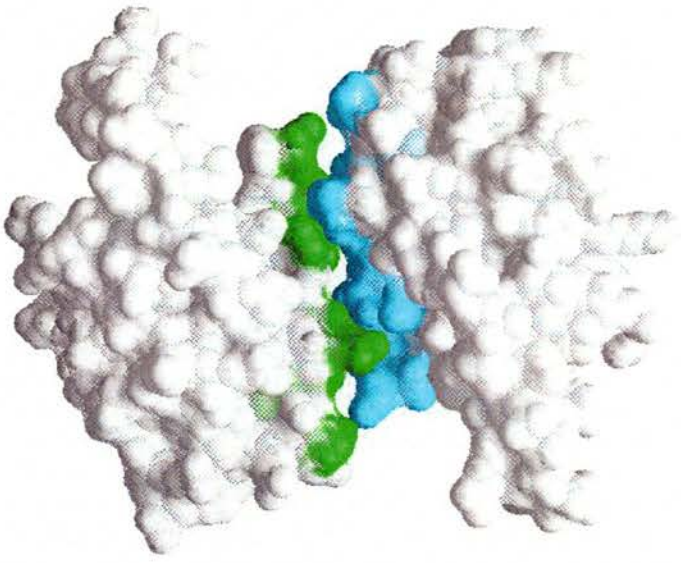

Fig.3.8.2 Molecular surfaces of the two monomers in the asymmetric unit forming the non-biological dimer. Interface residues in monomer A are green and in monomer B are blue

*Sso*SSB possesses a single OB fold, as predicted, that differs slightly from the fold template of five β-strands capped by an α-helix, as defined by Murzin (Murzin 1993).

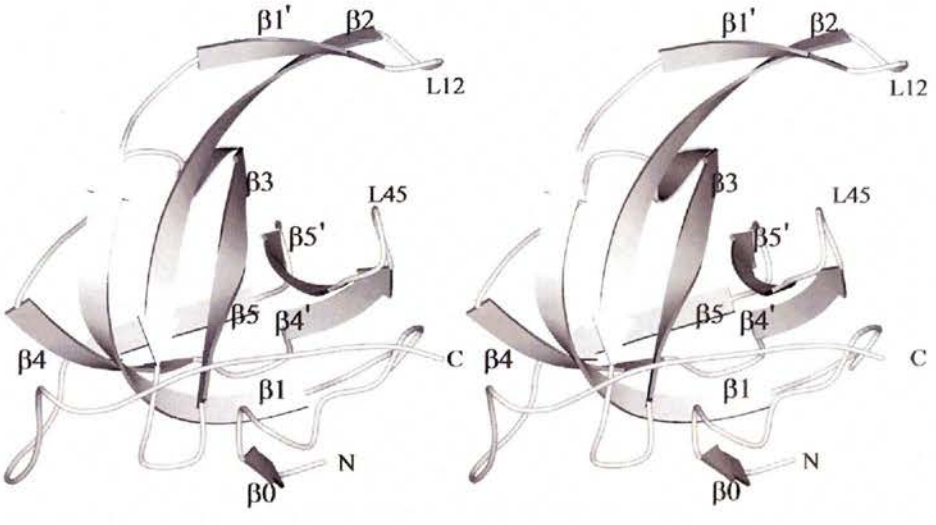

Fig.3.8.3 Stereo-images of the *Sso*SSB monomer. Secondary structure is labelled as described above

*Sso*SSB possesses an additional β-strand at the N-terminus (β0) and the α-helix commonly found between β3 and β4 is replaced by a short helical region (residues 57-59) and a turn (residues 60-62). The L_{12} and L_{45} loops (*Fig.3.8.3*) are important for ssDNA recognition in the OB fold, forming β-hairpins between β1'- β2 and β4'- β5' respectively.

3.9 Structural analysis

3.9.1 Structural homologues of *Sso*SSB

To determine the structural relationship between *Sso*SSB and its bacterial and eukaryal homologues, the coordinates were submitted to the DALI server (Holm and Sander 1993). The ssDNA-complexed form of human RPA70 was identified as the most similar protein (z=13.8). The remaining RPA subunits appeared further down the list, however surprisingly *E. coli* SSB was not flagged as a structural homologue. Over 55 homologous structures were identified, reflecting the ubiquitous nature of the OB fold.

Molecule	Rmsd over core†	Improved LSQ fit rmsd
apo-E.coli SSB	2.1Å	1.7Å, 66 α-carbons
E.coli-DNA SSB	2.2Å	1.8Å, 64 α-carbons
Human mitochondrial SSB	1.3Å	1.6Å, 66 α-carbons
RPA70:		
apo-DBD-A	1.0Å	1.6Å, 82 α-carbons
apo -DBD-B	0.9Å	1.4Å, 78 α-carbons
DBD-A - DNA	1.1Å	1.6Å, 84 α-carbons
DBD-B - DNA	0.8Å	1.6Å, 87 α-carbons

†The *Sso*SSB core is defined as residues 15-25, 39-36, 50-56, 66-71, 74-77, 83-88 and 91-95

Table 3.9.1.1 Similarities between *Sso*SSB and structural homologues

Using the EXPLICIT command in LSQMAN (Kleywegt 1996), two structures of interest were superimposed with *Sso*SSB over a set of residues deemed to form the core of the OB-fold (Table 3.9.1.1). The resulting root mean square deviations are as Table 3.9.1.1. The overall least-squares fit between molecules was determined by implementing the IMPROVE command in LSQMAN to iteratively add matching atoms outside the OB-fold core. It became clear that the RPA subunits used in this analysis showed a closer fit with *Sso*SSB in comparison with *E. coli* SSB. All the RPA70 models included in this study showed a better fit with *Sso*SSB than did *E. coli* SSB. Interestingly the Human mitochondrial SSB shows an LSQ fit similar to that of *E. coli* (1.2Å over 169 α-carbons between the two). This is consistent with the fact that the mitochondrial homolgue is almost identical in tertiary and quaternary structure to its bacterial counterpart (Yang, Curth et al. 1997).

The data show the OB-fold closest to *Sso*SSB in structure is DBD-B (*Fig.3.9.1.1*). Superposition of the Sulfolobus protein with apo-*E. coli* SSB matches 66 α-carbons with an rmsd of 1.6Å, whilst for apo-DBD-B, 78 α-carbons superimpose with an rmsd of 1.4Å. The structural differences between *E. coli* SSB and *Sso*SSB are reflected mainly in the loops (*Fig.3.9.1.1*). Of particular note are the L_{23} and L_{45} loops which are longer in *E. coli* SSB. These loops have aromatic residues that guide ssDNA around the tetramer through base stacking interactions. The L_{45} loop is proposed to be involved in cooperative assembly of tetramers along ssDNA (Raghunathan, Ricard et al. 1997; Raghunathan, Kozlov et al. 2000). In *Sso*SSB the much shorter L_{23} and L_{45} loops preclude the possibly that *Sso*SSB binds ssDNA by wrapping it round a tetrameric arrangement of OB folds in the same way as *E. coli* SSB

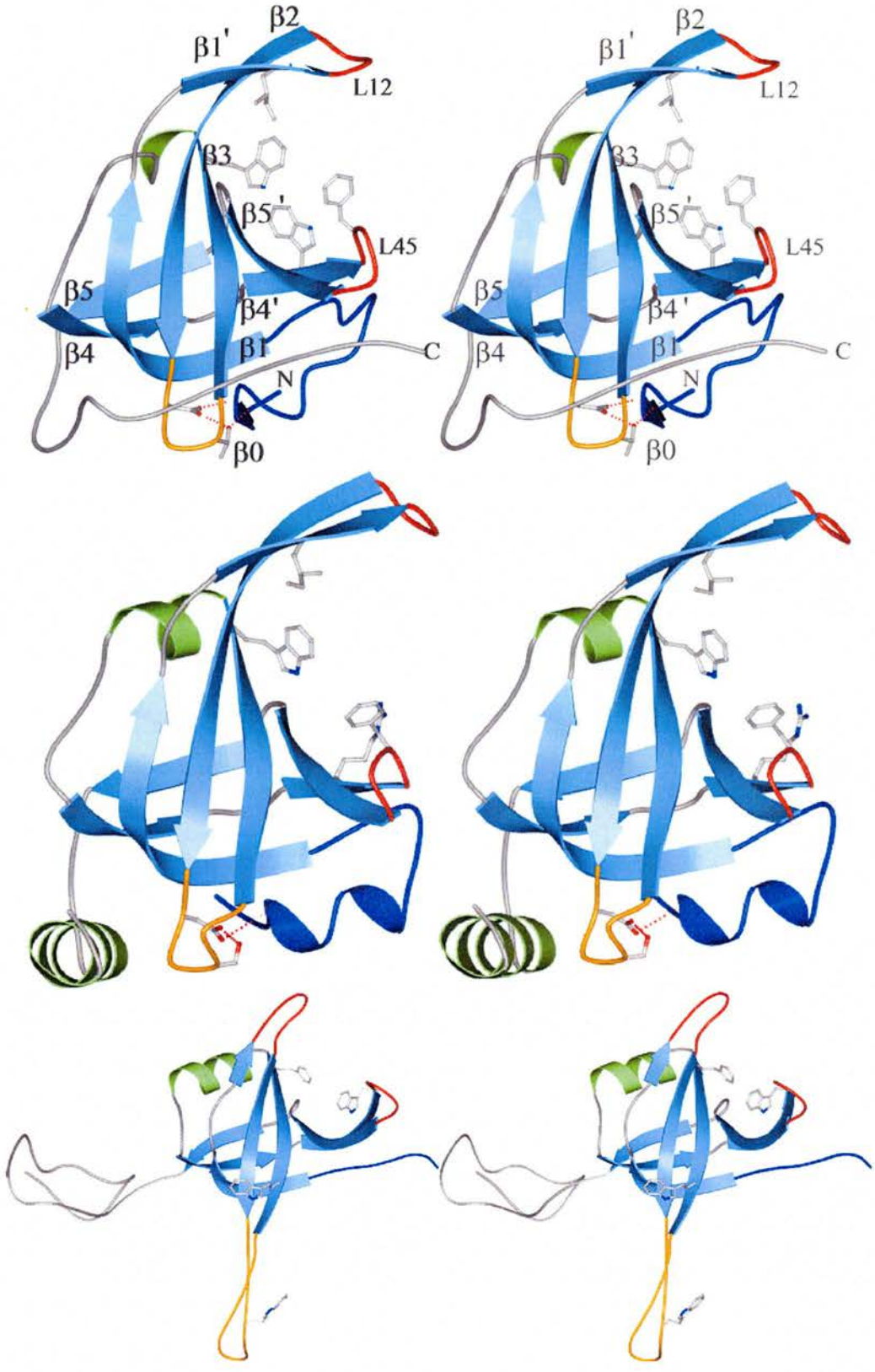

Fig.3.9.1.1 OB folds from *Sulfolobus solfataricus* (top), RPA70 DBD-B (middle) and *E.coli* (bottom). The L_{23} loop is coloured orange and the capping helix, green. Important DNA-binding residues are depicted in ball and stick.

The high degree of structural similarity facilitated structural alignment of *Sso*SSB with DBD-A and B (*Fig.3.9.1.2*).

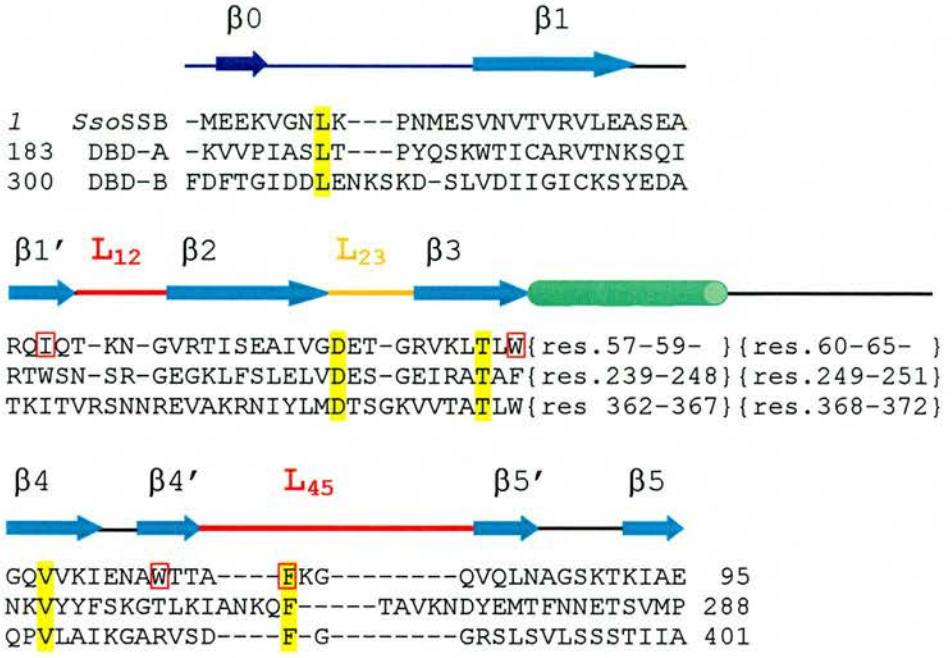

Fig.3.9.1.2 Structural alignment of *Sso*SSB with the two RPA70 OB folds, DBD-A and DBD-B. At the extreme N-terminus (blue), secondary structure is not conserved and therefore relates solely to SsoSSB in the alignment. The helical region and loop connecting β3 and β4 are not shown. These were difficult to align and do not constitute part of the superimposed cores. Absolutely conserved residues are shaded yellow, important DNA-binding residues are boxed in red.

The alignment highlights a conserved DxT/S motif first noted by Bochkareva et al (Bochkareva, Frappier et al. 1998) in the loop between β2 and β3 - the L_{23} loop (*Fig.3.9.1.1* and *3.9.1.2*). This sequence is absent in all the bacterial homologues, but present in the mitochondrial SSB where it plays an unknown function. In *Sso*SSB and RPA70 this region is remote from the DNA-binding site. The Asp and Thr residues (residues 46 and 48 respectively) form a hydrogen

bond network with the amide backbone at residues 5 and 6 (*Fig.3.9.1.1*). This results in the N-terminus of the protein folding into an initial turn of the β barrel. In DBD-B, Asp350 contacts the N-terminal amide backbone at residue 305 (*Fig.3.9.1.1*) and Ser352 makes a contact with the sidechain of Asp306 (not shown in *Fig.3.9.1.1*) forcing this region of structure to adopt a helical turn. In the *E. coli* structure, the N-terminus is an elongated strand which makes a β sheet interaction to stabilise the dimer. This N-terminal strand appears to be stabilised by contacts with the L_{23} hairpin from a neighbouring monomer. *Sso*SSB and RPA70 may be unable to achieve the oligomeric states observed in *E. coli* SSB because the N-terminus is folded back into the monomer through interactions with the L_{23} loop. This local region of structure, is perhaps the key feature defining the eukaryal and archaeal OB-folds and differentiating them from the bacterial and mitochondrial SSBs.

3.9.2 Single-stranded DNA-binding by *Sso*SSB

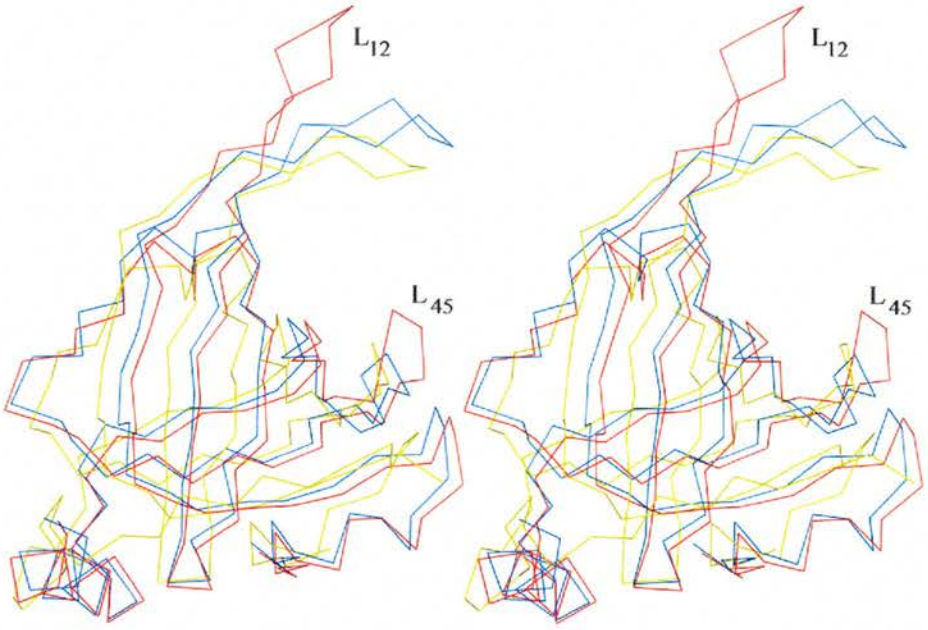

Fig.3.9.2.1 Superposition of *Sso*SSB (yellow), apo-DBDB (red) and DBDB-DNA (blue)

The OB fold of *Sulfolobus solfataricus* SSB is structurally very similar to DBD-B (*Fig.3.9.2.1*). The rmsd between the core residues of *Sso*SSB and DBD-B is only 0.9Å. Surprisingly the rmsd between the crenarchaeal protein and the DNA-complexed form of DBD-B is even lower (Table 3.9.1.1). Superimposition of the three models confirms that *Sso*SSB is closer in structure to DBD-B when ssDNA is bound (*Fig.3.9.2.1*). This is surprising, as crystallographic studies on SSB OB folds have shown that major conformational changes follow ssDNA-binding. A crystal-packing interface between two *Sso*SSB monomers shows that the L_{12} loops interlock via a complex set of interactions involving three sulphate ions (from the crystallisation conditions) and a network of water molecules

(*Fig.3.9.2.2*). Latterly it has been found that *Sso*SSB only crystallises in the presence of sulphate anions.

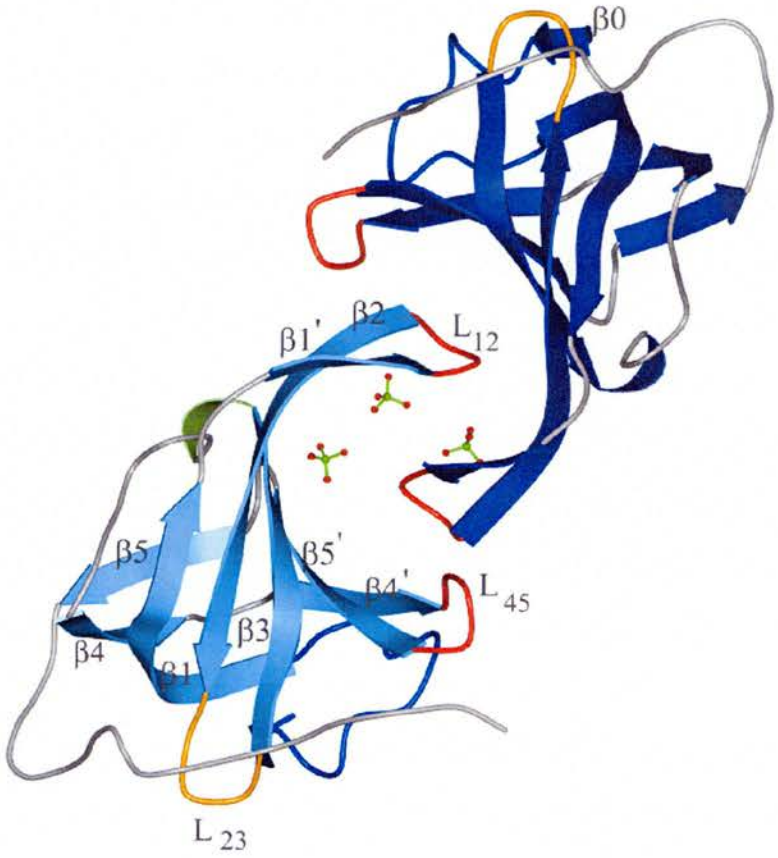

Fig.3.9.2.2 A crystal packing interface in SsoSSB involving sulphate ions (depicted in ball and stick)

These interactions force the L_{12} loop to fold down towards the same loop in the symmetry related molecule and the presence of sulphate ions at these positions may mimic ssDNA-binding by *Sso*SSB. The superposition of *Sso*SSB with DBD-B suggests that the two molecules share an almost identical ssDNA binding site that is located between the L_{12} loop, the L_{45} loop and the outside of the barrel. Inspection of the electrostatic potential of the molecular surfaces (*Fig.3.9.2.3*) shows a similar basic distribution of charge in the ssDNA binding cleft and around the DNA-binding loops. Outwith this region, the molecules

display dense electronegative patches that would repel the negatively charged phosphate backbone of DNA. In the RPA70- DNA complex, ssDNA tunnels through the DNA-binding site in a linear fashion with the OB folds arranged side by side in tandem.

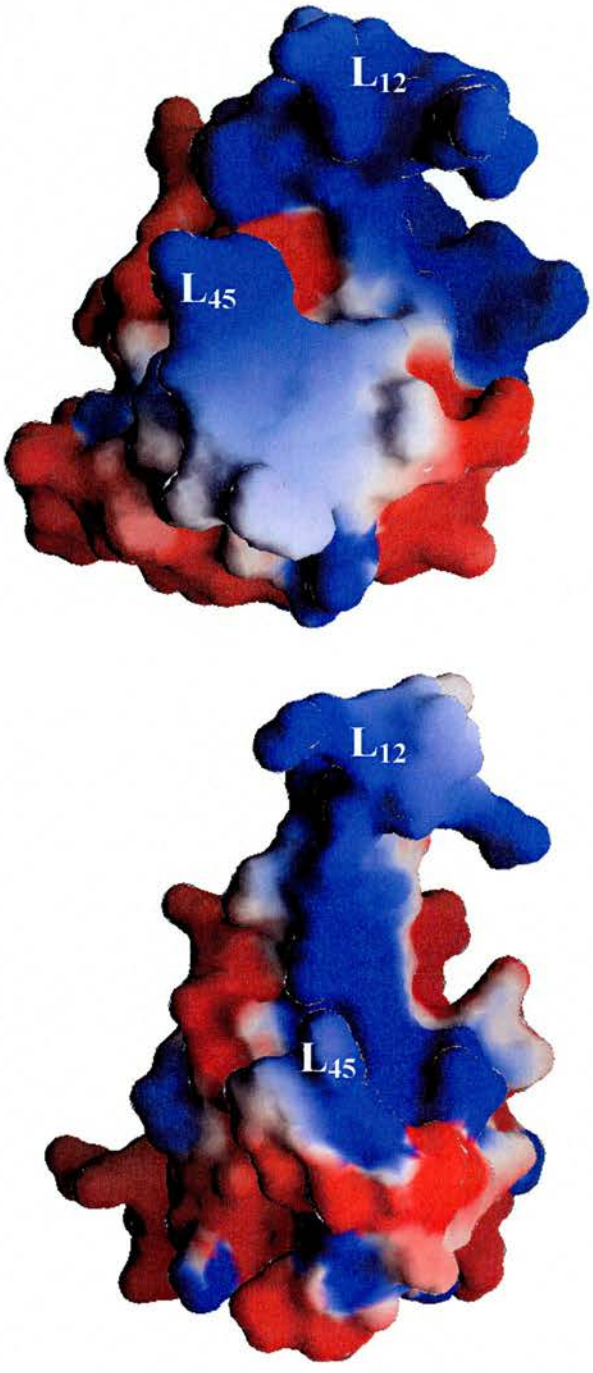

Fig.3.9.2.3 Electrostatic surfaces of Top - *Sso*SSB and Bottom - the apo DBD-B subunit of RPA70. The electrostatic surface is contoured from -3kT/e to +3kT/e

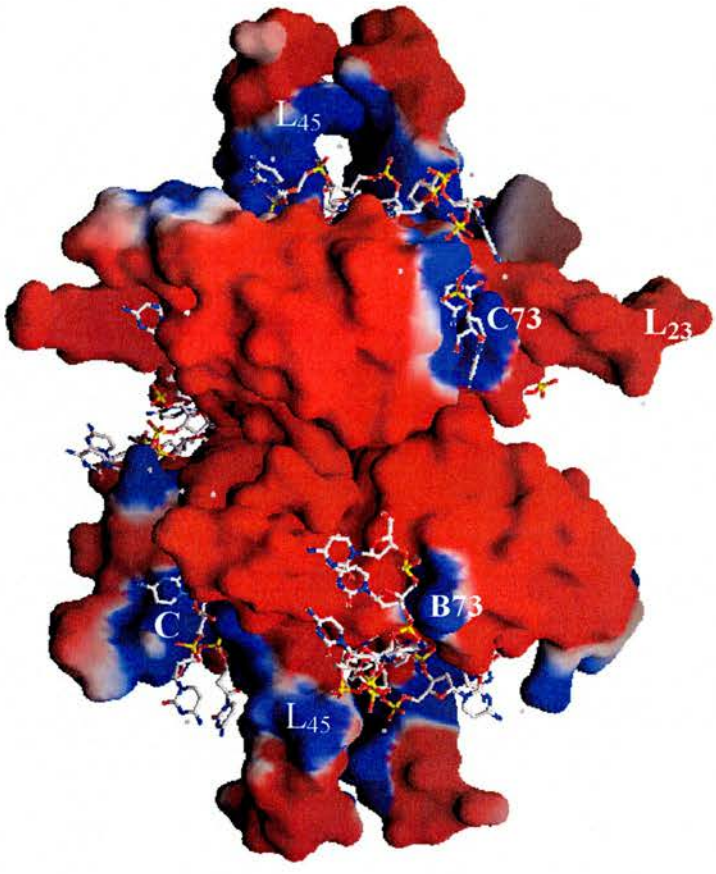

Fig.3.9.2.4 Electrostatic surface (contoured as the previous figure) of *E. coli* SSB. ssDNA is depicted in ball and stick, red - oxygen, blue - nitrogen, grey - carbon, yellow - phosphorus. C denotes a deep electropositive pocket analagous to the DNA-binding cleft in *Sso*SSB and the RPA70 OB folds.

In contrast, the electrostatic surface of *E. coli* SSB (*Fig.3.9.2.4*) is punctuated by pockets of dense electropostive charge that interact with the negatively charged ssDNA. DNA is guided round these basic patches in the tetramer by base stacking interactions with aromatic residues in the L_{23} and L_{45} loops. Taken together this evidence suggests that in *Sso*SSB, ssDNA follows a path similar to that in RPA70 without significant wrapping round the monomer as is seen in *E.coli* SSB. In fact, it seems likely that ssDNA-binding by the *Sulfolobus* protein follows the same RPA70 open/close model suggested by Bochkarev et al

(Bochkarev, Pfuetzner et al. 1997), involving large conformational changes in the DNA-binding loops. This model suggests that in the absence of DNA, the L_{12} loop is opened up from the outside of the barrel whilst closing of the L_{45} loop partially compacts the DNA-binding cleft. Upon ssDNA-binding it is proposed that the L_{45} loop opens out from the DNA binding cleft to allow the initial entry of ssDNA, followed by closing of the L_{12} loop to the cleft, as mimicked by the sulphate-binding interface in *Sso*SSB crystals, locking DNA in place.

A model for ssDNA-binding was generated from the *Sso*SSB/DNA-complexed DBD-B superimposition (*Fig.3.9.2.5*).

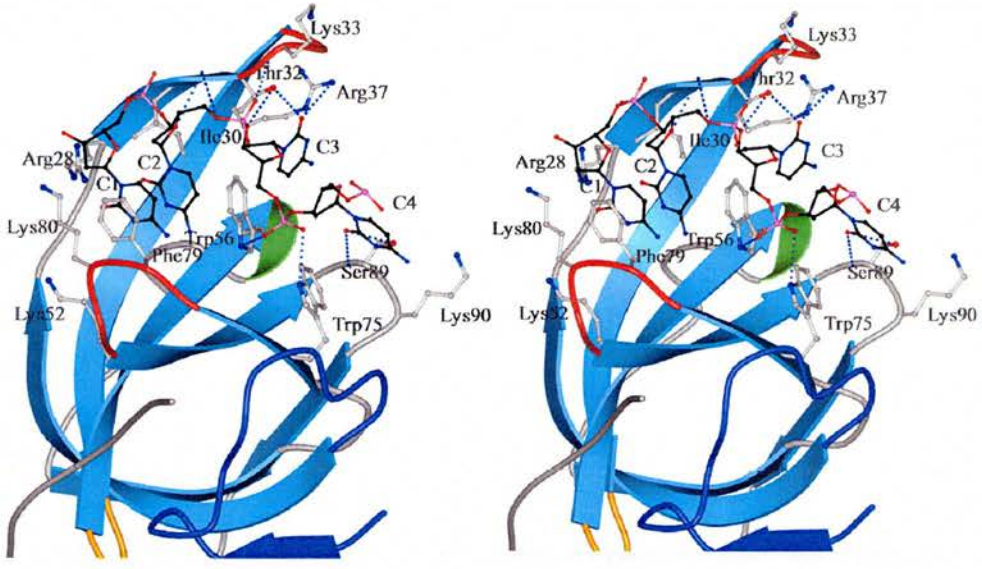

Fig.3.9.2.5 A model for ssDNA-binding in *Sso*SSB generated by superposition of the DBD-B - ssDNA complex on top of the Sulfolobus homologue. Hydrogen-bonds interactions are coloured blue

In the resulting model, only one close contact of 1.5Å is observed between the backbone carbonyl of Ile31 and C5 on nucleotide C2. All other protein-DNA contacts are 2.0Å or over. The basic residues (Arg28, Lys33, Arg37, Lys52, Lys

80 and Lys90) all have the potential to be involved in electrostatic interactions with ssDNA and are within 15Å of the phosphate backbone of the modelled DNA-fragment. Several potential hydrogen bonds are formed between nucleotides C2-C4 and either protein sidechains or the amide/carbonyl backbone.

The aromatic residues at positions 56 and 79 in *Sso*SSB are absolutely conserved in RPA70 where they make stacking interactions with ssDNA (Bochkarev, Pfuetzner et al. 1997). The model in Figure 3.0.4 suggests that Phe79 may be involved in a contiguous stacking arrangement with DNA-bases. Trp 56 may also base stack with nucleotides although this is not directly implied by the model. Mutation of Trp56 and Phe 79 results in 23-fold and 9 fold losses in ssDNA binding affinity respectively, as determined by isothermal titration calorimetry (Wadsworth and White, unpublished observations).

In the SsoSSB-ssDNA model, a third aromatic residue (Trp75) donates a proton from the indole nitrogen to the phosphate backbone to make a hydrogen bond. This interaction is supported by the fact that there is no equivalent hydrophobic residue at the same position in either DBD-A or DBD-B. Instead DBD-B contacts the phosphate backbone of ssDNA via a hydrogen bond from Arg382 (Bochkarev, Pfuetzner et al. 1997). Mutation of Trp75 in *Sso*SSB to alanine results in a 4-fold decrease in binding affinity, much smaller than for the conserved aromatics that are proposed to stack with ssDNA. The contribution of this residue to binding energy (assessed by changes in Gibbs-free energy) is also less (Wadsworth and White, unpublished observations). The importance of bulky, aromatic residues in DNA-binding by the OB-fold is underscored by the fact that *E.coli* SSB possesses aromatic residues at positions equivalent to 56 and 79 in *Sso*SSB, both of which are involved in base stacking interactions. The

binding site size observed in the model correlates with experimental evidence showing that ssDNA binding in *Sso*SSB is saturated upon by 4-5 nucleotides of DNA per monomer (Wadsworth and White 2001). This saturation point is preserved for all lengths of ssDNA studied. The high resolution structure of *Sso*SSB and subsequent DNA modelling into the OB-fold show that the crenarchaeal and eukaryal proteins are remarkably similar. This is astonishing in light of the fact that eukaryotes and the archaea are separated by billions of years of evolution. Bacteria, which are presumed to have appeared much closer to the archaea in the evolutionary timeline, possess single-stranded DNA binding proteins that have a C-terminal tail, present in the archaeal homologues but not in the eukaryal RPAs.

The evolutionary origins of the chimeric *Sso*SSB, with a eukaryotic-type DNA binding domain coupled to a bacterial-like protein-protein interaction domain, remain an intriguing puzzle. Nevertheless, ssDNA-binding in *Sso*SSB is far more similar to the mechanism achieved by the RPA70 subunit of the eukaryotic heterotrimer, where DNA tunnels through the binding site in a linear fashion. Conversely *E.coli* SSB wraps 65nt of DNA around a homotetrameric arrangement of OB folds by utilising longer DNA-binding loops that may also be involved in the formation of dimers within the tetramer and cooperative ssDNA-binding. The wealth of structural and biochemical data available for *Sso*SSB strongly suggests that this protein is structurally divergent from the bacterial homologues and achieves ssDNA-binding by a wholly different mechanism.

Future work

The structural analysis, backed up by biochemical data clearly show the expected ssDNA-binding mode of *Sulfolobus solfataricus* SSB. An SSB-ssDNA complex is essential to confirm the binding site size (4-5nt) and interactions involved in DNA recognition. The complex is being pursued by Andy McEwan in our lab. Initial screening with a 6-mer of cytosine (in 2-3 fold molar excess of SSB) has yielded crystals in the native, optimised conditions. It remains to be seen whether these crystals contain DNA, although it seems unlikely as the high ionic strength of sulphate in the screen can be expected to out-compete ssDNA for binding to the positively charged cleft. Screening should be carried out using a variety of lengths of ssDNA and crystal trials with a 4-mer are underway.

It was not possible to crystallise the full-length protein in this study most likely due to the flexible nature of the C-terminus. More extensive screening is required.

The identification of specific protein partners for SSB in DNA metabolism is an interesting question that remains to be answered. The large quantity of *Sso*SSB isolated from *Sulfolobus* cells suggests that this protein has many different roles. This is the ongoing work of our collaborators in the lab of Malcolm White, University of St.Andrews. Crystallographic analysis of protein-protein interactions poses a significant challenge but in this case can be facilitated by using the *Sso*SSB coordinates as a search model in molecular replacement.

The recent discovery of a novel human SSB in the lab of Malcolm White is an interesting discovery and the purity of the overexpressed protein is now sufficient to begin crystal trials

References

Abraham, E. P. and Chain, E. (1940). "An Enzyme from Bacteria Able To Destroy Penicillin." Nature **146**: 837.

Allard, S. T., Beis, K., Giraud, M. F., Hegeman, A. D., Gross, J. W., Wilmouth, R. C., Whitfield, C., Graninger, M., Messner, P., Allen, A. G., Maskell, D. J. and

Naismith, J. H. (2002). "Toward a structural understanding of the dehydratase mechanism." Structure (Camb) **10**(1): 81-92.

Allard, S. T., Giraud, M. F., Whitfield, C., Graninger, M., Messner, P. and Naismith, J. H. (2001). "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**(1): 283-95.

Allen, P. N. and Noller, H. F. (1989). "Mutations in ribosomal proteins S4 and S12 influence the higher order structure of 16 S ribosomal RNA." J Mol Biol **208**(3): 457-68.

Allsop, A. E. (1998). "New antibiotic discovery, novel screens, novel targets and impact of microbial genomics." Curr Opin Microbiol **1**(5): 530-4.

Allsopp, A. (1969). "Phylogenetic relationships of the Procaryota and the origin of the Eucaryotic Cell." New Phytol **68**: 591-612.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-402.

Augustin, M. A., Huber, R. and Kaiser, J. T. (2001). "Crystal structure of a DNA-dependent RNA polymerase (DNA primase)." Nat Struct Biol **8**(1): 57-61.

Baldock, C., Rafferty, J. B., Stuitje, A. R., Slabas, A. R. and Rice, D. W. (1998). "The X-ray structure of Escherichia coli enoyl reductase with bound NAD+ at 2.1 A resolution." J Mol Biol **284**(5): 1529-46.

Baluch, J., Chase, J. W. and Sussman, R. (1980). "Synthesis of recA protein and induction of bacteriophage lambda in single-strand deoxyribonucleic acid-binding protein mutants of Escherichia coli." J Bacteriol **144**(2): 489-98.

Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P. B. and Steitz, T. A. (1999). "Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit." Nature **400**(6747): 841-7.

Bastin-Shanower, S. A. and Brill, S. J. (2001). "Functional analysis of the four DNA binding domains of replication protein A. The role of RPA2 in ssDNA binding." J Biol Chem **276**(39): 36446-53.

Benach, J., Atrian, S., Gonzalez-Duarte, R. and Ladenstein, R. (1998). "The refined crystal structure of Drosophila lebanonensis alcohol dehydrogenase at 1.9 A resolution." J Mol Biol **282**(2): 383-99.

Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P. C., Nicolas, A. and Forterre, P. (1997). "An atypical topoisomerase II from Archaea with implications for meiotic recombination." Nature **386**(6623): 414-7.

Bergerat, A., Gadelle, D. and Forterre, P. (1994). "Purification of a DNA topoisomerase II from the hyperthermophilic archaeon Sulfolobus shibatae. A thermostable enzyme with both bacterial and eucaryal features." J Biol Chem **269**(44): 27663-9.

Bilgin, N., Richter, A. A., Ehrenberg, M., Dahlberg, A. E. and Kurland, C. G. (1990). "Ribosomal RNA and protein mutants resistant to spectinomycin." Embo J **9**(3): 735-9.

Birnboim, H. C. and Doly, J. (1979). "A rapid alkaline lysis procedure for screening recombinant plasmid DNA." Nucleic Acids Res **7**: 21512-21522.

Blackwell, L. J. and Borowiec, J. A. (1994). "Human replication protein A binds single-stranded DNA in two distinct complexes." Mol Cell Biol **14**(6): 3993-4001.

Blackwell, L. J., Borowiec, J. A. and Masrangelo, I. A. (1996). "Single-stranded-DNA binding alters human replication protein A structure and facilitates interaction with DNA-dependent protein kinase." Mol Cell Biol **16**(9): 4798-807.

Blankenfeldt, W., Kerr, I. D., Giraud, M. F., McMiken, H. J., Leonard, G., Whitfield, C., Messner, P., Graninger, M. and Naismith, J. H. (2002). "Variation on a Theme of SDR. dTDP-6-Deoxy-L- lyxo-4-Hexulose Reductase (RmlD) Shows a New Mg(2+)-Dependent Dimerization Mode." Structure (Camb) **10**(6): 773-86.

Bochkarev, A., Bochkareva, E., Frappier, L. and Edwards, A. M. (1999). "The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding." Embo J **18**(16): 4498-504.

Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. and Frappier, L. (1997). "Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA." Nature **385**(6612): 176-81.

Bochkareva, E., Belegu, V., Korolev, S. and Bochkarev, A. (2001). "Structure of the major single-stranded DNA-binding domain of replication protein A suggests a dynamic mechanism for DNA binding." Embo J **20**(3): 612-8.

Bochkareva, E., Frappier, L., Edwards, A. M. and Bochkarev, A. (1998). "The RPA32 subunit of human replication protein A contains a single-stranded DNA-binding domain." J Biol Chem **273**(7): 3932-6.

Bochkareva, E., Korolev, S. and Bochkarev, A. (2000). "The role for zinc in replication protein A." J Biol Chem **275**(35): 27332-8.

Bochkareva, E., Korolev, S., Lees-Miller, S. P. and Bochkarev, A. (2002). "Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA." Embo J **21**(7): 1855-63.

Borges-Walmsley, M. I. and Walmsley, A. R. (2001). "The structure and function of drug pumps." Trends Microbiol **9**(2): 71-9.

Breton, R., Housset, D., Mazza, C. and Fontecilla-Camps, J. C. (1996). "The structure of a complex of human 17beta-hydroxysteroid dehydrogenase with estradiol and NADP+ identifies two principal targets for the design of inhibitors." Structure **4**(8): 905-15.

Briggs, C. E. and Fratamico, P. M. (1999). "Molecular characterization of an antibiotic resistance gene cluster of Salmonella typhimurium DT104." Antimicrob Agents Chemother **43**(4): 846-9.

Brown, J. R. and Doolittle, W. F. (1995). "Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications." Proc Natl Acad Sci U S A **92**(7): 2441-5.

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

Brush, G. S., Anderson, C. W. and Kelly, T. J. (1994). "The DNA-activated protein kinase is required for the phosphorylation of replication protein A during simian virus 40 DNA replication." Proc Natl Acad Sci U S A **91**(26): 12520-4.

Brush, G. S., Morrow, D. M., Hieter, P. and Kelly, T. J. (1996). "The ATM homologue MEC1 is required for phosphorylation of replication protein A in yeast." Proc Natl Acad Sci U S A **93**(26): 15075-80.

Bujalowski, W. and Lohman, T. M. (1986). "Escherichia coli single-strand binding protein forms multiple, distinct complexes with single-stranded DNA." Biochemistry **25**(24): 7799-802.

Burns, S. M. and Hull, S. I. (1998). "Comparison of loss of serum resistance by defined lipopolysaccharide mutants and an acapsular mutant of uropathogenic Escherichia coli O75:K5." Infect Immun **66**(9): 4244-53.

CCP4 (2000). "Summaries of recent discussions on ccp4bb." CCP4 NEWSLETTER ON PROTEIN CRYSTALLOGRAPHY **38**.

Chandra, N., Acharya, K. R. and Moody, P. C. (1999). "Analysis and characterization of data from twinned crystals." Acta Crystallogr D Biol Crystallogr **55 (Pt 10)**: 1750-8.

Chatterjee, D. (1997). "The mycobacterial cell wall: structure, biosynthesis and sites of drug action." Curr Opin Chem Biol **1**(4): 579-88.

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**(13): 3342-6.

Chen, Z., Jiang, J. C., Lin, Z.-G., Lee, W. R., Baker, M. E. and Chang, S. H. (1993). "Site-specific mutagenesis of glycine-14 and two criticcal cysteinyl residues in Drosophila alcohol." Biochemistry **29**(5): 1112-1118.

Chiang, S. L. and Mekalanos, J. J. (1999). "rfb mutations in Vibrio cholerae do not affect surface production of toxin-coregulated pili but still inhibit intestinal colonization." Infect Immun **67**(2): 976-80.

Chittum, H. S. and Champney, W. S. (1994). "Ribosomal protein gene sequence changes in erythromycin-resistant mutants of Escherichia coli." J Bacteriol **176**(20): 6192-8.

Chrysogelos, S. and Griffith, J. (1982). "Escherichia coli single-strand binding protein organizes single-stranded DNA in nucleosome-like units." Proc Natl Acad Sci U S A **79**(19): 5803-7.

Cleaver, J. E. (1968). "Defective repair replication of DNA in xeroderma pigmentosum." Nature **218**(142): 652-6.

Cleland, W. W. (2000). "Low-barrier hydrogen bonds and enzymatic catalysis." Arch Biochem Biophys **382**(1): 1-5.

Cleland, W. W., Frey, P. A. and Gerlt, J. A. (1998). "The low barrier hydrogen bond in enzymatic catalysis." J Biol Chem **273**(40): 25529-32.

Copeland, H. F. (1939). "The kingdoms of organisms." Q.Rev.Biol **13**: 383-420.

Copeland, W. C. and Tan, X. (1995). "Active site mapping of the catalytic mouse primase subunit by alanine scanning mutagenesis." J Biol Chem **270**(8): 3905-13.

Coster, D., Knol, K. S. and Prins, J. A. (1933). "'Unterschiede in der Intensität der Röntgenstrahlenreflexion zu den beiden 111-Flächen der Zinkblende'." Z Phys **63**(345-69).

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

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. and McPherson, A. (1994). "Screening and optimization strategies for macromolecular crystal growth." Acta Crystallogr D Biol Crystallogr **50**(4): 414-423.

Dabrowski, S., Olszewski, M., Piatek, R., Brillowska-Dabrowska, A., Konopa, G. and Kur, J. (2002). "Identification and characterization of single-stranded-DNA-binding proteins from Thermus thermophilus and Thermus aquaticus - new arrangement of binding domains." Microbiology **148**(Pt 10): 3307-15.

Dale, J. (1994). Molecular genetics of bacteria, Wiley.

Diaz, R. L., Alcid, A. D., Berger, J. M. and Keeney, S. (2002). "Identification of residues in yeast Spo11p critical for meiotic DNA double-strand break formation." Mol Cell Biol **22**(4): 1106-15.

Drenth, J. (1999). Principles of protein x-ray crystallography. New York, Springer.

Dye. C, W. B. G., Espinal M. A., Raviglione M. C. (2002). "Erasing the World's Slow Stain: Strategies to Beat Multidrug-Resistant Tuberculosis." Science **295**: 2042-2046.

Edgell, D. R. and Doolittle, W. F. (1997). "Archaea and the origin(s) of DNA replication proteins." Cell **89**(7): 995-8.

Evans, P. R. (1997). ""SCALA"." Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography **33**: 22-24.

Finkel, T. and Holbrook, N. J. (2000). "Oxidants, oxidative stress and the biology of ageing." Nature **408**(6809): 239-47.

Fisher, R. G. and Sweet, R. M. (1980). "Treatment of diffraction data from protein crystals twinned by merohedry." Acta Crystallogr B **36**: 755-60.

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

Fox, G. E., Magrum, L. J., Balch, W. E., Wolfe, R. S. and Woese, C. R. (1977). "Classification of methanogenic bacteria by 16S ribosomal RNA chararterization." Proc Natl Acad Sci U S A **74**(10): 4537-4541.

Friedberg, E. C. (2001). "How nucleotide excision repair protects against cancer." Nat Rev Cancer **1**(1): 22-33.

Friedel, G. (1913). "Sur les symétries cristallines que peut révéler la diffraction des rayons Röntgen." Comptes Rendus Acad Sci (Paris) **157**: 1533-36.

Gallardo, F., Ruiz, J., Marco, F., Towner, K. J. and Vila, J. (1999). "Increase in incidence of resistance to ampicillin, chloramphenicol and trimethoprim in clinical isolates of Salmonella serotype Typhimurium with investigation of molecular epidemiology and mechanisms of resistance." J Med Microbiol **48**(4): 367-74.

Genschel, J., Curth, U. and Urbanke, C. (2000). "Interaction of E. coli single-stranded DNA binding protein (SSB) with exonuclease I. The carboxy-terminus of SSB is the recognition site for the nuclease." Biol Chem **381**(3): 183-92.

Ghosh, D., Sawicki, M., Pletnev, V., Erman, M., Ohno, S., Nakajin, S. and Duax, W. L. (2001). "Porcine carbonyl reductase. structural basis for a functional monomer in short chain dehydrogenases/reductases." J Biol Chem **276**(21): 18457-63.

Ghosh, D., Wawrzak, Z., Weeks, C. M., Duax, W. L. and Erman, M. (1994). "The refined three-dimensional structure of 3 alpha,20 beta-hydroxysteroid dehydrogenase and possible roles of the residues conserved in short-chain dehydrogenases." Structure **2**(7): 629-40.

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 3 alpha,20 beta-hydroxysteroid dehydrogenase: a member of a short-chain dehydrogenase family." Proc Natl Acad Sci U S A **88**(22): 10064-8.

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 (Pt 12)**: 2043-6.

Giraud, M. F. and Naismith, J. H. (2000). "The rhamnose pathway." Curr Opin Struct Biol **10**(6): 687-96.

Glaser, L. and Kornfeld, S. (1961). "The enzymatic synthesis of thymidine linked sugars. II. Thymidine Diphosphate L-Rhamnose." J Biol Chem **236**(6): 1795-1799.

Glassberg, J., Meyer, R. R. and Kornberg, A. (1979). "Mutant single-strand binding protein of Escherichia coli: genetic and physiological characterization." J Bacteriol **140**(1): 14-9.

Goldstein, J. L. and Brown, M. S. (1990). "Regulation of the mevalonate pathway." Nature **343**(6257): 425-30.

Gomes, X. V. and Wold, M. S. (1996). "Functional domains of the 70-kilodalton subunit of human replication protein A." Biochemistry **35**(32): 10558-68.

Grainger, C. T. (1969). "Pseudo-merohedral twinning: the treatment of overlapped data." Acta Crystallogr A **25**(427-34).

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**(35): 25069-77.

Griffith, J. D., Harris, L. D. and Register, J., 3rd (1984). "Visualization of SSB-ssDNA complexes active in the assembly of stable RecA-DNA filaments." Cold Spring Harb Symp Quant Biol **49**: 553-9.

Guillemot, D. (1999). "Antibiotic use in humans and bacterial resistance." Curr Opin Microbiol **2**(5): 494-8.

Gutierrez, C., Martin, G., Sogo, J. M. and Salas, M. (1991). "Mechanism of stimulation of DNA replication by bacteriophage phi 29 single-stranded DNA-binding protein p5." J Biol Chem **266**(4): 2104-11.

Guzder, S. N., Habraken, Y., Sung, P., Prakash, L. and Prakash, S. (1995). "Reconstitution of yeast nucleotide excision repair with purified Rad proteins, replication protein A, and transcription factor TFIIH." J Biol Chem **270**(22): 12973-6.

Haber, J. E. (1997). "A super new twist on the initiation of meiotic recombination." Cell **89**(2): 163-6.

Haeckel, E. (1866). Generelle Morphologie der Organismen. Berlin, Reimer.

Hanahan, D. (1983). "Studies on transformation of E.coli with plasmids." J Mol Biol **166**(4): 557-580.

Haseltine, C. A. and Kowalczykoski, S. C. (2002). "A distinctive single-strand DNA-binding protein from the Archaeon Sulfolobus solfataricus." Mol Microbiol **43**(6): 1505-15.

He, Z., Henricksen, L. A., Wold, M. S. and Ingles, C. J. (1995). "RPA involvement in the damage-recognition and incision steps of nucleotide excision repair." Nature **374**(6522): 566-9.

Hedges, S. B. (2002). "The origin and evolution of model organisms." Nat Rev Genet **3**(11): 838-49.

Hendrickson, W. A. (1991). "Determination of macromolecular structures from anomalous diffraction of synchrotron radiation." Science **254**(5028): 51-8.

Henricksen, L. A., Umbricht, C. B. and Wold, M. S. (1994). "Recombinant replication protein A: expression, complex formation, and functional characterization." J Biol Chem **269**(15): 11121-32.

Hol, W. G., van Duijnen, P. T. and Berendsen, H. J. (1978). "The alpha-helix dipole and the properties of proteins." Nature **273**(5662): 443-6.

Holm, L. and Sander, C. (1993). "Protein structure comparison by alignment of distance matrices." J Mol Biol **233**(1): 123-38.

Hori, H. and Osawa, S. (1979). "Evolutionary change in 5S RNA secondary structure and a phylogenic tree of 54 5S RNA species." Proc Natl Acad Sci U S A **76**(1): 381-5.

Hori, H. and Osawa, S. (1987). "Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences." Mol Biol Evol **4**(5): 445-72.

Howell, P. L. and Smith, G. D. (1992). "Identification of heavy-atom derivatives by normal probability methods." J Appl crystallogr **25**: 81-86.

Huet, J., Sentenac, A. and Fromageot, P. (1982). "Spot-immunodetection of conserved determinants in eukaryotic RNA polymerases. Study with antibodies to yeast RNA polymerases subunits." J Biol Chem **257**(5): 2613-8.

Ito, N., Komiyama, N. H. and Fermi, G. (1995). "Structure of deoxyhaemoglobin of the antarctic fish Pagothenia bernacchii with an analysis of the structural basis of the root effect by comparison of the liganded and unliganded haemoglobin structures." J Mol Biol **250**(5): 648-58.

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. and Miyata, T. (1989). "Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes." Proc Natl Acad Sci U S A **86**(23): 9355-9.

Jacobs, D. M., Lipton, A. S., Isern, N. G., Daughdrill, G. W., Lowry, D. F., Gomes, X. and Wold, M. S. (1999). "Human replication protein A: global fold of the N-terminal RPA-70 domain reveals a basic cleft and flexible C-terminal linker." J Biomol NMR **14**(4): 321-31.

Jameson, G. B. (1982). "On structure refinement using data from a twinned crystal." Acta Crystallogr A **38**: 817-820.

Jancarik, J. and Kim, S. H. (1991). "Sparse matrix sampling - a screening method for crystallisation of proteins." J Appl crystallogr **24**(4): 409-411.

Janeway, C. A., Travers, P., Walport, M. and Capra, J. D. (1999). Immunobiology - the immune system in health and disease., Elsevier science/Garland publishing.

Jarvis, F. G. and Johnson, M. J. (1949). "A glyco-lipid produced by Pseudomonas aeruginosa." J Am Chem Soc **71**: 4124-4126.

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 serovar typhimurium (strain LT2)." Mol Microbiol **5**(3): 695-713.

Jomma, H., Weisner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenhaler, H. K., Soldati, D. and Beck, E. (1999). "Inhibitors of the non-mevalonate pathway of isoprenoid biosynthesis as antimalarial drugs." Science **285**: 1573-1576.

Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard (1991). "Improved methods for building protein models in electron density maps and the location of errors in these models." Acta Crystallogr A **47 (Pt 2)**: 110-9.

Jornvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J. and Ghosh, D. (1995). "Short-chain dehydrogenases/reductases (SDR)." Biochemistry **34**(18): 6003-13.

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**(7): 4226-30.

Kakinuma, Y., Igarashi, K., Konishi, K. and Yamato, I. (1991). "Primary structure of the alpha-subunit of vacuolar-type Na(+)-ATPase in Enterococcus hirae. Amplification of a 1000-bp fragment by polymerase chain reaction." FEBS Lett **292**(1-2): 64-8.

Kandler, O. and Hippe, H. (1977). "Lack of peptidoglycan in the cell walls of Methanosarcina barkeri." Arch Microbiol **113**(1-2): 57-60.

Keeney, S., Giroux, C. N. and Kleckner, N. (1997). "Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family." Cell **88**(3): 375-84.

Keeney, S. and Kleckner, N. (1995). "Covalent protein-DNA complexes at the 5' strand termini of meiosis-specific double-strand breaks in yeast." Proc Natl Acad Sci U S A **92**(24): 11274-8.

Kelly, J. A., Dideberg, O., Charlier, P., Wery, J. P., Libert, M., Moews, P. C., Knox, J. R., Duez, C., Fraipont, C., Joris, B. and et al. (1986). "On the origin of bacterial resistance to penicillin: comparison of a beta-lactamase and a penicillin target." Science **231**(4744): 1429-31.

Kelly, T. J., Simancek, P. and Brush, G. S. (1998). "Identification and characterization of a single-stranded DNA-binding protein from the archaeon Methanococcus jannaschii." Proc Natl Acad Sci U S A **95**(25): 14634-9.

Kelman, Z., Pietrokovski, S. and Hurwitz, J. (1999). "Isolation and characterization of a split B-type DNA polymerase from the archaeon Methanobacterium thermoautotrophicum deltaH." J Biol Chem **274**(40): 28751-61.

Kenny, M. K., Schlegel, U., Furneaux, H. and Hurwitz, J. (1990). "The role of human single-stranded DNA binding protein and its individual subunits in simian virus 40 DNA replication." J Biol Chem **265**(13): 7693-700.

Kerr, I. D., Wadsworth, R. I., Blankenfeldt, W., Staines, A. G., White, M. F. and Naismith, J. H. (2001). "Overexpression, purification, crystallization and data collection of a single-stranded DNA-binding protein from Sulfolobus solfataricus." Acta Crystallogr D Biol Crystallogr **57**(Pt 9): 1290-2.

Kim, C. and Wold, M. S. (1995). "Recombinant human replication protein A binds to polynucleotides with low cooperativity." Biochemistry **34**(6): 2058-64.

Kirk, B. W. and Kuchta, R. D. (1999). "Arg304 of human DNA primase is a key contributor to catalysis and NTP binding: primase and the family X polymerases share significant sequence homology." Biochemistry **38**(24): 7727-36.

Kleywegt, G. J. (1996). "Use of non-crystallographic symmetry in protein structure refinement." Acta Cryst **D52**: 842-857.

Kleywegt, G. J. and Jones, T. A. (1996). "xdlMAPMAN and xdlDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets." Acta Crystallogr D Biol Crystallogr **52**: 826-828.

Kornfeld, S. and Glaser, L. (1961). "The enzymatic synthesis of thymidine-linked sugars. I Thymidine Diphosphate Glucose." J Biol Chem **236**(6): 1791-1794.

Kotra, L. P., Haddad, J. and Mobashery, S. (2000). "Aminoglycosides: perspectives on mechanisms of action and resistance and strategies to counter resistance." Antimicrob Agents Chemother **44**(12): 3249-56.

Krook, M., Marekov, L. and Jornvall, H. (1990). "Purification and structural characterization of placental NAD(+)-linked 15-hydroxyprostaglandin dehydrogenase. The primary structure reveals the enzyme to belong to the short-chain alcohol dehydrogenase family." Biochemistry **29**(3): 738-43.

Krulwich, T. A., Jin, J., Guffanti, A. A. and Bechhofer, H. (2001). "Functions of tetracycline efflux proteins that do not involve tetracycline." J Mol Microbiol Biotechnol **3**(2): 237-46.

Kushner, S. R. (1978). An improved method for transformation of Escherichia coli with ColE1 derived plasmids. Gen Eng, Elsevier. **17-23**.

Kuzuyama, T., Takahashi, S., Takagi, M. and Seto, H. (2000). "Characterization of 1-deoxy-D-xylulose 5-phosphate reductoisomerase, an enzyme involved in isopentenyl diphosphate biosynthesis, and identification of its catalytic amino acid residues." J Biol Chem **275**(26): 19928-32.

Labesse, G., Vidal-Cros, A., Chomilier, J., Gaudry, M. and Mornon, J. P. (1994). "Structural comparisons lead to the definition of a new superfamily of NAD(P)(H)-accepting oxidoreductases: the single-domain reductases/epimerases/dehydrogenases (the 'RED' family)." Biochem J **304 (Pt 1)**: 95-9.

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." Acta Crystallogr A **26**(Pt 2): 283-291.

Leslie, A. G. W. (1992). "Recent changes to the MOSFLM package for processing film and image plate data." Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography **26**.

Levy, C. W., Baldock, C., Wallace, A. J., Sedelnikova, S., Viner, R. C., Clough, J. M., Stuitje, A. R., Slabas, A. R., Rice, D. W. and Rafferty, J. B. (2001). "A study of the structure-activity relationship for diazaborine inhibition of Escherichia coli enoyl-ACP reductase." J Mol Biol **309**(1): 171-80.

Li, L., Lu, X., Peterson, C. A. and Legerski, R. J. (1995). "An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair." Mol Cell Biol **15**(10): 5396-402.

Lindahl, T. (1993). "Instability and decay of the primary structure of DNA." Nature **362**(6422): 709-15.

Lindahl, T. and Wood, R. D. (1999). "Quality control by DNA repair." Science **286**(5446): 1897-905.

Liu, J., Wu, T. C. and Lichten, M. (1995). "The location and structure of double-strand DNA breaks induced during yeast meiosis: evidence for a covalently linked DNA-protein intermediate." Embo J **14**(18): 4599-608.

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**(35): 10675-84.

Lohman, T. M. and Overman, L. B. (1985). "Two binding modes in Escherichia coli single strand binding protein-single stranded DNA complexes. Modulation by NaCl concentration." J Biol Chem **260**(6): 3594-603.

Luecke, H., Richter, H. T. and Lanyi, J. K. (1998). "Proton transfer pathways in bacteriorhodopsin at 2.3 angstrom resolution." Science **280**(5371): 1934-7.

Ma, Y., Pan, F. and McNeil, M. (2002). "Formation of dTDP-rhamnose is essential for growth of mycobacteria." J Bacteriol **184**(12): 3392-3395.

Madigan, M. T., Martinko, J. M. and Parker, J. (1997). Biology of Microorganisms. London, Prentice Hall.

Markham, P. N. and Neyfakh, A. A. (2001). "Efflux-mediated drug resistance in Gram-positive bacteria." Curr Opin Microbiol **4**(5): 509-14.

Massova, I. and Mobashery, S. (1998). "Kinship and diversification of bacterial penicillin-binding proteins and beta-lactamases." Antimicrob Agents Chemother **42**(1): 1-17.

Matsunaga, T., Park, C. H., Bessho, T., Mu, D. and Sancar, A. (1996). "Replication protein A confers structure-specific endonuclease activities to the XPF-ERCC1 and XPG subunits of human DNA repair excision nuclease." J Biol Chem **271**(19): 11047-50.

Matthews, B. W. (1968). "Solvent content of protein crystals." J Mol Biol **33**: 491-497.

McClelland, M. and Nelson, M. (1992). "Effect of site-specific methylation on DNA modification, methyltransferases and restriction endonulcleases." Nucleic Acids Res **20**: 2145-2157.

McMurry, L. M., Oethinger, M. and Levy, S. B. (1998). "Triclosan targets lipid synthesis." Nature **394**(6693): 531-2.

McNeil, M., Daffe, M. and Brennan, P. J. (1990). "Evidence for the nature of the link between the arabinogalactan and peptidoglycan components of mycobacterial cell walls." J Biol Chem **265**: 18200-18206.

McPherson, A. (1982). Preparation and analysis of protein crystals. New York, John Wiley.

Melo G., G. L. (1965). "The nucleotide specificity and feedback control of thymidine diphosphate d-glucose pyrophosphorylase." J Biol Chem **240**: 398-405.

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

Mims, C. A., Playfair, J. H. L., Roitt, I. M., Wakelin, D. and Williams, R. (1993). Medical Microbiology, Mosby.

Mitchison M., B. D. M., Vinh T., Rajakumur K., Faine S., Adler B. (1997). "Identification and characterisation of the dTDP-rhamnose biosynthesis and transfer of the lipoploysaccharide-ralted locus in Leptospira interrogans serovar copenhageni." J Bacteriol **179**(4): 1262-1267.

Morris, A., Kellner, J. D. and Low, D. E. (1998). "The superbugs: evolution, dissemination and fitness." Curr Opin Microbiol **1**(5): 524-9.

Morrison, D. C., Kline, F.L (1977). "Activation of the classical and properdin-pathways of complement by bacterial lipopolysaccharides (LPS)." J Immunol **118**: 362-368.

Mu, D., Park, C. H., Matsunaga, T., Hsu, D. S., Reardon, J. T. and Sancar, A. (1995). "Reconstitution of human DNA repair excision nuclease in a highly defined system." J Biol Chem **270**(6): 2415-8.

Murray, R. G. E. (1968). "Microbial structure as an aid to microbial classification and taxonomy." Spicy Prirodoved Fak. Univ. J. E. Purkyne Brne **43**: 249-252.

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

Murzin, A. G. (1993). "OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences." Embo J **12**(3): 861-7.

Nagelhus, T. A., Haug, T., Singh, K. K., Keshav, K. F., Skorpen, F., Otterlei, M., Bharati, S., Lindmo, T., Benichou, S., Benarous, R. and Krokan, H. E. (1997). "A sequence in the N-terminal region of human uracil-DNA glycosylase with homology to XPA interacts with the C-terminal part of the 34-kDa subunit of replication protein A." J Biol Chem **272**(10): 6561-6.

Neidhardt, F. C., Curtis III, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M. and Umbarger, H. E. (1996). Escherichia coli and Salmonella: cellular and molecular biology, 2nd ed. Washington, D.C, American Society for Microbiology.

Nichols, M. D., DeAngelis, K., Keck, J. L. and Berger, J. M. (1999). "Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11." Embo J **18**(21): 6177-88.

Obeid, J. and White, P. C. (1992). "Tyr-179 and Lys-183 are essential for enzymatic activity of 11 beta-hydroxysteroid dehydrogenase." Biochem Biophys Res Commun **188**(1): 222-7.

Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Persson, B. and Jornvall, H. (2003). "Short-chain dehydrogenases/reductases (SDR): the 2002 update." Chem Biol Interact **143-144**: 247-53.

Otwinowski, Z. and Minor, W. (1996). "Processing of X-ray diffraction data collected in oscillation mode." Methods Enzymol **276**: 307-326.

Patrick, S. M. and Turchi, J. J. (2002). "Xeroderma pigmentosum complementation group A protein (XPA) modulates RPA-DNA interactions via enhanced complex stability and inhibition of strand separation activity." J Biol Chem **277**(18): 16096-101.

Peltoketo, H., Isomaa, V., Maentausta, O. and Vihko, R. (1988). "Complete amino acid sequence of human placental 17 beta-hydroxysteroid dehydrogenase deduced from cDNA." FEBS Lett **239**(1): 73-7.

Pereira, P. J., Macedo-Ribeiro, S., Parraga, A., Perez-Luque, R., Cunningham, O., Darcy, K., Mantle, T. J. and Coll, M. (2001). "Structure of human biliverdin IXbeta reductase, an early fetal bilirubin IXbeta producing enzyme." Nat Struct Biol **8**(3): 215-20.

Perrakis, A., Morris, R. and Lamzin, V. S. (1999). "Automated protein model building combined with iterative structure refinement." Nat Struct Biol **6**(5): 458-63.

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

Philipova, D., Mullen, J. R., Maniar, H. S., Lu, J., Gu, C. and Brill, S. J. (1996). "A hierarchy of SSB protomers in replication protein A." Genes Dev **10**(17): 2222-33.

Centres for Disease Control and prevention (1997). "Interim guidelines for prevention and control of Staphylococcal infection associated with reduced susceptibility to vancomycin." MMWR Morb Mortal Wkly Rep . **46**: 626-628, 635.

Centres for Disease Control and prevention (1997). "Reduced susceptibility of Staphylococcus aureus to vancomycin-Japan, 1996." MMWR Morb Mortal Wkly Rep . **46**: 624-626.

Puhler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H. P., Lottspeich, F., Garrett, R. A. and Zillig, W. (1989). "Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome." Proc Natl Acad Sci U S A **86**(12): 4569-73.

Raghunathan, S., Kozlov, A. G., Lohman, T. M. and Waksman, G. (2000). "Structure of the DNA binding domain of E. coli SSB bound to ssDNA." Nat Struct Biol 7(8): 648-52.

Raghunathan, S., Ricard, C. S., Lohman, T. M. and Waksman, G. (1997). "Crystal structure of the homo-tetrameric DNA binding domain of Escherichia coli single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-A resolution." Proc Natl Acad Sci U S A **94**(13): 6652-7.

Rahim, R., Burrows, L. L., Monteiro, M. A., Perry, M. B. and Lam, J. S. (2000). "Involvement of the rml locus in core oligosaccharide and O polysaccharide assembly in Pseudomonas aeruginosa." Microbiology **146 (Pt 11)**: 2803-14.

Rees, D. C. (1980). "The influence of Twinning by Merohedry on Intensity Statistics." Acta Crystallogr A **36**: 578-581.

Rees, D. C. (1982). "A general theory of x-ray intensity statistics for twins by merohedry." Acta Crystallogr A **38**: 201-207.

Rees, D. C. and Lipscomb, W. N. (1980). "Structure of the potato inhibitor complex of carboxypeptidase A at 2.5-A resolution." Proc Natl Acad Sci U S A **77**(8): 4633-7.

Reeves, P. (1993). "Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale." Trends Genet **9**(1): 17-22.

Reusch, V. M., Jr. (1984). "Lipopolymers, isoprenoids, and the assembly of the gram-positive cell wall." Crit. Rev. Microbiol **11**: 129-155.

Reuter, K., Sanderbrand, S., Jomaa, H., Wiesner, J., Steinbrecher, I., Beck, E., Hintz, M., Klebe, G. and Stubbs, M. T. (2002). "Crystal structure of 1-deoxy-D-xylulose-5-phosphate reductoisomerase, a crucial enzyme in the non-mevalonate pathway of isoprenoid biosynthesis." J Biol Chem **277**(7): 5378-84.

Reynolds, R. A., Remington, S. J., Weaver, L. H., Fisher, R. G., Anderson, W. F., Ammon, H. L. and Matthews, B. W. (1985). "Structure of serine protease from rat mast cells determined from twinned crystals by isomorphous replacement and molecular replacement." Acta Crystallogr B **41**(139-47).

Richard, S. B., Bowman, M. E., Kwiatkowski, W., Kang, I., Chow, C., Lillo, A. M., Cane, D. E. and Noel, J. P. (2001). "Structure of 4-diphosphocytidyl-2-C-methylerythritol synthetase involved in mevalonate- independent isoprenoid biosynthesis." Nat Struct Biol **8**(7): 641-8.

Rocchetta, H. L., Burrows, L. L. and Lam, J. S. (1999). "Genetics of O-antigen biosynthesis in Pseudomonas aeruginosa." Microbiol Mol Biol Rev **63**(3).

Rossmann, M. G., Moras, D. and Olsen, K. W. (1974). "Chemical and biological evolution of nucleotide-binding protein." Nature **250**(463): 194-9.

Roujeinikova, A., Levy, C. W., Rowsell, S., Sedelnikova, S., Baker, P. J., Minshull, C. A., Mistry, A., Colls, J. G., Camble, R., Stuitje, A. R., Slabas, A. R., Rafferty, J. B., Pauptit, R. A., Viner, R. and Rice, D. W. (1999). "Crystallographic analysis of triclosan bound to enoyl reductase." J Mol Biol **294**(2): 527-35.

Rowlands, T., Baumann, P. and Jackson, S. P. (1994). "The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria." Science **264**(5163): 1326-9.

Rudolph, M. G., Kelker, M. S., Schneider, T. R., Yeates, T. O., Oseroff, V., Heidary, D. K., Jennings, P. A. and Wilson, I. A. (2003). "Use of multiple anomalous dispersion to phase highly merohedrally twinned crystals of interleukin-1beta." Acta Crystallogr D Biol Crystallogr **59**(Pt 2): 290-8.

Sambrook, j., Fritsch, E. J. and Maniatis, T. (1989). Molecular cloning : a laboratory manual, Cold Spring Harbor Laboratory Press.

Sancar, A. (1996). "DNA excision repair." Annu Rev Biochem **65**: 43-81.

Sancar, A., Williams, K. R., Chase, J. W. and Rupp, W. D. (1981). "Sequences of the ssb gene and protein." Proc Natl Acad Sci U S A **78**(7): 4274-8.

Sandigursky, M., Mendez, F., Bases, R. E., Matsumoto, T. and Franklin, W. A. (1996). "Protein-protein interactions between the Escherichia coli single-stranded DNA-binding protein and exonuclease I." Radiat Res **145**(5): 619-23.

Schnabel, R., Sonnenbichler, J. and Zillig, W. (1982). "Stimulation by silybin, a eukaryotic feature of archaebacterial RNA polymerases." Febs Lett **150**(2): 400-402.

Schnaitman C. A., K. J. D. (1993). "Genetics of Lipolysaccharide biosynthesis in enteric bacteria." Microbiol Rev **57**(3): 655-682.

Schneider, T. R. and Sheldrick, G. M. (2002). "Substructure solution with SHELXD." Acta Crystallogr D Biol Crystallogr **58**(Pt 10 Pt 2): 1772-9.

Sheldrick, G. M. and Schneider, T. R. (1997). "SHELXL: High-resolution refinement." Methods Enzymol **277**: 319-43.

Spurgeon, S. L. and Porter, J. W. (1981). Biosynthesis of isoprenoid compounds. New York, John Wiley & Sons, Inc.

Stammers, D. K., Ren, J., Leslie, K., Nichols, C. E., Lamb, H. K., Cocklin, S., Dodds, A. and Hawkins, A. R. (2001). "The structure of the negative transcriptional regulator NmrA reveals a structural superfamily which includes the short-chain dehydrogenase/reductases." Embo J **20**(23): 6619-26.

Stanier, R. Y. and van Neil, C. B. (1962). "The concept of a bacterium." Arch Microbiol **42**: 17-35.

Steitz, T. A., Smerdon, S. J., Jager, J. and Joyce, C. M. (1994). "A unified polymerase mechanism for nonhomologous DNA and RNA polymerases." Science **266**(5193): 2022-5.

Stewart, M. J., Parikh, S., Xiao, G., Tonge, P. J. and Kisker, C. (1999). "Structural basis and mechanism of enoyl reductase inhibition by triclosan." J Mol Biol **290**(4): 859-65.

Stinson, M. W., Nisengard, R. J. and Bergey, E. J. (1980). "Binding of Streptococcal antigens to muscle tissue in-vitro." Infect Immun **27**: 604-613.

Stryer, L. (1996). Biochemistry. New York, Freeman.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1990). "Use of T7 RNA polymerase to direct expression of cloned genes." Methods Enzymol **185**: 60-89.

Sutcliffe, J., Tait-Kamradt, A. and Wondrack, L. (1996). "Streptococcus pneumoniae and Streptococcus pyogenes resistant to macrolides but sensitive to clindamycin: a common resistance pattern mediated by an efflux system." Antimicrob Agents Chemother **40**(8): 1817-24.

Szabo, M., Bronner, D. and Whitfield, C. (1995). "Relationships between rfb gene clusters required for biosynthesis of identical D-galactose-containing O antigens in Klebsiella pneumoniae serotype O1 and Serratia marcescens serotype O16." J Bacteriol **177**(6): 1544-53.

Taylor, H. O. and Leslie, A. G. W. (1998). "A Program to Detwin Merohedrally Twinned Data." CCP4 newsletter **35**: 9.

Taylor, P. W. (1983). "Bactericidal and bacteriolytic activity of serum against gram-negative bacteria." Microbiol Rev **47**(1): 46-83.

Terwilliger, T. C. (2000). "Maximum-likelihood density modification." Acta Crystallogr D Biol Crystallogr **56 (Pt 8)**: 965-72.

Terwilliger, T. C. and Berendzen, J. (1999). "Automated MAD and MIR structure solution." Acta Crystallogr D Biol Crystallogr **55 (Pt 4)**: 849-61.

Terwisscha van Scheltinga, A. C., Valegard, K., Ramaswamy, S., Hajdu, J. and Andersson, I. (2001). "Multiple isomorphous replacement on merohedral twins: structure determination of deacetoxycephalosporin C synthase." Acta Crystallogr D Biol Crystallogr **57**(Pt 12): 1776-85.

Thatcher, D. R. and Sawyer, L. (1980). "Secondary-structure prediction from the sequence of Drosophila melanogaster (fruitfly) alcohol dehydrogenase." Biochem J **187**(3): 884-6.

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**(19): 5691-701.

Treuner, K., Findeisen, M., Strausfeld, U. and Knippers, R. (1999). "Phosphorylation of replication protein A middle subunit (RPA32) leads to a disassembly of the RPA heterotrimer." J Biol Chem **274**(22): 15556-61.

Tsukioka, Y., Yamashita, Y., Nakano, Y., Oho, T. and Koga, T. (1997). "Identification of a fourth gene involved in dTDP-rhamnose synthesis in Streptococcus mutans." J Bacteriol **179**(13): 4411-4.

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**(4): 1126-34.

Tsutsumi, S., Denda, K., Yokoyama, K., Oshima, T., Date, T. and Yoshida, M. (1991). "Molecular cloning of genes encoding major two subunits of a eubacterial V-type ATPase from Thermus thermophilus." Biochim Biophys Acta **1098**(1): 13-20.

Umbricht, C. B., Erdile, L. F., Jabs, E. W. and Kelly, T. J. (1993). "Cloning, overexpression, and genomic mapping of the 14-kDa subunit of human replication protein A." J Biol Chem **268**(9): 6131-8.

Vales, L. D., Chase, J. W. and Murphy, J. B. (1980). "Effect of ssbA1 and lexC113 mutations on lambda prophage induction, bacteriophage growth, and cell survival." J Bacteriol **143**(2): 887-96.

Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. and Clardy, J. (1993). "Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin." J Mol Biol **229**(1): 105-24.

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**(12): 5582-6.

Vriend, G. (1990). "WHAT IF: a molecular modeling and drug design program." J Mol Graph **8**(1): 52-6, 29.

Wadsworth, R. I. and White, M. F. (2001). "Identification and properties of the crenarchaeal single-stranded DNA binding protein from Sulfolobus solfataricus." Nucleic Acids Res **29**(4): 914-20.

Walther, A. P., Bjerke, M. P. and Wold, M. S. (1999). "A novel assay for examining the molecular reactions at the eukaryotic replication fork: activities of replication protein A required during elongation." Nucleic Acids Res **27**(2): 656-64.

Weiner, J. H., Bertsch, L. L. and Kornberg, A. (1975). "The deoxyribonucleic acid unwinding protein of Escherichia coli. Properties and functions in replication." J Biol Chem **250**(6): 1972-80.

Wettach, J., Gohl, H. P., Tschochner, H. and Thomm, M. (1995). "Functional interaction of yeast and human TATA-binding proteins with an archaeal RNA polymerase and promoter." Proc Natl Acad Sci U S A **92**(2): 472-6.

Whittaker, R. H. (1959). "On the Broad Classification of Organisms." Q.Rev.Biol **43**: 210-226.

WHO (2001). The World Health report.

WHO (2001). Worlh Health Organisation global strategy for containment of antimicrobial resistance.

WHO/CDS (2002). Scaling up the response to infectious to infectioous diseases. Report on Infectioous diseases 2002.

Wierenga, R. K., De Maeyer, C. H. and Hol, W. G. J. (1985). "Interaction of pyrophophate moeities with alpha-helices in dinucleotide binding proteins." Biochemistry **24**: 1346-1357.

Wierenga, R. K. and Hol, W. G. (1983). "Predicted nucleotide-binding properties of p21 protein and its cancer-associated variant." Nature **302**(5911): 842-4.

Wilcox, S. K., Cavey, G. S. and Pearson, J. D. (2001). "Single ribosomal protein mutations in antibiotic-resistant bacteria analyzed by mass spectrometry." Antimicrob Agents Chemother **45**(11): 3046-55.

Williams, K. R., Murphy, J. B. and Chase, J. W. (1984). "Characterization of the structural and functional defect in the Escherichia coli single-stranded DNA binding protein encoded by the ssb-1 mutant gene. Expression of the ssb-1 gene under lambda pL regulation." J Biol Chem **259**(19): 11804-11.

Williams, K. R., Spicer, E. K., LoPresti, M. B., Guggenheimer, R. A. and Chase, J. W. (1983). "Limited proteolysis studies on the Escherichia coli single-stranded DNA binding protein. Evidence for a functionally homologous domain in both the Escherichia coli and T4 DNA binding proteins." J Biol Chem **258**(5): 3346-55.

Wilson, A. J. C. (1949). "The probability distribution of X-ray intensities." Acta Crystallogr D Biol Crystallogr **2**(5): 318-321.

Woese, C. R. and Fox, G. E. (1977). "Phylogenetic structure of the prokaryotic domain: the primary kingdoms." Proc Natl Acad Sci U S A **74**(11): 5088-90.

Woese, C. R., Kandler, O. and Wheelis, M. L. (1990). "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya." Proc Natl Acad Sci U S A **87**(12): 4576-9.

Wold, M. S. (1997). "Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism." Annu Rev Biochem **66**: 61-92.

Wolfe, R. S. (1972). "Microbial formation of methane." Adv Microb Physiol **6**: 107-146.

Yajima, S., Nonaka, T., Kuzuyama, T., Seto, H. and Ohsawa, K. (2002). "Crystal structure of 1-deoxy-D-xylulose 5-phosphate reductoisomerase complexed with cofactors: implications of a flexible loop movement upon substrate binding." J Biochem (Tokyo) **131**(3): 313-7.

Yamashita, Y., Tomihisa, K., Nakano, Y., Shimazaki, Y., Oho, T. and Koga, T. (1999). "Recombination between gtfB and gtfC is required for survival of a dTDP-rhamnose synthesis-deficient Mutant of Streptococcus mutans in the presence of Sucrose." Infect Immun **3693-3697**.

Yang, C., Curth, U., Urbanke, C. and Kang, C. (1997). "Crystal structure of human mitochondrial single-stranded DNA binding protein at 2.4 A resolution." Nat Struct Biol **4**(2): 153-7.

Yang, F., Dauter, Z. and Wlodawer, A. (2000). "Effects of crystal twinning on the ability to solve a macromolecular structure using multiwavelength anomalous diffraction." Acta Crystallogr D Biol Crystallogr **56 (Pt 8)**: 959-64.

Yang, F., Forrer, P., Dauter, Z., Conway, J. F., Cheng, N., Cerritelli, M. E., Steven, A. C., Pluckthun, A. and Wlodawer, A. (2000). "Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD." Nat Struct Biol 7(3): 230-7.

Yeates, T. O. (1997). "Detecting and overcoming crystal twinning." Methods Enzymol **276**: 344-58.

Yeates, T. O. and Fam, B. C. (1999). "Protein crystals and their evil twins." Structure Fold Des 7(2): R25-9.

Yeates, T. O. and Rees, D. C. (1987). "An isomorphous replacement method for phasing twinned structures." Acta Crystallogr A **43**: 30-6.

Zeikus, J. G. (1977). "The biology of methanogenic bacteria." Bacteriol Rev **41**(2): 514-41.

Zillig, W., Stetter, K. O., Prangishvilli, D., Schäfer, W., Wunderl, S., Janekovic, D., Holz, I. and Palm, P. (1982). "Desulfurococcaceae, the second family of the extremely thermophilic, anaerobic, sulfur-respiring Thermoproteales." Zbl Bakt Hyg., I Abt Orig. **C3**: 304-317.

Zillig, W., Stetter, K. O., Schäfer, W., Janekovic, D., Wunderl, S., Holz, I. and Palm, P. (1981). "Thermoproteales: a novel type of extremely thermoacidophilic anaerobic archaebacteria isolated from Icelandic solfataras." Zbl Bakt Hyg., I Abt Orig. **C2**: 205-227.