Structural and Biochemical Studies of Carbohydrate Binding Proteins.

Stephen Woods Breckenridge Fullerton.

A thesis submitted to the University of St-Andrews for the degree of Doctor of Philosophy in the Faculty of science.

Department of Chemistry,
University of St-Andrews.

September 2003

Chapter Four

Table of figures.

Abstract.

Protein / carbohydrate interactions are involved in innumerable biological processes throughout all known forms of life. The specificity of these interactions allow control of processes such as embryogenesis, fertilisation, sugar transportation, immunological response, cancer, infection, cell to cell signalling and targeting of host cell glycoproteins, making protein / carbohydrate interactions attractive therapeutic targets. The ubiquitous nature of complex carbohydrate structures presents a serious problem for rational drug design. Many of these complex carbohydrate molecules are presented on cell surfaces and interact with a host of carbohydrate binding proteins including, lectins, antibodies, cell receptor proteins, some enzymes and bacterial adhesion proteins, with varying degrees of specificity. A greater understanding of the physical nature of protein carbohydrate interactions may prove to be the key for future design of therapeutic agents. In the past two decades, X-ray crystallographic study of protein / carbohydrate interactions has began to elucidate these specific interactions at an atomic level, providing a powerful insight into the molecular processes involved in protein carbohydrate interactions.

This thesis discusses some of the common features of protein / carbohydrate interactions, in particular the stabilisation of protein carbohydrate interactions via hydrogen bonding, van der Waals forces, covalent intermediates, ion bridges and hydrophobic interactions.

I, Stephen Fullerton, hereby certify that this thesis, which is approximately 30,000 words in length, has been written by me, and that it is a record of work carried out by me and that it has not been submitted in any other previous application for a higher degree.

Signature Date 28/4/04

I was admitted as a research student in September 1999 and as a candidate for the higher degree of PhD in September 1999 at the University of St-Andrews; the higher study for which this is a record was carried out in the University of St-andrews between September 1999 and September 2003.

Signature Date 28/4/04

I hereby certify that the Candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of the PhD in the University of St-Andrews and that Candidate is qualifiedto submit this thesis in application for that degree.

Signature Date 5/5/04

In submitting this thesis to the University of St-Andrews I understand that I am giving permission for it to be made available for the 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 *bons fide* Library or research worker.

Signature Date 28/4/04

Acknowledgements.

I would like to thank all of my colleagues at St-Andrews University for their invaluable help and support throughout my PhD, not least of all the technical support staff without which the science would not have been as forthcoming as it was. A great deal of thanks must also go to Dr John Ingledew for his help with the EPR experiments and Professor Steve Chapman and Dr Simon Daff for their help and input in the redox potentiometry experiments. I would especially like to thank the members of the Taylor-Naismith group for their input, advice and occasional parties, which made my PhD all the more bearable. I would like to thank my supervisor Professor Jim Naismith for his support and his patience and understanding at a particularly difficult time. I would also like to thank all my friends and especially my family and extended family who throughout my years at St-Andrews kept me grounded and supported me through some of the toughest times, and made all the happier times so much more joyful.

I would especially like to thank my parents who supported me on this crazy science adventure when many people including myself were less optimistic, their love and support has made everything possible and it is with great gratitude that I dedicate this thesis to them.

Chapter one

Introduction to protein carbohydrate interactions.

1.00 Summary.

Protein-carbohydrate interactions are ubiquitous in biological systems, controlling a sophisticated array of functions including cell-cell signalling, immunological responses and gene expression. The importance of these controlling protein-carbohydrate interactions have made them attractive therapeutic targets, however the large number of variations in carbohydrate structures presents a huge challenge for the design of rational therapeutics. Understanding the physical basis of protein-carbohydrate interactions requires an atomic level description of the contacts between the protein and the carbohydrate ligand. X-ray crystallography in combination with biochemical binding assays has begun to provide an insight into protein-carbohydrate interactions. This chapter will summarise the current information available from biochemical analysis and X-ray structures, identifying common features of protein-carbohydrate interactions.

1.01 Introduction.

Proteins that recognise carbohydrates are almost as ubiquitous as the complex carbohydrate molecules with which they interact. These proteins perform a diverse variety of biological functions including cell signalling, defence and metabolism (Lis, 1986). They interact with a number of carbohydrates with varying degrees of specificity and affinity. Protein-carbohydrate interactions such as cell to cell signalling and recognition are essential for the proper control required to manage homeostatic responses within higher organisms, such as glucose-glycogen control in diabetes. The failure of some of these interactions has led to expensive medical problems for both human and other higher mammals, including diabetes (Christesen, 2002), Andersen's disease (Guerra, 1986), Mc Ardle's disease (Bartram, 1995) and mutations in the human Mannose binding lectin are associated with increased incidence of general infection and further associations with autoimmune disease and rheumatoid arthritis (Wallis, 1999).

1.02 Protein-Carbohydrate interactions: putative therapeutic targets.

With the ubiquitous nature of carbohydrates throughout all known forms of life, the metabolism of essential pathogen specific carbohydrates and carbohydrate adducts provides potential therapeutic drug targets. As increasing strains of resistant bacteria emerge, it is important that new treatments are continually produced to keep bacterial infection under control. Currently the two main strategies employed in antibacterial action are inhibition of replication and transcription of the bacterial genome, or the disruption / destruction of the bacterial cell membrane, peptidoglycan layer or cell wall.

Drug therapies such as fluoroquinolones, microlide antibiotics, rifamycin antibiotics streptomycin antibiotics and capreomycin, all target the bacterial genetic replication, transcription and translation machinery (Gale, 1981; Winder, 1982; Blanchard, 1996; Pan,

1996; Russel, 1996). Antibacterial drugs such as β-lactams (Mims, 1998), Isoniazid, Ethambutol and cycloserines, target the cell wall or membrane surrounding bacteria. Isoniazid, Ethambutol, and Cycloserines are used to combat *Mycobacterium tuberculosis* (*M. tuberculosis*) (Chopra, 1998), acting upon the production of complex carbohydrates and lipids involved in the production of the bacterial cell membrane and surrounding peptidoglycan cell wall. If we are to continue to treat bacterial infections we must continue the search for new antibacterial agents. In all the afore-mentioned treatments there are drug resistant strains of bacteria and increasingly, multi-drug therapies have become necessary to control bacterial infections (Chopra, 1998).

1.03 Origins of resistance.

Bacteria have been around for approximately 3 billion years. In that time bacteria have encountered innumerable evolutionary challenges including early natural antibiotics from competing microbes and environmental pressures such as temperature, radiation, chemical toxins, high salt, acidic and basic conditions. In any bacterial colony (approximately 10^7 cells) a spontaneous mutation in the bacterial chromosome can occur in every three generations (Walsh, 2000). Therefore the ability of a particular bacterial strain to produce mutants conferring survival characteristics is high, considering that most bacterial population's double in size every 15-30 minutes. Bacteria are excellent at adapting to further challenges, both natural and synthetic. For every antibiotic produced in the last 50 years there is at least one strain of resistant bacteria (Medeiros, 1997).

1.04 Multi-drug resistance.

The emergence of new multi-drug resistant pathogens continually challenges the ability of modern therapeutics to control the spread of disease throughout the human and animal

population. Bacteria have the ability to transfer genetic information allowing them to share resistance with other bacterial species. This transfer of genetic information can occur in three main ways: transduction, conjugation and transformation. Transduction involves viral transfer of genetic information between mating bacteria. Conjugation is a simpler transfer of DNA across the sex pilus from donor bacterium to the acceptor bacterium. Transformation is the direct acquisition and absorbtion of DNA from the surrounding environment. The ability of bacteria to pass antibiotic resistance across a spectrum of pathogens allows commonly abundant sources of bacteria such as *Escherichia coli* and *Staphylococcus aureas* to act as a reservoir for drug resistant genes, leading to multi-drug resistant bacteria such as methicillan resistant *Staphylococcus aureas* (MRSA), vancomycin resistant enterococci, (VRE) including *Enterococci faecium* and *Escherichia coli* 0157 (Walsh, 2000; Coates, 2002).

While production of new antibacterial compounds may be effective against current resistant bacteria, unless a change in the management of antibiotic prescription occurs, adaptation of bacteria to new evolutionary pressures will rapidly render new antibiotics ineffective.

1.05 Protein-Carbohydrate interactions.

Two thirds of the planet's carbon resources exist as carbohydrates (mainly as cellulose and hemicellulose). Carbohydrates are one of the most diverse groups of active biomolecules, with new carbohydrate structures and functions being elucidated and assigned continually. They are present in all known forms of life and play important roles in energy storage, recognition processes and structural frameworks, (cellulose and lignin in plants and chitin in insect exoskeletons). Structural roles involve mainly large chains of oligosaccharides to produce strong complex polymers whose functions include cell adhesion and immune response recognition (Mitchell, 2002). Non-structural carbohydrate roles can include simple monosaccharide carbohydrates and as well as larger, more complex oligsaccharides,

generally requiring the participation of an enzyme or carbohydrate binding protein. Proteins bind various specific carbohydrates from small three carbon molecules such as pyruvate to large complex oligosaccharides such as starch. The increased understanding of key protein-carbohydrate interactions may lead to new therapeutic treatments for inborn errors of metabolism such as McArdles disease, cancer treatments (Lis, 1986) and new antibiotic drug discoveries (Coates, 2002). Structural analysis of these interactions allow an insight into the molecular processes controlling the specificity and affinity of carbohydrate binding. The recent increase in protein-carbohydrate structures in the Brookhaven PDB (Protein Data Base) database (http://www.rcsb.org/pdb/), have supplied some general rules to carbohydrate recognition and binding. Although reversible protein-carbohydrate binding can occur covalently, aldolases (Koeller, 2001), transferases (Davies, 1997), most reversible molecular interactions in biochemical systems such as, binding of substrates to enzymes, and receptor protein-carbohydrate interactions, are predominately non-covalent.

Investigations into the biological information relayed, from initial inter-molecular protein-carbohydrate interaction to the desired response has established the study of carbohydrate interactions, with obvious medicinal and commercial benefits.

1.06 Non-covalent Protein-Carbohydrate Binding.

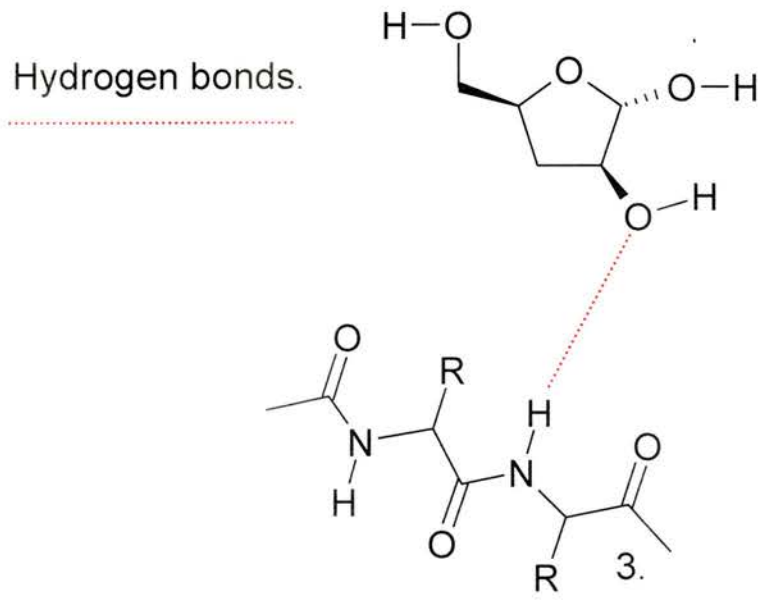
There are four different association forces: hydrophobic effects, hydrogen bonds, electrostatic bonds and van der Waals forces, which mediate the non-covalent binding of carbohydrates by proteins. These non-covalent forces differ in their geometry, strength and specificity, and can effect protein-carbohydrate interactions directly between active site residues and the carbohydrate or indirectly by incorporating bridging through metal ions and water molecules, further influencing the protein-carbohydrate interaction.

1.07 Hydrogen bonding.

Carbohydrate recognition and binding is predominantly controlled by hydrogen bonding from protein donor residues to specific sugar hydroxyl acceptors. The different stereogenic centers within different carbohydrates determine the relative positions of the acceptor hydroxyls and the protein donor residues, increasing or decreasing the number of hydrogen bonds and thereby increase or decrease binding affinity and specificity of the protein-ligand interaction. Hydrogen bonds occur when a hydrogen atom is shared between two other atoms. The atom to which the hydrogen is covalently linked is called the hydrogen donor, whereas the other atom is called the hydrogen acceptor. The acceptor atom possesses a partial negative charge, which attracts the partially positively charged hydrogen atom, and this is the basis of the interaction. The donor atom in a hydrogen bond is an oxygen atom or a nitrogen atom that has a covalently attached hydrogen atom. The acceptor atom can be either oxygen or nitrogen. The bond energies associated with these bonds range from 10 to 30 kJ/mole, so hydrogen bonding is much weaker than covalent or ionic bonds. Importantly hydrogen bonds are not as constricted as covalent bonds by bond lengths and geometry and can form a greater variety of interactions over wider or narrower bond angles, allowing a greater degree of acceptable contacts than covalent interactions. Hydrogen bonds play very important roles in protein, carbohydrate and nucleic acid structure as well as participating in protein-substrate binding mechanisms. Hydrogen bonding in protein-carbohydrate interactions can take several forms, direct, bidentate, cooperative and network hydrogen bonds have all been identified in high resolution protein-carbohydrate crystallographic structures.

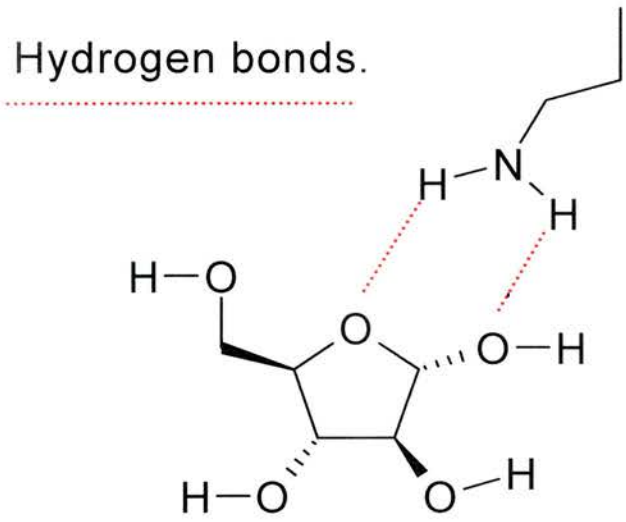
Direct hydrogen bonding occurs between the slightly positive charged proton of a hydrogen bond donator i.e. lysine or arginine and the slightly negatively charged acceptor atom i.e. a carbonyl oxygen or a water oxygen atom (Figure 1.01).

Figure 1.01 Direct hydrogen bond.



Bidentate hydrogen bonds form between two adjacent oxygens on the carbohydrate and an amine group on the protein. (Figure 1.02) The bidentate hydrogen bond is more stable than two monodentate interactions. The chelating effect (Page, 1971) is more stable as two hydrogen bonds are formed whilst only one side chain is immobilised, therefore the entropic penalty is much less.

Figure 1.02 Bidentate hydrogen bond.

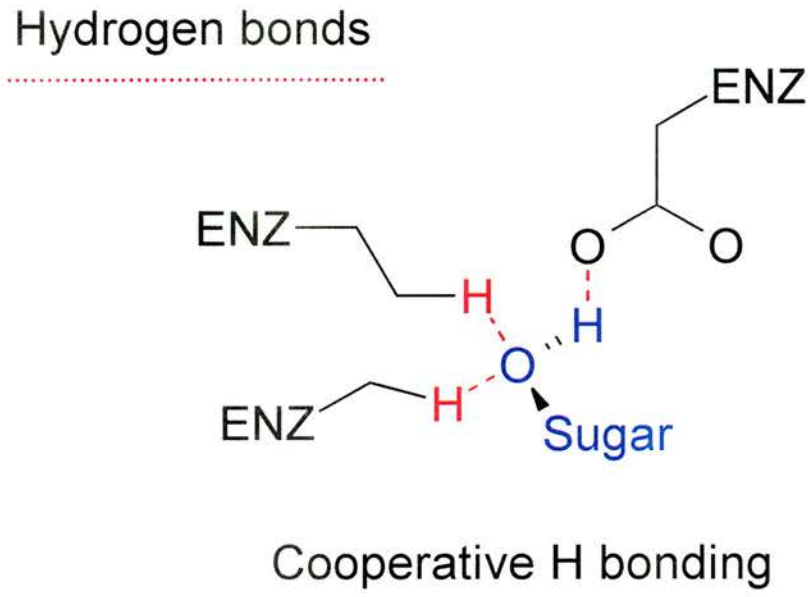


Cooperative hydrogen bonding occurs at the hydroxyl oxygen on the carbohydrate. The oxygen is Sp3 hybridised with two free lone pairs of electrons allowing a tetrahedral conformation (Figure 1.03).

Figure 1.03 Tetrahedral conformation of the hydroxyl oxygen.

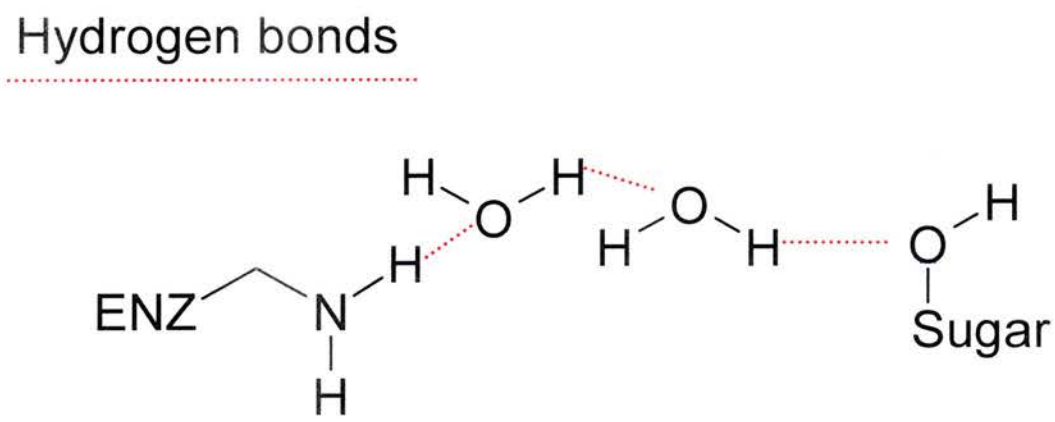
The free lone pairs of electrons on the Sp3 hybridised oxygen provide the necessary negative charge to allow Cooperative hydrogen bonding to occur. (Figure 1.04.) The torsional freedom of the proton and the two lone pairs of electrons on the hydroxyl group allow optimisation of hydrogen bonding, with a small entropic cost due to the fixing of the hydroxyl rotomer.

Figure 1.04 Cooperative hydrogen bonding.



Network hydrogen bonding contains interactions between several atoms, often involving water molecules. (Figure 1.05.)

Figure 1.05 Network hydrogen bonds.



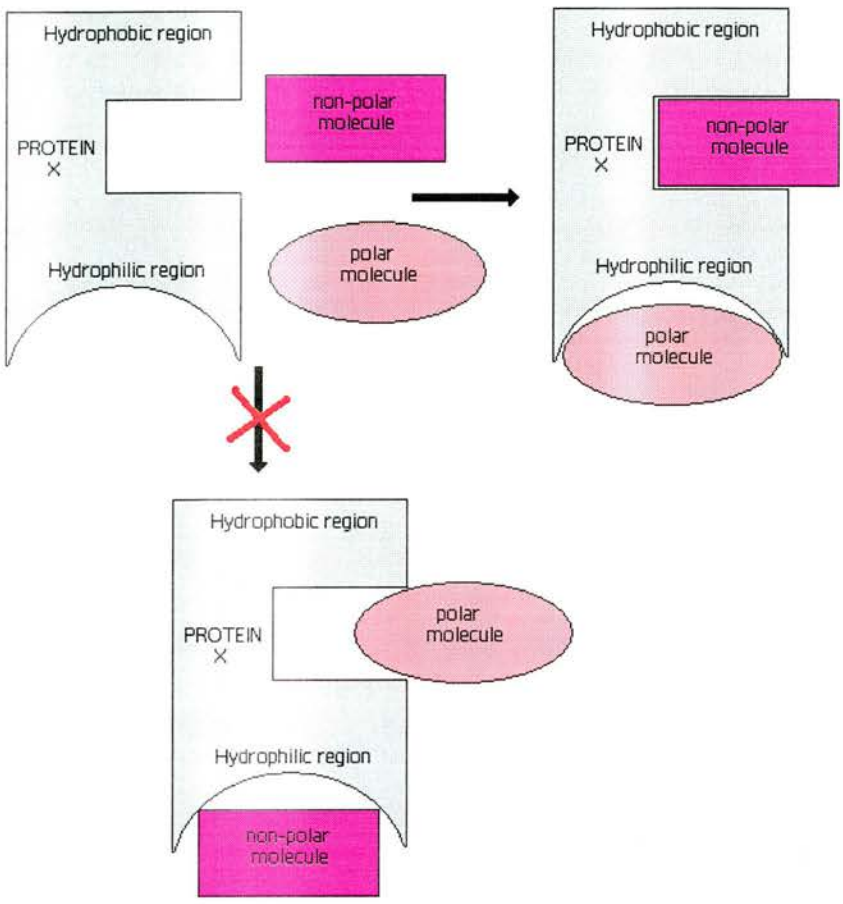
Water mediated hydrogen bonding between a protein and ligand can be as strong as a direct protein ligand interaction (Chervernak, 1994). It is therefore no surprise that proteins have made use of this abundant resource for increased binding stability. Conscription of water molecules to aid ligand binding has been reported in the structure of the isolectin I – trisaccharide Man α1-3Man β1-4GlcNAc co-complex (Bourne & Cambillau, 1990). Direct

contact with the carbohydrate ligand is present only at the monosaccharide binding site. The α1-3 terminal mannose sits within the monosaccharide binding site and the two remaining sugar moeties contact the protein through a series of hydrogen bond networks involving a total of twenty water molecules.

1.08 Hydrophobic effects.

The extent of hydrophobic effects is determined by the extent to which the substrate, protein or peptide involved is apolar. In protein folding, hydrophobic effects are the main driving force of the folding of the protein core. The hydrophobic effect is a result of the entropy-enthalpy effects on the order/disorder of the water in the system. Entropy is described as the measure of the disorder of a closed system, whereas enthalpy is the sum of the internal energy of a closed system. An exposed apolar side chain orders the surrounding water molecules decreasing the waters entropy and increasing the systems enthalpy. When these apolar side chains bury themselves in the centre of the protein structure the ordered water molecules are expelled to freely diffuse into the surrounding solvent, resulting in an increase in the waters entropy and a decrease in the systems enthalpy. The expulsion of water molecules from these apolar regions provides the positive entropy for protein core stability. This is termed the hydrophobic effect. As with protein folding the hydrophobic effect has varying influence on protein-protein, protein-ligand and protein-substrate interactions (Figure 1.06).

Figure 1.06 Hydrophobic interactions on binding, favourable and unfavourable.



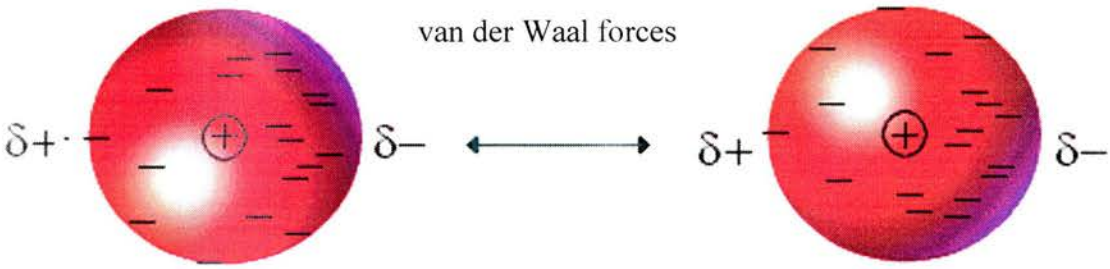
A protein with a large hydrophobic pocket binding a ligand with a complimentary hydrophobic region will have a similar effect on the surrounding water molecules, expelling the ordered waters around the hydrophobic areas of the ligand and binding pocket providing a positive entropy for ligand binding. The binding of the *Escherichia coli* chaperone protein GroEL with an unfolded protein in neutral solution has been studied thermodynamically by isothermal titration calorimetry (Lin, 1994). A single unfolded substilisin BPN molecule binds non-Cooperatively with the GroEL chaperone, exhibiting a positive enthalpy change. Protein-ligand and enzyme-substrate interactions can also be affected by the hydrophobic effects. Thyroid hormones, despite being derived from amino acids are inherently hydrophobic. It has been speculated that the hydrophobic nature of thyroid hormones is necessary for them to pass through the cell membrane and nuclear membrane before

association with the thyroid hormone receptor proteins, which can then directly bind DNA. In a recent rat hormone receptor-ligand complex (Wagner, 1995), the structure has shown that the hormone analogue is buried within the hydrophobic protein core indicating that a considerable force would have to be present to bury the ligand at the central binding site. In carbohydrate binding proteins the hydrophobic forces have a lesser effect. The binding sites of the proteins have invariably a hydrophobic pocket or residue, which stacks against the hydrophobic patch of a carbohydrate ligand. Specific orientations of carbohydrate non-polar CH groups can create a hydrophobic patch, which interacts with a hydrophobic pocket or residue at the receptor site on the protein, providing a favourable hydrophobic interaction for correct ligand orientation to facilitate optimum binding. However sugars are primarily hydrophilic in nature, reducing the hydrophobic binding effects. Once the hydrophobic forces have orientated the ligand, the ligand-protein interactions are governed by hydrogen bonds and van der Waals forces.

1.09 van der Waals Forces.

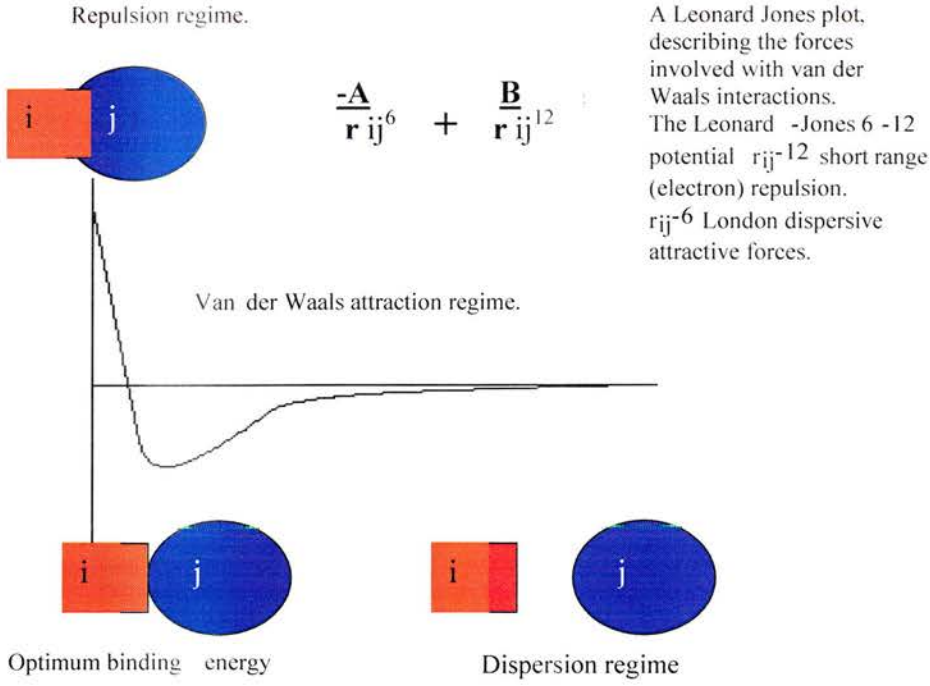
There is a non-specific attractive force between any two atoms when they are about three to four Å apart. This attractive force, called a van der Waals interaction, is weaker and less specific than hydrogen bonds. The force arises from the transient attraction between the nucleus of one atom and the electron cloud of the neighbour atom (Figure 1.07).

Figure 1.07 van der Waals forces.



If the interaction becomes too close then severe repulsion between the two nuclei of the atoms would occur. If the atoms move further apart the weak attractions will be broken (Figure 1.08). The van der Waals bond energy is about 2-4 kJ/mole and singly they contribute very little. van der Waals forces become significant in binding only when numerous bonds are formed e.g. in protein structure.

Figure 1.08 Leonard Jones Plot.



In enzymes van der Waals forces are significant when the substrate and active site of the enzyme are sterically complimentary to provide suitable points for interaction. Van der Waals forces can have considerable effects in solution between polar carbohydrate and water molecules.

1.10 Electrostatic bonding.

A charged group on a substrate can interact with an oppositely charged group on an enzyme by an electrostatic interaction. Electrostatic interactions are governed by Coulomb's law, which states that " the force that one particle exerts on another particle is directly proportional to the product of their charges and inversely proportional to the square of their separation". In the polyomavirus, the virus particle 1 protein structure has been determined to 1.9 Å (a recombinant pentamer), a lysine residue at the active site is positively charged and interacts with the negatively charged carboxylate group on the sialic acid component of the branched disialylated hexasaccharide receptor fragment (Stehle, 1997). The distance between these oppositely charged groups is 2.9 Å. This kind of interaction is called an ionic bond, salt linkage, salt bridge or ion pair, indicating an electrostatic interaction between oppositely charged groups. If a substrate is negatively charged it can form an electrostatic bond with the positively charged amino side groups of lysine, histidine or arginine residues in the protein. If the substrate is positively charged then the potential binding sites on the enzyme are the negatively charged carboxyl groups of aspartate and glutamate.

1.11 Lectins.

The study of carbohydrate binding in *vitro* has been well documented for over 100 years, with the first protein-carbohydrate interaction involving the plant lectin ricin, being recorded in 1888 (Stillmark, 1888). Due to their ability to bind specifically to a wide range of complex

carbohydrate structures, lectins, especially the legume lectins have proved to be a valuable scientific aid in elucidating protein-carbohydrate interactions and the study of immunology and glycobiology. Carbohydrate binding proteins although ubiquitous, can be separated into two distinct structural classes. Class I carbohydrate binding proteins, such as, maltose binding protein (Klein, 1991) and cyclodextrin glycosyltransferase (Hall, 1997) all encompass the bound carbohydrate with the active site buried within the centre of the protein molecule. Class II carbohydrate binding proteins have the carbohydrate binding site on a shallow cleft or indentation on the surface of the protein molecule, such as Con-A and peanut lectin (Naismith, 1996; Ravishankar, 1998). The lectin proteins are the most studied group of non-covalent carbohydrate binding proteins and are found in all types of organism from the influenza viral haemoglutinin lectin (Weis, 1988) (4HMG.pdb) to the human Ym1 TIM like lectin (Sun, 2001) (1E9L.pdb). Lectins bind carbohydrates with high specificity and they do so reversibly with generally no catalytic activity (Goldstein, 1986).

1.12 Legume lectins.

The legume lectins are the most highly studied family of carbohydrate binding proteins. The legume lectins are a highly conserved family of carbohydrate binding protein, found mostly in the seeds of legume plants. Although the legume lectin monomers contain a high level of primary sequence similarity and structural homology, they have widely variable carbohydrate specificities and structures. Con-A (1972; Hardman, 1972), pea (Einspahr, 1986), lentil (Loris, 1993, 1994) *Griffania simplicifolia* lectin 4 (Delbaere, 1993), soybean (Dessen, 1995), peanut (Banerjee, 1994, 1996), coral tree (Shaanan, 1991) and *lathyrus ochrus* (Bourne, 1990), have all recently had x-ray crystallographic structures published. Analysis of these structures reveal a conserved lectin, jelly-roll like fold. The lectin fold contains three β-sheets, two large anti-parallel β sheets, a seven stranded β-sheet (front sheet) and a six

stranded β-sheet (back sheet), primarily linked by a shorter five strand β-sheet (Banerjee, 1996). The two major sheets stack back-to-back sandwiching their hydrophobic residues at the core of the lectin monomer. These regions contain 50% of the total residues within the monomer. There are no α helices within legume lectin monomers. However some lectins contain small numbers of α helices, such as the C-type lectins (Poget, 1999). The residues contained in the linker-loop regions are structurally and functionally relevant, containing five conserved amino acid residues required for binding affinity, specificity and providing a second hydrophobic core for carbohydrate binding on the concave side of the front β-sheet.

1.13 Carbohydrate Interactions.

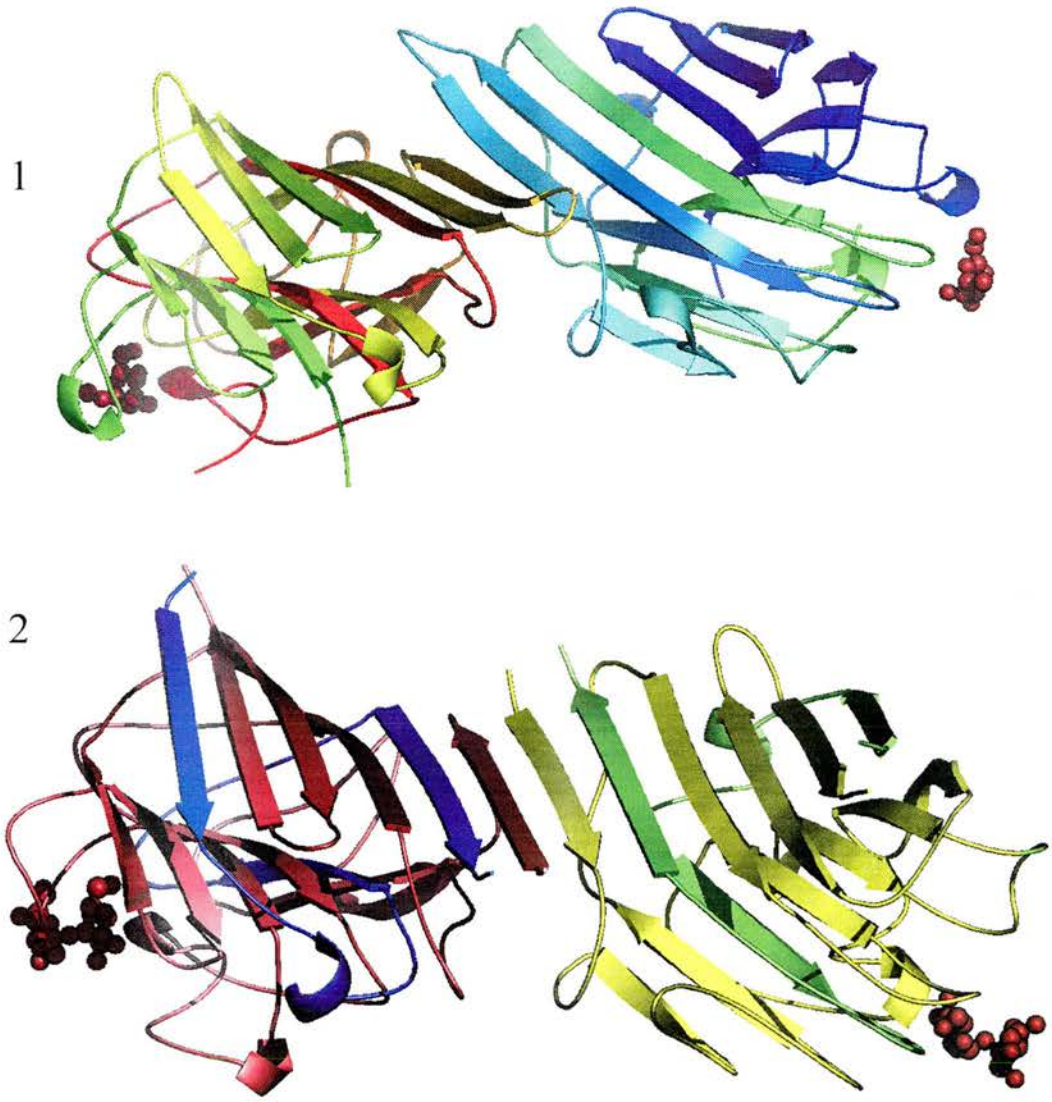
Carbohydrate interactions in legume lectins, consist of a primary monosaccharide binding site, with secondary binding sites situated outwith the monosaccharide binding site, to bind further oligosaccharide subunits. The lectin-binding site is produced by one of the linker loop regions folding over the front β-sheet forming a small carbohydrate binding pocket. The binding site contains four conserved residues involved in direct contact with the carbohydrate, and a fifth residue, which is situated upon the variable loop domain, associated with ligand specificity. Three of the four conserved binding residues, glycine, asparagine and aspartate, all hydrogen bond with the carbohydrate ligand. The fourth residue (either phenylalanine or tyrosine) is hydrophobic and stacks against the ligand's hydrophobic patch. These four conserved residues are held in place by two essential structural divalent cations, Ca^{2+} and a single transition metal ion. Con-A reversibly binds carbohydrate ligands only when the calcium binding S2 site is occupied. The presence of calcium is itself conditional of a second divalent ion binding event at the lectins S1 metal binding site which can bind a variety of divalent ions (Cd^{2+}, Co^{2+}, Mn^{2+}, Ni^{2+} or Zn^{2+}). The metal ions are only 4.2A apart and connected via two aspartate residues (Asp10 and Asp19). The Ca^{2+} ion at the S2 site

ligates the residues Asn14 and Asp 208 whose side chains bind the carbohydrate ligand directly (Asn14) or via a water molecule (Asp 208). The binding of the two divalent ions at the S1 and S2 binding sites induce large structural changes that pull the ligand binding residuesof the saccharide binding site into position and produce the carbohydrate binding of the lectin (Bouckaert,1996). Loss of these ions, results in the loss of carbohydrate binding. The fifth conserved residue confers binding specificity. This residue is found on the variable loop region defining the lectin's carbohydrate specificity. The variable loop region sterically hinders binding of non-specific carbohydrates by making it impossible for the carbohydrate to orientate itself towards the conserved binding residues (Derewenda, 1989; Harrop, 1996; Kanellopoulos, 1996).

1.14 Quaternary Structure.

Lectin quartenary structure is more diverse than would be expected, when considering primary sequence analysis. Lectin structures deposited to date form dimers or tetramers. Most lectins form a dimer known as the canonical legume lectin dimer. This dimer is characterised by a twelve–stranded anti-parallel β-sheet made up of the association of the two six-stranded monomer back sheets (Einspahr, 1986; Naismith, 1996) (Figure 1.09).

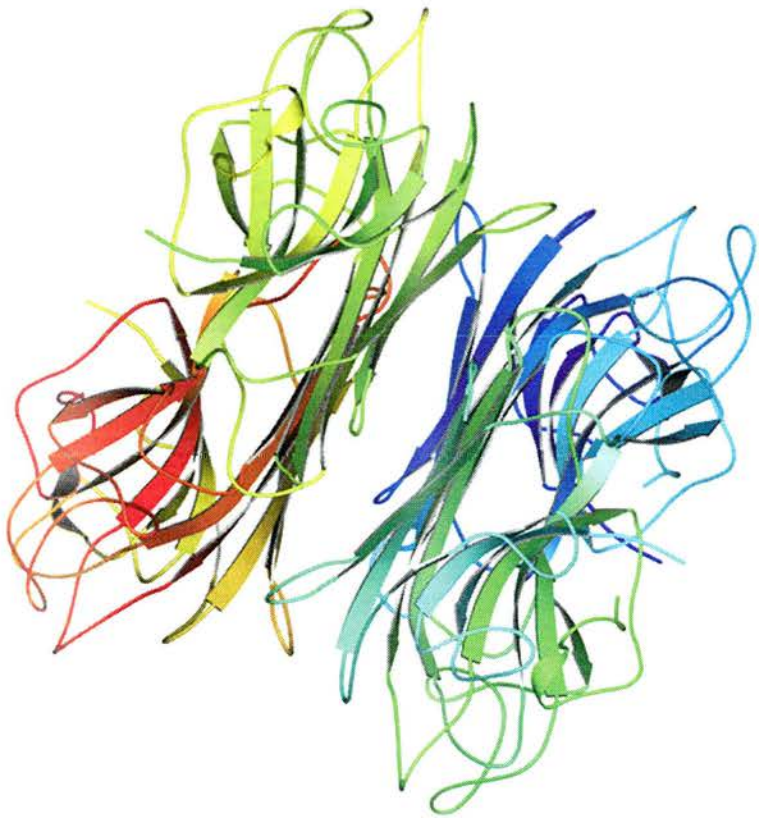
Figure 1.09 Con-A (1) and lentil lectin (2) dimers.



In the pea lectin and lentil lectin, the dimer interface is more hydrophilic in comparison with the Con-A dimer interface. In both pea and lentil lectins, the dimer-interface contains conserved water molecules. There are two structurally distinct lectins, which do not form the canonical legume lectin dimer, coral tree lectin and *Griffonia simplicifolica* lectin 4 (GS-4). Coral tree lectin cannot form the canonical dimer due to glycosylation at the residue Asp 17. The *Griffonia simplicifolica* lectin 4, forms a heterodimer of monomers which differ only in their post-translational glycosylation. The structure of the Con-A lectin has been solved only

as a tetrameric form. However the Con-A lectin quarternary structure is believed to be pH dependent, where protein incubated at pH 7.0 will give a tetrameric structure and protein incubated at pH 5.0 or lower will give dimeric and monomeric structures. The low pH dimer conformation is in the same canonical dimer form, however the ability to form the tetrameric structure at higher pH levels is believed to be the action of different protonation states of aspartates and glutamates at the dimer-dimer interface. Crystallographic data from the Con-A tetramer in native form and bound to mannose, α-D-methyl-mannopyrannoside and α-D-methyl-glucopyrannoside, have indicated small changes in conformation upon ligand binding. This suggests that the Con-A tetramer has some degree of flexibility in solution (Figure 1.10).

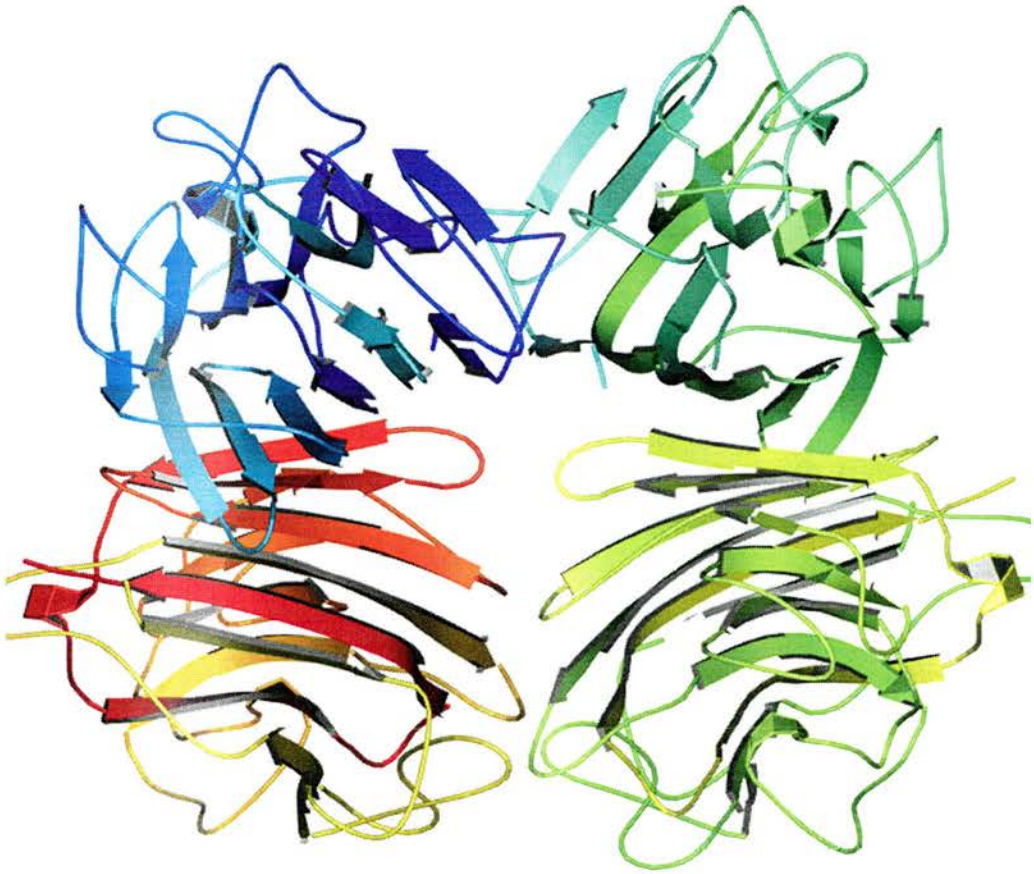
Figure 1.10 The Con-A tetramer.



The peanut lectin structure was determined in 1996 (Banerjee, 1996). It forms a tetrameric structure unlike any other homo-tetramer, where it contains neither 4-fold or 222-symmetry. The tetramer is formed from two GS-4 like dimers. A monomer from each dimer forms an

interface not unlike the canonical dimers. However no 12-stranded β-sheet is formed along the dimer interface, instead the N-terminal strands of the monomers are connected by 6 water bridges forming this unusual tetramer (Banerjee, 1996) (Figure 1.11).

Figure 1.11 The Peanut lectin tetramer.



Soybean agglutinin and *Phaseolus vulgaris* leuko-agglutanin lectins also have tetrameric (Dessen, 1995; Hamelryck T.W., 1996) structures made up of two classical canonical dimers. The soybean agglutinin and *Phaseolus vulgaris* leuko-agglutinin tetramers are however structurally distinct. Tetramer formation of soybean agglutanin and *Phaseolus vulgaris* leuko-agglutanin lectins is not pH dependent. The two 12-stranded β-sheets of the dimers associate with their outermost strands forming a large central void or pore in the centre of the

tetramer. This channel has no function as yet but may contain the disordered C-terminus of these lectins protecting them from proteolytic degradation (Loris, 1998).

1.15 Concanavalin A from *Canavalia ensiformis*.

Concanavalin-A (con-A) was first isolated from the Jack Bean in 1919 (Sumner, 1919). Since that time several con-A structures have been solved, the native structure was solved in 1972 (Edelman, 1972; Hardman, 1972), and the first monosaccharide bound structure was solved in 1976 (Hardman, 1976) to 6 Å resolution, characterising the first lectin-carbohydrate complex structure. A structure solved in 1989 (Derewenda, 1989) to 2.9 Å resolution described the protein-carbohydrate interactions in more detail. Further ligand complexes with methyl α-D-mannopyranoside (Naismith, 1994) and methyl α-D-glucopyranoside (Harrop, 1996) at higher resolutions increased the understanding of monosaccharide binding. The sugars are stabilised within the active site by several direct hydrogen bonds, van der Waals and polar contacts, burying 75% of the sugar's accessible surface area. In comparison with other carbohydrate binding proteins, L-arabinose binding protein (ABP), glucose / galactose binding protein (GGBP) and maltose binding protein (MBP) there is little discernable change in main chain conformation upon binding, allowing greater exposure of the carbohydrate to the surrounding solvent. This indicates that the full binding potential of the sugar is not realised (Quiocho, 1993).

Structures of Con-A have been solved containing larger complex carbohydrate ligands. These oligosaccharide structures (Loris, 1996; Naismith, 1996) revealed the specifics of the oligosaccharide binding site, with the oligosaccharide binding centered around the trimannoside core of biantennary N-linked glycans.

The 2.3 Å resolution structure (Naismith, 1996) of the complexed trimannoside showed that all three of the oligosaccharide residues were involved with direct interactions with the

protein, whereas a previous structure of the pea lectin trimannoside complex (Rini, 1993) had shown only direct contacts at the monosaccharide binding site. It is important to note that in both the monosaccharide and oligosaccharide structures the terminal sugar residue bound at the monosaccharide site has conserved direct contacts with the protein. Binding of the oligosaccharide (pentasaccharide Ka = 1.41 X 10^6 M^{-1}) exhibits an increased affinity for the lectin con-A over its monosaccharide counterpart (monosaccharide Ka = 0.82 X 104 M^{-1}) suggesting that the increased number of direct contacts and van der Waals contacts involved with the sugar rings and the extended binding site stabilise carbohydrate binding (Mandal 1994). Originally water meditated contacts where believed to aid ligand-protein interaction from waters recruited to the binding site to stabilise the interactions between sugars and proteins. However Naismith *et al* (Naismith, 1996) suggest only conserved waters already present within the binding site stabilise sugar binding.

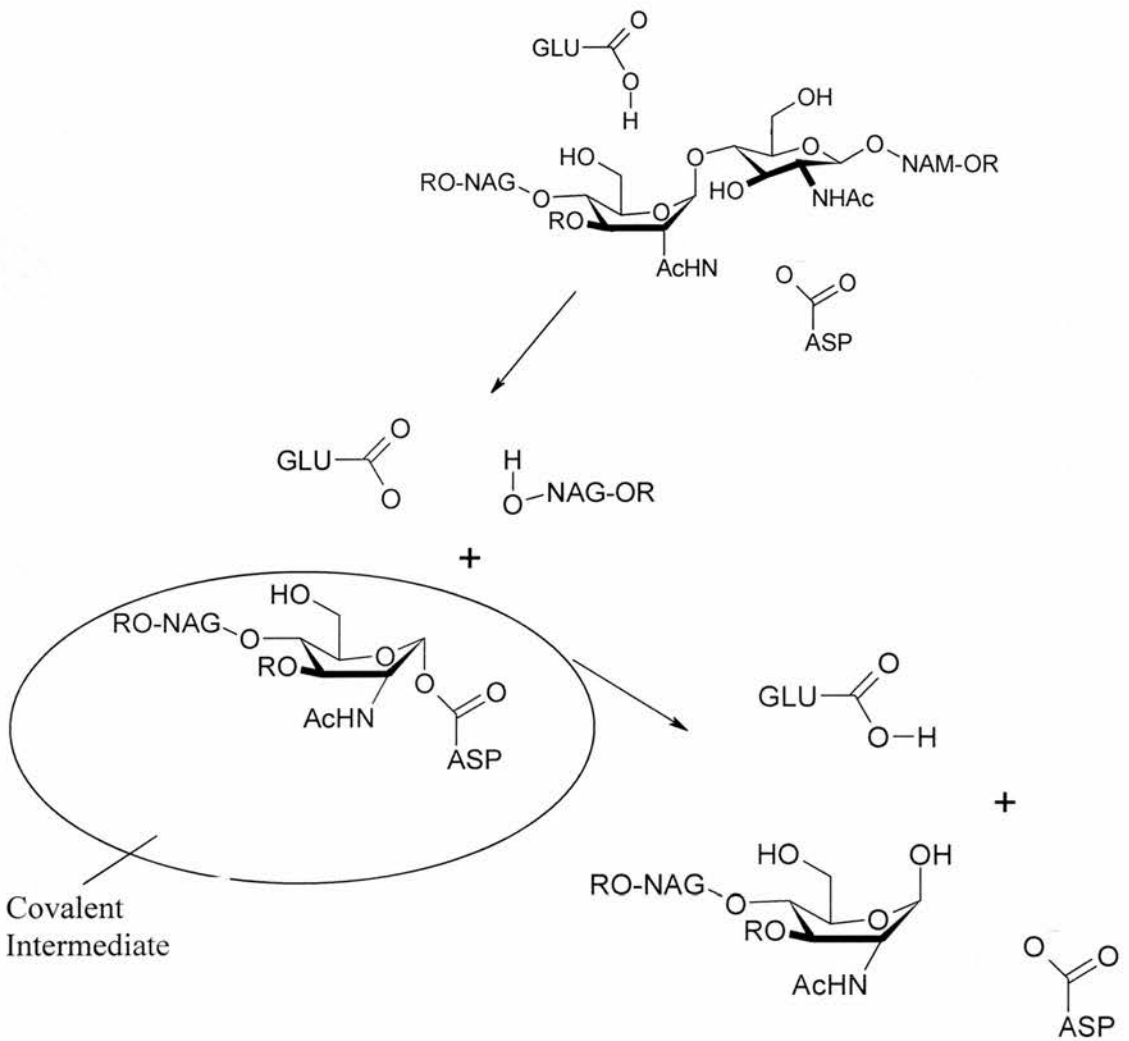
In the following chapter the crystalisation, structure solution and refinement of the 1.9Å resolution structure of the L-arabinose / con-A complex will be discussed.

1.16 Covalent Interactions.

Covalent intermediates can be found in proteases, lipases, lyases, hydrolases, transferases and glycosidases. The most abundant covalent protein-carbohydrate interactions concern the glycoproteins. Glyco-proteins are permanently glycosylated on the outer surface of the protein, for use in specific recognition in receptor mediated adhesion, biological signalling, etc. The vast range and versatility of biological glyco-proteins are too many to examine and discuss in this thesis, for an introductory review see Kobata 1992 and Grogan *et al* 2002 (Kobata, 1992; Grogan, 2002). This thesis will discuss the non-permanent interactions of carbohydrates and proteins at specific carbohydrate binding sites. Most proteins which bind carbohydrates covalently at the active site or carbohydrate binding sites are enzymes.

However there has been some speculation as to whether the animal lectin L-selectin forms a transient Schiff base covalent intermediates with the sialic acid ligand aldehyde group (Puri, 1996). The formation of covalent bonds by the Schiff base mechanism, is proposed by the authors in response to enhanced kinetics of L-selectin binding to the sialic acid ligand under reduction with cyanoborohydride. However there has been no further evidence to support these findings. The key enzyme-carbohydrate covalent intermediate is part of the general enzyme mechanism and once completed the carbohydrate is released. Hen egg white lysozyme, Class I aldolases and cyclodextrin glycosyl transferase enzymes have all recently had X-ray crystallographic structures solved with trapped covalent intermediates (Joost C. M., 1999; Allard, 2001; Heine, 2001; Vocadlo, 2001) . Each enzyme is proposed to form covalent intermediates with one of the two substrates / products. The hen egg white lysozyme first binds its oligosaccharide substrate in a cleft, six saccharide units long. The glycosidic bond is broken by protonation of the glycosidic oxygen by an active site glutamate. The resultant C1 oxocarbonium ion undergoes nucleophilic attack from an active site aspartate forming the covalent intermediate (Figure 1.12). The covalent intermediate is then hydrolysed, releasing the sugar substrate.

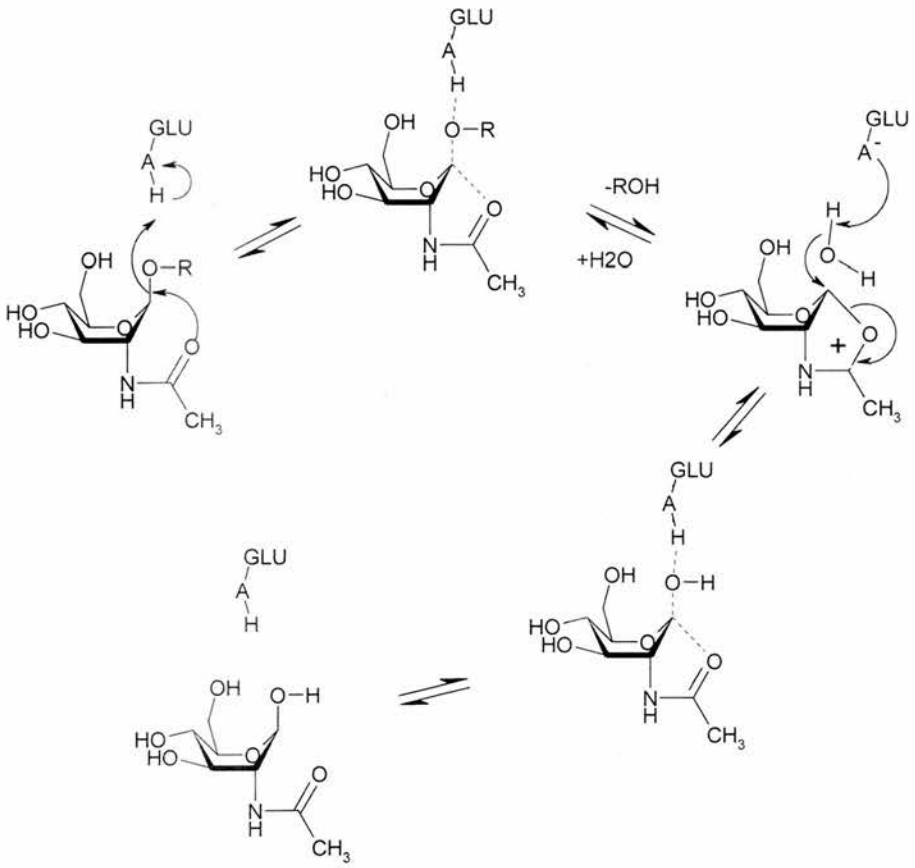
Figure 1.12 Action of hen egg white lysozyme.



Cyclodextrin glycosyl transferase enzyme has a similar mechanism also dependent on the dual acid system, of first glutamate protonation of a glycosidic bond and subsequent nucleophilic attack on the resultant oxocarbenium by an active site aspartate (Uitdehaag, 1999). These enzymes are dedicated to carbohydrate catabolism and belong to the extended glycosyl-hydrolyses family. The family has been classified into seventy-seven sub-families based on sequence similarity (Davies, 1997). Further crystallographic evidence for covalent bonding of carbohydrates from the structures of chitobiase mutants (Prag, 2000) and the crystal structure of myrosinase (Burmeister, 1997). The chitobiase enzyme would seem to follow the dual acid enzyme mechanism of glutamate and aspartate (Figure 1.13), whereas

the plant enzyme myrosinase is proposed to break the glycosidic bond by protonation of the glycosidic oxygen from a glutamine residue instead of the acid / base glutamate residue seen in previous glycosyl transferase structures. The resulting oxocarbenium C1 atom then undergoes nucleophilic attack by a glutamic acid residue, replacing the aspartate acid residue also present in previous glycosyl transferase structures. Many of the glycosyl transferases recognise and catabolise complex carbohydrates on the cell surface of important pathogens and as such their mechanisms may lead to a greater understanding of disease control.

Figure 1.13 The reaction mechanism of the chitobiase enzyme.



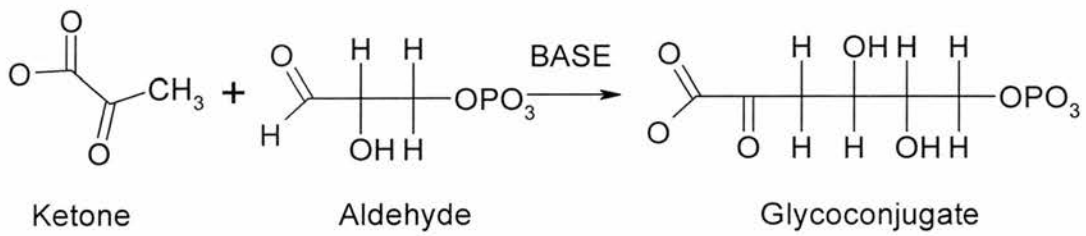
Covalent intermediates are also seen in the Lyase enzyme group. Within the carbohydrate lyases, the aldolase sub-group of enzymes have recently been of interest due to their ability to produce enantiomeric specific products for fine chemical production, and their essential roles

in the processing of small metabolic carbohydrates within Ebden Meyerhof Parnas (Searcy, 1984) and Entner Doudoroff glycolytic pathways (Entner, 1952).

1.17 Aldolases.

The structural study of protein-carbohydrate covalent intermediates has provided a molecular level insight into the mechanism of many of these carbohydrate processing proteins. In structural studies few of these proteins have been studied as extensively as the aldolase enzymes from the lyase family. The aldolase enzymes are a sub-group of the lyase family, which stereoselectively catalyses the Michael addition of a ketone to an aldehyde, forming various glycoconjugates (Figure 1.14).

Figure 1.14 The Aldol reaction.



There are over thirty identified aldolases, forming two distinct functional classes. Class I aldolases activate the nucleophile by a Schiff base at the catalytic lysine residue. Class II aldolases use a catalytic divalent metal ion and an active site base to activate the nucleophile. The Class I aldolases are found in all forms of life from protozoa to humans (Verlinde, 1999) The Class II aldolases (Figure 1.15) are found in the lower life forms such as bacteria, yeast and fungi, indicating that Class II aldolases may be good targets for anti bacterial / fungal therapies. The Class II aldolases are metallo-enzymes containing two metal ions at the active site, an essential catalytic divalent metal ion (Zn^{2+}) and a monovalent metal ion (K^{+}) thought to enable activation (Hall, 1999) (Figure 1.16).

Figure 1.15 Class II Fructose bis-phosphate (FBP) aldolase from *Escherichia coli*.

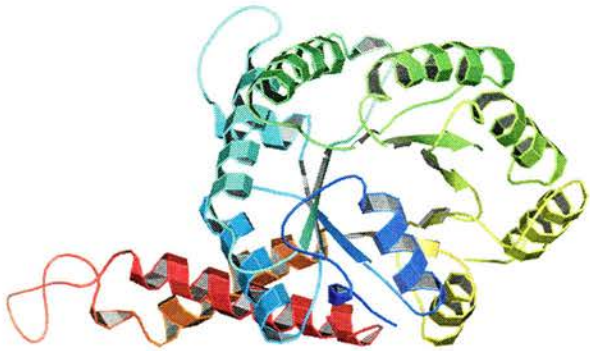
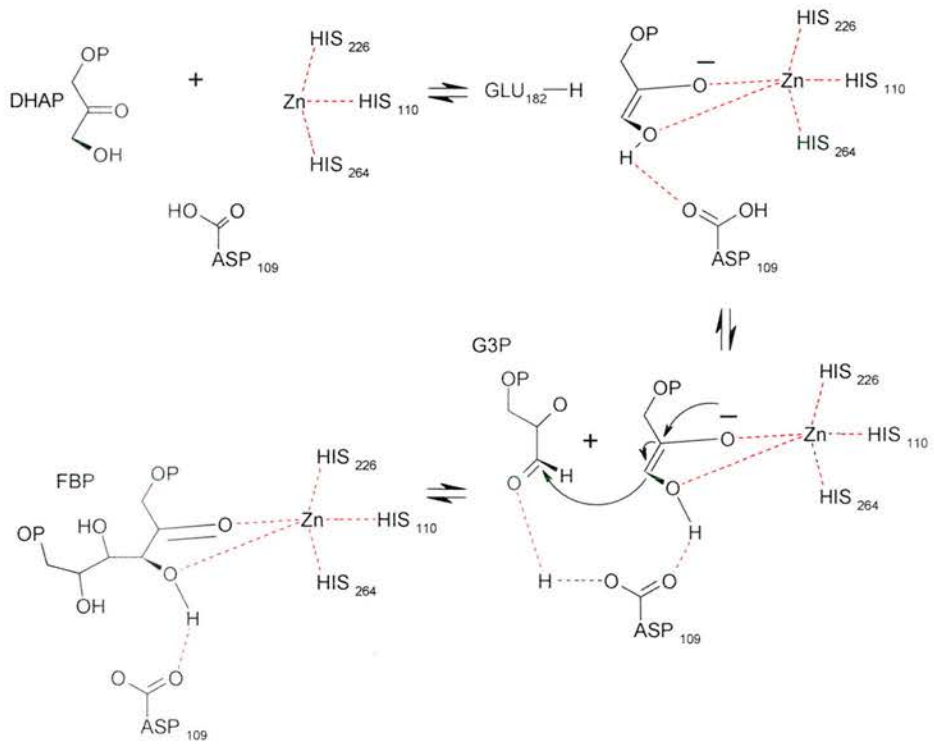


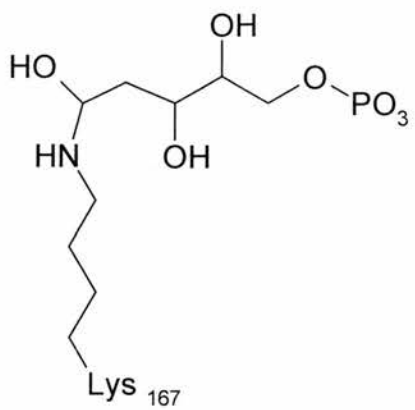
Figure 1.16 The proposed reaction mechanism of the Class II FBP aldolase.



Di-hydroxy-acetone-phosphate (DHAP) chelates with the catalytic zinc at the active site. DHAP is deprotonated by an active site base, possibly the glutamate 182 residue, producing the carbanion and providing the unsaturated linkage where the addition reaction takes place. Glyceraldehyde-3-phosphate (G3P) binds parallel to the ene-diolate nucleophile. The carbon-carbon bond is formed and FBP released (Hall, 1999).

The Class I aldolases catalyse the aldol reaction *via* a Schiff base mechanism involving the production of a covalent intermediate at the active site catalytic lysine residue. In the D-2-deoxyribose-5-phosphate (DERA) classified aldolase, the enzyme catalyses the production of 2-deoxyribose-5-phosphate from acetaldehyde and D-glyceraldehyde-3-phosphate. A covalent complex of the DERA enzyme from *Escherichia coli* was obtained by soaking the protein crystals with the substrate before crystallographic analysis. The resultant structure (Heine, 2001) showed a carbinolamine intermediate bound at the enzyme active site lysine (Figure 1.17).

Figure 1.17 DERA-2-deoxyribose-5-phosphate Schiff base covalent intermediate.



The Class I aldolases form a Schiff base intermediate with the catalytic lysine and the substrate as part of the catalytic mechanism. In Class I aldolases the reaction first forms a Schiff base with an active site lysine and the substrate nucleophile. The lysine Schiff base nucleophile forms a reactive enamine before condensation with the aldehyde to form the larger carbinolamine. The carbinolamine-Schiff base is reduced producing free product and the active site lysine. The crystal structure of the *Escherichia coli* Transaldolase B (a structural homologue of Class I aldolases) was determined to 1.87 Å resolution in 1996 (Jia, 1996). The further structures of human muscle (Gamblin, 1990), Rabbit muscle (Sygusch,

1987) and *Drosophilia melanogaster* FBP aldolases (Hester, 1992) have allowed the direct comparison between Class I FBP aldolases and the Class II FBP aldolase structure.

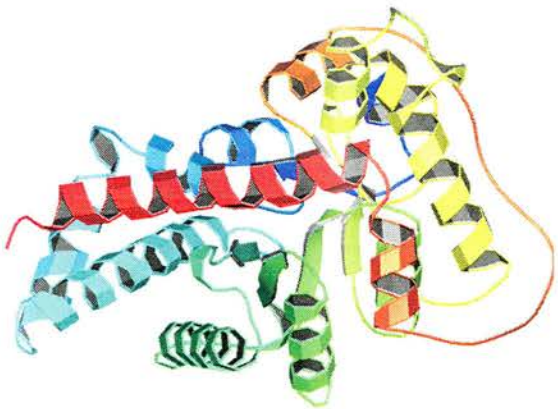
The Class II FBP aldolase from *Escherichia coli* functions as a homodimer with a molecular weight of 78 kDa. The monomer has the common α/β barrel conformation, with the central core consisting of eight parallel β-strands surrounded by eleven α helices. The catalytic site is positioned at the C-terminal end of the core β barrel strands. The active site contains two metal ions. One of the metal ions is buried within the active site, this ion is thought to be a monovalent ion, possibly potassium. The second metal ion is positioned on the surface of the active site and is believed to be the catalytic divalent Zn^{2+} ion. The zinc ion at the active site is believed to act as a Lewis acid electron sink, polarising the carbonyl bond of a ketone substrate for the cleavage / condensation reaction (Mildvan, 1971; Cooper, 1996).

The Class II aldolase functions as a dimer with an approximate length of 100 Å. The dimer interface consists of the two α helices ten and eleven from each monomer. The two helices interact forming a hydrophobic core at the dimer interface. The α-4 helices from each monomer also interact and align in an anti-parallel manner (Cooper, 1996; Zgiby, 2002).

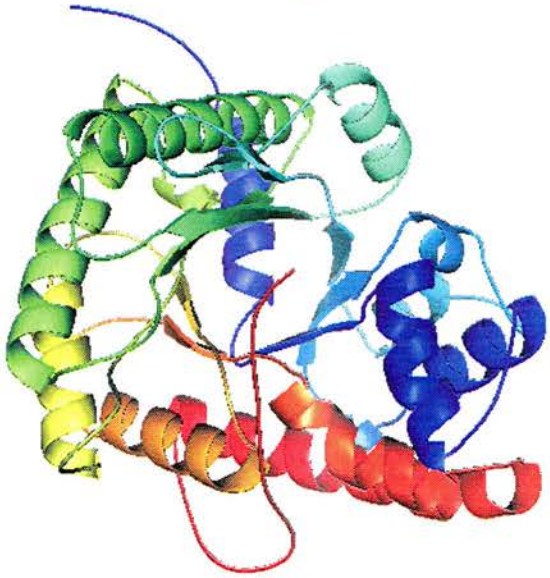
The Class I aldolase transaldolase B from *Escherichia coli*, has similar topology with the human, rabbit and fruit fly (*Drosophilia melanogaster*) FBP aldolases. Each of these enzymes has an eight stranded α/β barrel conformation, with an α helix situated at the N-terminus and three further α helices at the C-terminal (Figure 1.18).

Figure 1.18 Transaldolase B, Human FBP aldolase and Fruit fly FBP aldolase.

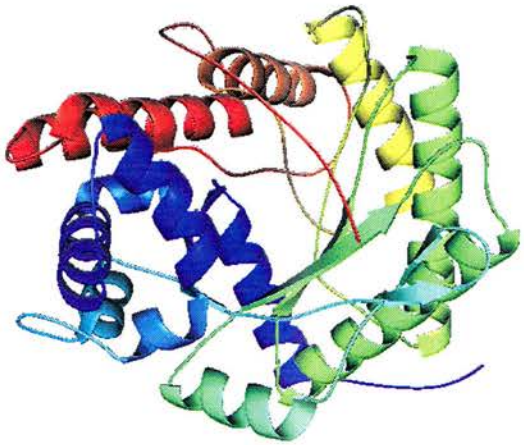
TransAldolase B



Human FBP aldolase



Fruit fly FBP aldolase



The transaldolase B enzyme contains two extra helices at the C-terminal. Superpositioning of the transaldolase structure and the human FBP aldolase by secondary structure element alignment produced 113 equivalent C α atoms with an rms fit of 2.53 Å (Jia, 1996). However this superpositioning does not align the active site lysine residues. Alignment by matching the β-strands containing the active site lysines, β-4 mapped to β-6, produced a better structural match. In this new alignment the two active site residues are at equivalent positions. The β-4 to β-6 alignment identifies several conserved polar residues at equivalent locations. Thirteen equivalent residues were conserved between the transaldolase B and the human FBP aldolase, most importantly, five conserved active site residues. The overall structural similarity of the transaldolase and the human FBP aldolase and the appearance of the five conserved active site residues, indicate both enzymes have evolved from a common ancestor with the Class I aldolases having cyclic permutation (Jia, 1996). Cooper *et al* (Cooper, 1996) superimposed the Fruit fly aldolase (Hester, 1992) and the Class II FBP aldolase from *Escherichia coli* (Cooper, 1996; Zgiby, 2002). The alignment of the eight stranded β cores, produced one hundred and twenty four equivalent C α atoms with an rms fit of 2.3 Å. Remarkably, the key catalytic residues from both the Class I and Class II aldolase structures were positioned at equivalent locations, with the ε-nitrogens of the Class I lysines (positions 146 and 229) positioned within 1 Å of the N-ε 2 atoms of the catalytic zinc binding residues histidines (positions 264 and 110). This structural similarity, although not conclusive strengthens the argument that the two classes evolved from a common ancestor.

Class I and Class II aldolases can be subdivided into four sub groups based upon their reaction nucleophile. The four sub-groups are 2-deoxyribose-5-phosphate (DERA) aldolases, dihydroxyacetone phosphate (DHAP) aldolases, the glycine dependent aldolases and the pyruvate / phosphoenol pyruvate dependent aldolases, each of which carry out the aldol reaction (Figure 1.14) using different specific substrate nucleophiles.

1.18 2-Deoxyribose-5-phosphate (DERA) aldolases.

The DERA aldolases (Figure 1.19) catalyse the addition of acetaldehyde and glyceraldehyde-3-phosphate via a Schiff base intermediate forming 2-deoxyribose-5-phosphate (Figure 1.21) It is unique in that it is the only known aldolase that can catalyse the addition of two aldehydes (Machajewski, 2000). These aldehyde products can then be used as substrates for another sub-group of aldolases, the DHAP aldolases. A combination of both the DERA and DHAP enzymes have been used as a one pot synthetic strategy to produce 5-deoxy ketoses for synthetic chemistry (Gijsen, 1995).

Figure 1.19 DERA aldolase.

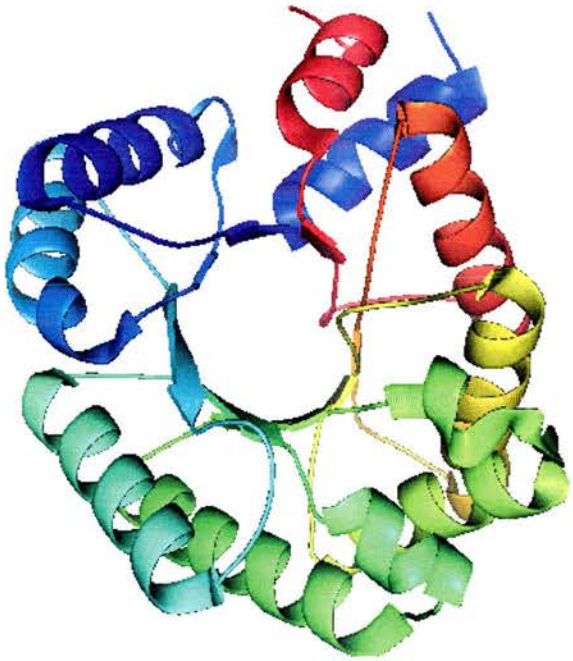
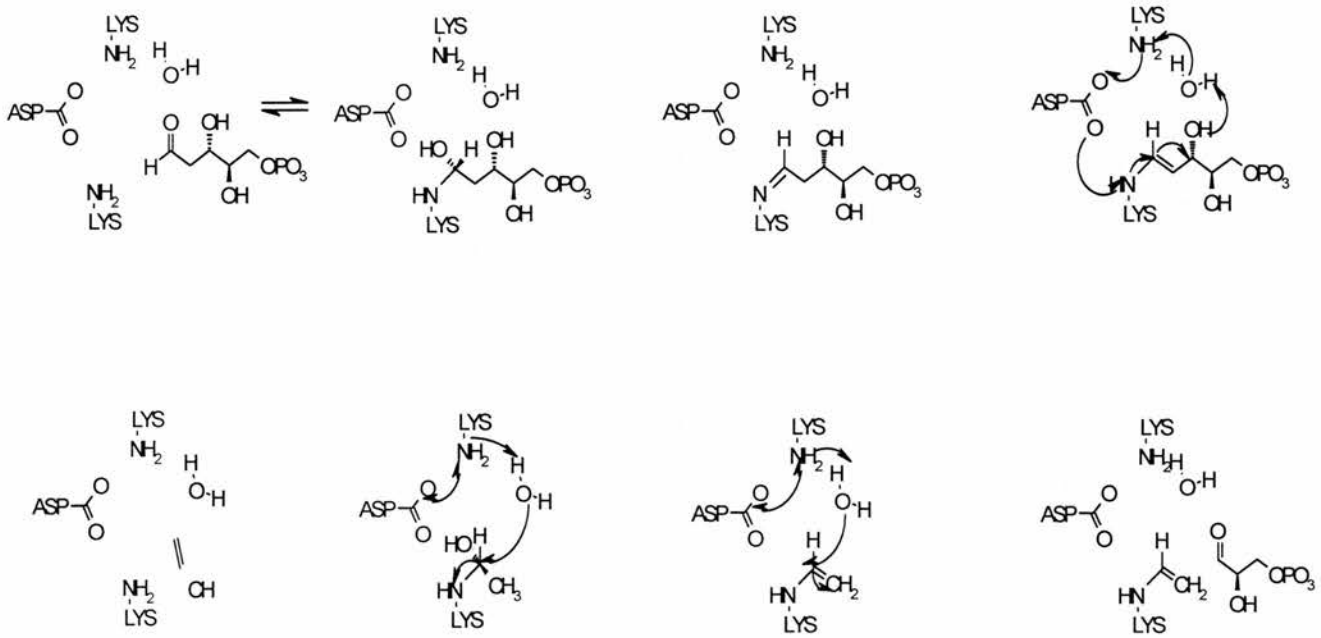


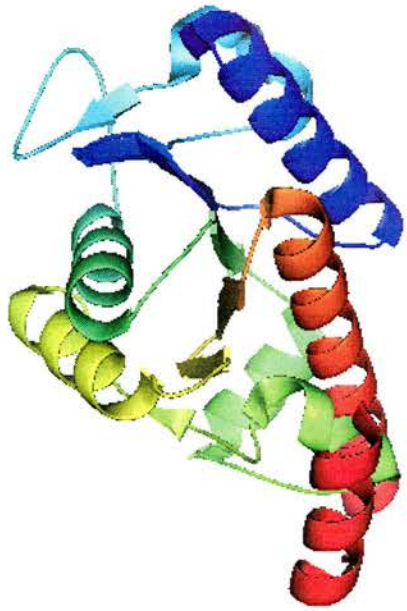
Figure 1.20 Proposed DERA mechanism.

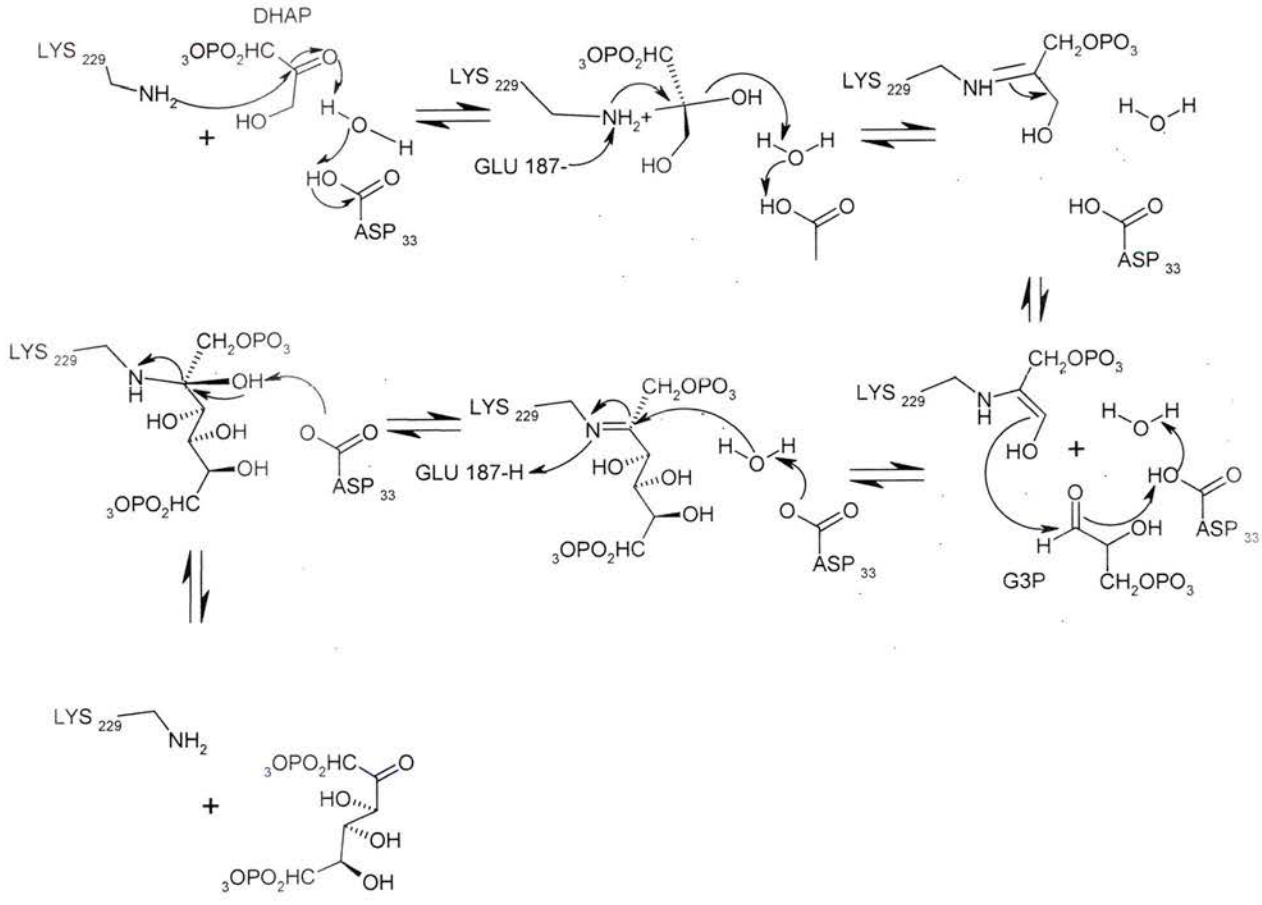


1.19 Dihydroxy acetone phosphate (DHAP) aldolases.

DHAP aldolases (Figure 1.21) catalyse the reversible aldol condensation reaction using the dihydroxy acetone substrate as the catalytic nucleophile. The fructose bis-phosphate (FBP) aldolase catalyses the reversible condensation of dihydroxy acetone and glyceraldehyde-3-phosphate to fructose bis-phosphate. These aldolases have been extensively used in the bio-organic synthesis of amine sugars and cyclitols (Chou, 1995). Further synthetic properties of the DHAP enzymes include the ability to form novel long chain sugars from the addition of DHAP and pentose / hexose substrates (Fessner, 1996).

Figure 1.21 DHAP aldolase monomer and proposed enzyme mechanism.

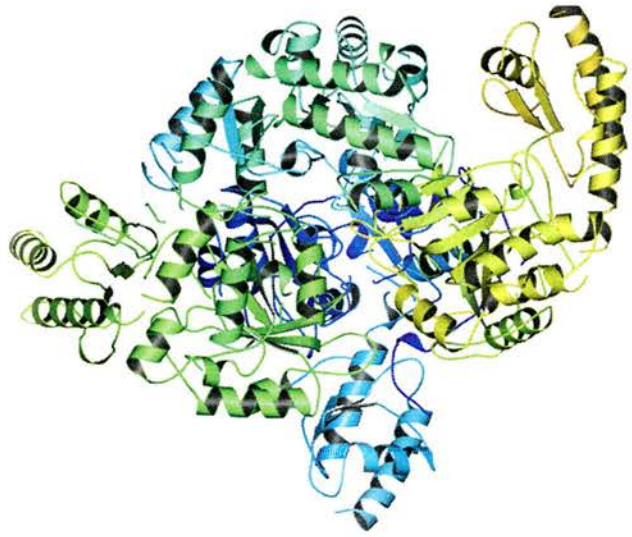




1.20 Glycine aldolases.

Glycine dependent aldolases (Figure 1.22) catalyse the reversible condensation of glycine with an aldehyde acceptor, producing a β-hydroxy-α-amino acid. These aldolases predominantly produce α amino acids specifically however L-threonine aldolase is known to accept a wide variety different substrates producing both erythro and threo amino acids. The ability of the L-threonine aldolases to produce varying products depending upon substrate has led to the use of L-threonine aldolase in synthesis of novel compounds such as the immunosuppressant mycesterin-D (Shibata, 1996).

Figure 1.22 Glycine dependent aldolase, L-threonine aldolase and enzyme reaction.



1.21 Pyruvate / Phosphoenol puyruvate aldolases.

The pyruvate / phosphoenol pyruvate enzymes (Figure 1.23) catalyse the reversible condensation of pyruvate / phosphoenol pyruvate with an aldehyde to produce the formation of a 4-hydroxy-2-ketobutyrate frame. The enzyme uses the pyruvate substrate as the catalytic nucleophile. Pyruvate-dependent aldolases have catabolic functions in *vivo*, whereas their counterparts employing phosphoenolpyruvate as the donor substrate are involved in the biosynthesis of keto acids. However, both classes of enzymes can be used to prepare similar keto acid products.

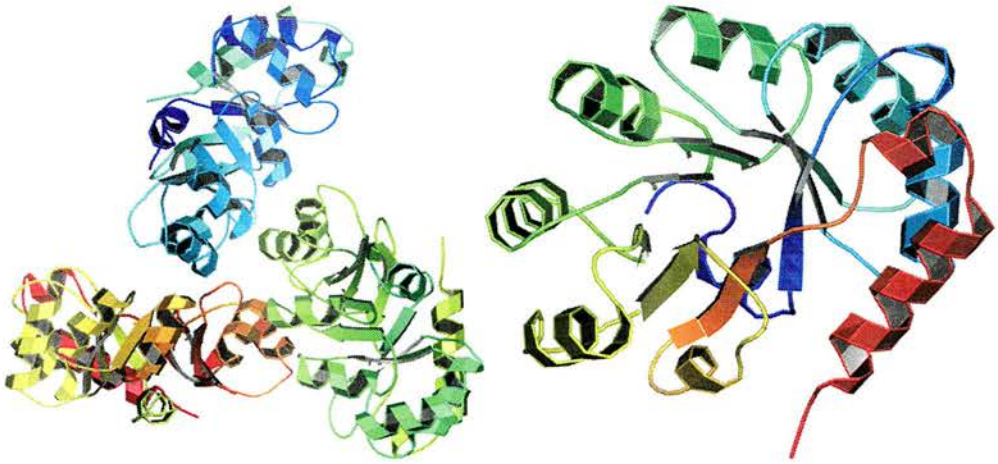
N-acetylneuraminic acid (Neu5Ac) aldolase or sialic acid aldolase is the most utilized member of this group (Fessner, 1996; Gijsen, 1996). Sialic acid aldolase catalyses the reversible aldol reaction of *N*-acetyl-D-mannosamine (ManNAc) and pyruvate to *N*-acetyl-D-neuraminic acid (Neu5Ac). Sialic acids are key cell-surface determinants of mammalian glycoconjugates, and are found at terminal positions of glycoproteins and glycolipids, important in biological recognition processes. Neu5Ac is the best-known member of this special Class of amino sugars. Sialic acid aldolase has been used to produce (*in vitro*) derivatives of Neu5Ac, (Lin, 1992) some of which have been found to act as inhibitors of sialidases (Murakami, 1996). Recently directed evolution of the *Escherichia coli* 2-keto-3deoxy-6-phophoglycerate (KDPG) aldolase (a pyruvate dependent aldolase) has been reported (Fong, 2000; Buchanan, 2001). The stereoselective way in which KDPG aldolase forms carbon-carbon bonds is of great interest to synthetic chemistry. The directed evolution of the *Escherichia coli* KDPG aldolase suggests that altering key amino acid residues around the active site alters substrate specificity. The possibility of protein engineering may provide enzymes for a variety of different substrates to be used to produce novel stereogenic compounds.

1.22 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolases.

The Entner Doudoroff pathway (Entner, 1952), oxidises glucose phosphate to 2-keto-3-deoxy-6-phosphogluconate (KDPG) which is cleaved by KDPG aldolase to produce pyruvate and G3P. The latter is oxidized to pyruvate by Ebden Mayerhof Parnas pathway enzymes wherein two ATP are produced by substrate level phosphorylations. KDPG aldolases have also been found in the pentose phosphate pathways, making the KDPG aldolase a key metabolic enzyme throughout all bacterial organisms. The KDPG aldolases are a member of the pyruvate dependent aldolases, and form a reactive nucleophile via a Schiff base intermediate with a catalytic lysine and the pyruvate / KDPG substrate. At present there are five KDPG aldolase structures deposited in the Brookhaven PDB data-base. Four of the structures are of the *Escherichia coli* KDPG aldolase, its mutant derivatives and a trapped covalent intermediate. The fifth structure is a suspect C-α trace of KDPG aldolase from *Pseudomonas putida* deposited in 1978. (Mavridis, 1982; Allard, 2001)

Both the *Escherichia coli* and *Pseudomonas putida* KDPG aldolase monomers form a TIM-barrel like eight stranded α/β fold. The tertiary structure of the *Escherichia coli* KDPG aldolases, all exhibit a propellor shaped trimer. (Figure 1.23)

Figure 1.23 Quaternary and monomeric structure of *Escherichia coli* KDPG aldolase.



The *Escherichia coli* KDPG aldolase structure, shows the hydrophobic side chains of the β sheet packing against the hydrophobic side chains of the surrounding amphipathic α helices. These α helices are adjacent to the β sheets with which they interact, providing the core barrel with a shell of hydrophobic residues originating from both the core β-strands and the surrounding α helices. Hydrophobic side chains from the β sheet point inside the barrel to form a central hydrophobic core. The catalytic lysine residues are located on the β-6 strand at the C-terminal end of the β barrel core in a freely solvent accessible depression.

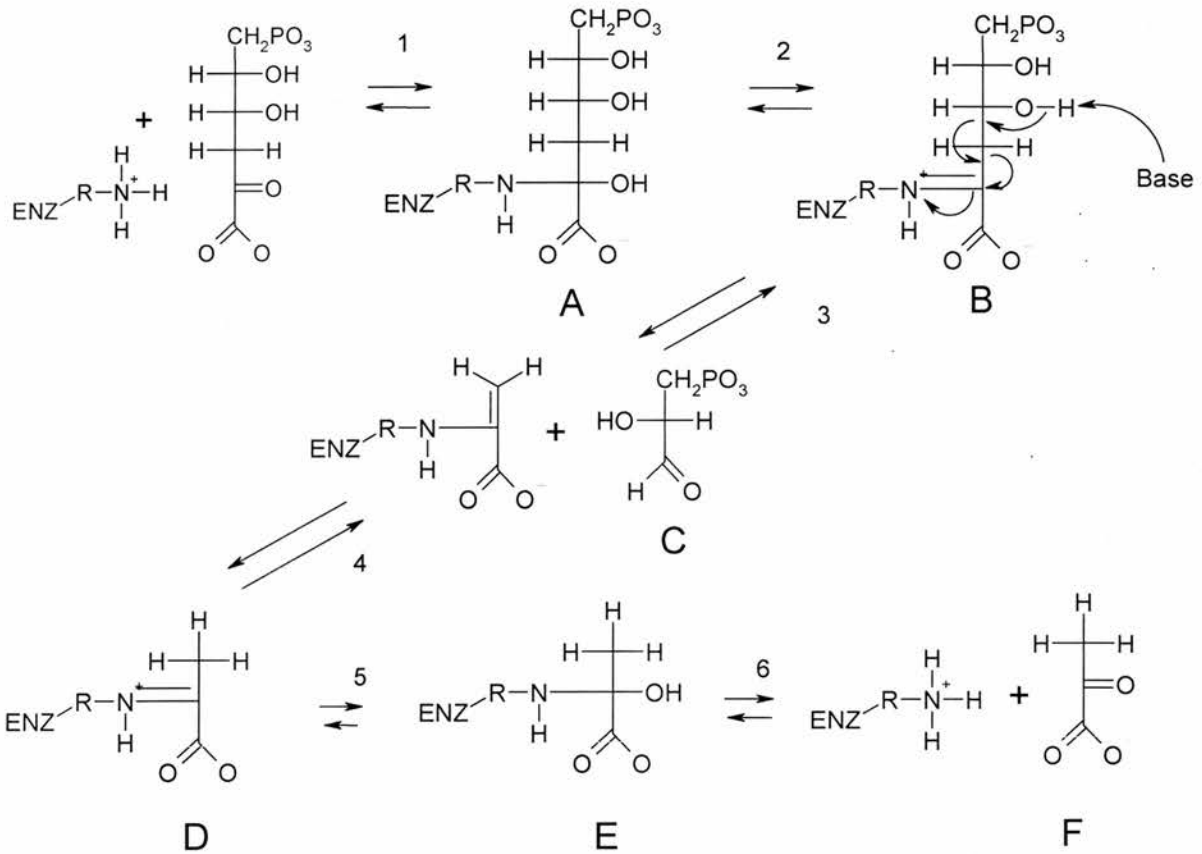
1.23 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase mechanism.

The proposed enzyme mechanism follows the generalised pyruvate dependent aldolase mechanism. The carbonyl carbon of KDPG undergoes nucleophilic attack from the catalytic lysine-133 nitrogen [1], which produces an enzyme-substrate carbinolamine covalent intermediate [A]. The carbinolamine intermediate undergoes dehydration [2] removing the C2 hydroxyl functional group producing the more stable Schiff base intermediate [B]. Deprotonation of the C4 hydroxyl is then believed to occur via an active site base residue,

collapsing the C3-C4 bond [3+4] forming free G3P [C] and a lysine-pyruvate Schiff base intermediate [D]

The Schiff base undergoes hydration [5] forming the 3-carbon enzyme-substrate carbinolamine intermediate [E]. The carbinolamine intermediate is then believed to be de-protonated *via* hydrogen transfer [6] releasing pyruvate [F] and free enzyme (Figure 1.24).

Figure 1.24 Proposed KDPG aldolase mechanism.



1.24 Synthetic possibilities of KDPG aldolase.

Toone *et al* first published work on the use of KDPG aldolase (Figure 1.24) as an enzyme for synthetic organic chemistry in 1992 (Allen, 1992). This initial study showed the enzyme could accept unnatural electrophiles at synthetically useful rates. Initially ten aldehydes were tested, the rates of reaction being reported relative to that for D-glyceraldehyde, itself

accepted at approximately 0.8% the rate of natural substrate. All of the substrates were utilised at much lower rates than the natural substrate; a phenomenon also noted for DERA (Barbas, 1990). The reactions did however proceed with the expected stereochemistry. KDPG aldolase was found not to utilise simple aliphatic aldehydes or benzaldehyde. The continued work on the *Escherichia coli* KDPG aldolase (Buchanan, 2001) involving the directed evolution of the active site has proven successful in providing a wider scope of potential substrates (Fong, 2000).

1.25 *Thermotoga maritima* KDPG aldolase.

The ability of the aldolases to form C-C bonds whilst providing control of the stereospecificity, have made the KDPG Class I aldolase enzymes important biocatalysts for organic synthesis.

The 3rd chapter in this thesis details the structural characterization of KDPG aldolase from *Thermotoga maritima*, a heat stable KDPG aldolase, allowing the direct comparison of the *Escherichia coli* non-heat stable aldolase and the thermostable *Thermotoga maritima* structures. Furthermore chapter 3 will report on the presence of a trapped covalent intermediate at the catalytic lysine, the presence of an occupied phosphate-binding site and through the modeling of the KDPG substrate, propose a possible mechanism for the KDPG aldolase aldol condensation reaction.

1.26 Galactofuranose synthesis: a possible drug target.

Both Gram-negative and Gram-positive bacteria have been found to contain galactofuranosyl (Gal*f*) residues within the mycolyl-arabinogalactan complex of bacterial cell walls and within cell surface O antigens. Gal*f* monosacharides are not found in human glycoconjugates and therefore the synthesis of Gal*f* has become a target for new antibacterial agents. Gal*f* is found

in a number of harmful pathogens including *Mycobacteria tuberculosis, Mycoplasma mycoide, Bacteriodes cellulosolvens, Clostridium thermocellum, Clostridium thermohydrosulfuricum. Escherichia coli* and *Klebsiella Pnuemoniae* (Koplin, 1997). Gal*f* residues are also found in fungal and eukaryotic cell glycoconjugates including *Penicilium spp, Aspergillus ssp, Neurospora crassa, Histoplasma capsulatum, Leishmania donovani, Leishmania major, Lieshmania mexicana, Crithidia spp* and in the lipopeptidophosphoglycans of *Trypanosoma cruzi* (McConville, 1993). As many of these micro-organisms are of serious risk to human health, the interest in the pathways involved in the production of Gal*f* have increased. Evidence from a temperature sensitive mutant of *Mycobacterium smegmatis* indicates a disruption of the arabino-galactan metabolic pathway inhibits cell viability (Vilchèze, 2000).

1.27 *Klebsiella pneumoniae.*

Klebsiella species are gram-negative bacteria found in both environmental and clinical sources. There are currently five known species in the Klebsiella genus, all sharing common osmotic (O) and capsular (K) antigens. *Klebsiella pneumonia* and *Klebsiella oxytoca* are commonly responsible for bacteraemia, pneumonia, meningitis and urinary tract infections (Kelly, 1996; Chen-Hsien, 1997). *Klebsiella* species are regarded as opportunistic pathogens, with increased incidence of infection with patients having underlying medical problems, for example liver cirrhosis and diabetes mellitus, and are identified as a serious noscomial infectious agent (Severon, 1996; Chen-Hsien, 1997).The cell wall of the *Klebsiella* species contains relatively little peptidoglycan but does contain an outer membrane composed of lipopolysaccharide (LPS), lipoprotein and other complex macromolecules including porins and phospholipids. The external face of the outer membrane is covered with LPS bound to lipid A in the phospholipid layer. The LPS molecules consist of the outer O polysaccharide

and the core polysaccharide, with the core polysaccharide bound to the lipid A coated outer membrane. The outer membrane is bound *via* lipoprotein molecules to a thin peptidoglycan layer surrounding the internal cytoplasmic membrane (Figure 1.25). The *Klebsiella* species express a structurally related family of lipopolysaccharide O antigens, each of these O antigens contain a common backbone of D-galactan 1. The antigenic serotype specificity results from the modification of the backbone side groups, altered domains, O-acetyl substitutions or D-galactopyranose side groups with various stoichiometries and linkages (Kelly, 1996). The LPS of the *Klebsiella* species have been reported as important virulence factors in clinical infections. Increased levels of virulence have been reported in infected mice models injected with *Klebsiella pneumoniae* LPS (Domenico, 1985). LPS is also reported as toxic in its protein complex capsular polysaccharide form, this is believed to cause lung tissue damage in cases of pneumonia (Severon, 1996). Previous experimental data from *Escherichia coli* K12 have already determined that the Lipid –A anchor within the LPS is a strong endotoxin responsible for cell damage and inhibition of the reticuloendothelial response (Wyckoff, 1998).

Figure 1.25 The Membrane of a Gram Negative Bacterium.

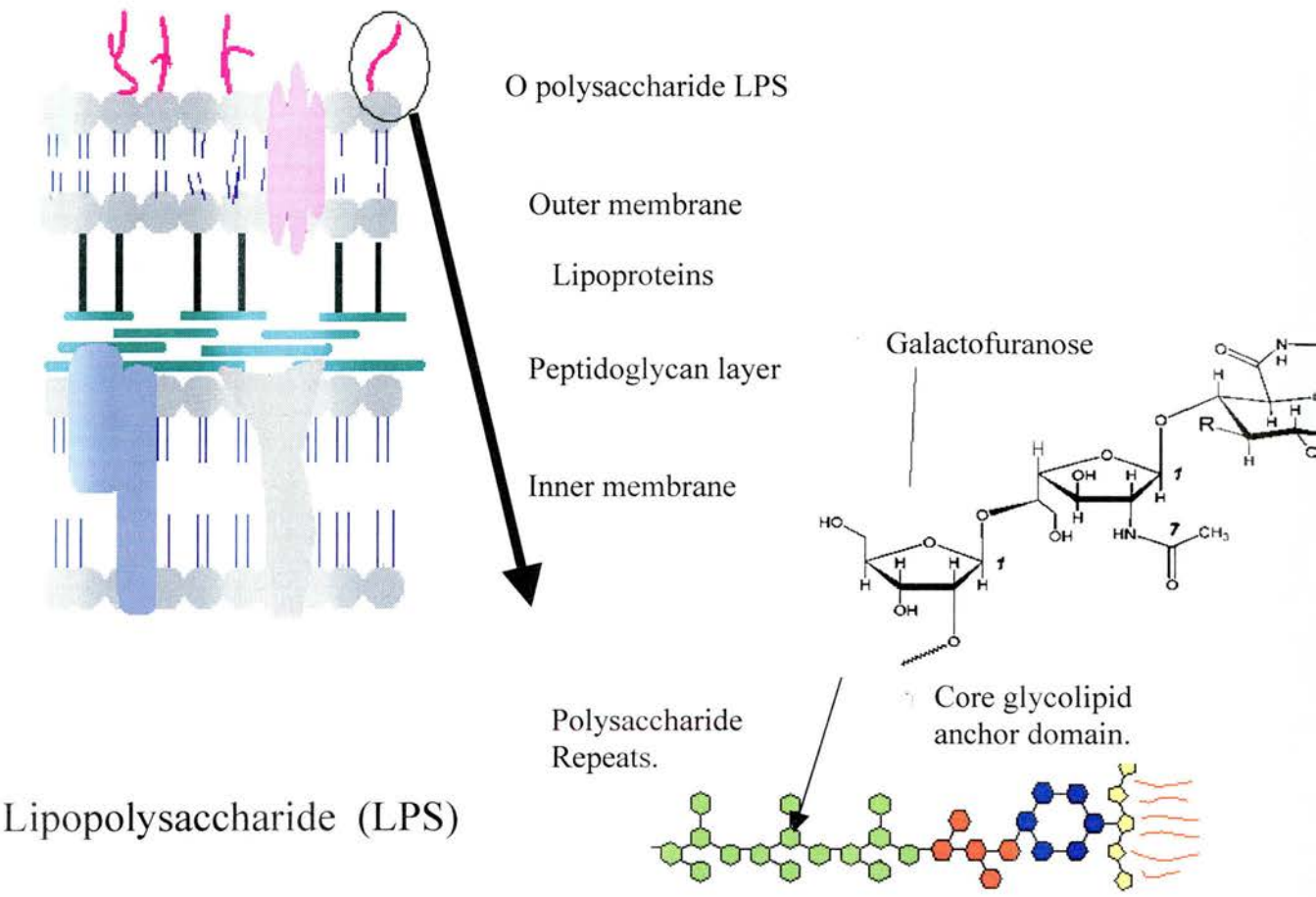


Figure 1.25 Diagram of the Gram negative cell envelope with the inner phospholipid bilayer protected by a thin peptidoglycan layer that is connected to the outer membrane by lipoproteins. The outer membrane is coated with the O polysaccharide LPS.

The *Klebsiella pneumoniae* LPS has also been implicated in the production of a temporary decrease in the reticuloendothelial response to early stage infection, thus increasing the levels of bacterial proliferation at the early stage (Domenico, 1985).

1.28 *Klebsiella pneumoniae* UDP-galactofuranose pathway.

Details of the Gal*f* metabolic pathway in bacteria first emerged in 1994 when a gene present in *Escherichia coli* was sequenced and found to encode a UDP-galactopyranose / UDP-galactofuranose mutase (UDP- Gal*p* mutase) (Stevenson, 1994).

The UDP- Gal*p* mutase from *Klebsiella pneumoniae* is translated from one of six genes in the clonally diverse Rfb gene cluster. The gene products of the cluster are responsible for the formation of the unmodified D-galactan I and the LPS backbone. The Rfb gene cluster contains an ATP binding cassette transporter, UDP- Gal*p* mutase, an initiating galactosyl-transferase enzyme, UDP-GlcNac: undecaprenylphosphate GlcNac-1-phosphate transferase and one further protein believed to be necessary for galactosyl-transferase activity (Kelly, 1996; Koplin, 1997). The six gene products perform four known separate functions. RfbA and RfbB encode the ATP binding cassette transporter proteins required for export of the completed polymerised D-galactose I LPS backbone molecules across the cell membrane to the periplasmic surface, where the O antigen LPS binds to the membrane inserted lipid A molecule (Bronner, 1992). The RfbA gene encodes the trans-membrane domain of the transporter, whereas RfbB encodes the cytoplasmic ATP binding region (Koplin, 1997). The exact function of RfbC gene product is at present unknown, however it has been reported as necessary for the completion of the D-galactan I polymer along with the initiating galactosyl-transferase enzyme encoded by the RfbF gene (Kelly, 1996). The RfbD gene product encodes the UDP- Gal*p* mutase, responsible for production of the galactofuranose precursors for D-galactan I synthesis.

The RfbE gene encodes an UDP-GlcNac: undecaprenylphosphate GlcNac-1-phosphate transferase (Koplin, 1997).

The UDP- Gal*p* substrate is passed from the galactose-4-epimerase to the UDP- Gal*p* mutase. The substrate then undergoes reversible isomerisation under the action of the RfbD gene

product (UDP- Gal*p* mutase) (Barlow, 1999b). The RfbF bifunctional galactosyl-transferase forms the D-galactan I disaccharide from UDP- Gal*f* and UDP- Gal*p*. The D-galactan I then forms an undecaprenyl-pyrophosphoryl N-acetyl-glucosamine primer under the action of the RfbE gene product, UDP-GlcNac: undecaprenylphosphate GlcNac-1-phosphate transferase (Koplin, 1997). Further polymerisations take place, producing the LPS backbone before the gene products of Rfba and RfbB (transporter cassette) transport the LPS backbone to the extracellular surface of the outer membrane (Bronner, 1992).

1.29 UDP-Galactopyranose mutase: A potential drug target.

Knockout of the UDP- Gal*p* mutase in *Escherichia coli* DH5 (non-LPS producing) and K12 Glf knockout serotype cells has shown loss of bacterial virulence. When plasmids containing the *Escherichia coli* Glf (Orf6) genes and *Klebsiella pneumoniae* RfbD genes were inserted into the *Escherichia coli* DH5α cells, the LPS phenotype is restored (Koplin, 1997). Indicating not only a high structural similarity but also a functional equivalence. UDP- Gal*p* mutase from *Escherichia coli*, *Mycobacterium tuberculosis* and *Klebsiella pneumoniae* have a 40 % sequence homology. The *Mycobacterium tuberculosis* UDP- Gal*p* mutase has 42.5 % sequence identity with *Escherichia coli* UDP- Gal*p* mutase and a 40.9 % sequence identity with the *Klebsiella pneumonia* UDP- Gal*p* mutase. A recent BLAST search has identified approximately thirteen separate mutases enzymes from several bacterial genus: *Mycobacterium leprae, Salmonella typhi, Agrobacterium tumefaciens, Ruegeria sp. Corneybacterium glutamicum, Xanthomonas axonopodis, Xanthomonas campestris, Streptococcus pneumoniae, Clostridium acetobutylicum, Campylobacter jejuni*, including the aforementioned *Klebsiella pneumoniae, Escherichia coli,* and *Mycobacterium tuberculosis.* With several putative mutases in *Salmonella typhimurium, Bifidobacterium longum, Corneybacterium efficiens, Streptococcus thermophilus and Bacteriodes fragilius*

(Appendix1). The high sequence homology and functional equivalence between the *Escherichia coli, Klebsiella pneumoniae, Mycobacterium tuberculosis* and other putative mutases may increase the biological range of inhibitors.

The 4th chapter in this thesis discusses the attempt to crystallise the *Klebsiella pneumoniae* UDP- Gal*p* mutase, and reports on the thermodynamic analysis of the enzyme in which we demonstrate that the presence of substrate stabilises a neutral flavo-semiquinone radical at 50% reduced and that the fully reduced flavin is anionic. The anionic flavo-hydroquinone would allow rapid single electron transfer in a crypto-redox reaction. In the discussion these findings are incorporated with previous reports, into considerations of the enzyme mechanism.

Chapter 2

1.94 Å resolution structure of Concanavalin A complexed with the furanose form of L-arabinose.

2.00 Summary.

The 3-dimensional structure of concanavalin-A (Con-A) and the furanose form of L-arabinose has been determined to 1.94Å resolution. The structure was determined by molecular replacement and refined to 1.94 Å resolution using the CCP4i Refmac refinement package. The final model has good geometry with an R_{factor} of 18.3% and a free R_{factor} of 22.3%. The asymmetric unit contains one tetramer with density within the A subunit to unambiguously model one furanose form of L-arabinose at the binding site. The L-arabinose forms a bidentate hydrogen bond with the Asp 208 residue via the O5 ring oxygen and the C2 hydroxyl oxygen. The O5 ring oxygen forms 3 hydrogen bonds: one to the Asp 208 and two further hydrogen bonds to the Asn 14 residue. The C2 hyroxyl oxygen makes 2 more hydrogen bonds: one to the Leu 99 nitrogen atom and one more to the Asp 208 forming the bidentate bond. The C3 hydroxyl points out of the active site making no direct contact with the protein. The C5 hydroxyl makes the final direct hydrogen bond to the nitrogen atom of Arg 228. The relatively hydrophobic C1 location on the carbohydrate ring is positioned adjacent to Tyr 12 forming a weakly hydrophobic / van der Waals interaction.

2.01 Introduction.

Con-A is a saccharide binding legume lectin from the Jack bean (*Canavalia ensiformis*). It belongs to one of the larger families of plant lectins (Sharon, 1993) whose ability to bind specifically to a variety of complex cell surface carbohydrate structures has produced a sizeable quantity of research (Sharon, 1972; Lis, 1986; Sharon, 1993; Naismith, 1994, 1996; Loris, 1998; Qasba, 2000; Dodd, 2001; Wormald, 2002) over the past 3 decades. The Con-A lectin is a tetramer, with each monomer consisting of 237 residues. Each monomer contains two metal binding sites and one saccharide binding site. Hardman and Ainsworth (Hardman, 1976) crystallised the first Con-A co-complex with mannoside in 1976. The co-complex crystalised in the space group $C222_1$, however the crystals diffracted poorly to no more than 6 Å. Although the structure had poor resolution the authors located what is now known to be the saccharide binding site. The two metal binding sites (S1 and S2) are located near and within the saccharide binding site, the S1 metal binding site binds transition metals (Ni, Zn, Mn, Co, Cd) whereas the S2 binding site binds only the divalent Ca and Cd ions. Both S1 and S2 metal binding sites must be occupied for saccharide binding to occur (Kalb, 1968) with the S2 binding site having an absolute requirement for divalent Ca or Cd ions (Sumner and Howell 1936). Hardman and Ainsworth (Hardman, 1976) also identified several of the key residues around the active site, including Tyr 12, Asn 14, Asp 16, Leu 99, Tyr 100, Asp 208, and Arg 228.

Goldstein *et al* 1974 assessed each sugar's atoms effect upon the binding affinity of the Con-A-carbohydrate interaction at the monosaccharide binding site. The Goldstein examination of carbohydrate binding found that atoms C4, O4, C5, O5, C6, and O6 must retain the same equatorial orientation as mannose and be un-substituted to facilitate binding. However, the O3 atom, although it must remain equatorial, can be modified by either a methyl or phenyl substituent and maintain some binding affinity (Goldstein, 1974).

In 1989 Derewenda *et al* (Derewenda, 1989) solved the structure of an α-methyl-D-mannopyranoside Con-A complex to 2.9 Å resolution. This structure identified more specific interactions between the protein and the ligand. The structure contained a tetramer within the asymmetric unit, with each monomer containing electron density for one molecule of α-methyl-D-mannopyranoside at each active site. A molecule of α-methyl-D-mannopyranoside could be unambiguously positioned at each binding site.

In 1994 Naismith *et al* (Naismith, 1994) reported the first high-resolution structure of a Con-A saccharide complex. The structure was solved and refined to 2.00 Å. The report highlighted the roles of active site residues Tyr 100, Tyr 12, Asp, 208 and Arg 228, indicating their role in providing steric requirements for saccharide binding. Furthermore the structure positioned the divalent cation Ca^{2+} as an integral component of the saccharide-binding site and identified a conserved water ligand to the calcium ion, necessary for the stabilisation of the *cis* peptide between Ala 207 and Asp 208.

Further structure solutions on mannopyranoside and glucopyranoside derivatives (Harrop, 1996; Kanellopoulos, 1996) and other larger oligosaccharides (Methyl-3,6-di-O-(α-D-mannopyranosyl)- α-D-mannopyranoside, (Methyl-3,6-di-O-(α-D-mannopyranosyl)- α-D-mannopyranoside, β-GlcNAc-(1-->2)-Man-α-(1-->3)-[β- GlcNAc-(1-->2)--Man-α- (1-->6)]-Man, Man α1-2 Man -OMe) (Loris, 1996; Moothoo, 1998, 1999) have confirmed the position of the calcium ion, and the structural aspects of the monosaccharide-binding site which confer binding specificity.

In all of the pyranoside binding structures, superimposition of the proteins with the docked ligands, have shown that the monosaccharide and oligosaccharide ligands follow the Goldstein rules for binding at the monosaccharide binding site.

All of the pyranose ligands bound at the monermeric site have the same orientation within the shallow binding cleft and identical contacts with the lectin. Trimannoside, β-D-Glcnac-(1,2)α-D-Man-(1,3)[β-D-Glcnac(1,2)α-D-Man (1,6)]α-D-Man, and α-Methyl-D-Mannopyranoside binding to the Con-A tetramer have been studied by Dimick et al 1999 (Dimick, 1999). The binding of the three structurally distinct ligands have shown that although the ligands have altered extended domains the core monosaccharide binding remains essentially unaltered with the binding cleft of the three different co-complex structures retaining an almost identical shape and binding affinity. Therefore the Con-A – terminal pyranose binding would seem to be unaffected whether there is an extended domain or not as would be expected in conjuction with the Goldstein rules (Goldstein, 1986).

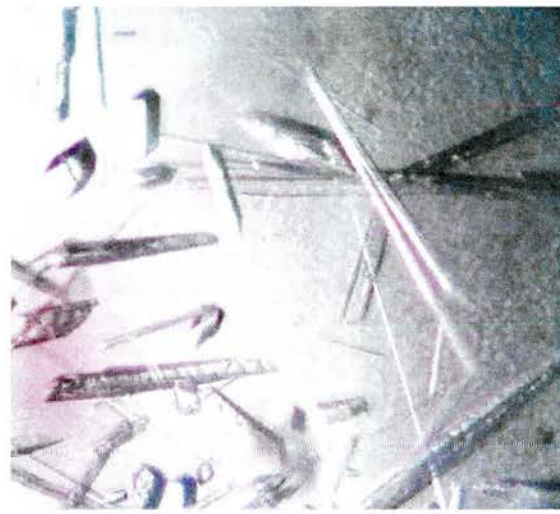
This chapter describes the crystalisation and structure solution of the first five membered monosaccharide ring bound to the Con-A monosaccharide-binding site. This chapter describes the co-crystalisation of Con-A with the furanose form of L-arabinose, the data collection, structure solution, refinement, ligand modelling and describes the protein-ligand contacts.

2.02 Experimental.

2.03 Crystalisation and data collection.

Concanavalin A protein was supplied by Sigma chemicals. Co-crystals of the Con-A-L-arabinose complex were obtained after an initial screen consisting of 10, 15, 20, and 25 % polyethylene glycol (PEG) 4000,Tris HCl buffer pH 5.0, 6.0, 7.0 and 8.0, 25, 50, 75, 100 mM L-arabinose and 15 mM Con-A. Optimised conditions (10mM Con-A, 1mM $CaCl_2$, 75 mM NaCl, 20 mM Tris pH 7.0, 15% PEG 4K and 25, 50, 75, 100 mM L-arabinose) produced larger crystals, with dimensions of approximately 0.8 x 0.1 x 0.1 mm^3 (Figure 2.01).

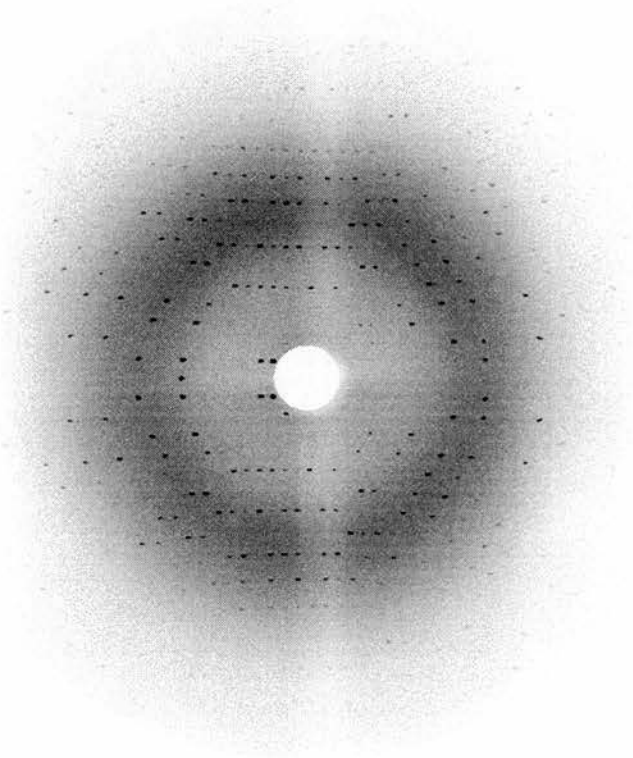
Figure 2.01 Concanavalin A / L-arabinofuranose co-complex crystals.



In house diffraction was undertaken on an ENRAF FR-591 Nonius X-ray generator with a Dip-2000 detector and osmic mirrors. Crystals were cryo-protected in 50% glycerol before being looped and mounted under a nitrogen cryo-stream, with a crystal to detector distance of 200 cm, and diffraction data collected from a 1.54 Å X-ray source from a rotating copper anode. The data collected for the 25, 50, and 75 mM L-arabinose crystal trials diffracted to 2.5, 2.3 and 2.8 Å respectively. The data indexed by MOSFLM (Powell, 1999) and scaled by the CCP4 scala programme (Evans, 1993). Data collected from the 100mM L-arabinose

crystals diffracted to 2.6Å in house. The electron density maps produced at 1 σ indicated some unexplained density at the A subunit binding site, however the experimental density was not sufficient to model a L-arabinose molecule. Further crystals were produced in the presence of 150mM L-arabinose, the crystals were cryo-protected as before and diffraction data collected on a MAR-CCD detector at DESY Hamburg Synchrotron (Station BW7A). The crystals diffracted to 1.94Å resolution and 120 images were collected at 1° non-overlapping oscillations (Figure 2.02).

Figure 2.02 The Diffraction Image from a 1.0 ° 10 minute exposure.



The data were indexed using the MOSFLM programme (Powell, 1999) and scaled using the scala programme from the CCP4 programme suite (Evans, 1993). The crystals had $P2_12_12_1$ symmetry with a unit cell of a = 64.12 **Å** b = 113.04 **Å** c = 119.37 **Å**, and contained 1 tetramer in the asymmetric unit.

Table 2.01. Crystal diffraction data.

Resolution Å	Number of unique reflections.	Completeness %	Rmerge %	Redundancy
8.69	824	98.4	0.023	4.5
6.14	1417	98.4	0.020	5.7
5.02	1774	99.4	0.020	5.8
4.34	2090	99.4	0.019	5.9
3.89	2349	99.7	0.021	5.9
3.55	2579	99.7	0.023	5.9
3.28	2802	99.8	0.026	6.0
3.07	2968	99.8	0.030	6.0
2.90	3158	99.9	0.033	6.0
2.75	3348	99.9	0.038	6.0
2.62	3535	100	0.045	6.0
2.51	3647	100	0.051	6.0
2.41	3813	100	0.062	6.0
2.32	3947	100	0.071	6.0
2.24	4106	100	0.076	6.0
2.17	4221	100	0.087	6.0
2.11	4356	100	0.098	6.0
2.05	4487	100	0.112	6.0
1.94	4026	100	0.136	6.0
43.85-1.94	59447	100	0.035	5.9

2.04 Structure solution.

The structure of Con-A from *C.ensiformis* (Figure 2.03) was determined by molecular replacement using a previously solved Con-A tetramer from the Brookhaven PDB Database (1CVN.pdb) as the model for the CCP4i programme MOLREP (Vagin, 1997).

Table 2.02. Molecular replacement solutions.

Solution	α	β	Γ	Correlation coefficient	R_{Factor} %
1	75.63	59.47	17.88	0.711	32.6

The Molrep programme identified one tetramer within the asymmetric unit (Molrep solution is shown in table 2.02).

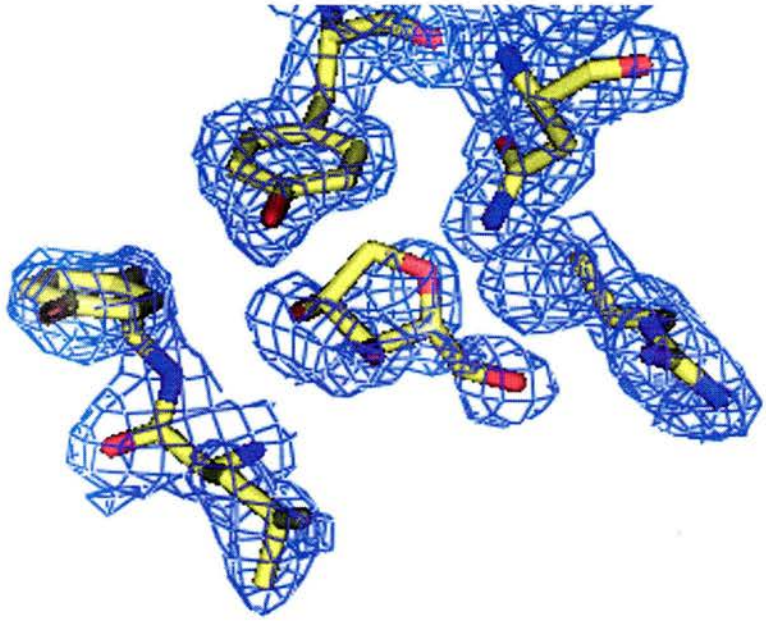
Figure 2 .03 ***Canavalia ensiformis*** **Con-A homo-tetramer.**

2.05 Refinement.

Prior to formal refinement 5% of the data was separated using the CCP4 "Uniqueify" programme to set 5% of the reflection data free from refinement calculation (A.T. Bruenger, 1995), but which has an R-factor calculated throughout the refinement. This gives a Free-R $_{factor}$ allowing the refinement quality to be statistically monitored (i.e. large deviations between the behaviour of R-$_{Factor}$ and the un-biased FreeR-$_{Factor}$ indicate: problems in refinement / quality of model or over-refinement). The model was fitted and adjusted to the experimental density maps manually in conjunction with the "O" graphics package (Jones, 1991). In areas of poorer electron density, side chains or other atoms that were positioned in areas without density, had their occupancy levels set to 0.01.

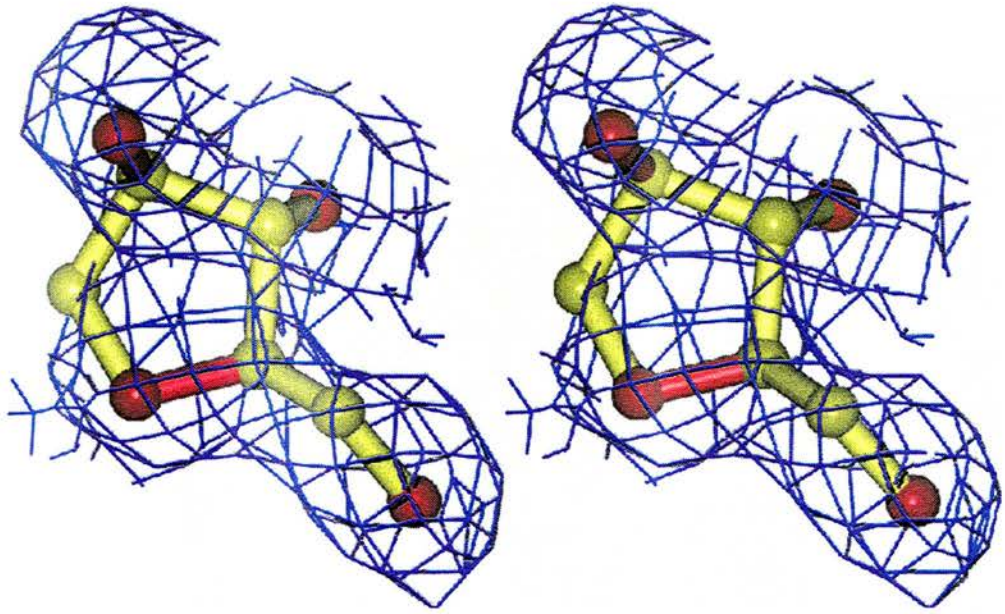
The model was then refined by the CCP4 programme Refmac 5 (Murshudov, 1997). The model was first refined with Rigid Body refinement and then subjected to restrained overall isotropic B-factor refinement, before further manual correction. Several cycles of restrained refinement and manual correction were carrier out until a Free-R_{factor} of 20 %. It was clear from the experimental density maps that on one of the four monomers an additional region of density was present at the binding site (Figure 2.04).

Figure 2.04 The additional 2Fo-Fc electron density at 1.0 σ found at the sub-unit A binding site with the l-arabinose modelled into place.



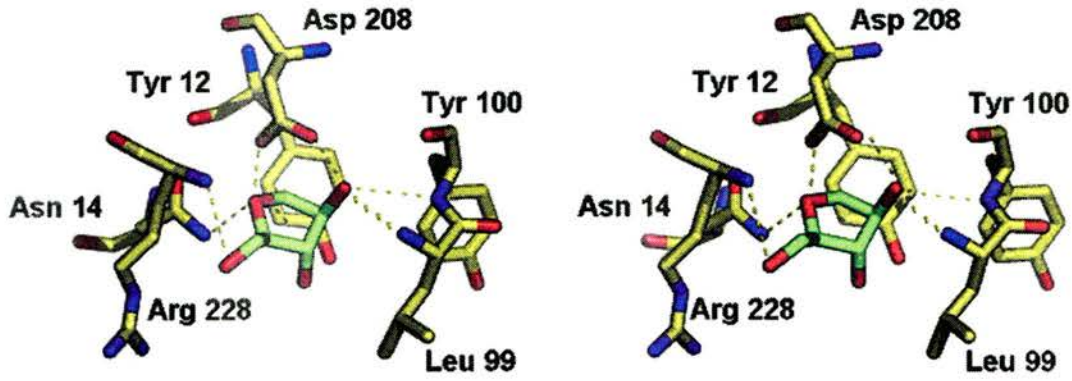
Analysis of the anomalous density allowed modelling of a single molecule of L-arabinose at the A chain active site binding cleft. An L-arabinose molecule PDB file was taken from the CCP4 ligand library. A single copy of the molecule was placed into the Con-A PDB file. Using the graphical package "O" the L-arabinose molecule was manoeuvred and orientated along the X, Y and Z coordinates of the model PDB file until positioned within the appropriate density (Figure 2.05).

Figure 2.05 The L-arabinofuranose ligand modelled into the additional 2Fo-Fc electron density in the omit map at 1.0 σ.



The L-arabinose molecule sits within the shallow binding cleft of the A subunit of the Con-A tetramer. The C2 and C5 Hydroxyls form contacts with active site residues Asp 208, Tyr 100, Leu 99, and Asn 14 respectively, with the furanose ring oxygen contacting Asn14 and making up one member of a bi-dentate hydrogen bond with Asp 208. The C3 hydroxyl oxygen points out of the active site cleft and the relatively hydrophobic patch at the C1 position sits adjacent to the hydrophobic Tyr 12 residue (Figure 2.06).

Figure 2.06 A stereo view of the L-arabinofuranose Ligand positioned within the binding site with the key binding site residues.



The addition of the l-arabinose aided the decrease in the R-$_{Factor}$ and FreeR-$_{Factor}$ in the following successive rounds of refinement (no NCS refinement was used throughout). Once the R-$_{Factor}$ and FreeR-$_{Factor}$ had stopped decreasing at 19.8% and 23.5% respectively. Waters were added using the ccp4 programme ARP-WARP (Perrakis 1999) and further cycles of refinement reduced both the FreeR-$_{Factor}$ and the R-$_{Factor}$ to 18.3% and 22.3% respectively.

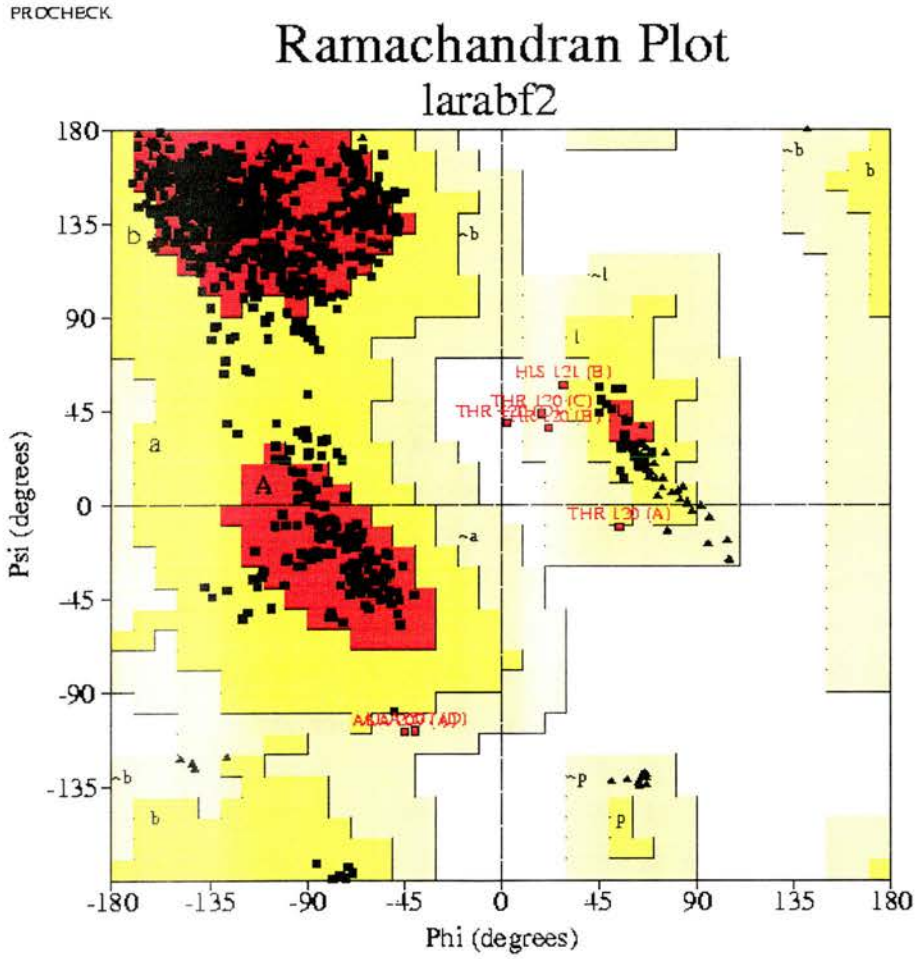
Table 2.03. Structural statistics.

Statistics	Concanavalin A 1.94 Å
R_{Factor}	18.3%
$FreeR_{Factor}$	22.3%
Average B-factor	14.9
RMSD bonds	0.02 Å^2
RMSD angles	1.81°
Residues in most favoured region of Ramachandran plot	86.5%
Number of non-Hydrogen atoms	7723
Number of water molecules	421

Table 2.04. Ramachandran statistics.

Residues in most favoured regions	720	86.5%
Residues in additional allowed	112	13.4%
Residues in disallowed regions	0	0.0%
Number of non glycine; non proline residues	832	100%
Number of glycine residues	64	N/A
Number of proline residues	44	N/A
Total number of residues	957	N/A

Figure 2.07 Ramachandran plot of the Con-A tetramer.



Glycine residues (not restricted to any particular region of the plot) are represented by a ▲, non glycine residues are represented by a ■. The darker the shaded region, the more favourable is the φ (Phi) ψ (Psi) combination.

2.05 Discussion.

2.07 Con-A structures (native and co-complex).

Con-A was the first legume lectin to be isolated characterised and have its molecular structure determined by X ray crystallography. There are 19 co-complex structures of Con-A deposited in the Brookhaven PDB database containing complexes with monosaccharide, oligosaccharides and peptide mimics. Of these co-complex structures the carbohydrate bound models contain varying ligand structures with a prerequisite terminal pyranose ring that binds singly into the monosaccharide-binding site. In 1995 Kalb *et al* (Kalb (Gilboa), 1995) reported the production of co-complex crystals containing the methyl-α-D arabinofuranoside molecule, however as yet there are no structures made available at present describing the binding of a furanose carbohydrate ligand at the Con-A binding site.

Superimposition of the Con-A co-complex structures from the protein structure database indicates nearly identical protein geometry, suggesting that although these models contain mono, di, tri and oligo-saccharides little movement of the protein is required to accommodate the different ligands.

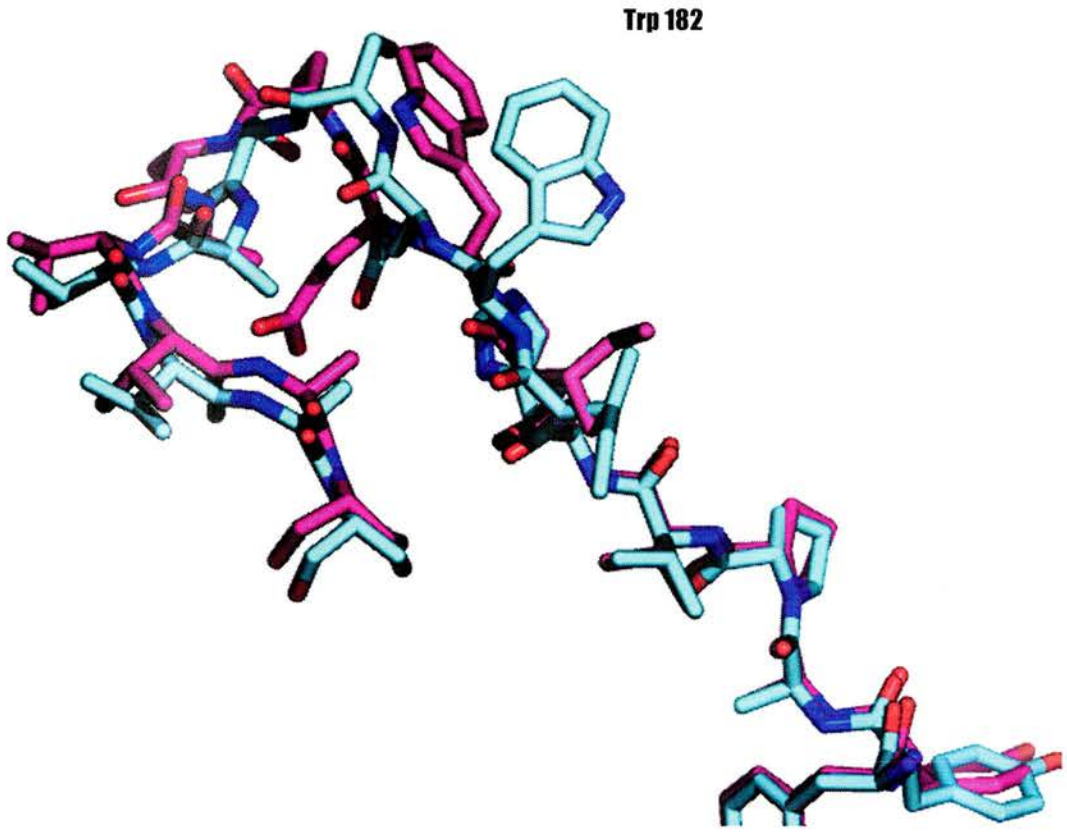
The native structure differs in that the binding of ligands separates the inter-subunit domains resulting in fewer subunit interactions. Binding at the monosaccharide binding site by methyl α-D-mannopyranoside (Naismith 1994), 4-nitrophenyl-α-D-mannopyranoside and 4-nitrophenyl-α-D-glucopyranoside (Kanellopoulos 1996) all have near identical effects on the lectin structure. Although the inter subunit interactions become weaker the fold of the Con-A is essentially unchanged. Superimposition of ligand bound structures and native Con-A have essentially the same Cα backbone. However ligand bound structures such as the methyl α-D-mannopyranoside and the 4-nitrophenyl gluco and manno pyranoside co-complexes have the conserved changes in specific amino acid side chain conformations. At the binding site there

is little change with several of the key residue side chains perhaps moving slightly toward the bound ligands forming hydrogen bonds and van der Waals interactions between the lectin and ligand. The most dramatic change occurs in a residue far removed from the active site. Superimposing several saccharide bound Con-A structures it was clear that in each structure the tryptophan at position 182 has changed its side chain conformation dramatically affecting the surrounding protein structure. Naismith et al 1994 state that when comparing the native and ligand co-complex structures of concanavalin A, the lectin saccharide complex results in a perturbation of the tetramer association separating the monomers by 0.9 Å decreasing the inter subunit interactions of the native Con-A structure by 35% (Naismith 1994).

2.07 Con-A / L-arabinofuranose Binding.

The overall structure of the Con-A lectin upon L-arabinofuranose binding, when compared to the mannose/gluco-pyranoside complexes is essentially very similar. There is no significant change in the conformation and position of the residues at the binding site. The altered conformation of the tryptophan at position 182 from the native structure is essentially identical. All four of the subunits in the L-arabinofuranose co-complex structure have the same altered Trp 182 conformation indicating that although the experimental electron density maps were too poor to model the ligand at the B, C and D sunbunit binding sites, there was sufficient interactions between these binding sites and the ligands to dramatically change the Trp 182 conformation (Figure 2.08).

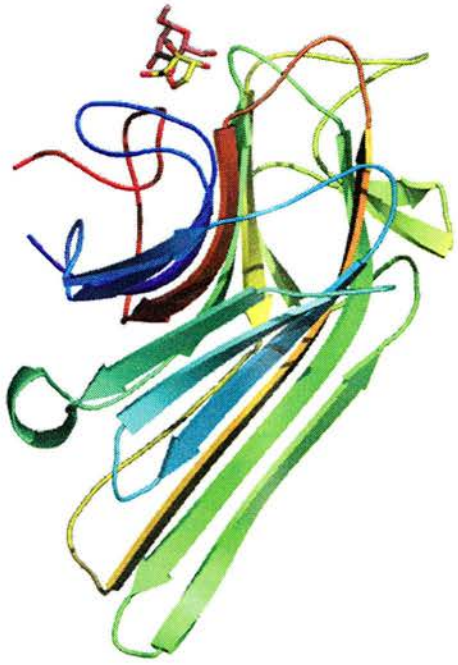
Figure 2.08 Overlay of the native (violet) and substrate bound (blue)Con-A loop containing residues 175- 192, showing the changing positions of the Trp 182 residue side chain.



2.09 The saccharide binding site.

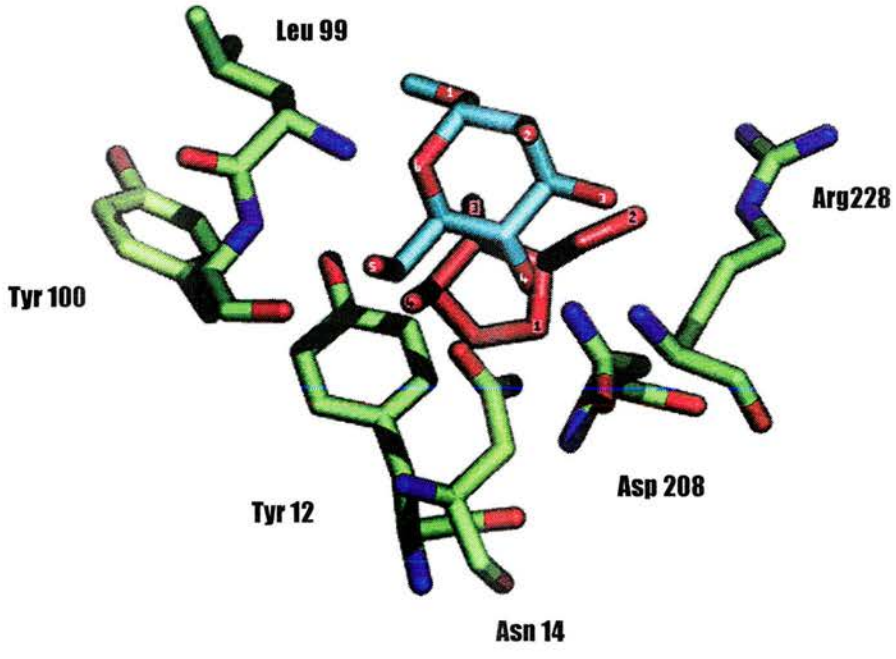
Each monomer has a saccharide binding site located in a shallow cleft at the surface of the protein (Figure 2.10). The experimental electron density map allowed the modelling of one L-arabinofuranose molecule into the binding cleft on the A subunit. The ligand molecule is positioned into a binding pocket consisting of residues Tyr 12, Asn 14, Leu 99, Tyr 100, Asp 208, and Arg 208.The L-arabinofuranose ligand sits in a completely different orientation within the binding cleft when compared to the multiple structures of Con-A pyranose ligand structures previously solved.

Figure 2.09 The A subunit with the modelled L-arabinofuranose molecule in superposition with the methyl-α-D-mannopyranoside molecule within the Con-A binding cleft.



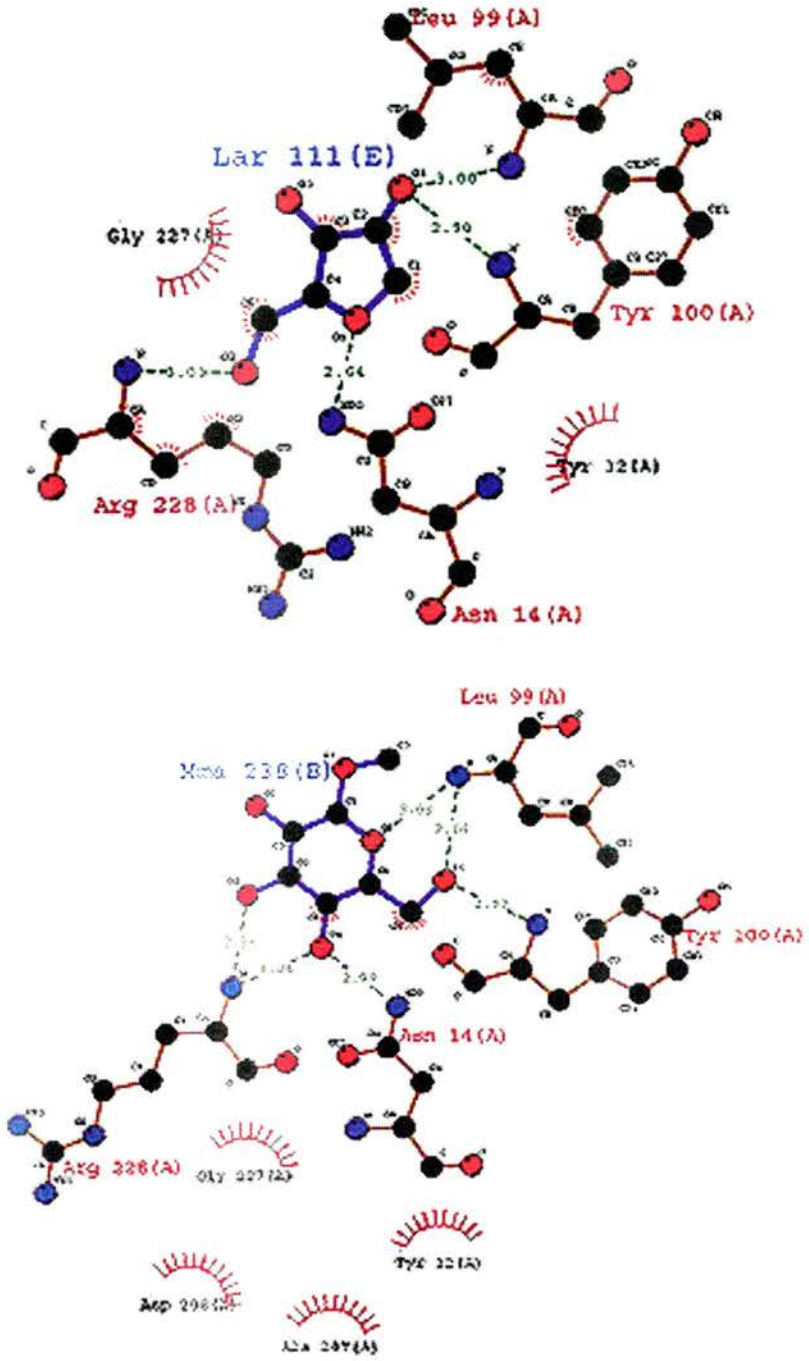
The smaller furanose ring is located deeper within the cleft with the O3 ring oxygen pointing towards the Asp 208 residue (Figure 2.10).

Figure 2.10 Overlay of methyl-a-D-mannopyranoside (blue) and l-arabinofuranose (violet) binding orientation.



The orientation of the furanose ring within the binding cleft is completely unexpected, with the ring oxygen (O1) mimicking the O4 contact of the pyranose bound structures, the O2 mimicing the O3 contacts and the O4 of L-arabinose is positioned to mimic the contacts of the O5 pyranose model (Figure 2.10).

Figure 2.11 Ligplot of both the pyranose (Mma) and furanose (Lar) models of con-A binding.



Key

Ligand bond

Non-ligand bond

Hydrogen bond and its length

His 33 Non-ligand residues involved in hydrophobic contact(s)

Corresponding atoms involved in hydrophobic contact(s)

The O3 oxygen points out of the binding cleft making no close contacts and the C1 hydrophobic patch is orientated towards the Tyr 12 residue (Figure 2.10).

The complete reorientation of the furanose ring within the binding cleft automatically disrupts the Goldstein rules on Con-A saccharide binding (Goldstein, 1974). The new orientation, although it may be somewhat forced with an excessive concentration of substrate, buries the saccharide deeper within the binding cleft with several close contacts made with the binding site key residues (Table 2.05).

Table 2.05 Contacts between L-arabinofuranose ligand and binding site residues.

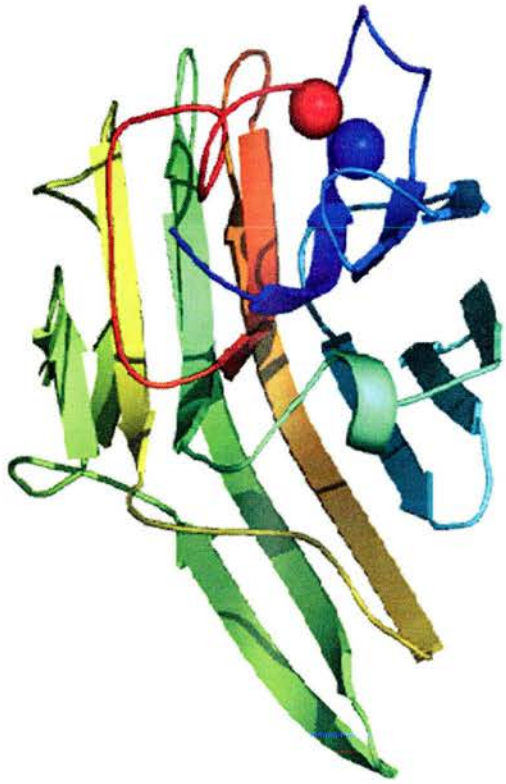
C1	TYR 12 CD2	2.74 Å
C1	TYR 12 CE2	2.73 Å
C1	TYR 12 2CG	3.13 Å
C1	TYR 12 CZ	3.13 Å
C4	ASN 14 ND2	3.09 Å
C5	ARG 228 N	3.00 Å
O1	ASP 208 OD2	2.99 Å
O4	LEU 99 N	3.00 Å
O4	TYR 100 N	2.90 Å
O4	ASP 208 OD1	2.71 Å
O1	ASN 14 ND2	2.64 Å
O2	ARG 228 N	3.01 Å

From table 2.05 we can identify the hydrophobic /van der waals interactions of the C1 atom and the Tyr 12 residue. The O1 and O4 members of the ligand provide a bidentate hydrogen bond between the ligand and the Asp 208. O4 makes further hydrogen bonds with the Leu 99 and the Tyr 100 residues making the same number of hydrogen bonds as the pyranose ligand O6 position. O1 also mimics the O4 oxygen on the pyranose ligands and makes two identical hydrogen bonds to Asn 14 and Asp 208. The O2 group makes 1 hydrogen bond to Arg 228 and a further close contact at Arg 228 further mimicking the O3 of the pyranose ring structures.

2.10 The metal binding sites.

Each Con-A monomer contains 2 metal sites named S1 (red) and S2 (blue). In the l-arabinofuranose structure the S1 site contains the Mn^{2+} and the S2 site contains the Ca^{2+} ion. The Manganese ion has an octahedral coordination and the Ca^{2+} ion has a 6 coordinate geometry. Both metal sites are as described in all of the previously published Con-A saccharide complexes (Kalb, 1968; Edelman, 1972; Hardman, 1972; Hardman, 1976; Derewenda, 1989; Naismith, 1994a; Harrop, 1996; Kanellopoulos, 1996; Loris, 1996; Naismith, 1996; Moothoo, 1998, 1999).

Figure 2.12 The Con-A monomer with the S1 (red) and S2 (blue) metal binding sites highlighted as coloured spheres.



2.10 Conclusions.

The comparisons between the 5 and 6 membered ring formations in ligand binding have produced some surprising results, the author would like to note that at 150 Mm l-arabinofuranose, the binding of the ligand is more likely the result of the extremely high concentration of the l-arabinose and should not be considered of any physiological relevance. The Gibbs free energy equation describes the amount of the free energy in a system as the sum of changes in enthalpy minus the temperature times the sum of the changes in entropy ($\Delta G = \Delta H - T\Delta S$). Gibbs free energy can also be described as $\Delta G = R\ T\ \ln Kd$ or $\Delta G = -R\ T\ \ln Ka$. Where R is the ideal gas constant, T the absolute temperature, ln Kd the natural log of the dissociation constant and ln Ka the log of the association constant. Thus allowing us to measure most association/dissociation constants for a variety of ligands and proteins. Most lectin - carbohydrate interactions are entropically unfavourable in water (Toone, 1994). Isothermal titration calorimetry studies have indicated that a significant proportion of the of the enthalpy arises from the transfer of solvent molecules between the bulk of the solvent and the solvation shell of the protein and ligand (Toone, 1994), with the favourable entropic contribution being provided by the expulsion of waters from the binding site and exclusion of the ligand from the bulk solvent.

The methyl-α-D-mannopyranoside structure indicates that binding of the pyranose ligand produces 8 direct hydrogen bonds, with 3 further hydrogen bonds provided via networked water molecules and 15 van der Waals contacts stabilising each ligand upon binding. In the furanose model there are only 6 direct hydrogen bonds with 1 network hydrogen bond from a water molecule and 7 van der Waals contacts, suggesting that the ligand is not as effectively shielded from the bulk solvent decreasing the favourable entropic effect. The loss of 2 direct hydrogen bonds and the additional water networked hydrogen bond, further decreases the stability of the furanose ligand, reducing the bond energies by 30-90 kJ/mole. The Con-A l-

arabinose structure suggests a considerably weaker binding affinity than the approximate Ka = 0.8 x10^4 M^{-1} for the pyranose ligands. A theoretical binding association constant could be worked out by using the equation

$$Ka = \frac{[EL]}{[E] \times [L]}$$

and therefore a theoretical ΔG using the concentration of the ligand (150 mM) and assuming from the crystal structure that only 25% of the enzyme in solution is bound. Therefore the Ka for Con-A and l-arabinose could be 0.0025/0.010 x 0.150 = 1.67 M^{-1} and the theoretical ΔG could then be written as the sum of (-R T ln Ka) which is -1.27 kJ mol^{-1}. Both of these theoretical values are far from being scientifically relevant, but would support the structural data in suggesting l-arabinose would have a weaker association constant. However there is no kinetic evidence to support this. The possibility of a weaker binding affinity may go some way to account for the poorer density regions at the B, C, and D subunits, and explain why it was necessary to co-crystallise the protein with 150 mM l-arabinofuranose before crystals with bound substrate could be produced.

Chapter 3

The 1.90 Å resolution structure of 2-keto-3-deoxy-6-phosphogluconate aldolase from *Thermotoga maritima*.

3.00 Summary.

The *Thermotoga maritima* KDPG aldolase employs a Schiff base mechanism *via* an active site catalytic lysine. Structural studies performed on this enzyme show that the *Thermotoga maritima* KDPG-aldolase has the classical α/β barrel structure of its monomer and the trimeric propeller conformation seen in *Escherichia coli* KDPG aldolase. Structural studies also identify the active site lysine, a putative phosphate / sulphate binding site and a covalently trapped Schiff base intermediate. The *Thermotoga maritima* KDPG-aldolase has potential in enantio-selective bio-catalysis, it has almost all the active site residues completely conserved in relation to its close relative, the *Escherichia coli* KDPG-aldolase. Thus it should be possible to repeat the biocatalytic / molecular biological approaches previously reported, allowing not only the enantio-selective synthesis of products from non chiral starting materials but also the directed evolution of the enzyme to allow a greater variety of substrates for synthetic organic chemistry. The *Thermotoga maritima* KDPG aldolase has an optimum reaction temperature of 70-80°C. The increased thermal stability of the *Thermotoga maritima* KDPG aldolase enables the comparison of changes in protein structure between two closely related enzymes, which function at temperatures 40 °C apart.

This chapter describes the expression purification and crystallisation of the *Thermotoga maritima* KDPG aldolase. We report on the collection and processing of crystallographic data, the solution of the protein structure with molecular replacement and refinement of the final model, and report on the aldolase enzyme, its structure / function relationship, and discuss substrate binding and structural thermostability.

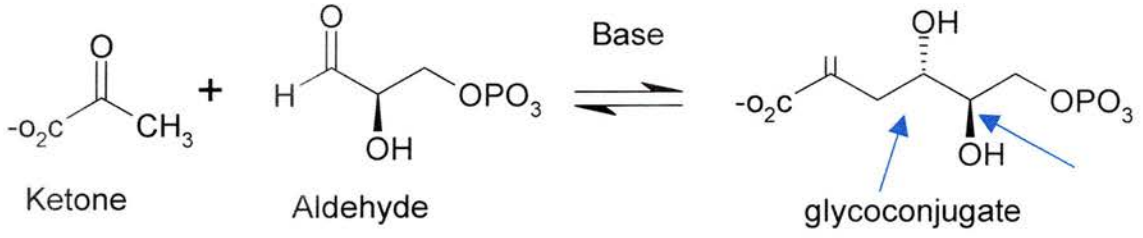
3.01 Introduction.

This chapter reports the protein structure of a new KDPG aldolase enzyme. The enzyme was isolated from bacteria found in heated aquatic sub-soils; *Thermotoga maritima* (Nelson, 1999; Griffiths, 2001). Several bio-catalysts have been recently studied with the hope of increasing the efficiency and scope of current chemical synthetic techniques (Gijsen, 1996; Takayama, 1997; Hoffmann, 1998; Zhong, 1998; Zimmermann, 1999; Koeller, 2000; Cordova, 2002). Included in these new bio-catalytic molecules, the aldolases have shown much promise, not only in their ability to enhance the efficiency of traditional synthetic chemical steps, but also in the reported re-engineering of the active site allowing a larger range of substrates (Buchanan, 2001). Aldolase enzymes are concerned with the production / degradation of various glyco-conjugates from aldehydes and ketones (Gijsen, 1996).

3.02 The Aldol Reaction.

The aldol reaction creates / breaks a bond between the α carbon atom of an aldehyde or ketone, to the carbonyl of another. The aldol reaction can create two adjacent stereogenic centres (⟶) , resulting in 4 stereo-isomers (Figure 3.01).

Figure 3.01. The aldol condensation.



3.03 Biocatalysis.

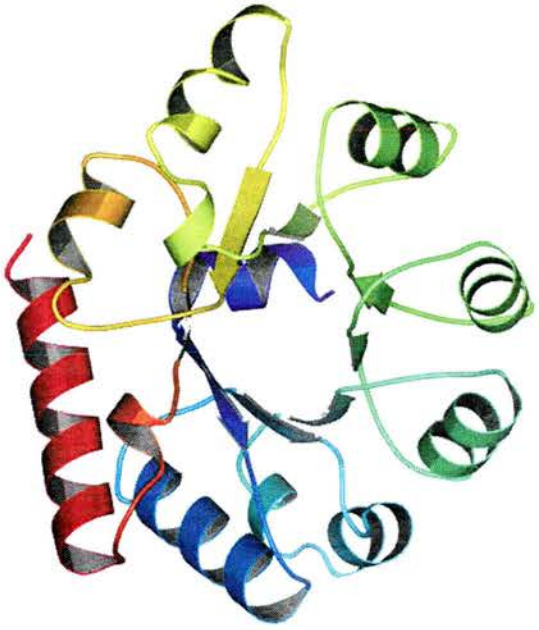
Biocatalysts have the potential to allow the chemist to synthesise a single desired enantiomeric product. Previous enzymatic catalysis of chiral and stereo-specific products have for the most part, been researched in the production of small quantities of fine chemicals. The use of biocatalysts in aldol reactions centers around the stereo-specific formation of the C-C bond. In traditional synthetic chemistry C-C bond formations are only enantio-selective when stoichiometric quantities of the desired chiral starting molecules are used (Mukaiyama, 1982). Regio-selectivity increases further traditional synthetic chemistry problems when handling the presence of reactive side groups such as alcohols, aldehydes, and ketones. The presence of such side groups increase the need for the synthetic chemist to protect and de-protect specific side groups throughout the synthetic steps to provide the desired regio-selectivity. The use of aldolases as biocatalysts is not a new idea (Gijsen, 1996; Koeller, 2000; Koeller, 2001) and results so far have been encouraging. The use of aldolase enzymes in synthetic processes as bio-catalysts have three major benefits to traditional synthetic chemistry. Aldolase catalysed aldol reactions are remarkably enantio-selective. Enantiomeric products can be produced by the addition of non-chiral materials reducing the cost of purified chiral starting materials. Aldolase bio-catalysed aldol reactions are completely regio-selective, removing the need for expensive protective / de-protective synthetic chemistry steps, and aldolase enzymes are active in relatively mild conditions ranging from 30-80°C with pH from 4-8 in water.

3.04 Aldolase enzymes.

The aldolase enzymes fall into two distinct categories, the structural differences are believed to be the product of a very early evolutionary split. The class I enzymes are present in all

living organisms and carry out complex carbohydrate chemistry in two major metabolic pathways: Entner-Dodouroff and Ebden-Mayerhoff-Parnas (Entner, 1952; Conway, 1992). Each of these pathways are centrally involved in cell glycolysis to provide ATP and building materials for cell survival (Wang, 2001). The *Esherichia coli* and *Pseudomonas putida* KDPG aldolases have an (α/β) TIM barrel structure (Walden, 2001). The structure contains a core of distorted β-strands arranged in a tightly packed cylindrical manner, closely resembling the shape of a barrel. The central "barrel" β-strands are connected by α-helices positioned around the circumference of the internal β strands (Figure 3.02).

Figure 3.02 ***Escherichia coli*** **KDPG aldolase monomer.**



The ability of aldolase enzymes to perform stereo-specific C-C bond addition, makes the understanding of aldolase enzyme mechanisms necessary to continue exploring enzyme mediated organic synthesis of chirally pure products. The remainder of this chapter discusses the experimental procedures used to solve the solution of a new class 1 KDPG aldolase

structure, discusses the results and proposes a reaction mechanism based upon a trapped covalent intermediate and substrate modelling.

3.05 Experimental.

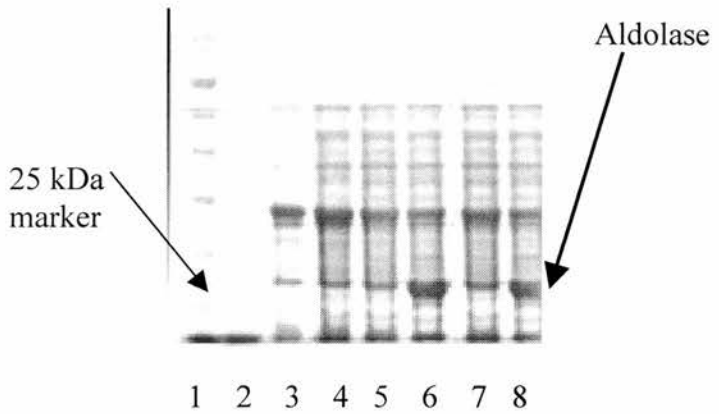
3.06 Protein expression.

The pET28-b plasmid (Griffiths, 2001) containing the Tm-0066 (Nelson, 1999) aldolase gene from *Thermotoga maritima* was supplied in DH5α *Escherichia coli* cells from E.J.Toone (Duke University). The DH5α cells were cultured in 10 mls Luria Bertani broth overnight at 37°C under 70μg/ml kanomycin resistance selectivity. The cells were harvested by centrifugation at 3000 rpm for 30 minutes. The cell pellet was washed with 10 mM Tris pH 7.0 and the Plasmid extracted *via* Quiagen miniprep kit (Quiagen, 2002).

The plasmid was inserted into competent BL21 DE3 cells *via* heat / cold shock method (Maniatis, 1989). 1μl of plasmid was added to 500 μls of competent BL21 DE3 cells and left for 30 minutes at 4°C. The cells were transferred to a 42°C heated water bath for 1 minute before returning to 4°C for a further 30 minutes. 200 μl of the cells were plated onto 2 ml agar plates containing 100 ug/ml Kanomycin overnight and viable colonies were then chosen for expression trials.

10 ml overnight growth cultures of the BL21 DE3 cells in Luria Bertani broth (Maniatis, 1989) containing 100 μg/ml kanomycin, were used to inoculate 10x 0.5 litres of Luria Bertani broth containing 100 μg/ml kanomycin (1ml / 500mls). The inoculated flasks were incubated at 37°C in shaker incubators at 250 rpm until an optical density (OD 600nm) of 0.8. The expression of KDPG aldolase from *Thermotoga maritima* was induced by the addition of 0.5 / 1.0 mM IPTG. The cells were harvested after three hours by centrifugation at 16,000g for 15 minutes producing cell pellets of approximately 15 grams wet weight. Initial SDS-PAGE gel analysis confirmed over expression of a band at 23 kDa (Figure 3.03).

Figure 3.03 Over-expression of KDPG Aldolase in BL21 DE3 cells.



Lane 1 contains the molecular weight markers. Lane 2 was run as a blank. Lane 3, 4 and 5 contain cell lysate from non-induced cell lines. Lane 6 contains Induced cell line (0.5 mM IPTG). Lane 7 contains non-induced cell line. Lane 8 contains induced cell line (1mM IPTG).

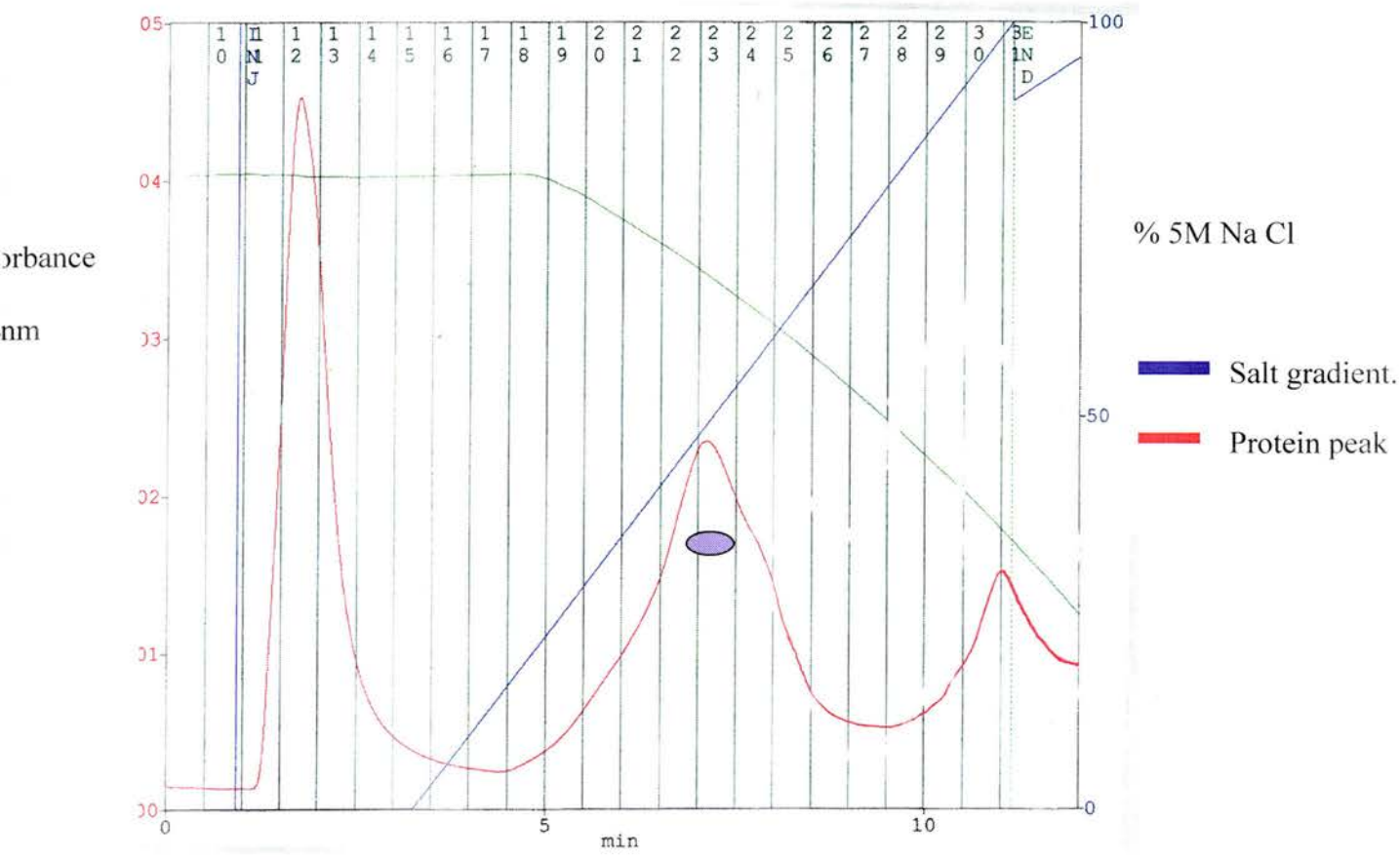
3.07 Protein purification.

The cell pellet was resuspended in 40 mls of 20 mM Tris pH 7.0, 0.1mM PMSF before sonication for 6, 1 minute bursts at 4°C. The cell debris was removed *via* centrifugation at 40,000g for 30 minutes. The clarified supernatant was separated into 3 ml aliquots before being incubated in a water bath at 75°C for 30 minutes. The resultant precipitant was removed by centrifugation at 40,000g for 30 minutes. The clarified supernatant was then dialysed in 3 changes of 4 litres of 20 mM Tris pH 7.0. The dialysed solution was loaded onto a Poros high pressure anion exchange column.

The column was pre-equilibrated with 15 column volumes of 20 mM Tris-HCl pH 7.0. The dialysed aldolase solution was loaded onto the column in 5 ml aliquots and washed with a further 15 column volumes of 20 mM Tris pH 7.0. The aldolase peak was eluted with a high

salt gradient from 0 % 5 M NaCl, 20 mM Tris pH 7.0 to 100 % 5 M NaCl, 20 mM Tris pH 7.0. The Aldolase peak eluted at approximately 40-50 % 5 M NaCl, 20 Mm Tris pH 7.0 (Figure 3.04).

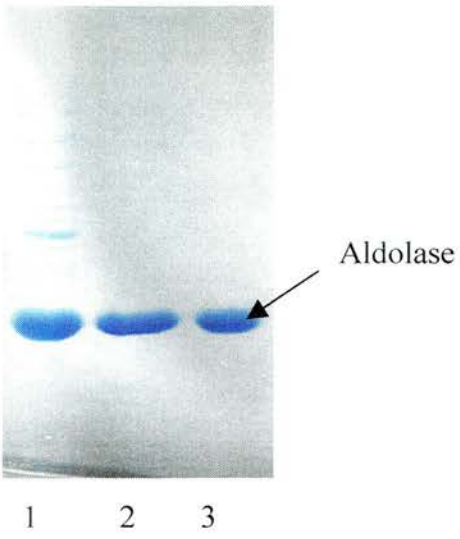
Figure 3.04 HPLC anion exchange chromatography purification of KDPG aldolase.



The KDPG aldolase peak elutated from the column at approximately 40-50% 5M NaCl.

The peak fractions ( ) 22 and 23 were analysed on SDS-PAGE to confirm molecular weight and purity (Figure 3.05).

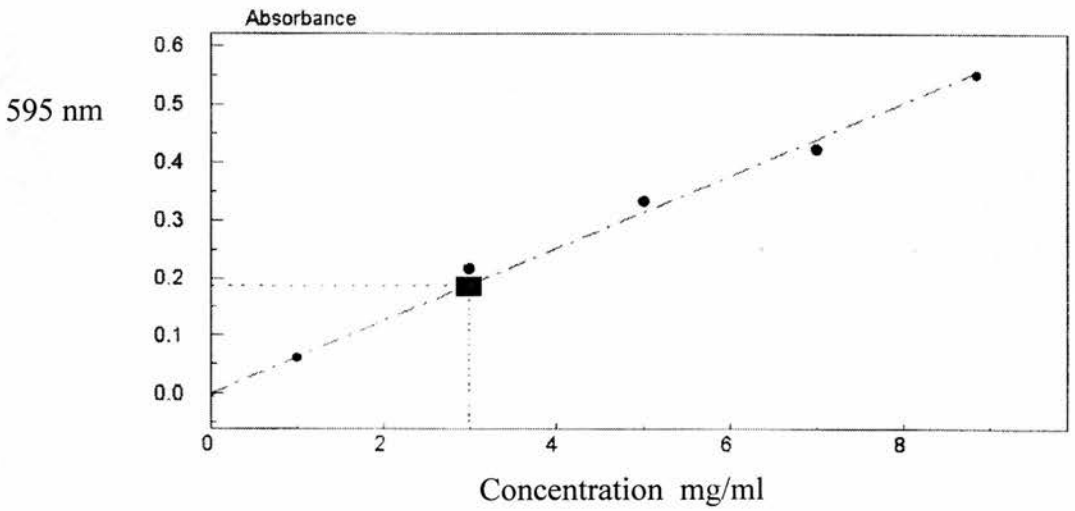
Figure 3.05. SDS-PAGE analysis of KDPG aldolase purification.



Lane 1 contains the heat-treated supernatant. Lanes 2 and 3 contain peak fractions 22 and 23 from the anion exchange purification step.

The protein samples from lanes 2 and 3 (fractions 22 and 23) were pooled and dialysed against 4 changes of 1 litre of 50 mM Tris pH 7.0 for 40 minutes at 4°C. The dialysed protein sample was concentrated into 3, 5, and 7 mg/ml. The protein concentration was determined by Bradford assay (Figure 3.06). 64 mg of pure protein was produced from the 5 litre fermentation process. The identity of the KDPG Aldolase was unequivocally determined by trypsin digest MALDI TOF spectrometry. The protein samples were flash frozen in liquid nitrogen in 100 µl aliquots before storage at –20 °C.

Figure 3.06. Bradford assay determination of protein sample concentration.



1, 3, 5, 7, and 9 mg/ml BSA standards. ●

Aldolase absorbance. ■

From the comparison of the absorbance of the aldolase sample and the absorbance of the BSA (bovine serum albumin) standards the approximate protein concentration can be determined. In the above case the aldolase sample had an approximate concentration of 2.98 mg/ml.

3.08 Crystallisation procedure.

3, 5, and 7 mg/ml aldolase solutions were screened on Hampton I and II screening conditions, using the sitting drop method at 22°C. Initial screens at 5 and 7 mg/ml concentrations produced small needle crystals after three days in two Hampton's I conditions.

Condition 20: 0.2 M ammonium sulphate, 0.1 M sodium acetate trihydrate pH 4.6, 25% w/v PEG 4000.

Condition 47: 0.1 M sodium acetate pH 4.6, 2.0 M ammonium sulphate.

Optimisation of crystal trials produced larger hollow rod shaped crystals after 3-4 days using the hanging drop method with 2µl mother liquor and 2µl of 5 mg/ml protein solution drops at 22°C, in 0.075 M sodium acetate trihydrate, 0.1 M ammonium sulphate, 28% w/v PEG 4000 (Figure 3.07).

Figure 3.07 A picture of the Aldolase crystals grown at 22°C.

A photograph of the hollow rod shaped crystals grown at 22°C. Each rod is approximately 2 mm long. Larger solid rods also appeared in the same conditions at 37°C (Figure 3.08).

Figure 3.08 A picture of the Aldolase crystals grown at 37°C.

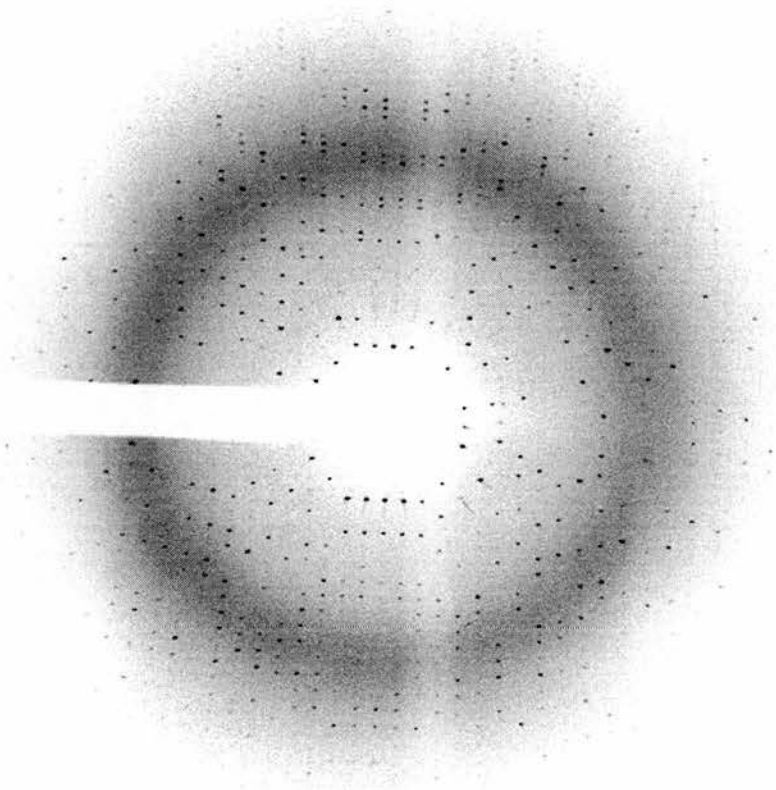
In house diffraction was undertaken on an ENRAF FR-591 Nonius X-ray generator with a Dip-2000 detector and osmic mirrors. Crystals were cryo-protected in 50% PEG 600 before being looped and mounted under a nitrogen cryo-stream, with a crystal to detector distance of 200 cm, and diffraction data collected from a 1.54 Å X-ray source from a rotating copper anode. Diffraction of the smaller hollow rods from the 22°C trial produced a 4 Å resolution diffraction pattern. The 37°C crystals failed to diffract.

3.09 Data Collection.

Diffraction data were collected on station ID14-2 using an ADSC Quantum-4 CCD detector, at the ESRF synchrotron, Grenoble, France. Crystals were cryo-protected in 50% PEG 600

prior to data collection at 100K. After several failed attempts to collect high-resolution diffraction data, a comparable crystal with no visually discernable differences was mounted on ID14-2 (λ=0.932) and diffracted to 1.9 Å with a crystal to detector distance of 183 mm (Figure 3.09).

Figure 3.09. The 1.9 Å diffraction pattern from the *Thermotoga maritima* KDPG Aldolase.



The crystal diffracted to 1.90 Å with a $P2_1$ space group and unit cell parameters of a= 42Å b= 101Å c=124 Å with $\alpha = 90°$ $\beta = 97°$ $\gamma = 90°$. Data were recorded as 300 images with a 10 second exposure and 0.5 degree oscillations. The data were indexed and integrated with the

data processing programme MOSFLM (Leslie, 1992). The resultant structure factor file was merged with Scala (Evans, 1993) to a resolution of 1.90Å.

Table 3.01. Crystal diffraction data.

Resolution Å	No. of measured Reflections.	Completeness %	R-merge %	Redundancy
6.01	9395	95.3	4.3	3.7
4.25	17982	95.5	4.2	3.9
3.47	23301	95.2	4.6	3.9
3.00	27932	95.1	5.7	4.0
2.69	31814	95.0	7.1	4.0
2.45	35016	94.9	9.0	4.0
2.27	37825	94.7	11.2	4.0
2.12	40243	94.6	15	4.0
2.00	42297	94.4	19.8	4.0
1.90	43947	94.3	29.6	3.9
32.97-1.90	309752	94.3	7.2	4.0

3.10 Structure solution.

The structure of the KDPG aldolase from *Thermotoga maritima* was determined by molecular replacement using the previously solved *Escherichia coli* KDPG aldolase trimer (Buchanan, 2001) as the model for the CCP4i, programme MOLREP (Vagin, 1997). The model trimer was converted into a poly-alanine model before structure solution.

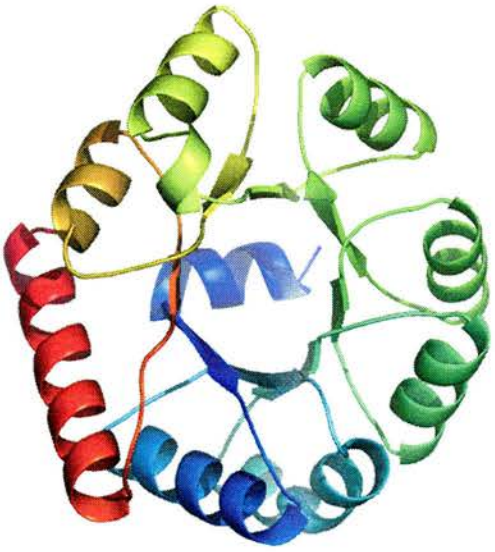
Table 3.02. Molecular replacement solutions.

Solution	α	β	γ	Correlation coefficient	R_{Factor} %
1	204.14	63.64	261.66	0.606	39.6
2	178.18	86.48	28.34	0.385	49.6

The Molrep programme identified two trimers within the asymmetric unit (Molrep solutions are shown in table 3.02).

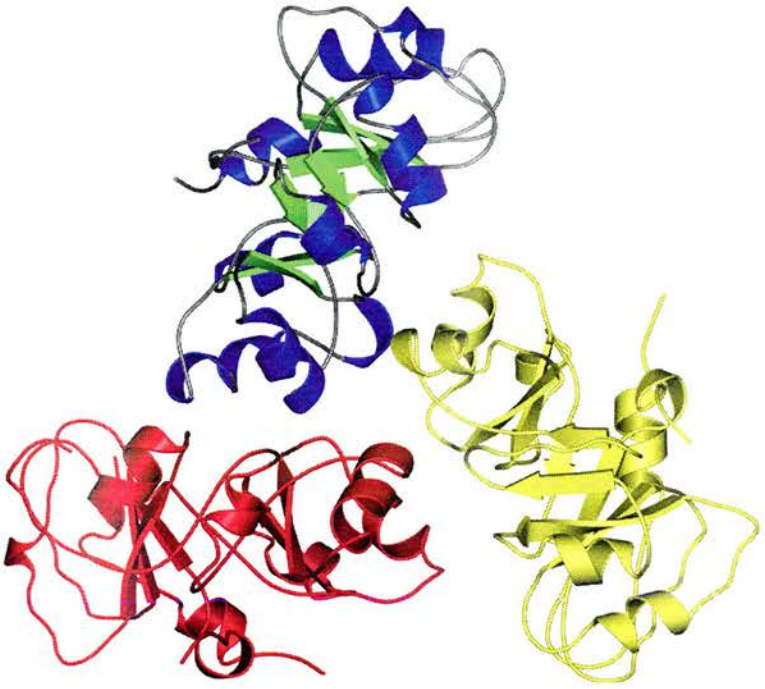
The solved structure indicated that the *Thermotoga maritima* KDPG aldolase was a trimer (Figure 3.11) consisting of three TIM barrel monomers (Figure 3.10). With each monomer bound to two others forming a propeller shape.

Figure 3.10 *Thermotoga maritima* KDPG aldolase monomer; exhibiting the classical TIM barrel conformation.



Each monomer contains 10 alpha helices and 8 beta strands in a TIM barrel conformation.

Figure 3.11 The propeller trimer as seen in the *Escherchia coli* KDPG aldolase structures.



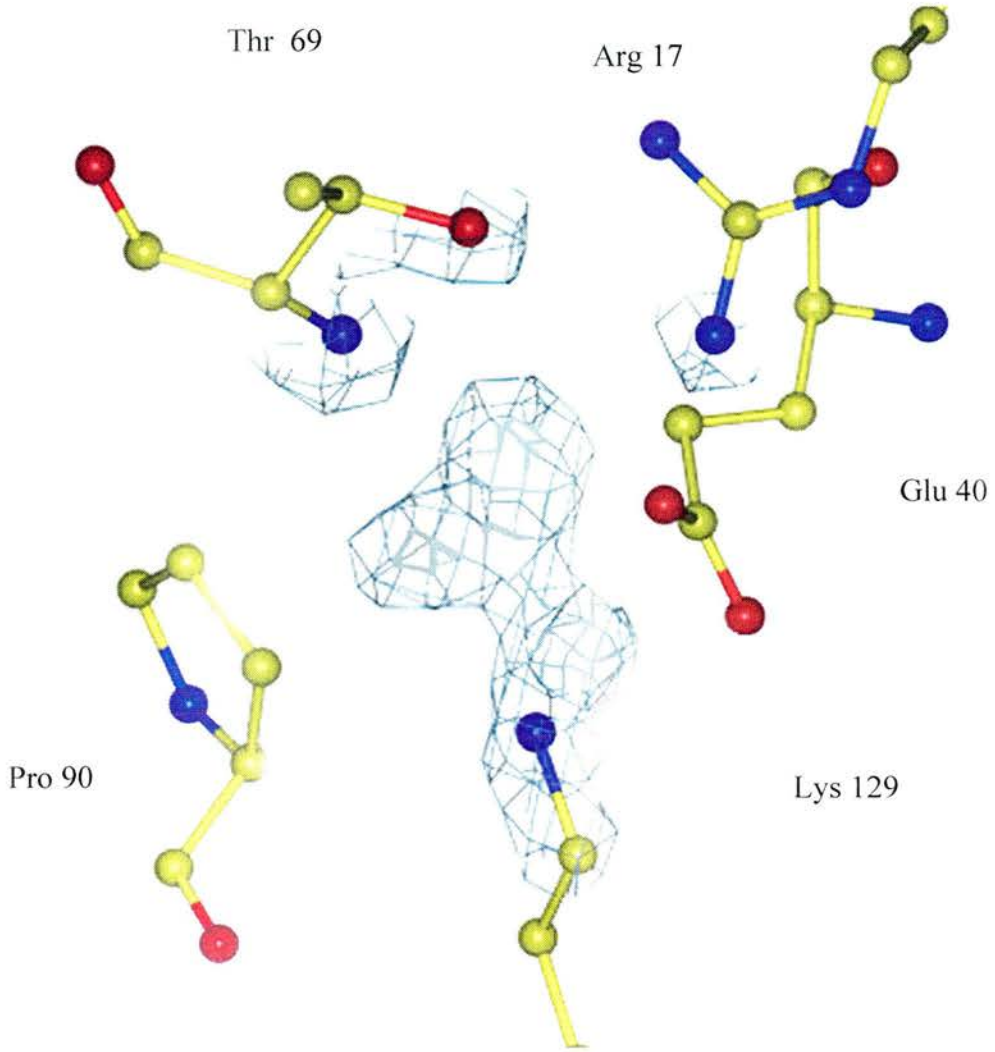
As can be seen each monomer makes contact to both of the other components in the trimer.

3.11 Refinement.

Prior to formal refinement 5% of the data was separated using the CCP4 "Uniqueify" programme to set 5% of the reflection data free from refinement calculation (Bruenger, 1995), but which has an $R\text{-}_{factor}$ calculated throughout the refinement. This gives a Free-R_{factor} allowing the refinement quality to be statistically monitored. The poly-alanine models were fitted and adjusted to the experimental density maps manually in conjunction with the "O" graphics package (Jones, 1991). The amino acid sequence and side chains were positioned in regions of positive density. Those side chains or other atoms that were positioned in areas without density, had their occupancy levels set to 0.01.

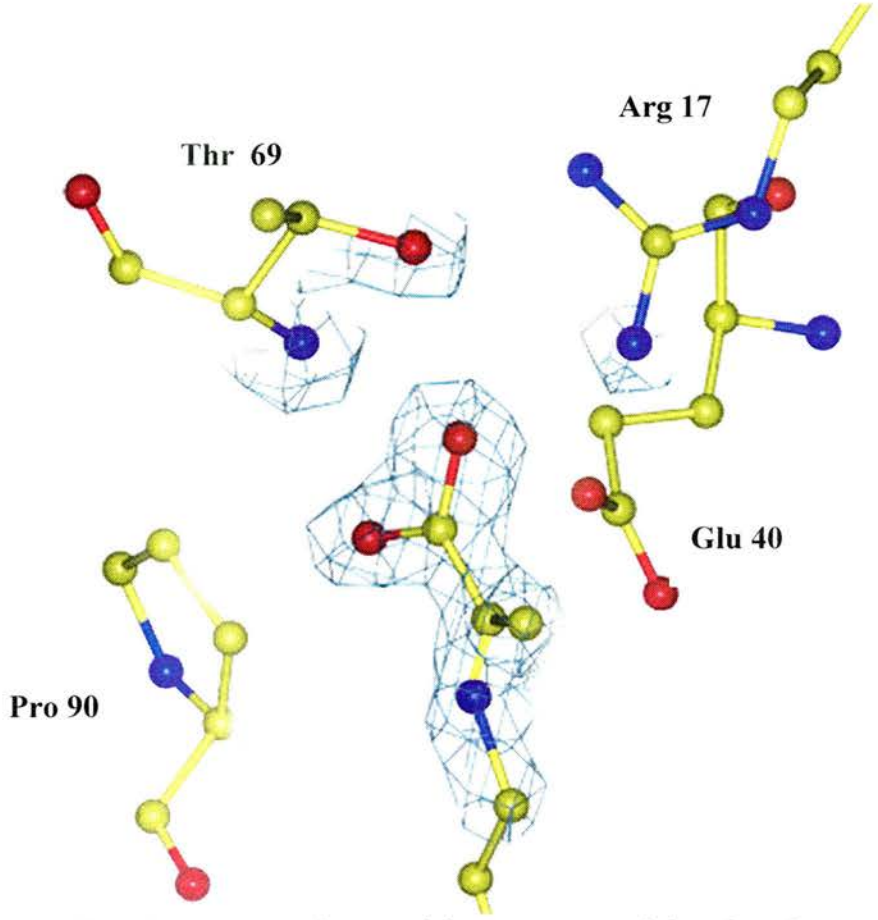
The models were then refined by the CCP4 programme Refmac 5 (Murshudov, 1997). The models were first refined with Rigid Body refinement. The models were then subjected to restrained overall isotropic B-factor refinement, before further manual correction (No NCS refinement was carried out throughout refinement). Several cycles of restrained refinement and manual correction were carrier out until a Free-R_{factor} of 27 %. It was clear from the experimental density maps that on each of the six putative active site lysines (lys 129) an extended region of density was present. (Figure 3.12)

Figure 3.12 A schematic of the active site residues showing an 2Fo-Fc map density at 1.0 σ contour level, with an extended area of positive density protruding from the putative active site lysine.



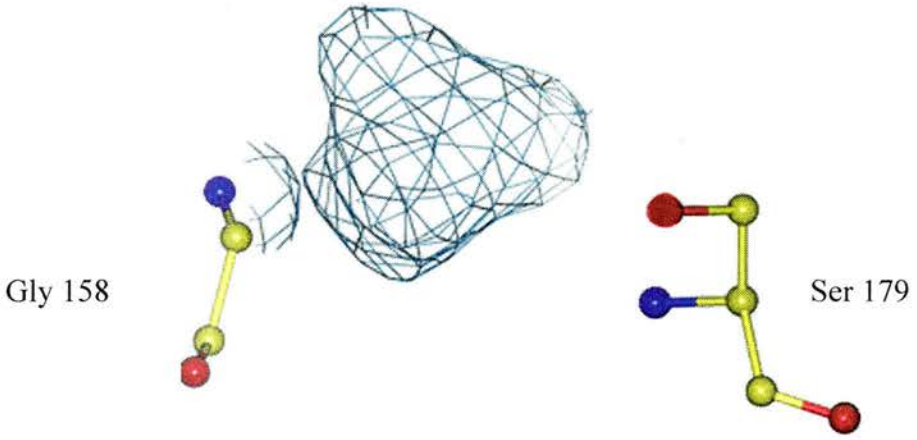
Further analysis of the density proposed a bound compound to the active site lysine. At this point ligand modelling took place. A pyruvate molecule PDB file was taken from the CCP4 ligand library. 6 copies of the molecule were then placed into the KDPG aldolase PDB file. Using the graphical package "O" each of the 6 pyruvate molecules were manoeuvred along the X, Y and Z coordinates of the model PDB file until positioned within the appropriate density (Figure 3.13).

Figure 3.13 A schematic of the active site residues with the pyruvate molecule modelled into the appropriate 2Fo-Fcdensity region at 1.0 σ contour level.



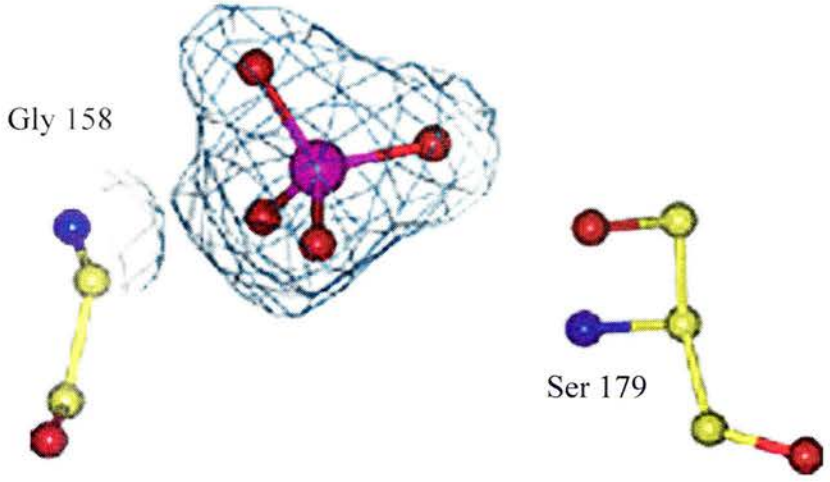
After modelling the pyruvate into positi gnising that there was a covalent bond between the Pyruvate C 2 and the Lysine NZ atom. Further refinement with CCP4 Refmac-5 restrained isotropic refinement continued. The addition of the pyruvate and the covalent bond aided the decrease in the R-$_{Factor}$ and FreeR-$_{Factor}$ in the following successive rounds of refinement. Once the R-$_{Factor}$ and FreeR-$_{Factor}$ had converged at 20.6% and 25.3% respectively, waters were added using the EMBL programme ARP-wARP (Perrakis 1999). Further refinement reduced both the Free R-$_{Factor}$ and the R-Factor to 19.2% and 24.1% respectively. Manual inspection of the water molecules discovered an area of density out-with the active site, which contained 4 water molecules. Removal of the waters proposed a density fit for a phosphate or sulphate ion (Figure 3.14).

Figure 3.14. The unexpected density with the 4 waters molecule removed.



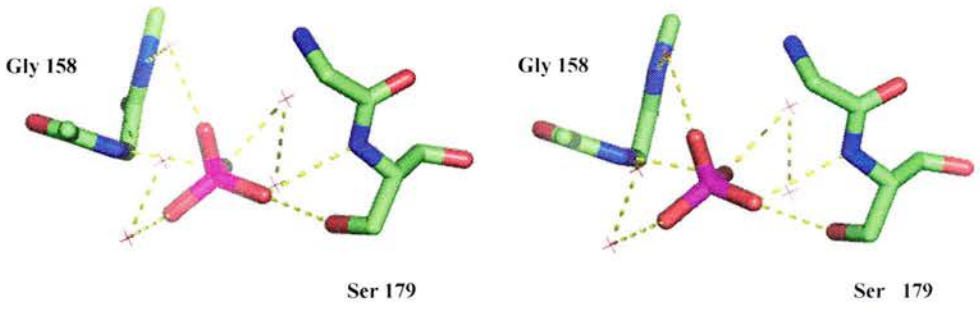
The unexpected density sits outside of the active site adjacent to residues A 158 and A 179. Each of the six monomers within the 2 trimers contained the same density in the same position. It was obvious this was not the result of some errant data. A sulphate molecule was modelled into the experimental density map for all 6 monomers (Figure 3.15) because of the high concentration of sulphate in the crystallisation conditions.

Figure 3.15. The sulphate model.



The phosphate ion is held in place by three direct hysrogen bonds, 1 to the Gly 158 N atom another to the Ser 179 N atom and the third contact to the Ser 179 hydroxyl group. A further 3 hydrogen bonds are constructed between networked water molecules (Figure 3.16).

Figure 3.16 A Stereo Representation of the Phosphate Ion and its Contacts with the Waters, Glycine 158 and Serine 179.



The removal of errant water molecules, the positioning of the sulphate ion and further cycles of refinement and manual manipulation of the KDPG aldolase model produced a final model with an R_{Factor} of 17.9% and a $FreeR_{Factor}$ of 22.9%. The model was statistically analysed with the CCP4 programme Procheck (Laskowski, 1993) before the structure was submitted to the Brookhaven PDB database. The final model statistics can be seen in Table 3.03.

The statistics from the Ramachandran plot (Ramachrisnan, 1965) are contained in Figure 3.17 and Table 3.04.

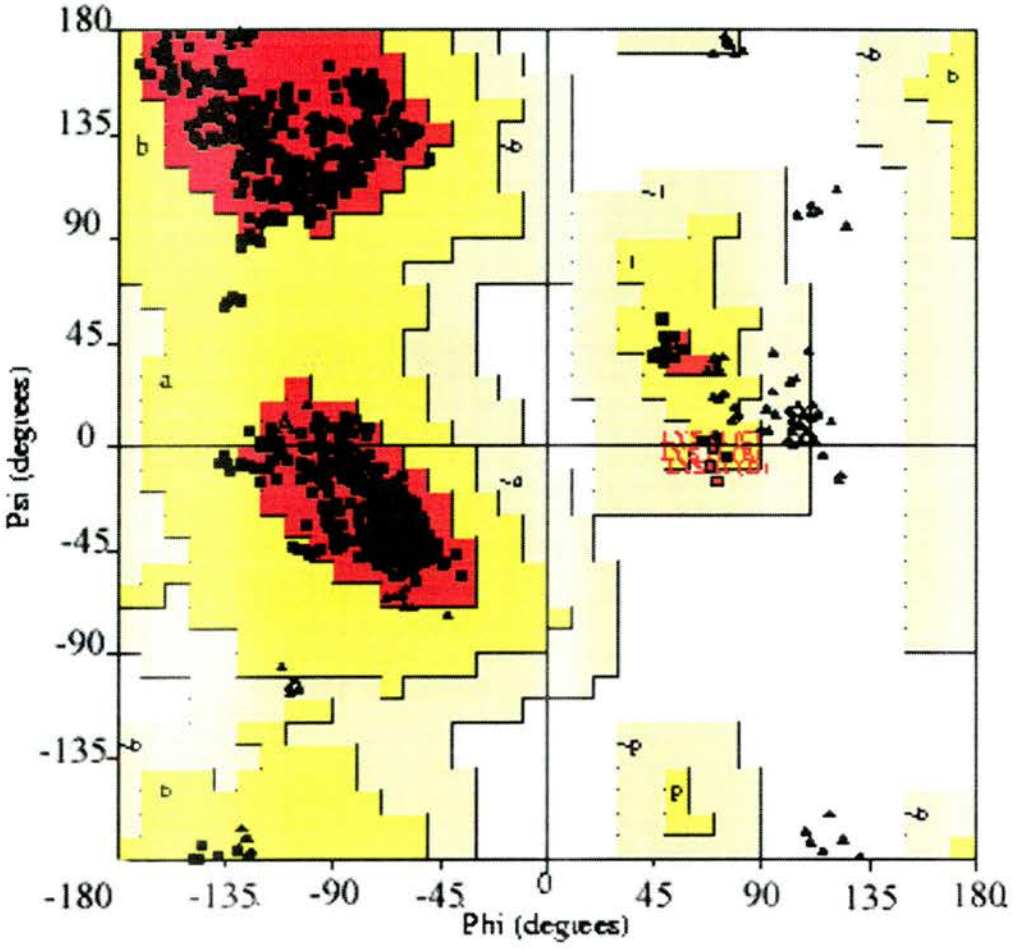
Table 3.03. Structural statistics.

Statistics	*Thermotoga maritima* KDPG Aldolase at 1.9 Å
R_{Factor}	17.9%
$FreeR_{Factor}$	22.9%
Average B-factor	18.4 $Å^2$
RMSD* bonds	0.021 Å
RMSD* angles	1.81°
Residues in most favoured region of Ramachandran plot	95.2%
Number of non-Hydrogen atoms	9156
Number of water molecules	369

Table 3.04. Ramachandran statistics.

Residues in most favoured regions	976	95.2%
Residues in additional allowed	3	0.3%
Residues in disallowed regions	0	0.0%
Number of non glycine; non proline residues	1044	100%
Number of glycine residues	144	N/A
Number of proline residues	60	N/A
Total number of residues	1218	N/A

Figure 3.17. Ramachandran plot.

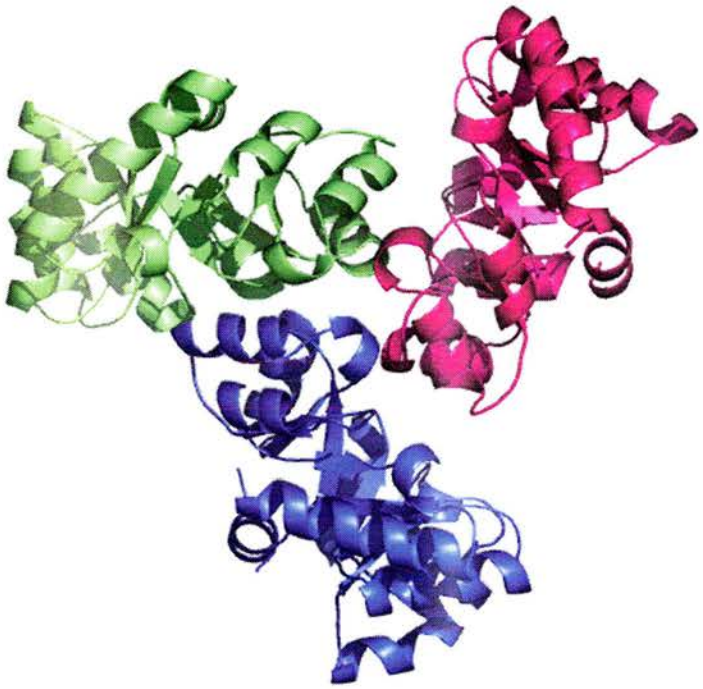


Glycine residues (not restricted to any particular region of the plot) are represented by a ▲, non glycine residues are represented by a ■. The darker the shaded region, the more favourable is the φ (Phi) ψ (Psi) combination.

3.12 Discussion.

As previously stated the *Thermotoga maritima* KDPG aldolase has a trimeric quartenery structure where each monomer makes contact with each of the other 2 monomers (Figure 3.18).

Figure 3.18 Model of the *Thermotoga maritima* KDPG aldolase trimer.



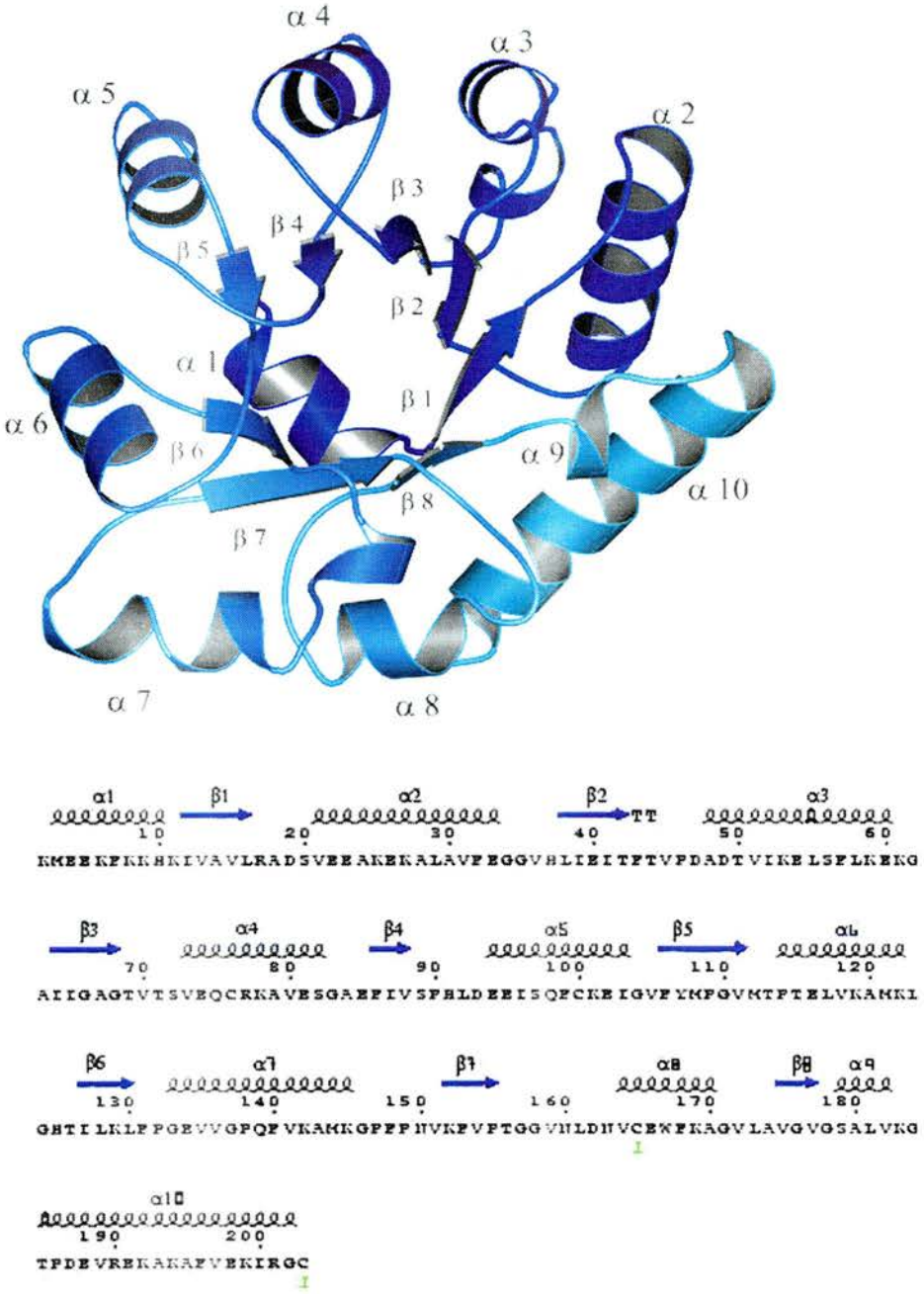
Both the monomer and trimer were analysed using the SC CCP4I programme (Lawrence 1993). Each trimer has 2163 surface atoms and 2436 buried atoms, with a solvent accessible area of 9899Å^2 out of a total 23598 Å^2 and each monomer has 752 surface atoms and 784 buried atoms with a solvent accessible area of 3485.67 Å^2 out of a total area of 8970.3 Å^2. The total area for the trimer is approximately 3500 Å^2 less than the sum of the three monomers. Therefore the formation of the trimer buries over 3000 Å^2, approximately 1100 Å^2 per monomer. The completed model of *Thermotoga maritima* KDPG aldolase trimer was further analysed on the Protein-Protein Interaction server, based at University College London (Figure 3.19).

Figure 3.19 Protein Interface statistics.

Protein Interface Parameter	Value
Interface Accessible Surface Area	610.93 $Å^2$
% Interface Accessible Surface Area	6.55
Planarity	1.30
Length & Breadth	26.78 & 20.61 Å
Length/Breadth Ratio	0.87
Interface Residue Segments	3
% Polar Atoms in Interface	31.25
% Non-Polar Atoms in Interface	68.70
Secondary Structure	Alpha
Hydrogen Bonds	3
Salt Bridges	0
Disulphide Bonds	0
Gap Volume	3889.96 $Å^2$
Gap Volume Index	3.14

The results of the analysis were similar to that of the *Escherichia coli* aldolase model. The low number of hydrogen bonds between the monomers and the larger proportion of hydrophobic residues situated at the interface, suggest that hydrophobic interactions are the main force for trimeric adhesion. The trimer consists of 3 identical TIM barrel (α/β barrel) fold monomers, (Figure 3.18) with each monomer consisting of 10 α-helices and 8 β-strands, linked together by 17 turns (Figure3.20).

Figure 3.20 Secondary structure elements.



Each of the monomers contains an N-terminal α helix that caps each α/β barrel as previously reported in the *Escherichia coli* KDPG aldolase, (Allard, 2000; Buchanan, 2001) the α1 helix figure 3.19. The beginning of the N-terminus and the end of the C-terminus have poor experimental density and therefore the model has been built with the first and last 2 amino acid residues missing. The loop region between residues 182-189 had poor experimental

density in all six monomers found in the asymmetric unit, although modelling of the α-carbon backbone was possible. The same loop region in the *Escherichia coli* model was reported to have poor experimental density, suggesting that this region may be flexible throughout the KDPG aldolase family, allowing the C-terminal α helix some flexibility.

3.13 *Escherichia coli /Thermotoga maritima* KDPG Aldolase: The basis of thermostability.

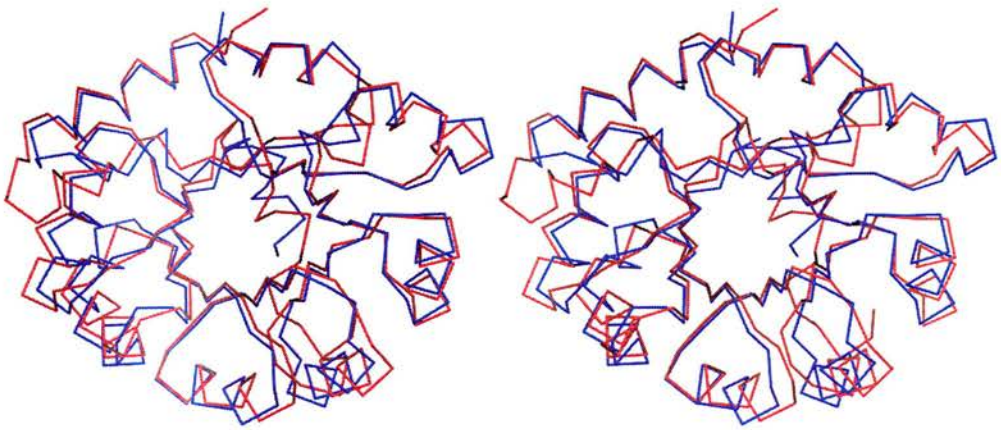
The high sequence similarity between the *Escherichia coli* KDPG aldolase and the *Thermotoga maritima* KDPG aldolase (34% identity: 54% homology) indicates a high level of structural and functional equivalence (Figure 3.21).

Figure 3.21 Sequence alignment of *Thermotoga maritima* and *Escherichia coli* (1eua.pdb) KDPG aldolase's.

```
T.m.  MKNWKTSAESILTTGPVVPVIVVKKLEHAVPMAKALVAGGVRVLEVTLRTECAVDAIRAI
      ::      : ..    .: :. ....:.:   : :.  :::...:.:. .  :  .:. .
E.c.  MK-----MEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKEL
             10        20        30        40        50

               70        80        90       100       110
T.m.  A-KEVPEAIVGAGTVLNPQQLAEVTEAGAQFAISPGLTEPLLKAATEGTIPLIPGISTVS
      .  .   ::.::::: . .:  ...:.::.: .:: : : . .   :  .  .::. : .
E.c.  SFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPT
         60        70        80        90       100       110

      120       130       140       150       160       170
T.m.  ELMLGMDYGLKEFKFFPAEANGGVKALQAIAGPFSQVRFCPTGGISPANYRDYLALKSVL
      ::. .:  :   .:.::.:. :  . ..:. ::: .:.: :::::..  :  ...   .::
E.c.  ELVKAMKLGHTILKLFPGEVVGP-QFVKAMKGPFPNVKFVPTGGVNLDNVCEWFK-AGVL
        120       130       140       150       160       170

       180       190       200       210
T.m.  CIG-GSWLVPADALEAGDYDRITKLAREAVEGAKL
      .: :: :: .   :. .    .:   : ..:
E.c.  AVGVGSALVKGTPDEVREK---AKAFVEKIRGCTE
                 180       190          200
```

Super-positioning of the *Escherichia coli* KDPG aldolase and the *Thermotoga maritima* KDPG aldolase shows a very close structural similarity (Figure 3.22).

Figure 3.22 A stereo super-imposition of *Thermotoga maritima* and *Escherichia coli* KDPG aldolase.



Thermotoga maritima KDPG aldolase. *Escherichia coli* KDPG aldolase.

A visual examination of the two models revealed only one obvious difference: the change in the length of the extended helix-loop region between the final β strand and the C-terminal α helix. In the *Escherichia coli* aldolase model there are 13 residues separating the two secondary structure elements, 7 residues in the *Thermotoga maritima* aldolase model replace this. This may be the result of the differing environments under which each enzyme functions, with the reduced loop region in the *Thermotoga maritima* aldolase conferring increased thermal stability. The super-positioning of the two structures by sequence gave a residue by residue rms deviation of 1.91Å for the amino acid residues 2-203. Super-positioning of the secondary structure regions (i.e. α-helix to α-helix and β-strand to β-strand), produced an rms deviation of 1.5 Å for the secondary structural elements Cα atoms

(i.e. α-helices and β-strand). The superpositioning was performed by the DALI programme (Holm 1996).

The active site of the *Escherichia coli* KDPG aldolase contains an active site lysine at position 133, involved in Schiff base catalysis (Allard, 2000; Buchanan, 2001). The active site also contains residues Val20, Glu 45, Arg 49, Thr 73, Phe 135 and Thr 161 (Buchanan, 2001). These residues are conserved in the *Thermotoga maritima* KDPG aldolase and are present as the equivalent residues; Val 15, Glu 40, Arg 17, Thr 69, Phe 131 and Thr 156. The equivalent active site lysine is found at position 129 (Lys129). All but one of the active site residues are perfectly conserved, with the exception of the *Escherichia coli* aldolase Arg 49. The Arg 49 residue is replaced in the *Thermotoga maritima* aldolase model by the residue Arg 17. Although super-positioning of the two models puts the guanine functional groups of the Arg 17 of the *Thermotoga maritima* aldolase and the Arg 49 of the *Escherichia coli* aldolase in close proximity to the active site, their position is relatively distant sequentially. As reported (Buchanan, 2001) the *Escherichia coli* KDPG aldolase contains a large hydrophobic pocket adjacent to the active site. The hydrophobic pocket contains the amino acid residues, Val 18, Val 20, Phe 90, Ile 92 Ile 112 and Cys 159. These residues produce a hydrophobic pocket formed by the inner walls of the β-barrel.

A manual search produced a similar candidate in the *Thermotoga maritima* aldolase model. A search through the model pdb file using the Voidoo programme from the Uppsala software factory revealed a comparative hydrophobic pocket containing the residues; Val 13, Val 15, Glu 40, Val 88, Met 108, Lys 129, Val 154, Thr 156, and Gly 176. The hydrophobic pocket has a volume of 84.78 $Å^3$, is adjacent to the active site and contains three active site residues Val 15, Lys 129 and Thr 156.

The *Thermotoga maritima* aldolase is a thermophilic enzyme that has an optimum reaction temperature of approximately 80°C. There have been numerous theories on the mechanisms

of protein thermo-stability. Protein folding is predominantly driven from a free unravelled peptide to a more rigid condensed structure *via* hydrophobic interactions (Dill, 1990). The expulsion of water produces a condensed structure, which is reinforced and stabilised by a contribution of smaller more defined interactions, including hydrogen bonds, ions pairs and van der Waals interactions, producing clearer secondary structure. The hydrophobic effect is by far the greater force in protein folding and is considered the predominant stabilising effect. In the last 20 years sequence and structural data from related meso/thermophiles (Vielle, 2001) have given us some insight into protein stability. A direct comparison of the meso / thermophiles have shown that hydrophobic forces and core residues involved in secondary structure are better conserved than surface residues (Szilagyi, 2000). Therefore any additional forces of thermo-stabilisation are found in less conserved areas of the protein sequence nearer the protein surface (Karishikoff, 1998).

In general the core regions of both mesophilic and thermophilic proteins are essentially very similar. Therefore protein stability in thermophiles must derive from changes in the outer structure of the protein. A comparison on the sequence data of the mesophilic and thermophilic proteins has indicated that certain residues are substituted to increase thermostability. In general thermophilic proteins contain more alanines and less glycines, fewer lysines and cysteines and more arginines and tyrosines. Thermophilic proteins have a greater number of charged residues than mesophiles, and also contain more hydrophobic residues (Kumar, 2000), although there are some exceptions to these general rules.

Hydrophobic interactions (Pace, 1992), disulphide bridges (Matsumara, 1989; Cacciapouti, 1994), aromatic interactions (Teplyakov, 1990) and hydrogen bonding (Tanner, 1996) have all been examined for thermo-stabilisation activity. The general trend suggests that increased hydrophobic interactions and hydrogen bonds, increases protein stability. However from the thermophilic structures deposited the evidence on disulphide bridges, aromatic interactions

and hydrogen bonding are inconclusive. Previous work highlighted that although ion pairs were present in proteins there was insufficient force supplied to create protein folding (Dill, 1990). The electrostatic interactions within a protein must therefore have a stabilisation effect after the initial folding event. It has been suggested (Perutz, 1978) that ions pairs are stronger in proteins than solvents. Although ion pairs had been thought of as destabilising in mesophilic proteins, increased numbers of ion pairs on the surface of thermophilic proteins have been reported (Elcock, 1998; de Bakker, 1999; Walden, 2001).

The structural analysis of the KDPG aldolase, revealed only one major structural anomaly between the *Thermotoga maritima* aldolase and its mesophilic cousin the *Escherichia coli* aldolase. The loss of 6 amino acid residues, from the C-terminal loop / turn region. The loss of these 6 residues would seem to be necessary for increased stability as previous crystallographic studies of both the *Escherichia coli* aldolase and *Thermotoga maritima* aldolase have outlined this area as a more flexible region. Comparative sequence analysis of the two enzymes, have listed the different (Table 3.06) residue composition for both the mesophilic *Escherichia coli* aldolase and the thermophilic *Thermotoga maritima* aldolase.

From the results in Table 3.06, it is clear that the general rules for amino acid composition and thermostability do not all apply to the KDPG aldolase comparisons. *Thermotoga maritima* KDPG aldolase has fewer hydrophobic residues than the *Escherichia coli* KDPG aldolase. The thermophilic aldolase contains similar quantities of glycine and cysteines and contains more lysines and less arginines. A comparative study of the structural PDB files for both the *Escherichia coli* and *Thermotoga maritima* aldolases using the hb-plus programme (Mc Donald, 1994) has identified 177 hydrogen bonds per monomer in the *Thermotoga maritima* aldolase and 178 hydrogen bonds per monomer in the *Escherichia coli* aldolase.

Interestingly 29% of the thermophilic *Thermotoga maritima* KDPG aldolase sequence contains charged residues whereas the mesophilic *Escherichia coli* KDPG aldolase sequence

contains only 22.% charged residues. In total the thermophilic aldolase contains 12 more charged residues than the mesophilic aldolase. The *Thermotoga maritima* KDPG also contains 18 salt bridges compared to 8 salt bridges contained within the *Escherichia coli* KDPG aldolase structure. Smaller ionic interactions contained within the structures were assessed by the ACT programme (CCP4: Wojtek Rypniewski and Howard Terry).

Table 3.05. Amino acid composition of *Thermotoga maritima* and *Escherichia coli* KDPG aldolase.

Thermotoga maritima KDPG Aldolase				*Eshcherchia coli* KDPG Aldolase			
Ala	(A)	17	8.00 %	Ala	(A)	30	14.0 %
Arg	(R)	4	2.00 %	Arg	(R)	7	3.0 %
Asn	(N)	4	2.00 %	Asn	(N)	4	2.0 %
Asp	(D)	5	2.5 %	Asp	(D)	6	3.00 %
Cys	(C)	4	2.00 %	Cys	(C)	3	1.50 %
Gln	(Q)	3	1.5 %	Gln	(Q)	5	2.20 %
Glu	(E)	22	10.0 %	Glu	(E)	16	7.50 %
Gly	(G)	19	10.0 %	Gly	(G)	20	9.4 %
His	(H)	4	2.00 %	His	(H)	1	0.50 %
Ile	(I)	10	5.00 %	Ile	(I)	13	6.0 %
Leu	(L)	14	7.00 %	Leu	(L)	22	10.0 %
Lys	(K)	23	11.0 %	Lys	(K)	13	6.0 %
Met	(M)	6	3.0 %	Met	(M)	4	2.0 %
Phe	(F)	13	6.0 %	Phe	(F)	6	3.0 %
Pro	(P)	10	5.0 %	Pro	(P)	14	6.50 %
Ser	(S)	7	3.4 %	Ser	(S)	9	4.0 %
Thr	(T)	11	5.0 %	Thr	(T)	13	6.00 %
Trp	(W)	1	0.5 %	Trp	(W)	2	1.00 %
Tyr	(Y)	1	0.5 %	Tyr	(Y)	4	2.00 %
Val	(V)	27	13.0 %	Val	(V)	21	10.00 %

Although similar the thermophilic protein contained about 10% more ionic interactions. The *Escherichia coli* aldolase contained 412 ionic interactions whereas the *Thermotoga maritima* Aldolase contained 461 ionic interactions.

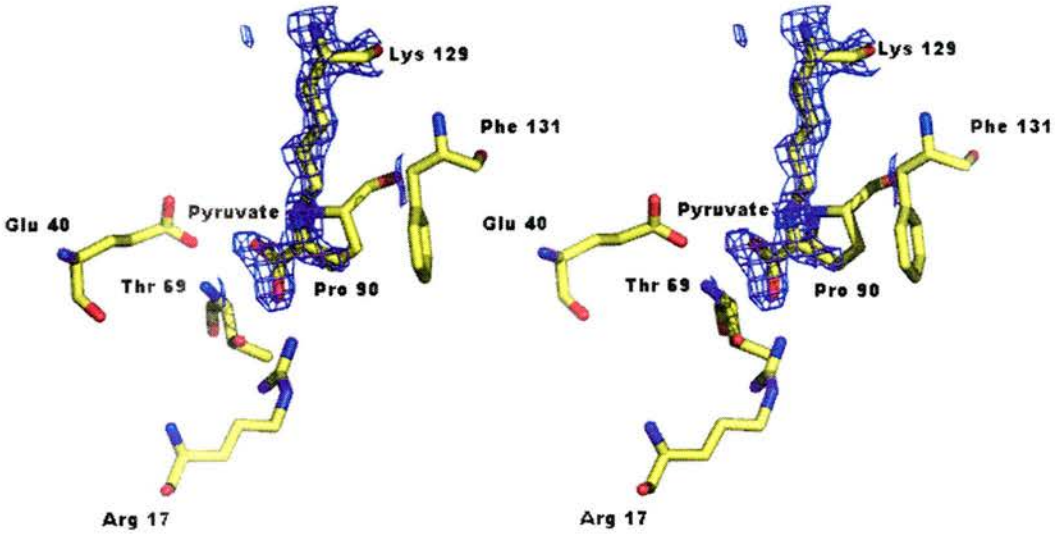
From the data presented, it may be reasonable to assume that an increased content of charged residues within the thermophilic, *Thermotoga maritima* KDPG aldolase protein, has led to increased ionic interactions and ion pair formation. Previous literature, (Tanner, 1996; Elcock, 1998; Walden, 2001) indicates the importance of ionic interactions and ion pairs / salt bridges in protein thermo-stability. Most of the previous studies in thermo-stability have taken place with extreme thermophiles and hyper-thermophiles, each showing a greater degree of increased electrostatic interactions between the surface or outer residues of the globular protein. The results presented in this chapter, indicate an increased number of ionic interactions, charged residues and ion pairs/ salt bridges throughout the thermophilic *Thermotoga maritima* KDPG aldolase. Although there is a small (10%) increase in ionic interactions, which in some small way may help increase protein stability, the 50% increase in salt bridges and the loss of the 6 residues in the flexible C-terminal loop region, are the most likely key differences conferring protein thermo-stability.

3.14 The active site.

The active site is located within the central barrel structure, with the active site lysine (Lys 129) side chain pointing into the center of the barrel from its location upon the β 6 strand. This is a common arrangement found in all but two of the Class I Schiff base forming aldolases (Transaldolase (Jia, 1996) and 5-aminolaevulinic acid dehydratase (Erskine, 1997)). Lysine 129 was confirmed as the catalytic lysine after structure solution, when during model refinement excess density regions were found at all six of the active site lysines. These excess density regions fitted perfectly the model of a pyruvate molecule (Figure 3.20). The

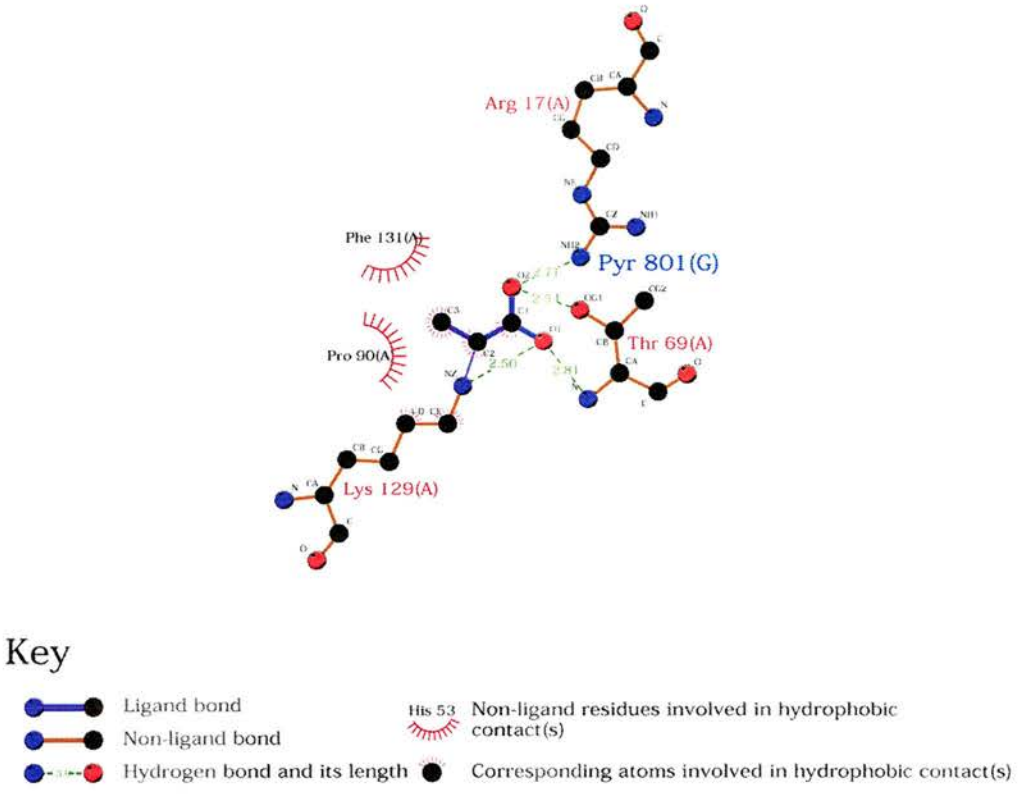
pyruvate ligand was un-expected as neither the protein solution nor the crystals had ever been treated with pyruvate, leading to the conclusion that the pyruvate must have become trapped during the expression and purification of the recombinant protein.

Figure 3.23 Stereo view of the catalytic lysine bound to the pyruvate Schiff base covalent intermediate. This figure depicts the active site lysine bound to the pyruvate covalent intermediate, surrounded by the active site residues Arg 17, Pro 90, Phe 131, Glu 40 and Thr 69.



The pyruvate ligand is covalently bound to the active site lysine forming the Schiffs base intermediate. The pyruvate molecule sits co-planar to the A90 proline ring making a hydrophobic contact with the proline A90 and the Phe A 131. Hydrogen bonds with the imine NH_2 of Arg 17, and the N of Thr 69 from the 2 oxygens at the C1 carboxyl terminus stabilise the covalent intermediate (Figure 3.24).

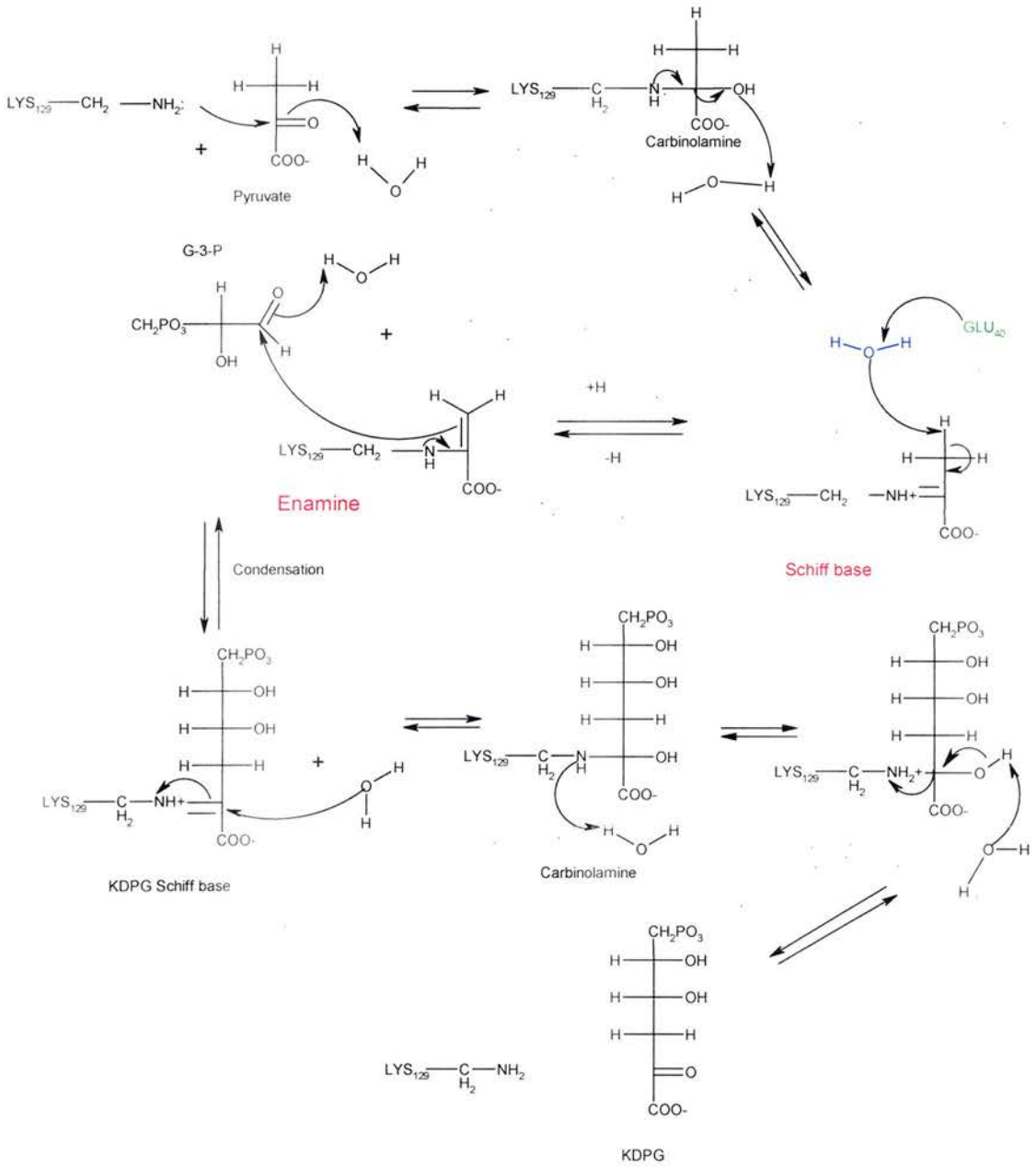
Figure 3.24 Ligplot view indicating the interactions between the pyruvate Schiff base intermediate and the active site environment.



The glutamic acid Glu 40 sits 4.56 Å from the Schiff base C_2 and contacts the C_2 of the pyruvate ligand via a water bridge (W 17). The water molecule is found at all six of the active sites in the asymmetric unit, indicating that the water may be structurally or functionally relevant. In the previously published *Escherichia coli* KDPG aldolase structure (Allard, 2001) the 3 carbon carbinolamine intermediate is trapped with the hydroxyl group of the carbinolamine forming a tetrahedral conformation around the central C_2 carbon. The hydroxyl group protrudes out towards the conserved active site glutamate residue. The carbinolamine structure also has an equivalent water molecule at a comparable position allowing the interaction of the active site glutamate with the carbinolamine, hydroxyl group or the lysine NH_2 group *via* a water bridge. The proposed mechanism (Figure 3.24) requires the dehydration of the carbinolamine to the Schiffs base before proton abstraction creates an

enamine, which upon condensation with glyceraldehyde-3-phosphate produces the 2-keto-3-deoxy-6-phosphogluconate Schiff base.

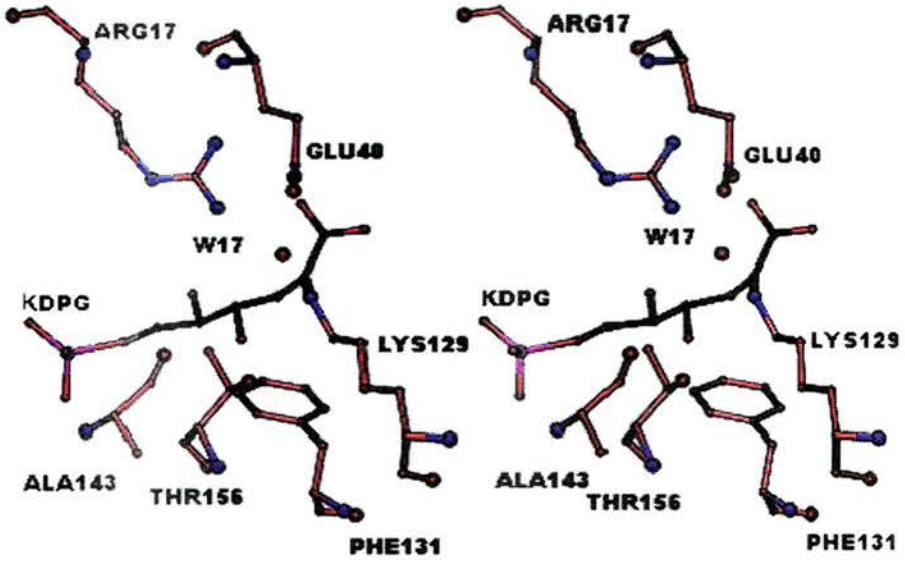
Figure 3.25 Proposed aldolase mechanism.



The solution of the KDPG aldolase to 1.90 Å resolution clearly shows the covalently bound Schiff base intermediate trapped before de-protonation to the reactive enamine (Figure 3.25). The conserved water molecule (H2O) acts in conjunction with the active site glutamate (GLU_{40}) to remove the C3 proton.

Modeling of KDPG into the active site (Figure 3.26) has allowed us to propose a preliminary catalytic mechanism. The KDPG molecule was modelled with the C1, C2 and C3 carbons situated in an approximate position with the bound pyruvate molecule. The C4, C5 and C6 carbons are then modelled between the C3 carbon and the sulphate ion bound at the entrance to the active site cleft, with the KDPG phosphate group modelled in an approximate position to the bound sulphate molecule. In this model the C4 hydroxyl is positioned within hydrogen bonding distance of the conserved active site threonine (Thr 156) and the conserved water residue.

Figure 3.26 Stereo-image of the modelled KDPG surrounded by active site residues, Phe 131, Thr 156, Ala 143, Arg 17, Glu 40 Lys 129 and the conserved water molecule.



The C5 hydroxyl is positioned within hydrogen bonding distance of the active site alanine oxygen from the B subunit. The Phe 131 residue positioned below the KDPG produces a slight kink in the molecules shape at the C3-C4 bond. The mechanism proposed involves the binding of the KDPG at the active site. The C4 hydroxyl, hydrogen bonds to the active site

threonine and the C5 hydroxyl hydrogen bonds to the alanine (Ala 143) oxygen from the B subunit and the active site arginine (Arg 17) stabilising the molecule in the active site and conferring stereo-selectivity of the substrate. The KDPG C2 carbon atom undergoes nucleophilic attack from the active site lysine, forming the carbinolamine intermediate. The carbinolamine intermediate undergoes dehydration forming the KDPG Schiff base. The active site glutamate acts as a base and abstracts the C4 hydroxyl proton *via* the conserved catalytic water(W17), forming the reactive enamine. The enamine undergoes condensation forming G3P and the pyruvate form of the Schiff base intermediate. The pyruvate Schiff base then undergoes hydration and deprotonation to produce free pyruvate and enzyme.

Chapter 4

UDP-Galactopyranose Mutase; a Structural and Biochemical Investigation.

4.00 Summary.

UDP-galactopyranose mutase (UDP-Gal*p*) mutase is a flavo-protein which catalyses the inter-conversion of the six membered UDP-Gal*p* sugar to the five membered UDP-galactofuranose (UDP-Gal*f*) sugar. The enzyme catalyses the synthesis of a key cell wall component precursor of many bacterial pathogens: the activated biosynthetic precursor, UDP-Gal*f*. The enzyme is crucial for bacterial cell viability in a number of important human pathogens including *Escherichia coli, Klebsiella pnuemoniae* and most importantly *Mycobacterium tuberculosis*.

Mycobacterium tuberculosis is estimated to be present in one third of the world's population and to be responsible for 8 million acute infections each year, resulting in 2 million deaths per annum. Due to the high numbers of infected individuals providing a large pathogen reservoir and the vast increase in multi-drug resistant strains of bacteria, there is an increased importance to the discovery of new therapeutic methods.

No equivalent enzyme has been found within higher mammals, making the UDP-Gal*p* mutase a potential drug target. The enzyme mechanism is of basic scientific interest also; it must proceed by a ring breakage and closure as the glycosidic bond is formed by the anomeric oxygen, a mechanism as yet without precedent. The structure of the *Escherichia coli* mutase is known but by itself does not provide enough information to resolve the mechanism. The overall reaction is electrically neutral, but a crypto-redox reaction is suggested because the enzyme's flavin co-factor must be reduced for activity. This section reports upon the crystallisation attempts of the *Klebsiella pneumoniae* UDP-Gal*p* mutase and on the results of a thermodynamic analysis of the flavin co-factor with the objective of defining the chemistry at the active site.

4.01 Introduction.

4.02 UDP-galactopyranose mutase.

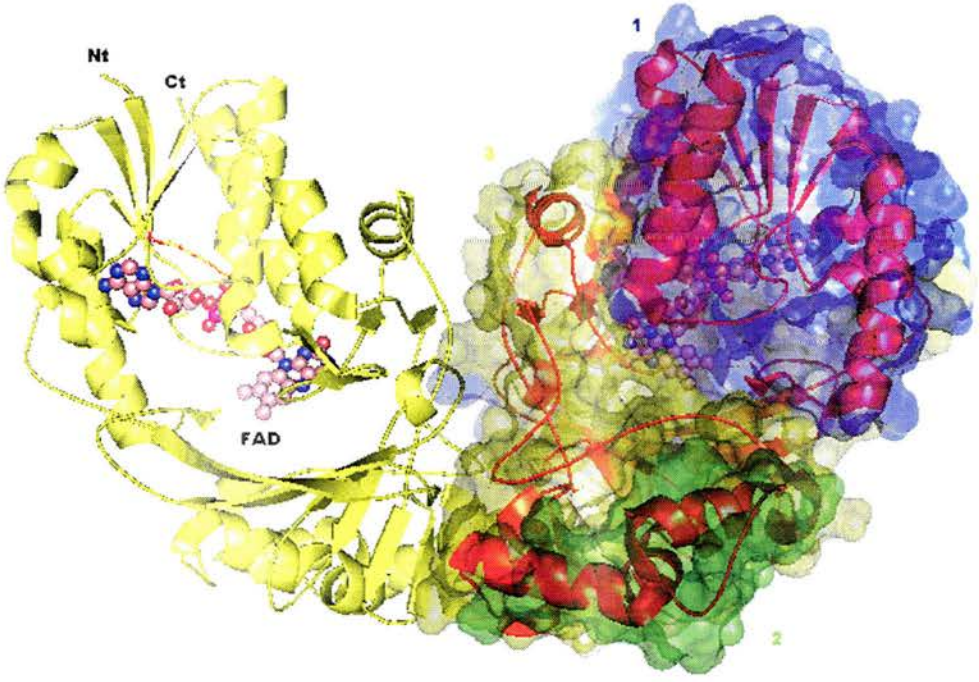
The *Klebsiella pneumoniae* UDP-Gal*p* mutase was chosen as an experimental model after previous structural studies upon the *Escherichia coli* UDP-Gal*p* mutase proved moderately successful. The *Escherichia coli* UDP- Gal*p* mutase structure could not provide a co-crystal structure to aid in illuminating the enzyme mechanism. *Klebsiella pneumoniae* UDP-Gal*p* was introduced into the study in an attempt to provide enzyme-ligand co-complexes

The *Klebsiella pneumoniae* UDP- Gal*p* mutase is a bright yellow flavo-protein, from a 1152 nucleotide reading frame, encoding the 384 amino acid protein. It has a theoretical pI of 6.06 and a molecular weight of 44,454 Da. A BLAST search of the N-terminal sequence of the RfbD gene product recognised similar sequences in flavin containing oxidases and dehydrogenases (Koplin, 1997). The N-terminus had been identified as a flavin co-factor binding site, because the N-terminus between Lys 5 and Asp 33 contains a GxGxxG signature for a Rossman fold (Koplin, 1997). The Rossman fold is situated at the N-terminus of all three of the UDP- Gal*p* mutases previously discussed in the introductory chapter, which suggested NADPH/NADH/FAD co-factors were necessary for enzyme activity. Although it was suspected that FAD would be the co-factor that was bound to the Rossman fold at the N-terminus, a reported requirement for the presence of NADH/NADPH for enzyme activity (Koplin, 1997) suggested that the Rossman fold was necessary for NADH/NADPH binding, and that the FAD cofactor may be bound to an alternate site on the protein.

4.03 *Escherichia coli* UDP-Gal*p* mutase structure.

Crystallographic studies on the *Escherichia coli* UDP- Gal*p* mutase reported a 2.4 Å resolution structure (Figure 4.01) identifying several key amino acids and the position of the non-covalently bound flavin co-factor (Sanders, 2001). The *Escherichia coli* UDP- Gal*p* mutase structure is a mixed α/β protein structure with each monomer having a cupped palm shape. Each mononer consists of three domains. Domain I (1) consists of 4 parallel beta strands containing the FAD cofactor non-covalently bound at the αβα Rossman fold motif. Domain II (2) a 5-helix bundle is connected to Domain I via the six anti-parallel beta strands of Domain III (3). The dimer interface consists of Domains II and III with contacts between monomers occurring mainly between non-conserved residues. The active site is situated in a cleft between Domains I and II with Domain III forming the palm of the cupped hand.

Figure 4.01 *Escherichia coli* UDP-galactopyranose mutase.



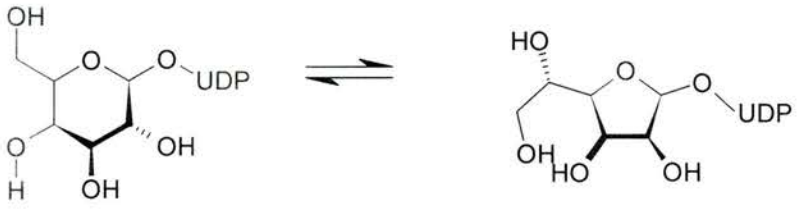
Nt* = *N terminus
Ct* = *C terminus

The active site cleft contains the FAD cofactor with the re-face facing outwards into the cleft, a conserved tryptophan residue at position 156, four conserved tyrosine residues and 2 conserved arginine residues. Computer modelling has suggested that the UDP moiety of the substrate stacks with the tryptophan residue, with the conserved arginines binding the phosphate bridges of the nucleotide, and the flavin co-factor and the four tyrosines making contacts with the sugar ring (Sanders, 2001).

4.04 The UDP-Galactopyranose mutase Mechanism.

Until recently, little was known about the UDP- Gal*p* mutase enzyme activity and (Figure 4.02) mechanism (Barlow, 1999). Enzyme activity studies had shown a 1-2% conversion of UDP- Gal*p* to UDP- Gal*f* over a 1 hour period rising to an 8% conversion upon the addition of 200mM NADH/NADPH (Barlow, 1999). Discussion of the requirement of oxidised/reduced enzyme or presence of NADH/NADPH for efficient UDP- Gal*f* production (Barlow, 1999; Zhang, 2000) continued until 2001, when enzymatic conversion of UDP-Gal*p* to UDP- Gal*f* was reported to occur in reduced enzyme only, without the requirement of NADH/NADPH co-factors (Sanders, 2001). The reports that UDP- Gal*p* mutase required NADH/NADPH for activity seemed strange given that there was no net transfer of electrons (Nassau, 1996; Koplin, 1997; Barlow, 1999). Furthermore studies on the enzyme indicated that transfer of hydride from the R face of NADH was rate limiting in aerobic conditions (Barlow, 1999) and enzyme activity could be activated and deactivated upon reduction with dithionite and oxidation with O_2 or potassium ferricyanide (Zhang, 2000; Sanders, 2001). Although other reports suggested that oxidised enzyme retained some activity, (Zhang, 2000) it has been suggested that the activity may be the result of photo-reduction of the enzyme (Zhang, 2000; Sanders, 2001).

Figure 4.02 UDP-galactopyranose mutase mechanism.

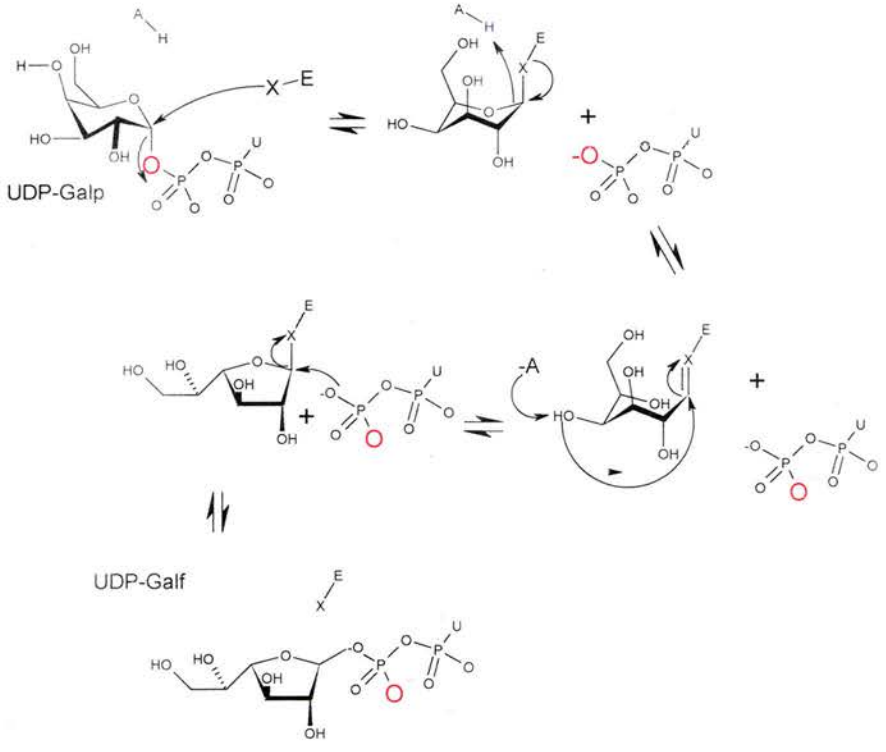


UDP-galactopyranose UDP-galactofuranose

Figure 4.02 The Interconversion of UDP-Galactopyranose to UDP-Galactofuranose under the action of UDP-Galactopyranose mutase.

A study demonstrating that the nucleotide diphosphate group bound to the anomeric position statistically scrambled a labelled oxygen atom with the same rate as the reaction itself, provided further insight into the catalytic mechanism (Barlow, 1999a). This observation led to the conclusion that during turnover the glycosidic bond must be broken as part of the mechanism (Figure 4.03).

Figure 4.03 Proposed mechanism with positional isotope exchange showing oxygen scrambling.



The same researchers also showed that fluoro-deoxy analogues of substrates (substituted at the 2 and 3 positions) were subtrates for the enzyme (although the 2-substituent was a poor substrate), furthermore neither eliminated HF (Barlow, 2000). These results were confirmed by a subsequent study on the *Escherichia coli* UDP- Gal*p* mutase, although in this case covalent intermediates were apparently formed between the substrate and enzyme (Zhang, 2000). An elegant mechanism was proposed by both groups, which accommodated the requirement for rupture of the glycosidic bond during catalysis. In this model the O4 of galactose acts as a nucleophile to displace the UDP group, forming a bicyclic sugar and UDP. The mechanism does not require any transfer of electrons, but in this scenario reduced flavin could promote the reaction by stabilising the oxocarbonium ion formed in the transition state. The determination of the *Escherichia coli* UDP- Gal*p* mutase enzyme has shown that the isoalloxazine ring of the flavin sits in an exposed position and modeling studies based on the structure suggested that the galactose ring of UDP- Gal*p* would be adjacent to this isoalloxazine ring (Sanders, 2001).

4.05 Experimental (Crystallography).

4.06 Protein expression trials.

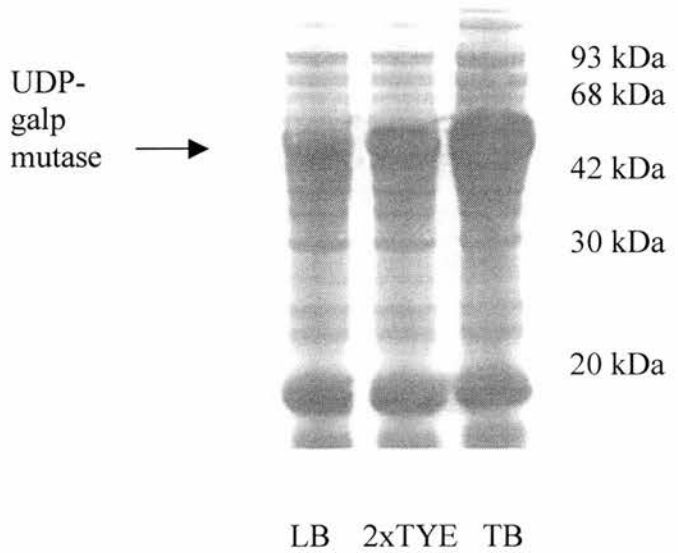
(Unless stated all chemicals were supplied by Sigma Chemicals, Poole, Dorset, U.K.)

500 mls of Terrific broth media (Maniatis, 1989) was inoculated with 1ml of BL21 DE3 cell culture, transfected with the pWQ66 plasmid containing the *Klebsiella pneumoniae* UDP-Gal*p* gene, from a 5 ml overnight growth culture in Luria Bertani broth. Cultures were grown, under 50μg/ml kanomycin, at different temperatures of 25°C, 30°C, and 37°C, until an optical density of 0.6 (600nm). The cell lines were induced with 1mM concentrations of IPTG for 3 hours.

The cultures were harvested by centrifugation at 8000 rpm for 10 minutes in a Beckman AVANTI J-25 centrifuge. 0.5g of wet cell pellet was resuspended in reducing solution and boiled for 10 minutes. 20 μl of the cell lysates were run on a 12 % SDS-PAGE gel to visually determine optimum conditions for UDP- Gal*p* mutase production. The above criteria were repeated for both 500 ml cultures in Luria Bertani broth and 2xTYE broth (Maniatis, 1989).

Optimum expression of UDP- Gal*p* mutase occurred in cells harvested from Terrific broth, incubated at 37°C. Terrific broth media visibly contained greater amounts of UDP- Gal*p* mutase than the 2xTYE and Luria Bertani broth counterparts (Figure 4.04).

Figure 4.04 SDS-PAGE gel indicating different levels of protein expression between different media.



Lane 3 (TB) contains the largest band at 45 kDa when compared to lanes 1 (LB) and 2 (2xTYE). This band corresponds to the molecular weight of the UDP- Gal*p* mutase monomer.

4.07 Production of recombinant UDP-galactopyranose mutase.

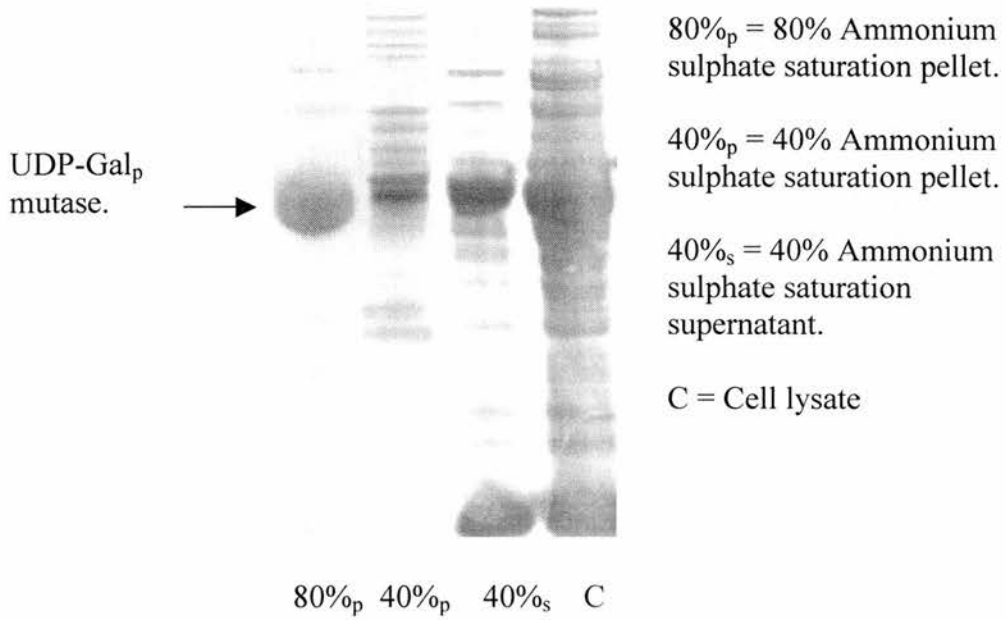
A stirred and aerated 9.5 litre fermentation of Terrific broth was allowed to reach 37°C overnight in a New Brunswick Microferm fermentor (during growth and induction of cells, the fermentor was stirring at approximately 300 rpm and aerated with 10 litres per minute air. The media was inoculated with 500 ml of BL21 DE3 cells containing the pWQ66 plasmid, and grown in the presence of 30μg/ml kanomycin until an optical density (600nm) of 0.6. The culture was induced with a final concentration of 1mM isothiolgalacto-pyranoside (IPTG) for three hours.

The cells were harvested via centrifugation at 16000g for 15 minutes. Cells from the 9.5 litre fermentation were harvested and weighed, approximately 40 grams of wet cell pellet was collected. The cell pellet was lysed by stirring for 30 minutes at 4°C in the presence of 20 μg/ml DNAase, 20 μg/ml lysozyme 1mM dithiothreitol (DTT) and 1 mM phenylmethylsulphonyl-fluoroside (PMSF) in 100 ml 50mM Tris pH 7.0. The cellular debris was removed by centrifugation of the lysis supernatant at 40,000g for 30 minutes at 4°C.

4.08 Protein purification.

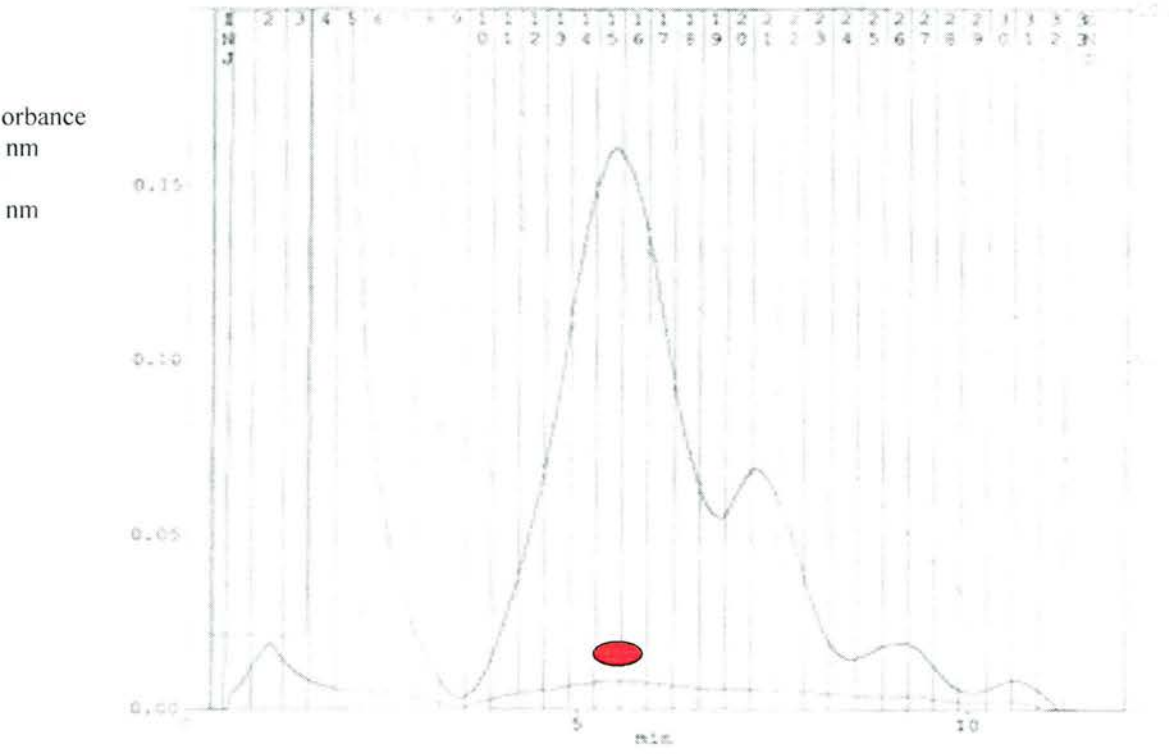
The lysis supernatant was brought up to 40 % ammonium sulphate saturation over one hour with stirring at 4°C. The precipitate was then removed with centrifugation at 40,000g for 30 minutes at 4°C. The precipitate pellet was re-suspended and washed down an Amersham Pharmacia PD10 gel filtration column to de-salt the sample before testing for UDP- Gal*p* mutase on SDS-PAGE (Maniatis, 1989). The remaining supernatant was brought up to 80 % ammonium sulphate concentration. The solution was clarified by centrifugation at 40,000g at 4°C for 30 minutes and the precipitate pellet collected and tested for presence of the enzyme as before. Once the presence of UDP- Gal*p* mutase was confirmed (Figure 4.05.), the pellet was re-suspended and dialyzed against 3 changes of 2 liters of 10 mM Hepes pH 7.0, 2mM DTT for 40 minute intervals at 4°C to remove salt. The dialysed supernatant was then concentrated to 20 ml using an Amicon nitrogen fed pressure concentrator, with a 30,000 Da molecular weight filter, with constant stirring at 4°C.

Figure 4.05 SDS-PAGE analysis of the ammonium sulphate precipitation.



Small quantities of mutase protein were found in both the 40 % ammonium sulphate precipitate and the 80% ammonium sulphate cut supernatant. The quantities were too small to warrant further purification. Figure 4.05 clearly shows most of the UDP-Gal*p* mutase is contained within the 80 % ammonium sulphate precipitate. The resuspended and dialyzed 40-80% ammonium sulphate cut was loaded onto a BioCAD 700E perfusion chromatography workstation, 8 ml anion exchange column. The column was equilibrated in 10 mM Hepes pH 7.0 at 40 ml/minute at a pressure of 1300 psi for 15 column volumes before loading 4 ml aliquots of the dialyzed supernatant. The column was then washed for a further 15 column volumes with 10 mM Hepes 7.0. The column was then eluted with a (NaCl) 500 mM salt gradient 0-100 %. The fraction collector was set to collect fractions over 0.02 absorbance at λ= 480nm. Fractions containing UDP- Gal*p* mutase (Figure 4.06) were collected and concentrated to 10 ml final volume.

Figure 4.06 A trace of the HPLC anionic column separation of the resuspended 40-80 % ammonium sulphate cut fraction.

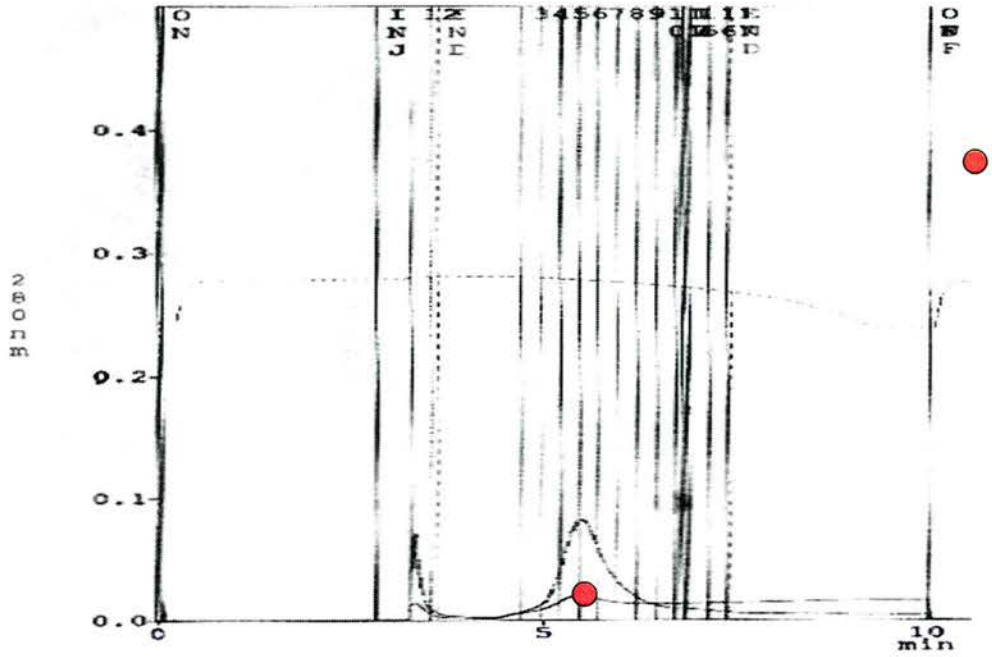


The protein peak was eluted from the column at approximately 30 % 500mM NaCl solution, and collected. Fractions 14-17 were pooled and concentrated to 20 ml final volume and the sample was dialyzed against 6 liters (3 changes of 2L) 10 mM Hepes, pH 7.0, 40% ammonium sulfate before hydrophobic interaction column purification.

The dialyzed sample was loaded onto a BioCAD 700E perfusion chromatography workstation 8 ml hydrophobic interaction column and run at 40 ml/minute at a pressure of 1300 psi. The column was equilibrated in 40 % ammonium sulphate, 10 mM Hepes pH 7.0 for 15 column volumes before loading 4 ml aliquots of the protein. The column was then washed for a further 15 column volumes with 40 % ammonium sulfate, 10 mM Hepes pH 7.0 before eluting with a 0-100 % 10 mM Hepes pH 7.0 gradient (Figure 4.07). The fractions

were collected as before and concentrated to a final concentration of 7mg/ml. The protein sample was stored at –20°C in 30 % glycerol at a final concentration of 5mg/ml in 500 µl aliquots.

Figure 4.07 A trace of the HPLC hydrophobic affinity column separation of the semi-purified protein sample.



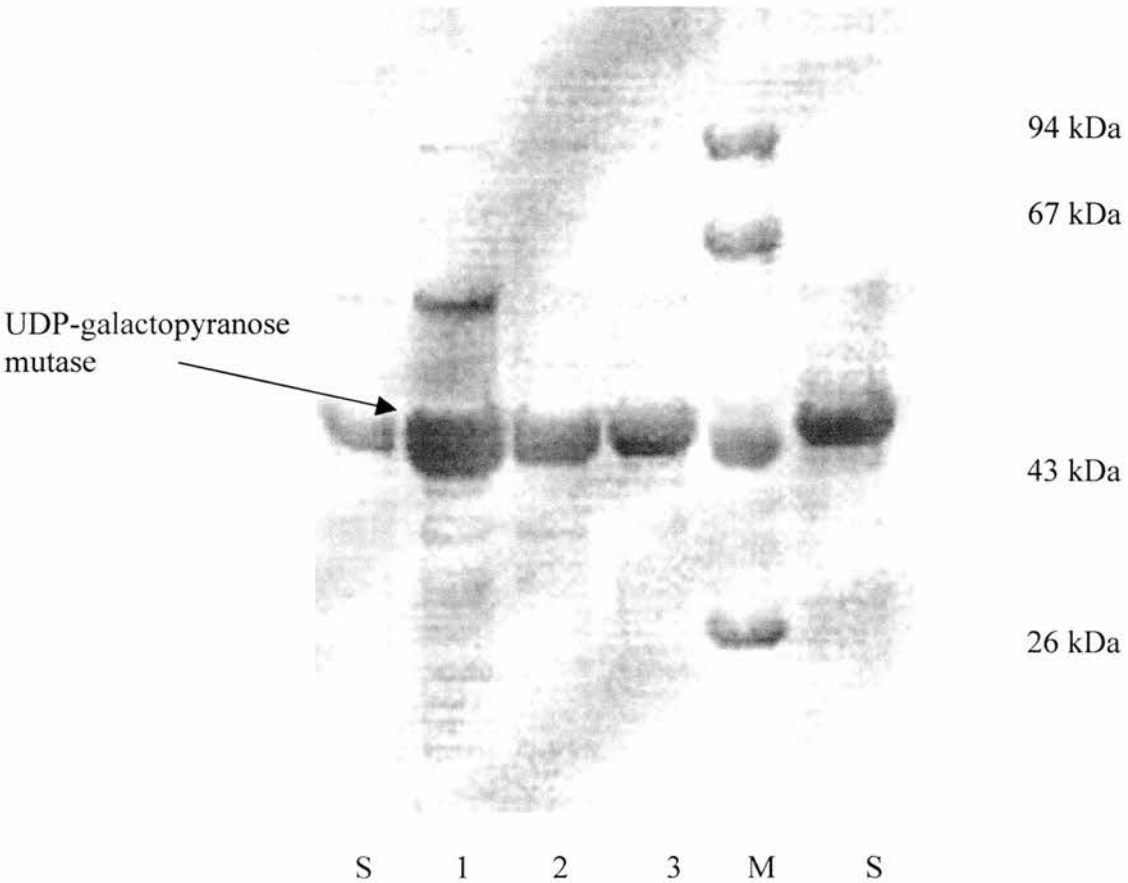
The protein peak eluted from the column at approximately 28 % ammonium sulphate saturation, 10mM Hepes pH 7.0 solution, fractions 4, 5, 6 and 7 were pooled and concentrated.

4.09 Assessing protein purity.

Assaying the purity of the protein sample requires removal of the ammonium salt and glycerol by running the sample on an Amersham Pharmacia PD 10 column and eluting with

10 mM Hepes pH 7.0 prior to SDS-PAGE analysis. SDS-PAGE analysis produced very faint contaminating bands after Coomassie blue staining (Figure 4.08.).

Figure 4.08 SDS-PAGE analysis of purification steps.



S= Sample of Escherichia coli UDP- Gal*p* mutase, *M= Molecular weight markers, 1-3 Lanes containing purification steps (1: 40-80% Ammonium sulphate precipitate, 2: After anion exchange purification, 3: after hydrophobic interaction purification.) Although lane 3 would seem to be almost clear, on the actual gel there were several very faint bands.*

The contaminating bands decreased through each sample, from sample 1: 40-80% Ammonium sulphate precipitate. 2: HPLC anionic exchange column. 3: Hydrophobic interaction column.

4.10 Assessing protein concentration.

Initial protein concentration was assessed by a Bradford assay. The flavin concentration was assessed by the 450 nm absorbance, using an experimental extinction coefficient 13500 M^{-1} cm^{-1}. The experimental extinction coefficient was determined by taking the 450nm absorbance reading of 900 µl of UDP- Gal*p* mutase (ABS 1) in a quartz micro-cuvette. The protein sample is transferred to an eppendorf and denatured with addition of 100 µl of 50% trichloroacetic acid. After 5 minutes on ice the solution was clarified by centrufugation at 40,000g for 2 minutes. The supernatant was transferred to a quartz micro-cuvette and a second reading taken at 450 nm (ABS 2).

The extinction coefficient can be determined by the equation.

ε Unknown = ε FAD x ABS 1 / ABS 2

Absorbance ABS 1 from 100 ul protein sample at 450 nm = 0.169

Absorbance ABS 2 from 100 ul denatured protein sample at 450 nm = 0.141

FAD extinction coefficient = 11,300 $M^{-1}cm^{-1}$.

ε Unknown = 11,300 $M^{-1}cm^{-1}$ x 0.169/ 0.141

ε Unknown = 13,543 $M^{-1}cm^{-1}$

A comparison of the flavin and protein concentrations:

Protein = 2.2 mg/ml = 0.048 mM.

FAD = 0.035 mM

Therefore 0.045 / 0.048 X 100 = 93.75 % of expressed protein has FAD bound.

The loss of the FAD may destabilize the protein structure (personal communication David Sanders, University of Sasketchewan, Sasketchewan Canada.) so that only protein with FAD bound remains after purification leading to incorrectly elevated levels of FAD incorporation.

For subsequent preparations the concentration of FAD containing protein was determined spectrophotometrically using the Beer Lambert law.

Abs = ε x UDP-Gal*p* mutase concentration x Path length.

If Path length remains 1cm, then UDP-Gal*p* mutase concentration = Abs / ε

Therefore 450 nm absorbance of 0.131 = 0.131/13543 Molar = 9.67×10^{-6} Molar.

Giving a final UDP-Gal*p* mutase concentration of

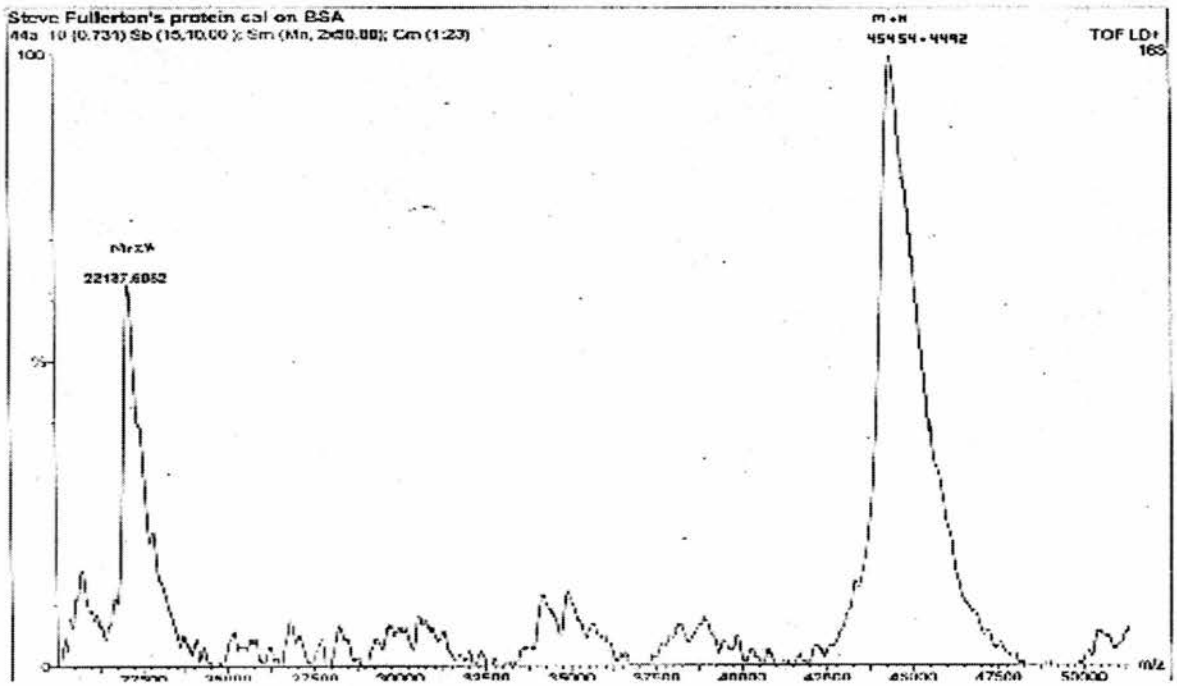
9.67×10^{-6} x 45545 (molecular weight for single subunit) = 0.44 mg/ml.

4.11 Protein identification and quality assessment.

Protein identification was carried out by N-terminal sequencing of 50µl of 4mg/ml protein solution. The sample was run on an automated Applied Biosystems Procise Protein sequencer. The automated N-terminal cleavage of amino acids was allowed to run for the first 5 amino acids in the N-terminal sequence. The first 5 amino acids of the N-terminus MKSKK matched the sequencing data produced from the sample.

Maldi-TOF mass spectrometry, confirmed the molecular weight of the protein and confirmed that a single species was present at 45 kDa (Figure 4.09).

Figure 4.09 Mass spectrograph of UDP-Galactopyranose mutase.



4.12 Tryptic digest characterisation.

Further investigation by trypsin, Endo-lysine-C and Endo-aspartate-N digests of the protein were carried out to investigate binding site of the flavin co-factor. From each of the three digests we can account for all but 1 residue (valine 214) within the UDP- Gal*p* mutase (Figure 4.10).

Figure 4.10 A map of total coverage of the peptide residues detected in the MALDI-TOF digests.

MKSKKILIVGAGFSGAVIGRQLAEKGHQVHIIDQRDHIGGNSYDARDSETNVMVHVYGPHIFHTDNET
VWNYVNKHAEMMPYVNRVKATVNGQVFSLPINLHTINQFFSKTCSPDEARALIAEKGDSTIADPQTFE
EQALRFIGKELYEAFFKGYTIKQWGMQPSELPASILKRLPVRFNYDDNYFNHKFQGMPKCGYTQMIKS
ILNHENIK

V Valine 214

DLQREFIVEERTHYDHVFYSGPLDAFYGYQYGRLGYRTLDFKKFTYQGDYQGCAVMNYCSVDVPYT
RITEHKYFSPWEQHDGSVCYKEYSRACEENDIPYYPIRQMGEMALLEKYLSLAENETNITFVGRLGTYR
YLDMDVTIAEALKTAEVYLNSLTENQPMPVFTVSVR

Only one residue Valine 214 was not detected in any of the 84 digest fragments.

The mapping of the UDP- Gal*p* mutase peptide fragment although inconclusive suggested that the FAD cofactor is not covalently bound to the protein.

4.13 Initial crystallisation screen.

Initial screening began in 96 well sitting drop plates allowing a greater variety of conditions to be tested. Initial screens involved testing 1,3,5,7, and 8 mg/ml protein samples in 10mM Hepes pH 7.0 against a range of crystallisation conditions, including Hamptons screens I + II, Hampton's ammonium sulphate screen, Hampton's lite screen and Hampton's PEG ion screen (for further information go to Hampton's research, 27632 El Lazo Road, Suite 100, Laguna Niguel,CA 92677-3913, or www.hamptonresearch.com/hrproducts/xtalscreens.html). Micro crystals formed after 3 days in Hamptons PEG ion screen, condition 5 (0.2 M $MgCl_2$, 20% w/v PEG 3350). Initial optimisation screens containing 10 % PEG 4K, 10 mM Hepes pH 7.0 to 30 % PEG 4K 10 mM Hepes pH 7.0 and then further screening at various PEG 4K concentrations in 10 mM Hepes with pH values 5,6,7,8,9,10, followed a common pattern where protein concentrations above 7 mg/ml in 10-20 % PEG 2-4k, 10 mM Hepes 0.1-0.2M

$MgCl_2$ solutions produced precipitate. Protein concentrations of 3mg/ml and below produced neither precipitate nor crystals before the drop dried out, and protein concentrations of 3-7 mg/ml in 10 mM Hepes showed varying precipitation/clear and extremely small needle like crystals in varying concentrations between 14-17 % PEG 4K.

4.14 Optimising crystallisation conditions.

Optimisation of the crystal trials was carried out with a constant 15 % PEG 4K concentration and a 5.5 mg/ml protein concentration with varying pH, Hepes, and salt concentrations.

Table 4.01 Optimisation of crystal trials.

All crystal trials were carried out in 10 mM Hepes at 293 K incubation for 24 hours.

	pH 6.5	pH 7	pH 7.5	pH 8	pH 8.5	pH 9	pH 9.5
0.1 M $MgCl_2$	C	C	C	C	C	C	X
0.15 M $MgCl_2$	C	C	C	C	C	X	X
0.2 M $MgCl_2$	C	C	C	C	C	X	X
0.25 M $MgCl_2$	C	C	C	C	P	X & P	X & P
0.3 M $MgCl_2$	C	C	C	P	P	P	P

C = clear. *P = precipitate.* *X = crystals.*

A pH of 6 or less produced no crystals and a pH above 9.5 precipitated the 15 % PEG 4K.

In Table 4.01, small needle crystals formed at higher pH values. Optimisation of crystallisation proceeded, concentrating on crystal size in conditions around 15 % PEG 4K, 10 mM Hepes pH 9-9.5 and 0.2 M $MgCl_2$.

Table 4.02 Further optimisation of crystal trials.

Crystallisation screens where carried out in 96 well plates containing 100μl well mother liquor and 2μl mother liquor + 2μl protein sample sitting drops.

	5.0 mg/ml	5..5 mg/ml	6.0 mg/ml	6.5 mg/ml	7.0 mg/ml	7.5 mg/ml	8.0 mg/ml
pH 9.0	C	C	C	...P	P	P	P
pH 9.1	C	C	X	X	P	P	P
pH 9.2	C	X	X	X	X	X &P	P
pH 9.3	C	X	X	X	X *	X &P	P
pH 9.4	X	X	X *	X *	X *	X & P	X & P
pH 9.5	X	X	X **	X **	X & P	X &P	X & P

*The * symbol indicates wells where the larger crystals formed. The ** symbol indicates wells where the largest crystals formed.*

Table 4.02, indicates the optimum crystallisation conditions, these are incubated at 4°C for 24 hours as 6-6.5 mg/ml protein concentration in 10 mM Hepes pH 9.5, with mother liquor containing 10 mM Hepes pH 9.5, 15 % PEG 4k and 0.2M $MgCl_2$.

Although there had been increases in crystal size, the crystals produced were not large enough for X-ray crystallography. Optimisation of the crystal trials continued with a constant 15 % PEG concentration, and 5.5 - 7 mg/ml protein concentration with varying pH, Hepes/Tris, and salt concentrations.

Table 4.03 Final crystal optimisation trial.

All crystal trials were carried out at 4°C incubation for 24 hours.

**** = considerable improvement in crystal size*

Mixed wells = protein and mother liquor was mixed thoroughly by repeated pipetting

Precipitant	Buffer/ pH	Additives	Protein concentration	Mixed wells	Crystals
PEG 4K	Hepes 0.1-0.3M pH 7-9.5	0.1-0.3M $MnCl_2$	5.5-7 mg/ml	Yes	Yes
PEG 4K	Tris 0.1-0.3M pH 7-9.5	0.1-0.3M $MnCl_2$	5.5-7 mg/ml	Yes	Yes ***
PEG 6K	Hepes 0.1-0.3M pH 7-9.5	0.1-0.3M $MnCl_2$	5.5-7 mg/ml	Yes	No
PEG 6K	Tris 0.1-0.3M pH 7-9.5	0.1-0.3M $MnCl_2$	5.5-7 mg/ml	Yes	No

The results in Table 4.03 indicate that the previous crystal trials had proven effective, however changing the buffer from Hepes to Tris and carrying out the same crystallisation conditions improved crystal size and quality (0.1-0.15mm long rods with 0.02-0.05 mm widths).

4.15 Crystal seeding.

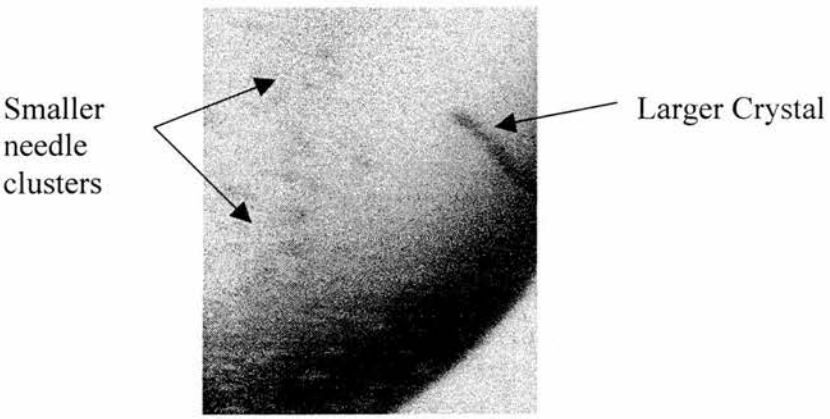
The process of seeding was carried out to increase crystal size. Small crystals unsuitable for X-ray diffraction were lifted from their incubated drops, washed in fresh mother liquor and transplanted into similar fresh crystallisation conditions to provide nucleation points for the

production of larger crystals. Seeding crystals were removed from their mother liquor by lifting them out on a thin human hair and transplanted into their new environment by dragging the hair through the crystallisation drop (Stura, 1991).

Seeding of crystallisation drops produced no improvement at the initial 22°C temperature. Any deviation from the optimised 6.5 mg/ml protein sample, and the 0.1M Tris pH 9.5, 15 % PEG 4K, 0.2M $MnCl_2$ crystallisation conditions had no beneficial effect in the seeded screens.

Further seeding experiments were carried out at 4°C, seeding at 4°C proved successful providing one crystal rod (0.2mm long, 0.07 mm wide) large enough to use for X-ray diffraction after seven days (Figure 4.11).

Figure 4.11 Photograph of the single UDP- Galp mutase crystal rod.



In the above picture the resolution and focussing is unfortunately poor, however, the large crystal rod on the upper right side of the well diffracted to less than 4Å.

4.16 X-Ray diffraction of UDP-Galactopyranose mutase crystal.

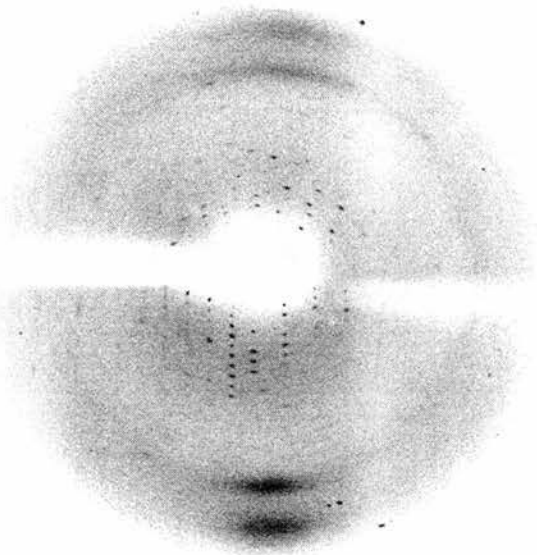
The crystal was removed from the mother liquor using the vacuum capillary technique. The crystal was transferred to mother liquor, then to mother liquor with 7.5 % glycerol and then

to mother liquor with 15 % glycerol before being looped out of the final solution of cryo-protectant using a 10 micron cryo-loop (Hampton research) and mounted under the nitrogen cryo-stream. The crystal was then mounted upon the ENRAF Nonius FR-591 X-ray generator, with a Dip-2000 detector and osmic mirrors. The crystal to detector distance was set at 200 cm, and diffraction data collected from a 1.54 Å X-ray source from a rotating copper anode.

4.17 Collection of X-ray crystallographic Diffraction data.

The UDP-Gal*p* mutase crystal diffracted to 3.7 Å in-house under a cryogenic nitrogen stream at 110 K. Although the diffraction of the crystal was poor, with a low resolution and high mosaicity (Figure 4.12), it was a very encouraging result. It was the first diffraction of *Klebsiella pneumoniae* UDP- Gal*p* mutase crystals, at the optimised crystallisation conditions.

Figure 4.12 Diffraction pattern of UDP-Galactopyranose mutase crystal.



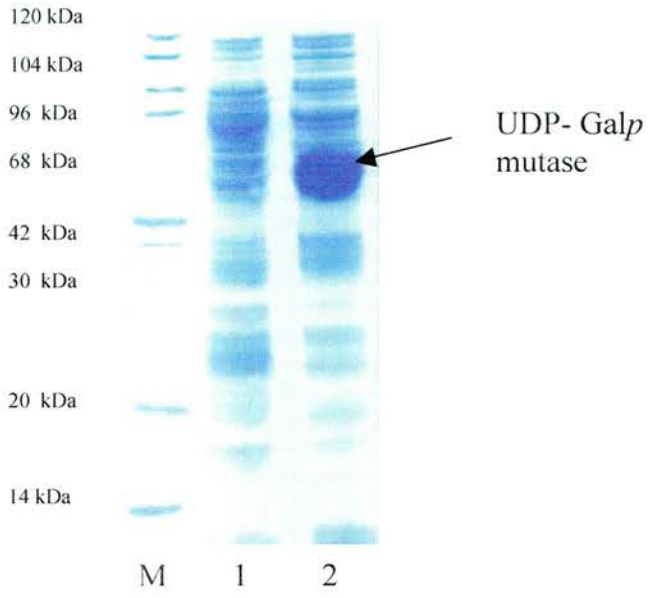
In Figure 4.12 we have the diffraction pattern from the rod shaped crystal from UDP- Gal*p* mutase. The high level of mosaicity is clearly seen in the smeared diffraction spots.

4.18 Experimental (Redox potentiometry and EPR).

4.19 His-tagged Protein Production.

Escherichia coli strain BL21 DE3 was used to express the histidine tagged *Klebsiella pneumoniae* UDP- Gal*p* mutase. The RfbD gene from the pWQ66 plasmid (Koplin, 1997) was cloned into a modified hexa His-tagged expression vector containing a Tobacco Etch Virus (TEV) protease cleavage site. The bacteria were grown aerobically at 37°C at pH 7.5 in Terrific Broth containing 50μg/ml kanomycin until an optical density of 0.6-0.8 at 600 nm, at which point RfbD expression was induced for 3 hours with 1 mM isothiogalacto-pyranoside (IPTG). Cells were harvested by centrifugation at 7500 rpm. 9.5 litres of cell culture was harvested by centrifugation at 7500 rpm for 15 minutes, collecting 54.58g of cell pellet (wet weight). Protein expression was determined by SDS-PAGE analysis (Figure 4.13).

Figure 4.13 SDS-PAGE analysis of His-tagged UDP-Galactopyranose mutase expression.



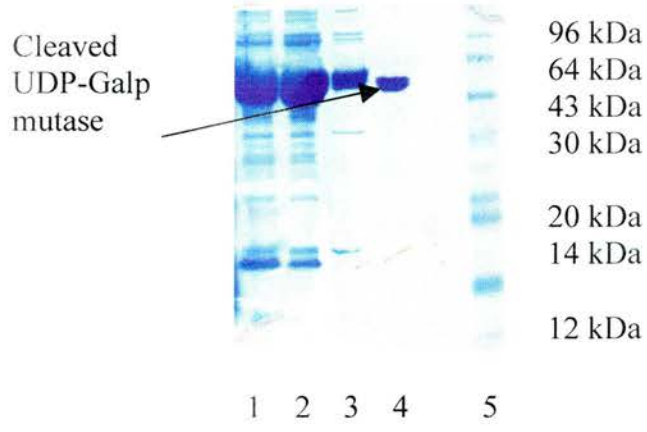
In figure 4.13, a protein band in the cell lysate (lane 2) at approximately 47 kDa has appeared in comparison with the non-induced sample (lane 1).

4.20 Protein purification.

A cell pellet (50g wet weight) was resuspended to 100ml in 50 mM HEPES buffer pH 7.0 containing 1 mM phenylmethylsulphonyl-fluoroside (PMSF) at 4°C. Lysozyme (0.1 $mg.ml^{-1}$) and DNAase I (0.1 $mg.ml^{-1}$) were added and the suspension stirred at room temperature for 30 minutes before sonication with a Sanyo soniprep 150, at 4°C for 10 x 30 seconds. The sonicate was clarified by centrifugation at 50,000g at 4°C for 30 mins. The supernatant was decanted and passed through a 0.45 μm filter (Acrodisc) before loading onto 10 ml Quiagen Ni-NTA metal chelate column at 4°C. The column was pre-charged with 10 column volumes of 500 mM nickel sulfate, and washed with 20 column volumes of 50 mM Hepes pH 7.0. The supernatant was loaded onto the column at 5 $ml.min^{-1}$ and the column washed with 20

column volumes of 50 mM Hepes, 25 mM imidazole pH 7.0 before elution of the protein with 50 mM HEPES, 250 mM imidazole pH 7.0. The eluate (approx. 60 ml) was dialyzed in 5 x 1litre batches at 4°C for 30 minutes in 50 mM Hepes pH 7.0. TEV (tobacco etch virus protease) cleavage was carried out by the addition of 30 μg/ml TEV protease and 1 mM DTT (dithiothreitol) to the sample. The cleavage reaction was left stirring at room temperature for 3 hours. The reaction mixture was dialyzed against 50 mM Hepes pH 7.0 to remove excess DTT before further purification. The cleaved protein sample was passed through a metal chelate column as before and the cleaved protein collected in the unbound fraction. SDS-PAGE analysis (Coomassie staining) shows a single band at the correct molecular weight (Figure 4.14). MALDI TOF spectrometry of trypsin digests unequivocally confirmed identity.15 – 20 mg of pure protein was obtained from 1 liter of culture. The protein sample was concentrated using an Amicon ultra filtration cell (10 kDa cut-off) to 5 $mg.ml^{-1}$ and then frozen at -80°C with 30% glycerol for storage. Concentration was determined using the absorbance of the flavin at 450 nm using extinction coefficient of 13500 M^{-1} cm^{-1} (see section 4.10).

Figure 4.14 SDS-PAGE analysis of UDP-Galactopyranose purification.



Lanes 1+2 contain the elutant supernatant from the first metal chelate column. Lane 3 contains the elutant from a second run of purification through the metal chelate column. Lane 4 contains the flow through from the metal chelate column after TEV protease cleavage of the His tagged linker. Further crystalisation trials were not pursued from this point due to constraints on protein quantity for redox potentiometry experiments.

4.21 Redox potentiometry.

Redox titrations were carried out at 25 ± 2 °C under a nitrogen atmosphere in a Belle technology glove-box. Oxygen levels maintained at less than 5 ppm as described previously (Daff, 1997). Aliquots of 1.5 ml of 5 $mg.ml^{-1}$ UDP- Gal*p* mutase were run through a Bio-Rad 10 DG column equilibrated with 100 mM MOPS pH 6.5, 7 and 7.5 to remove oxygen and glycerol. The resultant 3 ml aliquots of 50-100 μM UDP- Gal*p* mutase were titrated potentiometrically using sodium dithionite as reductant and potassium ferricyanide as oxidant in the presence of redox mediators. These included 1 μM pyocyanine, 10 μM hydroxynaphthaquinone, 2 μM FMN, 0.5 μM benzyl-viologen and 2 μM of methyl-viologen. Further titrations were conducted in the presence of UDP-galactose (200 μM), 2-fluoro-deoxy-UDP- Gal*p* (200 μM) and UDP (500 μM).

4.22 Manipulation and analysis of redox data.

The potentiometric titrations were analyzed by plotting absorbance at 450 nm and 580 nm against the electrode potential. Data were fitted simultaneously at both wavelengths to a Nernst function describing a two-electron redox process (Daff, 1997) using non-linear regression analysis, (Figure 4.15) equation 1 (Origin software, Microcal). Some analyses required subtraction of absorbance contributions caused by increases in turbidity. As turbidity causes greater scattering at lower wavelengths, it was possible to remove this effect

by using A_{600}-1.5 x A_{670} instead of the absorbance at 580 nm. In order to estimate the single electron reduction potentials effectively in this case, the absorbance coefficients A_{600} and A_{670} for the semiquinone were fixed during the fitting process to values determined using the substrate-bound enzyme.

Figure 4.15 Equation 1: data manipulation.

Equation (1)

$$A = (a10^{(E-E'_1)/59} + b + c10^{(E'_2-E)/59}) / (1 + 10^{(E-E'_1)/59} + 10^{(E'_2-E)/59})$$

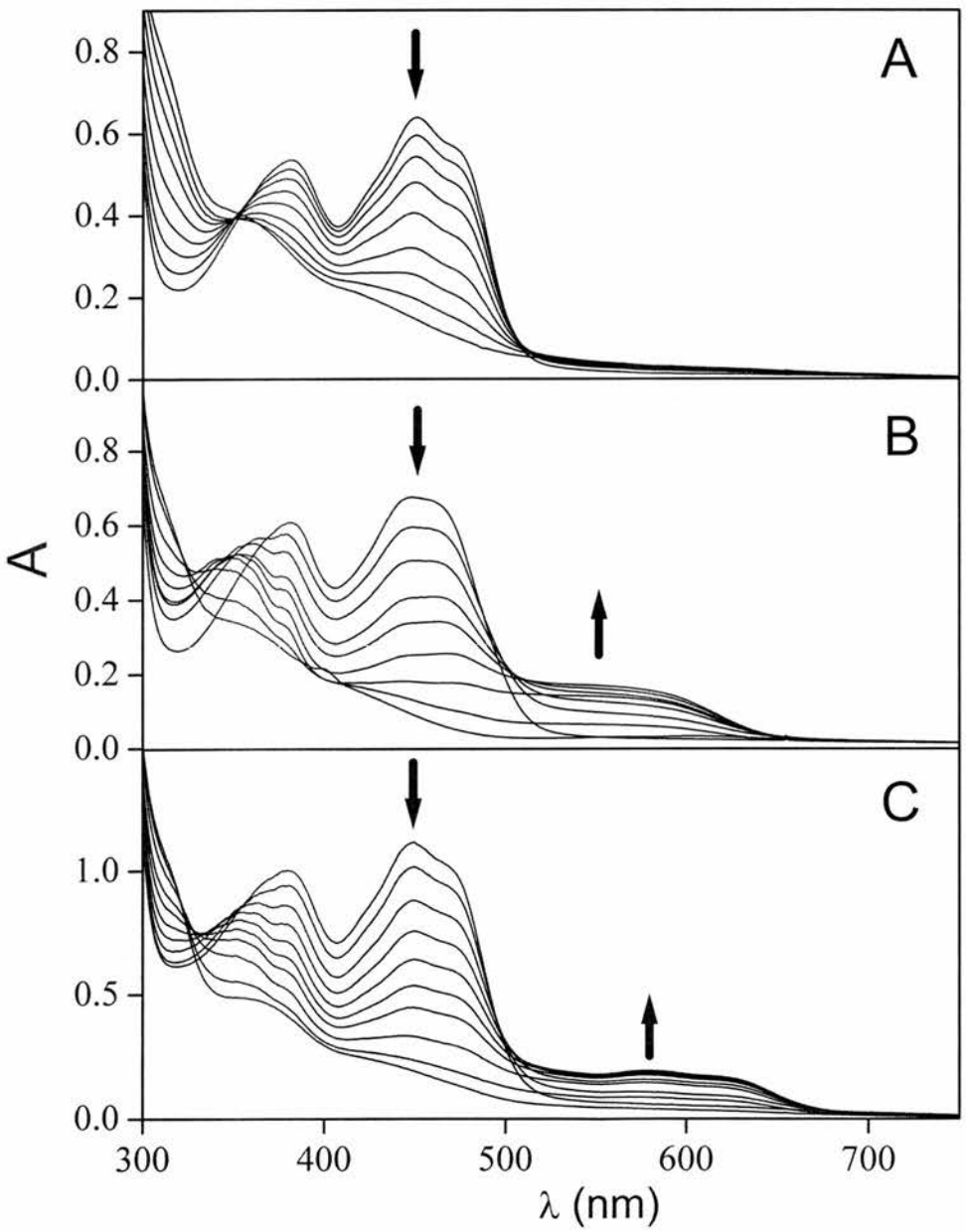
4.23 Redox potentiometry and ligand effects observed in the near ultra-violet and visible spectral region.

UDP- Gal*p* mutase is a typical flavo-protein with absorption maxima at 457 nm and 380 nm. Figure 4.16 shows spectra taken during progressive reduction of the enzyme during a redox titration. Three different conditions are shown, the absence of substrate (panel A), the presence of UDP (panel B) and the presence of UDP-galactose (panel C).

In the absence of substrate, reduction of the enzyme causes a decrease in the intensity of the visible absorption bands and no strong absorbance characteristic of either red anionic or blue neutral flavo-semiquinone. In the presence of 500μM UDP (Figure 4.16 panel B), the spectrum of the fully oxidized enzyme is altered. The most obvious effect is an apparent loss of fine structure in the band centered on 457 nm. Spectra obtained during reductive titration of the enzyme in the presence of UDP show the presence of a blue flavo-semiquinone as a redox intermediate, with an increase in absorption between 500 nm and 650 nm. In the presence of 200μM substrate (UDP-galactose), the spectrum of the fully oxidized enzyme is identical to that of the substrate-free enzyme (Figure 4.16 panel C). However, progressive reduction of the substrate-bound enzyme also leads to the formation of a stable blue flavo-

semiquinone, which bleaches on further reduction. The semiquinone formed during titration in the presence of UDP-galactose has a clear absorption maximum at 580 nm and extends to longer wavelengths than that formed with bound UDP, with more spectral fine-structure in the 457nm region.

Figure 4.16 UV/Visible absorption spectra taken during the reduction steps of the UDP-Gal*p* mutase potentiometric redox titration.



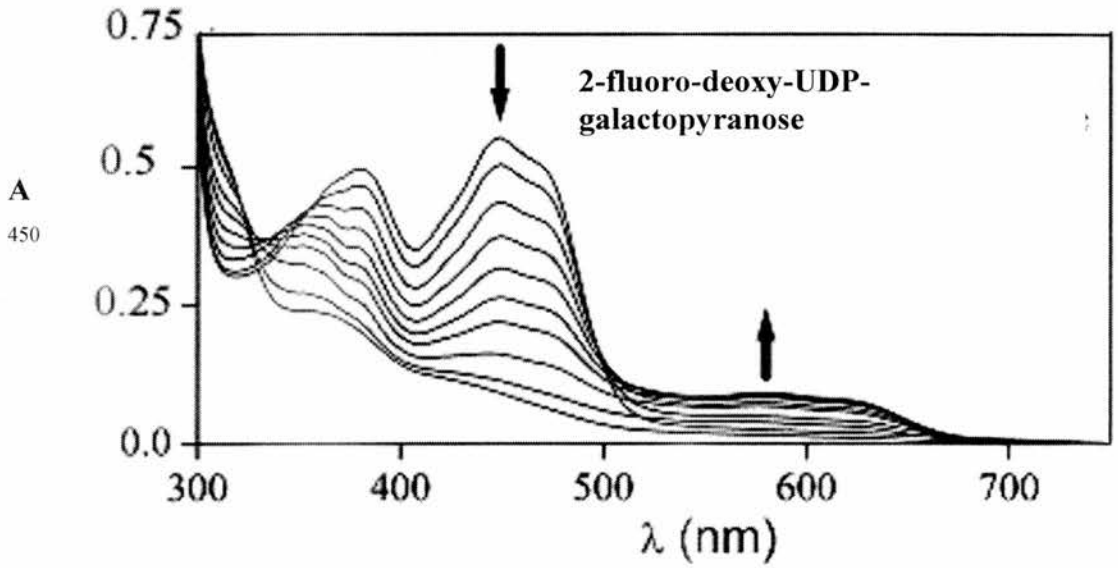
Spectra were recorded at different potentials over the reductive phase of a potentiometric redox titration of UDP- Gal*p* mutase: in the absence of ligand (A), in the presence of UDP (B) and in the presence of UDP-galactose (C). The enzyme concentration was 100μM in 100mM MOPS at pH 7.0 in A and B and pH 6.5 in C.

Figure 4.18 shows plots of the data taken from the redox titrations. In panel A of Figure 4.18 there is an analysis of a redox titration of the enzyme in the absence of substrate. Changes at 450 nm and 600 nm are plotted as a function of redox potential. The 450 nm data shows an n=2 midpoint potential (pH 7.0) of -32 mV (w.r.t the hydrogen half cell). Also shown in panel A of Figure 4.19 there is a plot of the change in absorbance at 600 nm (Δ A*) corrected for baseline drift (magnified 10-fold). Although only small, this effect indicates that a small amount of neutral blue semiquinone is formed during the titration. Using absorption coefficients calculated from the substrate-bound data, it was possible to estimate the potentials of the two one-electron redox couples by fitting the absorbance data at both wavelengths simultaneously. The two half potentials are separated by approximately 137 mV, with the midpoint potential of the oxidised/semiquinone FAD redox couple (E_1) at –100 mV (see Table 4.04).

Further redox titrations were conducted in the presence of 200 μM UDP-galactose at pH 6.5, pH 7.0 and pH 7.5. All titrations were reversible, generating the same spectral species at similar potentials in both oxidative and reductive directions. As shown in Figure 4.16 (Panel C), the presence of substrate results in the stabilization of the blue semiquinone form of the FAD with a broad absorption band appearing around 580 nm as a redox intermediate. Midpoint reduction potentials in the presence of ligands were determined for the two one n=1 transitions from analysis of the spectral changes at 450 nm and 580 nm (shown in Figure 4.18, Panel B). The datasets were fitted simultaneously at both wavelengths for each titration.

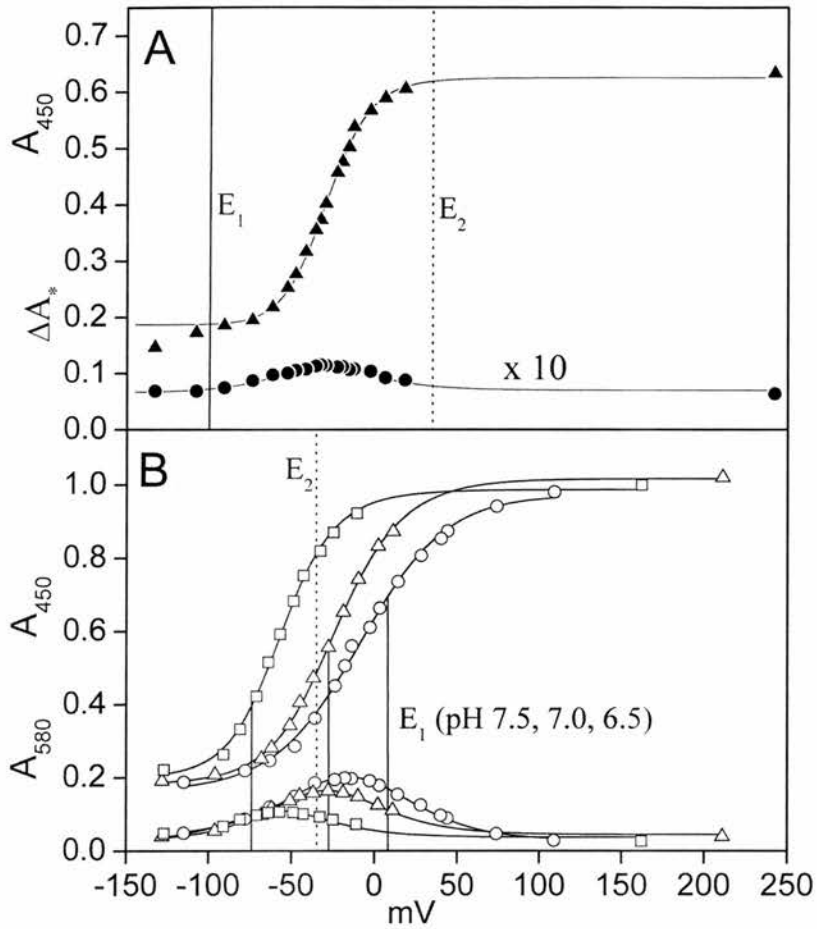
Redox titrations conducted in the presence of 2-fluoro-deoxy -galactopyranose (gift of Laura L. Kiessling, University of Wisconsin) were identical to those conducted with UDP-galactose (Figure 4.17).

Figure 4.17 Redox titration of 2-fluoro-deoxy-UDP-galactopyranose.



However, in the presence of UDP the titrations were not readily reversible, the hysteresis prevented accurate determination of reduction potentials. The comparison between the spectra for UDP, UDP-galactose and 2-fluoro-deoxy-UDP-galactopyranose (Figure 4.16 and Figure 4.17) shows that the UDP perturbs the spectrum of the oxidised enzyme, UDP-galactose and 2-fluoro-deoxy-UDP-galactopyranose does not. The most probable explanation of this is that UDP binds leaving the galactose binding site vacant which perturbs the flavin spectrum.

Figure 4.18. Analysis of the potentiometric titrations in the visible spectral region; determination of the reduction potentials.



Panel A in Figure 4.18 shows the absorbance as a function of the ambient redox potential at 450 nm ▲-▲, and at 600 nm ●-● where the absorption difference was calculated using: (A_{600}-1.5 x A_{670}) x 10 (ΔA_*) The titration was performed in 100 mM MOPS pH 7.0 as described previously (section 4.20).

Panel B shows normalized absorbance changes at 450nm (upper set) and 580 nm (lower set) for UDP-Gal*p* mutase in the presence of 200 μM UDP-galactose in 100 mM MOPS at pH 6.5 (o-o), pH 7.0 (Δ-Δ),pH 7.5 (□-□). The data points were fitted to a Nernst relationship as described in section 4.20. Values obtained for E_1, the quinone/semiquinone redox couple and E_2, the semiquinone/hydroquinone redox couple are given in Table 4.04.

Table 4.04 Midpoint reduction potentials (v NHE) for the FAD of UDP- Gal*p* mutase in the presence and absence of substrate.

	E_1 (mV)[a]	E_2 (mV)[b]	E_{2c} (mV)[c]
Mutase (pH 7.0)[d]	-100 ± 3	+37 ± 3	-32
Mutase + UDP-Gal (pH 6.5)[d]	+4 ± 3	-34 ± 4	-15
Mutase + UDP-Gal (pH 7.0)[d]	-27 ± 2	-34 ± 3	-31
Mutase + UDP-Gal (pH 7.5)[d]	-75 ± 3	-43 ± 3	-59

[a] Midpoint potential (n=1) of the oxidized/semiquinone FAD redox couple

[b] Midpoint potential (n=1) of the semiquinone/hydroquinone FAD redox couple

[c] Midpoint potential (n=2) of the oxidized/hydroquinone FAD redox couple

[d] Conducted in 100 mM MOPS pH 6.5, 7.0 or 7.5 (± 0.1) as indicated.

At pH 7.0 the reduction potential of the two-electron couple is the same in the absence of substrate (-31 mV, Table 4.04). However, the one-electron couples are shifted by 70 mV each in opposite directions leading to a stabilization of the neutral (blue) semiquinone. At pH 6.5 the oxidized/semiquinone (E1) potential is shifted up by 31 ± 5 mV relative to pH 7, whereas the semiquinone/hydroquinone (E2) potential is unchanged. For each proton coupled to an electron acquisition, a pH shift of 0.5 units should be accompanied by a reduction potential shift of 30 mV. This indicates that a single proton is accepted by the flavin on formation of the semiquinone, but no protons are associated with the subsequent formation of the hydroquinone,

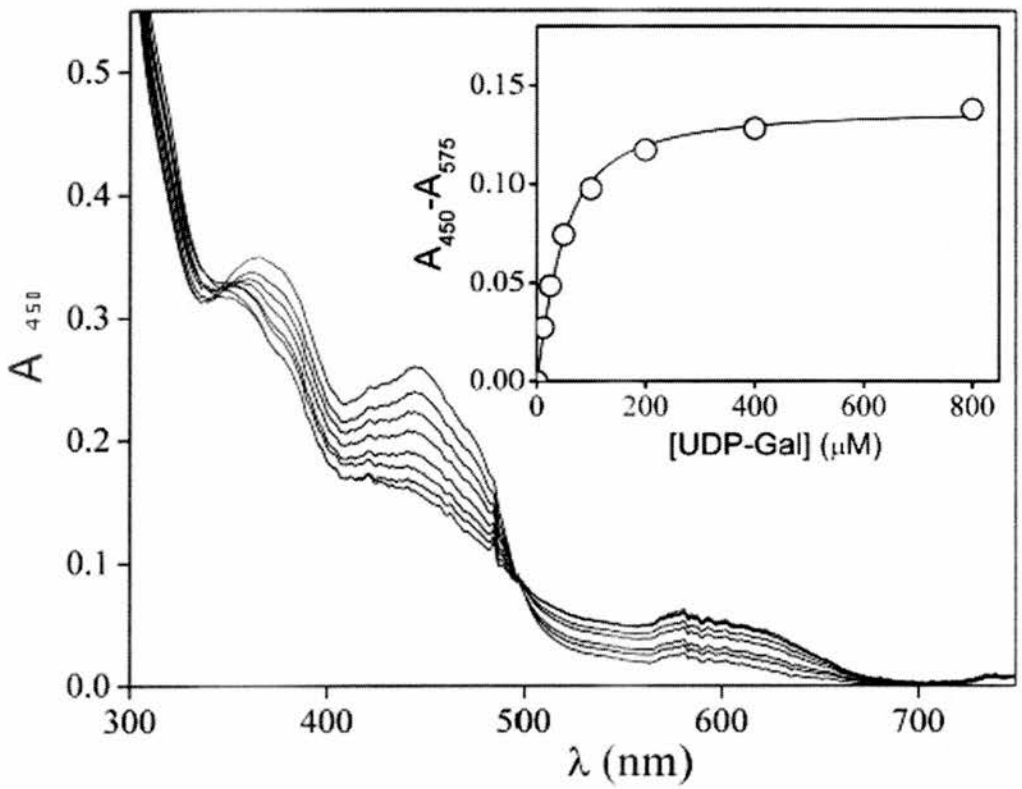
i.e. $FAD^{-} \rightarrow FADH^{\bullet} \rightarrow FADH^{-}$. The same pH dependency is observed to pH 7.5 there are similar shifts in the opposite direction of 48 ± 5 mV (E1) (Table 4.04).

4.24 Substrate-binding titrations.

UDP- Gal*p* mutase was prepared in an anaerobic glove box as described above at a concentration of approximately 35 μM. Aliquots of sodium dithionite were added until the enzyme was approximately 50% reduced (characterized by loss of half the absorbance at 450 nm). UDP-Gal and UDP were titrated into the solution stepwise. The absorbance change (A_{450} - A_{575}) was plotted against substrate concentration and fitted to a single-binding site model using non-linear regression analysis to determine an apparent dissociation constant (K_d).

In order to determine the dissociation constant of the substrate, the difference in the visible absorption spectrum caused by semiquinone stabilization was used as a quantitative measure of bound substrate (Figure 4.17). The enzyme was first reduced until only half of the 457 nm absorption band remained and was then titrated with substrate (UDP-gal). This caused a decrease in absorbance at 457 nm and a concomitant increase at 580 nm, characteristic of blue semiquinone formation (see Figure 4.16). Addition of excess substrate led to saturation and no further spectral change. The change in the absorbance difference A_{450} - A_{575} was fitted to a single binding-site model to give an apparent K_d of 28 ± 3 μM. for UDP-galactose. A similar titration with UDP was also conducted, but the data did not conform to any reasonable model.

Figure 4.19 Determination of the K_d for UDP-Gal with the UDP-Galactopyranose mutase.



UV/Visible absorption spectra of 50 %-reduced UDP- Gal*p* mutase (35 μM) titrated with aliquots of UDP-Gal up to 800 μM. Inset: plot of the absorbance difference at 450nm and 575 nm v the UDP-Gal concentration, fitted to a single-site saturation binding model (K_d = 28 ± 3 μM).

4.25 Electron paramagnetic resonance (EPR) studies.

EPR spectra were obtained using a Bruker X-band EPR spectrometer (EMX). The EPR experimental parameters were the following; temperature 120K, modulation frequency, 100 kHz; modulation amplitude, 0.2 mT; microwave frequency, 9.512GHz; microwave power, 0.2 mW. Sample temperature was regulated by a liquid nitrogen transfer system.

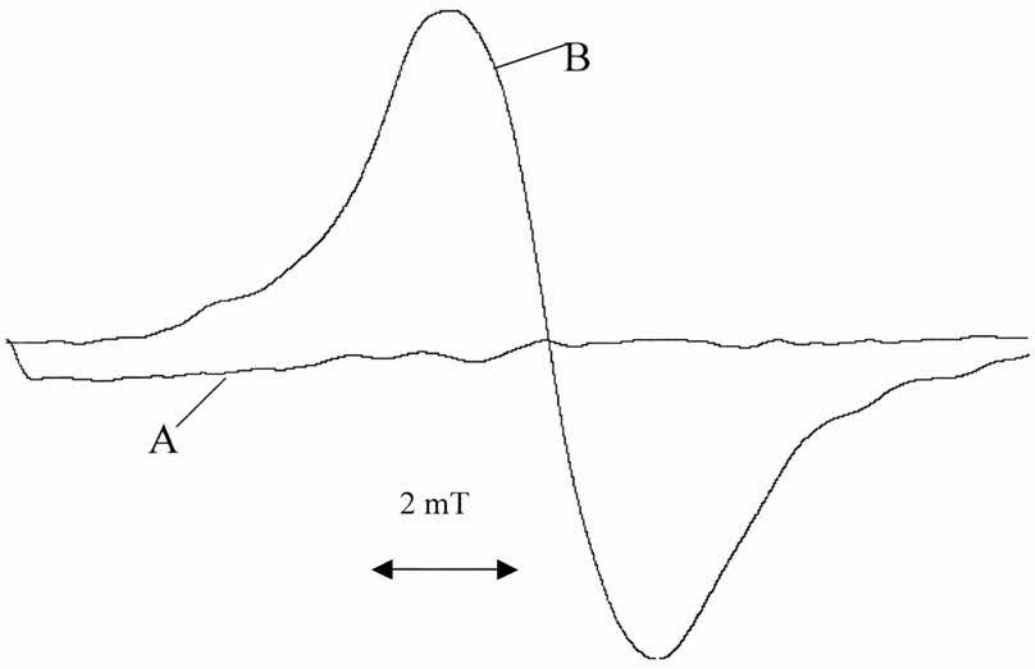
EPR spectra were obtained of UDP- Gal*p* mutase samples potentiometrically poised during redox titration and trapped by freezing. The titrations were performed according to (Dutton, 1978) and the mediating dye system developed to give minimum free radical interference (Hastings, 1998). The redox mediators included the following at 50 μM except 5-phenazine ethoxy-sulphate which was added to a final concentration of 0.5 μM: indigo tetrasulphonate ($E_{m7=}$ -47 mV), duraquinone ($E_{m7=}$ +10 mV), 2-hydroxy-napthaquinone ($E_{m7=}$ -146 mV), 5-phenazine ethoxy-sulphonate ($E_{m7=}$ +74 mV) and 1,4-napthaquinone ($E_{m7=}$ +60 mV). The UDP- Gal*p* mutase concentration was 40 μM and the buffer 100 mM Hepes pH 7.0 and when present, the substrate UDP-galactose was 150 μM. The ambient redox potential (E_h) was adjusted by titration of potassium ferricyanide and sodium dithionite and measured using a combination platinum (Ag/AgCl) reference electrode purchased from Russell pH Ltd (Auchtermuchty, Fife, U.K.). Redox titrations were conducted anaerobically; the redox titration vessel was continuously flushed with N_2 that had been passed through a Nil-Ox O_2-scrubbing apparatus (Jencons Scientific, Hemel Hempstead, Herts, U.K.). The samples were poised, transferred anaerobically to EPR tubes, quick frozen in an isopentane-cyclohexane freezing mixture and stored under liquid N_2 until analysed.

The semiquinone signals were quantitated by double integration of the signal against a 10 μM 1,1-diphenyl-2-picrylhydrazyl free radical standard and a 1 mM $CuSO_4$ standard and analysed as previously described (Hastings, 1998).

4.26 Redox potentiometry and ligand effects observed by electron paramagnetic resonance spectroscopy.

Potentiometric redox titrations were performed in the presence and absence of substrate and samples measured by EPR spectroscopy, as described in section 4.25. EPR can only detect the semiquinone radical form of the enzyme. In the absence of substrate no significant radical signal is observed. In the presence of 150 μM UDP-galactose a radical signal is observed and its magnitude follows a bell-shaped titration curve, which is similar to the optical data under similar conditions. EPR spectra poised at approximately –50 mV, in the presence (B) and absence (A) of substrate are shown in Figure 4.20.

Figure 4.20 EPR spectra of the flavo-semiquinone of UDP-galactopyranose mutase in the absence and presence of substrate.



The radical observed has the line-shape and line-width expected of a neutral flavo-semiquinone by comparison with known neutral and anionic flavo-semiquinones. The presence of a strongly coupled proton at the N(5) position causes broadening and a peak to trough line-width of 1.85-2.0 mT, while an anionic flavo-semiquinone gives a narrower peak to trough line-width of 1.5 mT (*7*). The zero crossing point of the UDP- Gal*p* mutase flavo-seminquinone spectrum is approximately g = 2.005 and the peak to trough line-width is approx. 1.85 mT. Quantitation of the EPR spectrum gives the maximal radical at concentration –50mV as 50-60% of the enzyme concentration.

4.27 Discussion.

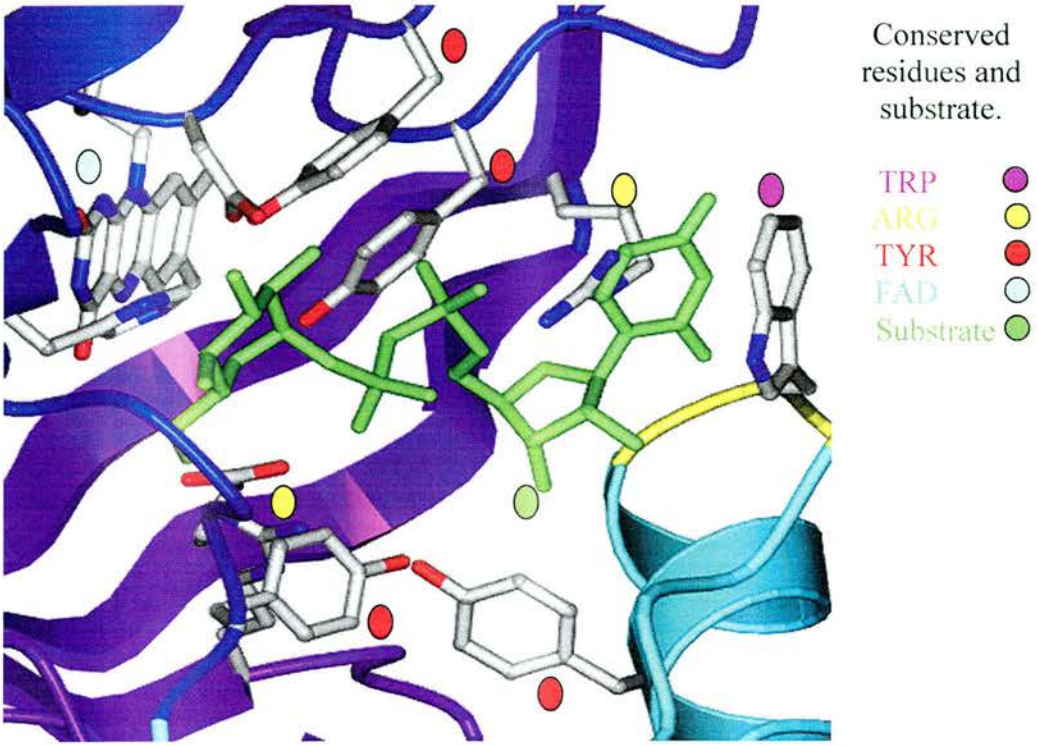
Over expression of UDP-Gal*p* mutase from *Klebsiella pneumoniae* produced consistently high yields of purified protein from both the pWQ66 vector and the new His-TEV modified vector (20mg/litre) throughout this study. The initial crystallisation results that produced the larger, seeded 0.6 mm UDP- Gal*p* mutase protein crystal rod, although encouraging, have failed to provide further UDP- Gal*p* mutase crystals of appropriate size for crystallographic study. Extensive attempts at seeding produced a single crystal of appropriate size, and therefore may be attributed more to good fortune than good laboratory practice. The failure to reproduce crystals of a suitable size may be attributed to a non-homologous protein sample with small protein impurities. Recent studies by co-workers using the new His-TEV vector, have produced crystals of a suitable size for X-ray crystallography in new conditions: 20-30% PEG mono-methyl-ether 550, 0.1 M Bis/Tris pH 6.5, 50 mM $CaCl_2$ (data not shown), indicating that initial trials may not have attained appropriate levels of protein purity.

The solution of the *Escherichia coli* UDP- Gal*p* mutase structure (Figure 4.01) has provided researchers with a set of structural definitions. The structure is a dimer in agreement with previously reported dynamic light scattering (M^{c}Mann, 1998) results, with 13 bridging water

molecules and 10 hydrogen bonds across the dimer interface. With each monomer consisting of three separate domains. The structure places the FAD cofactor at the N-terminal $\alpha\beta\alpha$ Rossman fold motif of domain I. Domain II consists of the dimer interface and domain III the flexible linker.

A large cleft is flanked by domain I and domain II and capped at one end by domain III. The FAD isoalloxazine ring is situated along one side of the cleft, with the re face facing outward, with the N5 of the FAD hydrogen bonding to the Alanine 55 carbonyl. The conserved tryptophan residue at position 156 sits opposite the FAD co-factor at the far side of the cleft. Four conserved tyrosines (positions 151, 181, 311, 346) sit around the cleft with their side chains pointing into the open cleft and two conserved arginines (247 and 278) are positioned at the capped end of the cleft. Modelling studies of UDP-Galactose binding (Figure 4.21) suggest the stacking of the UDP nucleotide at the tryptophan (156) with the two arginine residues binding the phosphate backbone and the galactose ring situated adjacent to the isoalloxazine ring. The four conserved tyrosines are positioned in close contact with the galactose ring.

Figure 4.21 A 2D representation of the substrate bound UDP- Gal*p* mutase model.



The stabilisation of a neutral semiquinone radical by substrate and the reduction of the flavin to an anionic hydroquinone, is interesting in terms of the potential involvement of the flavin in catalysis. While it is possible that the flavin has no direct role in catalysis, and serves only to stabilise the protein fold, this seems unlikely given the impact that its redox state has on the catalytic rate, (at least 10^5) (Sanders, 2001). The results obtained are the first detailed evaluation of the thermodynamic redox properties of the flavin in an enzyme of this type. These parameters limit what is and is not possible chemistry for the flavin during turnover. They establish that the reduced flavin is found in the anionic hydroquinone form, that substrate binding causes a stabilisation of the neutral, blue semiquinone form of the FAD and that the single electron oxidation of hydroquinone to semiquinone is not coupled to proton transfer in the pH range studied. The results also suggest that the carbohydrate part of the substrate is positioned close in space to the isoalloxazine ring of the FAD on binding. Any

proposed mechanism, requires that the glycosidic bond is broken, i.e. UDP is transiently separated from the galactose, to account for the observed scrambling of the oxygen atoms attached to the glycosidic phosphate group during turnover (Barlow, 1999). This seems to force the conclusion that a step in the reaction is the formation of an oxocarbonium anion, even if only at the transition state. Combining this requirement with the redox results may reveal possible mechanisms, involving the transfer of two, one or no electrons.

The published proposal that the turnover mechanism progresses through a bicyclic intermediate is attractive for its simplicity (Barlow, 1999, 2000; Zhang, 2000, 2001) and is consistent with the oxygen scrambling data (Barlow, 1999). The redox potentiometry data provide some support for this mechanism. The carbohydrate moiety of UDP-Gal*p* is close in space to the flavin and the fully reduced hydroquinone is anionic. Therefore the negatively charged isoalloxazine ring could stabilise the formation of a carbocation transition state on the substrate involved in forming the bicyclic sugar. Such an interpretation could explain the effect of redox state of the flavin on activity. However for simple charge stabilisation this is an extremely large effect. The mechanism requires an intermediate state consisting of two molecules: UDP and a bicyclic sugar (Barlow, 1999). The bicyclic sugar has been made and is kinetically stable in water (R.A. Field, University of East Anglia, personal communication). UDP is also chemically stable. Therefore, this mechanism appears to involve formation of two stable intermediates, neither of which have been observed as bi-products of catalytic turnover. Although not conclusively ruling out this mechanism, the presence of a kinetic trap coupled to a requirement for the flavin to be reduced does provide difficulty for this mechanism. Without structural data we cannot exclude the possibility of another nucleophile (other than O4) attacking the oxocarbonium anion and thus preventing formation of the stable bicyclic sugar cannot be excluded. However, unless $FADH^-$ is directly

involved in activating this nucleophile, it is difficult to see why the oxidation state of flavin is so important for activity.

An attractive alternative to this mechanism is a crypto-redox process involving transient transfer of electrons. However, previous studies (personal communication David Sanders, University of Sasketchewan, Sasketchewan Canada.) have been unable to observe any changes in the spectral properties of the flavin during stopped flow rapid turnover experiments. This negative result indicates that either electron transfer does not occur, or is not part of the rate-determining step. This is entirely reasonable if, for example, electron transfer results in the formation of an unstable species. For electron transfer to be feasible the potential of electron donor and electron acceptor must be closely matched, since the electron(s) require movement back and forth without forming stable intermediates. Secondly the substrate must be close enough in space for such a redox process to be feasible. The redox potentiometric data quantitates the driving force for electron transfer from the FAD hydroquinone for a one and two electron process. There are no estimates in the literature for the reduction potentials of oxocarbonium ions. However, they may have an appropriately accessible redox state. The results comparing the UDP and UDP-galactose titrations suggest that, indeed the carbohydrate ring is close in space to the isoalloxazine ring of the FAD. Flavins commonly transfer or accept 2 electrons during turnover. The fact that both 2 and 3 fluoro-deoxy UDP-galactose are viable substrates for the UDP-Gal*p* mutase presents formidable difficulties for any mechanism involving formation of a carbanion at C1, C2, C3 or C4. In each case, the β-elimination of F^- would be highly favored, but no fluoride ions were detected. We are unable to formulate a mechanism which involves two-electron transfer from FAD to substrate but does not involve carbanion formation at C1, C2, C3 or C4. A two electron process therefore seems unlikely.

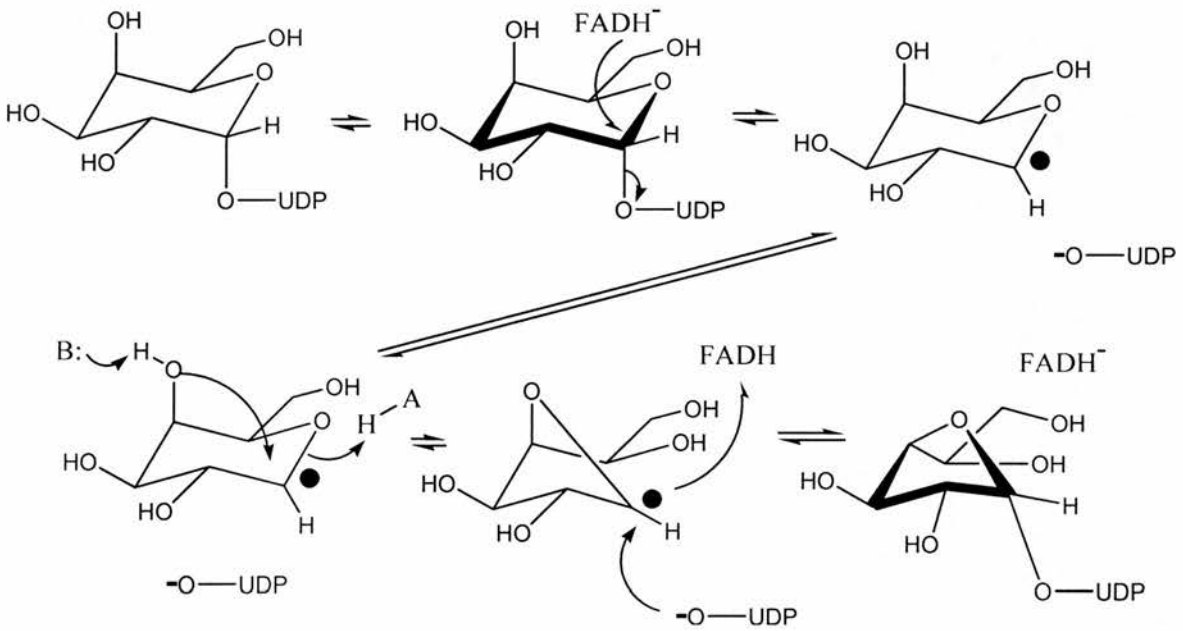
For one electron transfer to be viable, the semiquinone form of the donor must be thermodynamically accessible. The results of this study show that the semiquinone form of flavin is indeed significantly stabilised in the substrate bound enzyme. Given the available data, we propose a mechanism involving a single electron transferring to the oxocarbonium to produce an anomeric radical. Single electron transfer from the anionic hydroquinone results in formation of the neutral blue semiquinone so need not be coupled to proton transfer. This would allow the reaction to be both rapid and reversible. By generating a carbohydrate radical transition state in this direct manner, stable intermediates would also be avoided. The formation of a radical at the C1 position of sugars has recent precedent in the chemistry of both furanosyl and pyranosyl rings (Togo, 1998). Such anomeric radicals are highly reactive and have been harnessed synthetically by other laboratories (Groeninger, 1987; Togo, 1998). Radical formation would facilitate ring contraction by inducing nucleophilic attack by O4 at the C1 position, with cleavage of the C1 O5 bond. The electron could then be transferred from the anomeric position of the carbohydrate back to the FAD semiquinone. The probability of this mechanism depends on the accessibility of the carbohydrate radical transition state. It has the attractions of avoiding carbanion formation, which is largely ruled out on the basis of experiments with fluoro-substrates and of introducing a redox role for the flavin, which seems likely in view of the dependence of activity on its reduction state. The redox potentiometry results demonstrate that the FAD is ideally suited to fulfilling its role in this particular mechanism (Figure 4.22).

4.28 Conclusion.

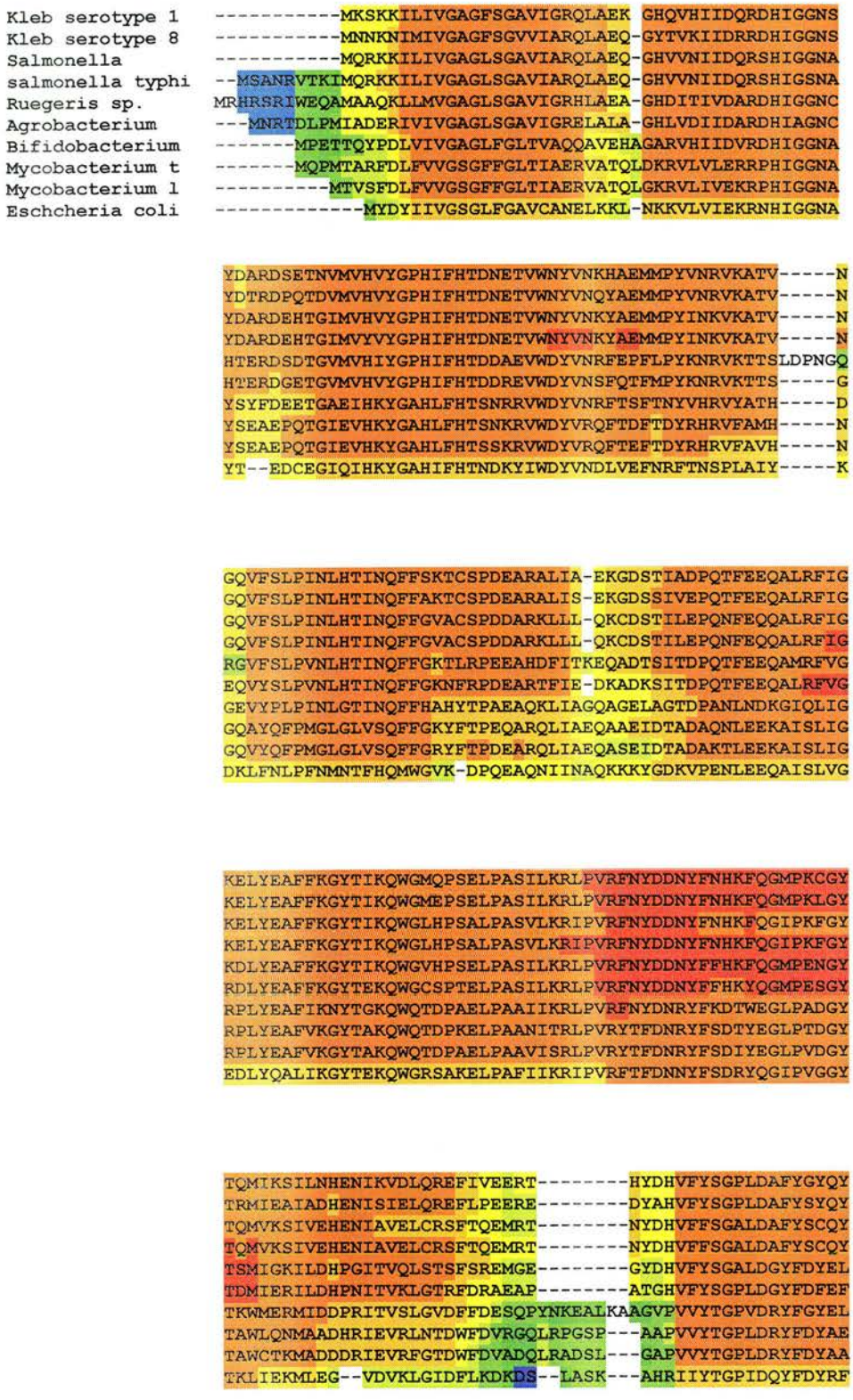
The mechanism of ring contraction performed by the UDP-Gal*p* mutase enzyme provides a formidable challenge to the chemist and biologist. Our work provides important quantitative measurements of the chemistry that is possible at the flavin cofactor of the enzyme. We have

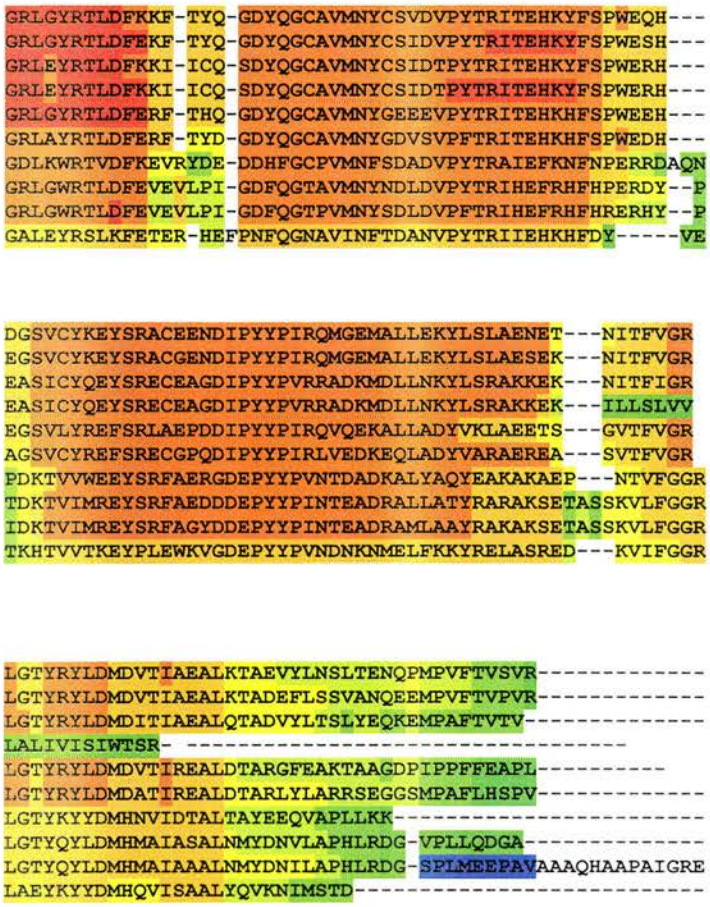
advanced a testable mechanism consistent with all the published data, which involves a single electron crypto redox process.

Figure 4.22 Proposal for the outline mechanism of ring contraction.



Appendix 1 UDP-Gal*p* mutase sequence alignment showing sequence similarity with several clinical infectious pathogens.




GRLGYRTLDFKKF-TYQ-GDYQGCAVMNYCSVDVPYTRITEHKYFSPWEQH---
GRLGYRTLDFEKF-TYQ-GDYQGCAVMNYCSIDVPYTRITEHKYFSPWESH---
GRLEYRTLDFKKI-ICQ-SDYQGCAVMNYCSIDTPYTRITEHKYFSPWERH---
GRLEYRTLDFKKI-ICQ-SDYQGCAVMNYCSIDTPYTRITEHKYFSPWERH---
GRLGYRTLDFERF-THQ-GDYQGCAVMNYGEEEVPYTRITEHKHFSPWEEH---
GRLAYRTLDFERF-TYD-GDYQGCAVMNYGDVSVPFTRITEHKHFSPWEDH---
GDLKWRTVDFKEVRYDE-DDHFGCPVMNFSDADVPYTRAIEFKNFNPERRDAQN
GRLGWRTLDFEVEVLPI-GDFQGTAVMNYNDLDVPYTRIHEFRHFHPERDY--P
GRLGWRTLDFEVEVLPI-GDFQGTPVMNYSDLDVPFTRIHEFRHFHRERHY--P
GALEYRSLKFETER-HEFPNFQGNAVINFTDANVPYTRIIEHKHFDY-----VE

DGSVCYKEYSRACEENDIPYYPIRQMGEMALLEKYLSLAENET---NITFVGR
EGSVCYKEYSRACGENDIPYYPIRQMGEMALLEKYLSLAESEK---NITFVGR
EASICYQEYSRECEAGDIPYYPVRRADKMDLLNKYLSRAKKEK---NITFIGR
EASICYQEYSRECEAGDIPYYPVRRADKMDLLNKYLSRAKKEK---ILLSLVV
EGSVLYREFSRLAEPDDIPYYPIRQVQEKALLADYVKLAEETS---GVTFVGR
AGSVCYREFSRECGPQDIPYYPIRLVEDKEQLADYVARAERFA---SVTFVGR
PDKTVVWEEYSRFAERGDEPYYPVNTDADKALYAQYEAKAKAEP---NTVFGGR
TDKTVIMREYSRFAEDDDEPYYPINTEADRALLATYRARAKSETASSKVLFGGR
IDKTVIMREYSRFAGYDDEPYYPINTEADRAMLAAYRAKAKSETASSKVLFGGR
TKHTVVTKEYPLEWKVGDEPYYPVNDNKNMELFKKYRELASRED---KVIFGGR

LGTYRYLDMDVTIAEALKTAEVYLNSLTENQPMPVFTVSVR----------
LGTYRYLDMDVTIAEALKTADEFLSSVANQEEMPVFTVPVR----------
LGTYRYLDMDITIAEALQTADVYLTSLYEQKEMPAFTVTV----------
LALIVISIWTSR- ----------
LGTYRYLDMDVTIREALDTARGFEAKTAAGDPIPPFFEAPL----------
LGTYRYLDMDATIREALDTARLYLARRSEGGSMPAFLHSPV----------
LGTYKYYDMHNVIDTALTAYEEQVAPLLKK----------
LGTYQYLDMHMAIASALNMYDNVLAPHLRDG-VPLLQDGA----------
LGTYQYLDMHMAIAAALNMYDNILAPHLRDG-SPLMEEPAVAAAQHAAPAIGRE
LAEYKYYDMHQVISAALYQVKNIMSTD----------

References.

Allard, J., Grochulski, P., Sygusch, J. (2001). Covalent Intermediate Trapped in 2-Keto-3-deoxy-6-phosphogluconate KDPG Aldolase Structure at 1.95 - Å Resolution. *PNAS* **98,** 3679-3684.

Allen, S. T., Heintzelman, G. R., Toone, E. J. (1992). Pyruvate Aldolases as Reagents for Stereospecific Aldol Condensation. *J. Org. Chem.* **57,** 426-427.

Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A., and Vijayan, M. (1996). Conformation, Protein-Carbohydrate Interactions and a Novel Subunit Association in the Refined Structure of Peanut Lectin-Lactose Complex. *J. Biol. Chem.* **259,** 281-296.

Banerjee, R., Mande, S. C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S .K., Suguna, K., Surolia, A., Vijayan, M. (1994). Crystal Structure of Peanut Lectin, a Protein with an Unusual Quaternary Structure. *Proc. Nat. Sci.* **91**.237-231

Barbas, C. F., Wang, Y-F., Wong, C-H. (1990). Deoxy-ribose-5-phosphate Aldolase as a Synthetic Catalyst. *J.A.C.S.* **112,** 2013-2014.

Barlow, J. N., and Blanchard, J.S. (2000). Enzymatic Synthesis Of UDP-(2-Deoxy-2-Fluoro)-D-Galactose and Substrate Activity with UDP-galactopyranose Mutase. *Carbohydrate. Res.* **328,** 473-480.

Barlow, J. N., Girvan, M.E., Blanchard, J.S. (1999). Positional Isotope Exchange Catalysed by UDP_Galactopyranose Mutase. *J.Am.Chem.Soc* **121,** 6968-6969.

Barlow, J. N., Marcinkeviciene, J. and Blanchard, J.S. (1999). *Enzymatic Mechanisms*. IOS Press.

Bartram, C., Edwards, R.H.T., Beynon, R.J. (1995). McArdle's Disease - Muscle Glycogen Phosphorylase Deficiency. *Biochimica. et. Biophys.* **1272,** 1-13.

Blanchard, J. S. (1996). Molecular Mechanisms of Drug Resistance in *Mycobacterium Tuberculosis. Annual Review Biochemistry* **65,** 215-239.

Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rougé, P., Fontecilla-Camps, J.C. (1990). X-ray Crystal Structure Determination and Refinement at 1.9 Å Resolution of Isolectin I from the Seeds of *Lathyrus ochrus*. *J. Mol. Biol.* **214,** 571-584.

Bourne, Y., Anguille, C., Rouge, P., Fontecilla-Camps, J-C., & Cambillau, C. (1990). Co-Crystallization and Preliminary X-Ray Diffraction Studies of *Lathyrus ochrus* Isolectin with Di- and Trisaccharides and a Biantennary Octa-saccharide. *J. Mol. Biol.* **213**.211-214

Bronner, D. C., B.R., Whitfield, C. (1992). Identification of an Atp Binding Cassette Transport Sytem Required for the Translocation of lipopolysaccharide O Antigen

Side Chains Across the Cytoplasmic Membrane of *Klebsiella pneumoniae* Serotype 1. *Mol. Microbiol.* **14,** 505-519.

Brunger, A. T. (1995). The free R Value: A More Objective Statistic for Crystallography.*Nature.***355**.472-474.

Bouckaert, J., Poortmans, F., Wyns, L., Loris, R. (1996).Sequential Structural Changes upon Zinc and alcium Binding to Metal-free Concanavalin A. *Journal of Biological Chemistry.* **271**.16144-16150.

Buchanan, L. V., Wymer, N., Henderson, D., Mehta, N., Botting, C.H., Pocivavsek, L., Fierke, C.A., Toone, E.J. & Naismith, J.H. (2001). Directed Evolution of a new catalytic site in 2-keto-3-deoxy-6-phosphogluconate Aldolase from *Eschericia coli.*" Structure folding & design. *Structure.* **9,** 1-9.

Burmeister, W. P., Cottaz, S., Driguez, H., Lori, R., Palmieri, S., Henrissat, B. (1997). The Crystal Structures of Sinapis alba Myrosinase and a Covalent Glycosyl-enzyme intermediate Provide Insights into the Substrate Recognition and Active-site Machinery of an S-glycosidase. *Structure* **5,** 664-676.

Cacciapouti, G. S., Porcelli, M., Bertoldo, M. De Rosa., Zappia, V. (1994). Purification and Characterisation of the Extremely Thermophilic and Thermostable 5-Methylthioadensosine Phosphorylase from the Archeon *Solfalobus sulfitaricus.* Purine Nucleoside Phosphorylase Activity and Evidence for Intersubunit Disulfide Bonds. *J.Biol.Chem.* **269,** 24762-24769.

Chen-Hsien, L., Chang, W.N., Wu, H.S. (1997). *Klebsiella Pneumoniae* Meningitus: Analysis on Clinical Features of Thirty Two Adult Patients. *Chin.Med.J.* **60,** 296-302.

Chervernak, M. C., Toone, E.J. (1994). A Direct Measure of the Contribution of Solvent Reorganisation to the Enthalpy of Ligand Binding. *J. Am. Chem. Soc.* **116,** 10533-10539.

Chopra, I., Brennan, P. (1998). Molecular Action Of Antimycobacterial Agents. *Tubercule and Lung Disease* **78,** 89-98.

Chou, W-C., Fotsch, C., Wong, C-H. (1995). Synthesis of Nitrocyclitols Based on Enzymatic Aldol Reaction and Intramolecular Nitroaldol Reaction. *J. Org. Chem.* **60,** 2916-2917.

Christesen, H. B. T., Jacobsen, B.B., Odili, S., Buettger, C., Cuesta-Munoz, A., Hansen, T., Brusgaard, K., Massa, O., Magnuson, M.A., Shiota, C., Matschinsky, F.M., Barbetti, F.,. (2002). The Second Activating Glucokinase Mutation (A456V) Implications for Glucose Homeostasis and Diabetes Therapy. *Diabetes* **51,** 1240-1246.

Coates, A., Yanmin, H., Bax, R., and Page, C. (2002). The Future Challenges Facing The Developement of New Antimicrobial Drugs. *Nature Reviews Drug Discovery* **1,** 895-910.

Conway, T. (1992). The Entner-Doudoroff Pathway: History, Physiology and Molecular Biology. *FEMS Microbiol Reviews* **103,** 1-28.

Cooper, S. J., Leonard, G. A., McSweeney, S. M., Thompson, A., Naismith, J. H., Qamar, S., Plater, A. R., Berry, A., Hunter, W. N. (1996). The Crystal Structure of a Class II Fructose-1,6-bisphosphate Aldolase shows a Novel Binuclear Metal-binding Active site Embedded in a Familiar Fold. *Structure* **4,** 1303-1315.

Cordova, A., Notz, W., Barbas III, C.F. (2002). Proline Catalysed One-Step Asymmetric Synthesis of 2-Hydroxy-(2E)-Hexenal from Acetaldehyde. *J.Org.Chem* **67,** 301-303.

Daff, S. N., Chapman, S.K., Turner, K.L., Holt, R.A., Govindaraj, S., Poulos, T.L., Munro, A.W. (1997). Redox Control of the Catalytic Cycle of Flavocytochrome P450-BM3. *Biochemistry* **36,** 13816-13823.

Davies, G., Withers, S.G., and Sinnot, M.L. (1997). Comprehensive biological catalysis. *Comprehensive biological catalysis.* Vol 1, 119-208. [Academic press London]).

de Bakker, P. L., Hunenberger, P.H., McCammon, J.A. (1999). Molecular Dynamics Simulations of the Hyperthermophilic protein Sac7d from *Sulfolobus acidocaldarias*: Contribution of Salt Bridges to Thermostability. *J.Mol.Biol.* **285,** 1811-1830.

Delbaere, L. T. J., Vandonselaar, M., Lata Prasad, Quail, J.W., Wilson, K.S. and Dauter, D. (1993). Structure of the Lectin IV of *Griffonia simplicifolia* and its Complex with the Lewis b Human Blood Group Determinant at 2·0 Å Resolution. *J. Mol. Biol.* **230,** 950-965.

Derewenda, Z., Yariv, J., Helliwell, J.R., Kalb, A.J., Dodson, E.J., Papiz, M.Z., Wan, T., Campbell, J. (1989). The Structure of the Saccharide Binding Site of Concanavalin-A. *EMBO* **8,** 2189-2193.

Dessen, A., Gupta, D., Sabesan, S.C., Brewer, F., Sacchettini, J.C.,. (1995). X-ray Crystal Structure of the Soybean Agglutinin Cross-Linked with a Biantennary Analog of the Blood Group I Carbohydrate Antigen. *Biochemistry* **34,** 4933-4942.

Dill, K. A. (1990). Dominant Forces in Protein Folding. *Biochemistry* **29,** 7133-7155.

Dimick, S. M., Powell, S.C., McMahon, S.A., Moothoo, D.N., Naismith, J.H., Toone, E.J. (1999). On the Meaning of Affinity: Cluster Glycoside Effects and Concanavalin A. *J.Am.Chem.Soc* **121,** 10286-10296.

Dodd, R. B., and Drickamer, K. (2001). Lectin-like Proteins in Model Organisms: Implications for Evolution of Carbohydrate-Binding activity. *Glycobiology* **11,** 71-79.

Domenico, P., Diedriech, D.D., Straus, D.C. (1985). Extracellular Polysaccharide Production by *Klebsiella Pneumoniae* and its Relationship to Virulence. *Can.J.Microbiol.* **31,** 472-478.

Dutton, P. L. Redox potentiometry: Determination of Midpoint Potentials of Oxidation-Reduction Components of Biological Electron-Transfer Systems. (1978). *Methods in Enzymology* **54,** 411-435.

Edelman, G. M., Cunningham, B. A., Reeke, G.N.Jr., Becker, J. W., Waxdal, M.J., Wang, J.L. (1972). The Covalent and Three-dimensional Structure of Concanavalin A. *Proc Nat Sci Am.* **69**.2580-2584

Einspahr, H., Parks, E.H., Suguna, K., Subramanian, E., Suddath, F.L. (1986). The Crystal Structure of Pea Lectin at 3.0-Å Resolution. *J.Biol.Chem.* **261,** 16518-16527.

Elcock, A. H. (1998). The Stability of Salt Bridges at High Temperatures: Implications for Hyperthermophilic proteins. *J. Mol. Biol.* **284,** 489-502.

Entner, N., and Doudoroff, M. (1952). Glucose and Gluconic Acid Oxidation of *Pseudomonas saccharophilia. J.Biol.Chem.* **196,** 853-862.

Erskine, P. T., Senior, N., Awan, S., Lambert, R., Lewis, G., Tickle, I.J., Sarwar, M., Spencer, P.,Thomas, P.,Warren, M.J., Shoolingin-jordan, P.M., Wood, S.P., Cooper, J.B. (1997). X-Ray Structure of 5-aminolaevulinate Dehydratase, a Hybrid Aldolase. *Nature structural Biology* **4,** 1025-1031.

Evans, P. R. (1993). Data Collection and Processing. *Proceedings of the CCP4 Study Weekend.* 114-122.

Fessner, W-D., Walter, C. (1996). Enzymatic C-C Bond Formation in Asymmetric Synthesis. *Topics in Current Chemistry* **184,** 97-194.

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

Gale, E. F., Cundliffe, E., Reynolds, PE., Richmond, M.H., and Waring, M.J., ed. (1981). *The Molecular Basis Of Antibiotic Action.* John Wiley & sons, London.

Gamblin, S. J., Davies, G. J., Grimes, J. M., Jackson, R. M., Littlechild, J. A., Watson, H. C. (1990). The Crystal Structure of Human Muscle Aldolase at 3.0 Å Resolution. *FEBS Letters* **262,** 282-286.

Gijsen, H. J. M., Qiao, L., Fitz, W., Wong, C-H. (1996). Recent Advances in the Chemoenzymatic Synthesis of Carbohydrates and Carbohydrate Mimetics. Chemical Reviews 96, 48-51. *Chemical Reviews* **96,** 48-51.

Gijsen, H. J. M., Wong, C-H. (1995). Sequential 1 Pot Aldol Reaction Catalysed by 2-deoxyribose-5-phosphate Aldolase and Fructose-1,6-bisphosphate Aldolase. *J.Am.Chem.Soc.* **117,** 2947-2948.

Goldstein, I. J., Portez, R.D. (1986). Isolation and Chemical Properties of Lectins.The Lectins: Properties, Functions and Applications. Academic Press, Orlando.

Goldstein, I. J., Reichert, C.M., Misaki, A. (1974). Interaction of Concanavalin-A with Model Substrates. *Ann. NY Acad. Sci.,* **234,** 283-295.

Griffiths, J. S., Wymer, N.J., Njolito, E., Niranjanakumari, S., Fierke, C.A., Toone, E.J. (2001). Cloning, Isolation and Characterisation of the *Thermotoga Maritima* KDPG Aldolase. *Bioorganic and Medicinal Chemistry* **10,** 545-550.

Groeninger, K. S., Jaeger, F., Giese, B. (1987).Rapid and Mild Generation of carbon radicals by an Anchimeric Approach. *Liebigs Ann. Chem.*, 731-732.

Grogan, M. J., Pratt, M.R., Marcaurelle, L.A., Bertozzi, C.R. (2002). Homogeneous Glycopeptides and Glycoproteins for Biological Investigation. *Annu Rev Biochem.* **71,** 593-634.

Guerra, A. S., van Diggelen, O.P., Carneiro, F., Tsou, R.M., Simoes, S., Santos, N.T. (1986). A Juvenile Variant of Glycogenosis IV (Andersen disease). *Eur J Pediatr* **145,** 179-181.

Hall, D. R., Leonard, G. A., Reed, C. D., Watt, C. I., Berry, A., Hunter, W. N. (1999). The Crystal Structure of *Eschericia coli* Class II Fructose-1,6-bisphosphate Aldolase in Complex with Phosphoglycolohydroxamate Reveals Details of Mechanism and Specificity. *J. Mol. Biol.* **287,** 383-394.

Hall, J. A., Thorgeirsson, T.E., Lu, Y.-K., Nikaido, Shin & H. (1997). Two Modes Of Ligand Binding in Maltose-Binding Protein of *Escherichia coli.* Electron Paramagnetic Resonance Study of Ligand-Induced Global Conformational Changes by Site-Directed Spin Labeling. *J.Biol.Chem.* **272,** 17610-17614.

Hamelryck T.W., D-T. M.-H., Poortmans F., Chrispeels M.J., Wyns L., Loris R. (1996). The Crystallographic Structure of Phytohemagglutinin-L. *J. Biol. Chem.* **271,** 20479-20485.

Hardman, K. D., Ainsworth, C. F. (1972). Structure of Concanavalin A at 2.4 Å Resolution. *Biochemistry* **11,** 4910-4919.

Hardman, K. D., Ainsworth, C. F. (1976). Structure of the Concanavalin A Methyl-alpha.-D-Mannopyranoside Complex at 6 Å Resolution. *Biochemistry* **15,** 1120-1128.

Harrop, S.J., Helliwell, J.R., Wan, T.C.M., Kalb, A.J., Tong, L., Yariv, J. (1996). Structure Solution of a Cubic Crystal of Concanavalin A Complexed with Methyl-alpha-D-Glucopyranoside. *Acta Cryst* **52,** 143-155.

Hastings, S. F., Kaysser, T.M., Jiang, F., Salerno, J.C., Gennis, R.B., Ingledew, W.J. (1998). Identification of a Stable Semiquinone Intermediate in the Purified and Membrane Bound Ubiquinol Oxidase-Cytochrome bd from *Escherichia coli. Eur.J. Biochem.* **225,** 317-323.

Heine, A., DeSantis, G., J.G. Luz, J.G., Mitchell, M., Wong, C-H., Wilson I.A. (2001). Observation of Covalent Intermediates in an Enzyme Mechanism at Atomic Resolution. *Science* **294,** 369-374.

Hester, G., Brenner-Holzach, O., Rossi, F. A., Struck-Donatz, M., Winterhalter, K. H., Smit, J. D., Piontek, K. (1992). The Crystal Structure of Fructose-1,6-bisphosphate Aldolase from *Drosophila melanogaster* at 2.5 Å Resolution. *FEBS Letters* **292,** 237.

Hoffmann, T., Zhong, G., List, B., Shabat, D., Andreson, J., Gramatikova, S., Lerner,R.A., Barbas III, C.F. (1998). Aldolase Antibodies of Remarkable Scope. *J.Am.Chem.soc.* **120,** 2768-2779.

Holm, L. Sander, C. Mapping the Protein Universe. (1996) *Science* **273** 595-602

Jia, J., Huang, W., Schorken, U., Sahm, H., Sprenger, G. A., Lindqvist, Y., Schnider, G. (1996a). Crystal Structure of Transaldolase B from *Eschericia coli* suggests a Circular Permutation of the α/β Barrel within the Class I Aldolase Family. *Structure* **4,** 715-724.

Jones, A., Zou, J., Cowan, S., Kjeldgaard, M.,. (1991). Improved Methods for Building Protein Models in Electron density Maps and the Location of Errors in these Models. *Acta Cryst.* **A 47,** 110-119.

Joost C. M. Uitdehaag, R. M., Kor H. Kalk, Bart A. van der Veen, Lubbert Dijkhuizen, Withers, S.G. & Bauke W. D. (1999). X-ray Structures along the Reaction Pathway of Cyclodextrin Glycosyltransferase Elucidate Catalysis in the α-Amylase Family. *Nature. Struct. Biol.* **6,** 432-436.

Kalb, A. J., Levitzki, A. (1968). Metal-binding Sites of Concanavalin A and their role in the binding of alpha-Methyl-D-glucopyranoside. *Biochem J.* **109,** 669-672.

Kalb, A. J., (Gilboa), Yariv, F. F. J., Eisenstein, M. (1995). Properties of a Crystal of the Complex of Methyl D-arabinofuranoside with Concanavalin A. *Acta Cryst* **D51,** 1077-1079.

Kanellopoulos, P. N., Pavlou, K., Perrakis, A., Agianian, B., Vorgias, C.E., Mavrommatis, C., Soufi, M., Tucker, P.A., Hamodrakas, S.J. (1996). The Crystal Structure of the Complexes of Concanavalin A with 4-Nitrophenyl-α-D-mannopyranoside and 4-Nitrophenyl-α-D-glucopyranoside. *J.Struct. Biol.* **116,** 345-355.

Karishikoff, A., Ladenstein, R. (1998). Proteins from Thermophilic and Mesophilic Organisms Essentially do not Differ in Packing. *Protein. Eng.* **11,** 867-872.

Kelly, R. F., Whitfield, C. (1996). Clonally Diverse Rfb Gene Clusters are Involved in Expression of a Family of Related D-Galactan O Antigens in Klebsiella Species. *J. Bacteriology* **17,** 5205-5214.

Klein, C., Schulz, G.E. (1991). Structure of Cyclodextrin Glycosyltransferase Refined at 2.0 Å Resolution. *J. Mol. Biol.* **20,** 737-750.

Kobata, A. (1992). Structures and Functions of the Sugar Chains of Glycoproteins. *Eur. J. Biochem.* **209,** 483-501.

Koeller, K. M. Wong. C.-H. (2001). Enzymes for Chemical Synthesis. *Nature* **409,** 232-240.

Koeller, K. M. Wong. C-H. (2000). Emerging Themes in Medicinal Glycoscience. *Nature Biotechnology* **18,** 835-841.

Koplin, R., Brison, J.R., Whitfield, C. (1997). The UDP-Galactofuranose Precursor Required for the Formation of the Lipopolysaccharide O Antigen of Klebsiella Pneumoniae Serotype O1 is Synthesised by the Product of the $RfbD_{kpo1}$ Gene. *Am.Soc.Biochem.Mol.Biol.* **272,** 4121-4128.

Kumar, S., Tsai, C.J., Nussinov, R. (2000). Factors Enhancing Protien Thermostability. *Protein. Eng.* **3,** 179-191.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993). PROCHECK: a Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Cryst.* **26,** 283-291.

Lawrence, M.C. Colman, P.M.(1993) Shape Complementarity At Protein/Protein Interfaces. *J.Mol.Biol.* **234**. 946-950

Leslie, A. G. W. (1992). Recent Changes to the MOSFLM Package for Processing Film and Image Plate Data. *In Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography.*

Lin, C-H., Sugai, T., Halcomb, R. L., Ichikawa, Y., Wong, C.-H. (1992). Unusual Stereoselectivity in Sialic Acid Aldolase Catalysed Aldol Condensations: Synthesis of both Enantiomers of High-Carbon Monosaccharides. *J. AM. Chem. Soc.* **114,** 10138-10145.

Lin, Z., Schwarz, F.P., Eisenstein, E. (1994). The Hydrophobic Nature of GroEL-Substrate Binding. *J.Biol.Chem.* **270,** 1011-1014.

Lis, H., N. Sharon. (1986). Lectins as Molecules and as Tools. *Ann. Rev. Biochem.* **55,** 35-67.

Loris, R., Hamelryck, T., Bouckaert, J., Wyns, L. (1998). Legume Lectin Structure. *Biochimica et Biophysica Acta* **1383**.9-36.

Loris, R., Maes, D., Poortmans, F., Wynsm, L., Boukaert, J. (1996). A Structure of the Complex between Concanavalin A and Methyl-3,6-di-O-(alpha-D-mannopyranosyl)-alpha-D-mannopyranoside Reveals Two Binding Modes. *J. Biol. Chem.* **271,** 30614-30618.

Loris, R., Steyaert, J., Maes, D., Lisgarten, J.,Pickersgill, R., Wyns, L. (1993). Crystal Structure Determination and Refinement at 2.3 Å Resolution of the Lentil Lectin. *.Biochemistry* **32,** 14229-14229.

Loris, R., Van Overberge, D., Dao-Thi, M.H., Poortmans, F., Maene, N., Wyns, L . (1994). Structural Analysis of two Crystal Forms of Lentil Lectin at 1.8 Å Resolution. *Proteins* **20,** 330-346.

Machajewski, T. D., Wong, C.-H. (2000). The Catalytic Asymmetric Aldol Reaction. *Angeward Chemistry International Edition* **39,** 1352-1374.

Maniatis, T., Fritsch, E.F., Sambrook, J.,. (1989). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbour, New York.

Matsumara, M., Signor, G., Mathews, B.W. (1989). Substantial increase of Protein Stability by Multiple Disulphide Bonds. *Nature* **342,** 291-293.

Mavridis, I. E., Hatada, M. H., Tulinsky, A., and Lebioda, L. (1982). Structure of 2-Keto-3-deoxy-6-phosphogluconate Aldolase at 2.8Å Resolution. *J. Mol. Biol.* **162,** 419-444.

Mc Donald, I. K., Thornton, J.M. (1994). Ion Pairs in Proteins. *J. Mol. Biol.* **168,** 867-885.

McConville, M. J., Collidge, T.A.C., Ferguson, M.A.T. Schnieder, P. (1993). The Glycoinositol Phopholipids of *Lieshmania mexicana* Promastigotes. Evidence for the Presence of Three Distinct Pathways of Glycolipid Biosynthesis. *J.Biol.Chem.* **268,** 15595-15604.

M^{c}Mann, S., Leonard, G.A., Buchannan, L.V., Giraud, M.F., Naismith, J.H. (1998). Initiating a Crystallographic Study of UDP-Galactopyranose mutase. *Acta. Crysta.D.* **D55,** 399-402.

Medeiros, A. A. (1997). Evolution and Dissemination Of ß-lactamases Accelerated by Generations of ß-lactam Antibiotics. *Clinical Infectious Disease* **24,** p19-45.

Mildvan, A. S., Kobes, R.D., Rutter; W.J. (1971). Magnetic Resonance Studies of the Role of the Divalent Cation in the Mechanism of Yeast Aldolase. *Biochemistry* **10,** 1191-1204.

Mims, C., Playfair, J., Roitt I., Wakelin, D., and Williams, R. (1998). *Medical Microbiology.* Mosby International Publishers, London.

Mitchell, E., Houles, C., Sudakevitz, D., Wimmerova, M., Gautier, C., Perez, S., Wu, A.M., Gilboa-Garber, N., Imberty, A. (2002). Structural Basis for Oligosaccharide-Mediated Adhesion of *Pseudomonas Aeruginosa* in the Lungs of Cystic Fibrosis Patients. *Nat Struct Biol.* . **12,** 918-921.

Moothoo, D. N., Canan, B., Field, R.A., Naismith, J.H. (1999). Man alpha-1-2 Mana-OMe Concanavalin A Complex Reveals Balance of Forces Involved in Carbohydrate Recognition. *Glycobiology* **9,** 539-545.

Moothoo, D. N., Naismith, J.H. (1998). Concanavalin A Distorts the beta-GlcNac-(1-2)-Man Linkage of beta-GlcNac-(1-2)alphs-Man-(1-3)-[beta-GlcNac-(1-2)-alpha-Man-(1-6)]-Man upon Binding. *Glycobiology* **8,** 173-181.

Mukaiyama, T. (1982). The Directed Aldol Reaction. *Organic Reactions* **28,** 203.

Murakami, M. I., K. Achiwa, K. (1996). Chemoenzymematic Synthesis of Neuraminic Acid Analogs Structurally Varied at C-5 and C-9 as Potential Inhibitors of the Sialydase of the Influenza Virus. *Carbohydrate Residue.* **280,** 101-110.

Murshudov, G. N., Vagin, A.A., Dodson, E.J.,. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst* **D53,** 240-255.

Naismith, J. H., Emmerich, C., Habash, J., Harrop,S.J., Helliwell, J.R., Hunter, W.N., Raftery, J., Kalb, A.J., Yariv, J. (1994). Refined Structure of Concanavalin-A Complexed with Methyl Alpha-D-Mannopyranoside at 2 Å Resolution and comparison with the Saccharide-Free structure. *Acta Cryst* **50,** 847-858.

Naismith, J. H., Field, R. A. (1996). Structural Basis of Trimannoside Recognition by Concanavalin A. *J. Biol. Chem.* **271,** 972.

Nassau, P. M., Martin, S.L., Brown, R.E., Weston, A. Monsey, D., M^{c}Niel, M.R., Duncan, K. (1996). Galactofuranose Biosynthesis in *Escherchia coli* K12. Identification And Cloning of UDP-Galactopyranose Mutase. *Am.Soc.Microbiol.*, 1047-1052.

Nelson, K. E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garret, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Philips, C.A., Richardson, D., Hiedleberg, J., Sutton, G.G., Fleischmann, R.D., Eisen J.A., White,O., Salzberg, S.L., Smith, H.O., Venter, J.C., Fraser, C.M. (1999). Evidence for Lateral Gene Transfer Between Archaea and Bacteria from Genome Sequence of *Thermotoga maritima. Nature* **399,** 323-329.

Pace, C. N. (1992). Contribution of Hydrophobic Effects on Globular Protein Stability. *J.Mol.Biol.* **226,** 29-35.

Page, M. I., Jencks, W.P. (1971). Entropic Contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and The Chelate Effect. *Proc. Natl. Acad. Sci.* **68,** 1678-1683.

Pan, X. S., Ambler, J., Mehtar,S., Fischer,M.L. (1996). Involvement of Topoisomerase and DNA Gyrase as Ciprofloxacin Targets in *Streptococcus pneumoniae. Antimicrob.Agents.Chemother.* **40,** 2321-2326.

Perrakis, A., Morris, R.M., Lamzin, V.S. (1999) Automated Protein Model Building Combined With Iterative Structure Refinement. *Nature Struct.Biol.***6**, 458-463.

Perutz, M. F. (1978). Electrostatic Effects in Proteins. *Science* **201,** 1187-1191.

Poget, S. F., Legge, G.B., Proctor, M.R., Jonathon, P., Butler, G., Bycroft, M., and Williams, R.L. (1999). The Structure of a Tunicate C-type Lectin from *Polyandrocarpa misakiensis* Complexed with D-Galactose. *J. Mol. Biol.* **290,** 867-879.

Powell, H. R. (1999). The Rossmann Fourier Autoindexing Algorithm in MOSFLM. *Acta Cryst* **D55,** 1690-1695.

Prag, G., Papanikolau, Y., Tavias, G., Vorgias C.E., Petratos, K., Oppenheim, A.B. (2000). Structures of Chitobiase Mutants Complexed with the Substrate Di-N-acetyl-D-glucosamine: the Catalytic Role of the Conserved Acidic Pair, Aspartate 539 and Glutameate 540. *J.Mol.Biol* **300,** 611-617.

Puri, K. D., Springer, T.A.,. (1996). A Schiff Base with Mildly Oxidised Carbohydrate Ligands Stabilises L-selectin and not P-selectin or E-selectin Rolling Adhesions in Shear Flow. *J.Biol.Chem.* **271,** 5404-5413.

Qasba, P. K. (2000). Involvement of Sugars in Protein-Protein Interactions. *Carbohydrate Polymers* **41,** 293-309.

Quiagen. (2002). Purification of Plasmid DNA . *QIA-prep Miniprep Handbook,* 22-23.

Quiocho, F. A. (1993). Probing the Atomic Interactions between Proteins and Carbohydrates. *Biochem. Soc. Trans.* **21,** 442-448.

Ramachrisnan, C., Ramachandran, G. (1965). Stereochemical Criteria for Polypeptide and Protein Chair Conformation. *Biophysical journal* **5,** 909-933.

Ravishankar, R., Surolia, A., Vijayan, M., Lim, S., Kishi, Y. (1998). Preferred conformation of C-lactose at the free and peanut lectin bound states. *J.Am.Chem.Soc.* **120,** 11297-11303.

Rini, J. M., Hardman, K.D., Einspahr, H., Suddath, F.L. Carver, J.P. (1993). X-Ray Crystal-Structure of a Pea Lectin Trimmanoside Complex at 2.6 Å Resolution. *J.Biol.Chem.* **268,** 10126-10132.

Russel, A. D., Chopra, I., ed. (1996). *Understanding Antibacterial Action of Resistance.* London: Ellis Horwood.

Sanders, D. A. R., Staines, A.G., McMahon, S.A., McNeil, M.R., Whitfield, C., Naismith, J.H. (2001). UDP-Galactopyranose Mutase has A Novel Structure and Mechanism. *Nat.Struct.Biol.* **8,** 858-863.

Searcy, D. G., & Whatley, F. R. (1984). *Thermoplasma acidophilum*: Glucose Degradative Pathways and Respiratory Activities. *Syst. Appl. Microbiol.* **5.**30-40.

Severon, W. B., Kelly, R.F., Richards, J.C., Whitfield, C. (1996). Structure of the Core Oligosaccharide in the Serotype O8 Lipopolysaccharide from *Klebsiella Pneumoniae. J.Bacteriology* **6,** 1731-1741.

Shaanan, B., Lis, H., Sharon, N. (1991). Structure of a Legume Lectin with an ordered N-linked Carbohydrate in Complex with Lactose. *Science* **254,** 862-866.

Sharon, N. (1993). Lectin-Carbohydrate Complexes of Plants and Animals: an Atomic View. *Trends in Biochemical Sciences* **18,** 221-226.

Sharon, N., Lis, H. (1972). Lectins: Cell-Agglutinating and Sugar Specific Proteins. *Science* **177,** 949.

Shibata, K., Shingu, K., Vassilev, V. P., Nishide, K., Fujita, T., Node, M., Kajimoto, T., Wong, C.-H. (1996). Kinetic and Thermodynamic Control of L-threonine Aldolase Catalysed Reaction and its Application to the Synthesis of Mycestericin D. *Tetrahedron Letters* **37,** 2791-2794.

Stehle, T., Harrison, S.C., (1997). High-Resolution Structure of a Polyomavirus VP1-Oligosaccharide Complex: Implications for Assembly and Recptor Binding. *EMBO* **16,** 5139-5148.

Stevenson, G., Neal, B., Lui, D., Hobbs, M., Packer, N.H., Bately, M., Redmond, J.W., Lindquist, L., Reeves, D.R. (1994). Structure of the O Antigen Of *Escherchia coli* K-12 and the Sequence of its Rfb Gene Cluster. *J. Bacteriology* **174,** 4144-4156.

Stillmark, H. (1888). Uber Ricin, ein giftiges Ferment aus den Samen von Ricinus communis L. und einigen anderen Euphorbiaceen. Inaug. Diss Dorpat.

Stura, E. A., Wilson, I.A. (1991). Application of the Streak Seeding Technique in Protein Crystallisation. *J.Crystal.Growth.* **110,** 270-282.

Sumner, J. B. (1919). The Globulins of the Jack Bean *Canavalia ensiformis. J.Biol.Chem.* **37,** 137-142.

Sun, Y. J., Chang, N.C., Hung, S.I., Chang, A.C., Chou, C.C., Hsiao, C.D. (2001). The Crystal Structure Of a Novel Mammalian Lectin, Ym1, Suggests a Saccharide Binding Site. *J.Biol.Chem.* **276,** 17507-17514.

Sygusch, J., Beaudry, D., Allaire, M. (1987). Molecular Architecture of Rabbit Skeletal Muscle Aldolase at 2.7-Å Resolution. *Proc Natl Acad Sci U S A* **84,** 7846-7850.

Szilagyi, A., Zavodszky, P. (2000). Structural Differences between Mesophilic, Moderately Therrmophilic and Extremely Thermophilic Subunits: Results of a Comprehensive Survey. *Structure* **8,** 493-504.

Takayama, S., Martin, R., Wu, J., Laslo, K., Siuzdak, G., Wong, C-H. (1997). Chemo-enzymatic Preparation of Novel Cyclic Imine Sugars and Rapid Biological Activity Evaluation Using Electrospray Mass Spectrometry and Kinetic analysis. *J.Am.Chem.soc.* **119,** 8146-8151.

Tanner, J. J., Hecht, R.M., Krause, K.L. (1996). Determinants of Enzyme Thermostability Observed in the Moleciular Structure of *Thermus aquaticus* D-glyceraldehyde-3-phophate Dehydrogenase at 2.5 Å Resolution. *Biochemistry* **35,** 2597-2609.

Teplyakov, A. V., Kuranova, I.P., Harutyunyan, E.H., Vainshtein, B.K., Frommel, C., Hohne, W.E., Wilson, K.S. (1990). Crystal Structure of Thermitase at 1.4 Å Resolution. *J. Mol. Biol.* **214,** 261-279.

Togo, H., He, W., Waki, Y., Yokoyama, M. (1998). C-Glycosidation Technology with Free Radical Reactions. *Synlett* **7,** 700-717.

Uitdehaag, J. C. M., Mosi, R., Kalk, K.H., van der Veen, B.A., Dijkhuizen, L., Withers, S.G., & Dijkstra, B.W. (1999). X-ray Structures along the Reaction Pathway of Cyclodextrin Glycosyltransferase Elucidate Catalysis in the α-Amylase Family. *Nature. Struct. Biol.* **6,** 432-436.

Vagin, A., Teplyakov, A. (1997). MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Cryst.* **30,** 1022-1025.

Verlinde, C. L. M. J., Quigley, P. M. (1999). Structure-based Re-evaluation of the Mechanism of Class I Fructose-1,6-bisphosphate Aldolase. *J. Mol. Mod.* **5,** 37-45.

Vielle, C., Ziekus, G.J.,. (2001). Hyperthermophilic Enzymes: Sources, Uses and Molecular Mechanisms for Thermostability. *Microbiol, Mol. Biol. Rev.* **65,** 1-43.

Vilchèze, C., Morbidoni, H.R., Weisbrod,T.R., Iwamoto, H., Kuo, M., Sacchettini, J.C., and Jacobs, W.R. Jr. (2000). Inactivation of the InhA-Encoded Fatty Acid Synthase II (FASII) Enoyl-Acyl Carrier Protein Reductase Induces Accumulation of the FASI End Products and Cell Lysis of *Mycobacterium Smegmatis*. *Bacteriology* **182,** 4059-4067.

Vocadlo, D. J., Davies, G.J., Laine, R. and Withers, S.G. (2001). Catalysis by Hen Egg White Lysozyme Proceeds by a Covalent Intermediate. *Nature* **412,** 835-838.

Wagner, R. L., Apriletti, J.W., McGrath, M.E., West, B.L., Baxter, J.D., Fletterick, R.J. (1995). A Structural Role for Hormone in the Thyroid Hormone Receptor. *Nature* **378,** 690-697.

Walden, H., Bell, G.S., Russell, R.J.M., Siebers, B., Hensel, R., Taylor, G.L. (2001). Tiny TIM: A Small Tetrameric Hyperthermostable Triosephosphate Isomerase. *J. Mol. Biol.* **306,** 745-757.

Wallis, R.,. Cheng, J.Y.T. (1999). Molecular Defects in Variant Forms of Mannose-Binding Protein Associated with Immunodeficiency. *J. Imunnology* **163,** 4953-4959.

Walsh, C. (2000). Molecular Mechanisms that Confer Antibacterial Drug Resistance. *Nature* **406,** 775-781.

Wang, I. N., Dykhuizen, D.E. (2001). Variation of Enzyme Activities at a Branched Pathway Involved in the Utilisation of Gluconate in *Escherchia coli. Evolution.* **55,** 897-908.

Weis, W., Brown, J.H., Cusack, S., Paulson, J.C., Skehel, J.J., Wiley, D.C . (1988). Structure of the Influenza Virus Haemagglutinin Complexed with its Receptor, Sialic acid. *Nature* **333,** 426-431.

Winder, F. G. (1982). Mode of Action of the Antimycobacterial Agents and Associated Aspects of the Molecular Biology of the Mycobacteria. Volume 1, pp. 353-438. Academic Press, London.

Wormald, M. R., Sharon, N. (2002). Carbohydrates and Glycoconjugates: Hot Roles for Glycosylation: Signal Transduction, Control of Cell Developement and Differentiation, and Innate Imunnity. *Current Opinion in Structural Biology.* **12,** 567-568.

Wyckoff, T. J. O., Raetz, C.R.H., and Jackman, J.E. (1998). Antibacterial and Antinflammatory Agents that Target Endotoxin. *Trends In Microbiology* **4,** 154-159.

Zgiby, S., Plater, A. R., Bates, M. A., Thomson, G. J., Berry, A. (2002). A Functional role for a Flexible Loop containing Glu182 in the Class II Fructose-1,6-bisphosphate Aldolase from *Escherichia coli. J. Mol. Biol.* **315,** 147-156.

Zhang, Q. B., Lee, H.W.,. (2001). Mechanistic Investigation Of UDP-Galactopyranose Mutase from *Escherichia coli* using 2- and 3-Fluorinated UDP-Galactofuranose as Probes. *J.Am.Chem.Soc.* **123,** 6756-6766.

Zhang, Q. B., Lui, H.W. (2000). Studies Of UDP-Galactopyranose Mutase From Escherichia coli: An Unusual Role of Reduced FAD in its Catalysis. *J.Am.Chem.Soc.* **122,** 9065-9070.

Zhong, G., Shabat, D., List, B., Anderson, J., Sinha, S.C., Lerner, R.A., Barbas III, C.F. (1998). Catalytic Enantioselective Retro-Aldol Reactions: Resolution of ß-Hydroxyketones with Aldolase antibodies. *Angew. Chem. Int. Ed.* **37,** 2481- 2484.

Zimmermann, F. T., Schneider, A., Schlorken, U., Sprenger, G.A., Fessner W.D. (1999). Efficient Mutli-Enzymatic Synthesis of D-Xylulose 5 Phosphate. *Tetrahedron Asymmetry* **10,** 1643-1646.

Abbreviations

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

A - Adenine, C – Cytosine, G – Guanine, T – Thymine

m	metre
g	gram
s	second
l	litre
Da	Dalton units
°C	degrees Celsius
M	molar
V	volt
Å	Angstrom
K	degrees Kelvin

Abs	Absorbance
BSA	Bovine serum albumin
Con-A	Concanavalin A
dH_2O	distilled water
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethlenediaminetetracetate

EPR	electron paramagnetic resonance
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
HEPES	N-2-hydroxyethylpiperazine-N-2-ethanesulfonic acid
HPLC	High Performance Liquid Chromatography
Hq	hydroquinone
IPTG	isopropyl thiogalactoside
kb	kilobase
K_d	binding constant
KDPG	2-keto-3-deoxy-6-phosphogluconate.
K.pneumoniae	*Klebsiella pneumoniae*.
LB	luria broth
M_w	molecular weight
M.tuberculosis	*Mycobacterium tuberculosis.*
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NHE	normal hydrogen electrode
OD	Optical density
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
PMSF	phenylmethylsulfonyl fluoride
SDS	sodium dodecyl sulphate
Sq	semiquinone
TB	Terrific broth
T.maritimas	*Thermatoga maritimas*
Tris	Tris(hydroxymethyl) aminomethane
UDP-gal_p	UDP-Galactopyranose
UDP-Gal_f	UDP- Galactofuranose
UV	ultra violet
Vis	visible