Thermostable Enzymes Important

For Industrial Biotechnology

Submitted by Aaron Charles Westlake to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Biological Science in September 2018

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Aaron Charles Westlake

[1]

Abstract

The use of enzymes in technology is of increasing commercial interest due to their high catalytic efficiency and specificity and the lowering of manufacturing costs. Enzymes are also becoming more widely utilised because they are more environmentally friendly compared to chemical methods. Firstly, they carry out their reactions at ambient temperatures requiring less energy to achieve the high temperatures and pressures that many chemical methods require. Secondly, they can substitute for toxic chemical catalysts which need careful disposal. In this project two classes of enzymes of industrial interest from thermophiles were investigated, lactonase enzymes and 1-deoxy-D-xylulose 5-phosphate (DXP) synthases.

A quorum sensing lactonase from *Vulcanisaeta moutnovskia*, a thermoacidophilic anaerobic crenarchaeon, was expressed in high levels in an *Escherichia coli* host, then purified and characterised with a range of industrially relevant substrates. These enzymes are of industrial interest for water treatment and bioreactors for their ability to prevent biofilm formation in bacteria. This enzyme showed different specificity to another well characterised quorum sensing lactonase from a thermophilic crenarchaeon, *Sulfolobus solfataricus*. Crystals of the native enzyme were obtained. Structural examination revealed that the *V. moutnovskia* lactonase possesses an α-helix obstructing a hydrophobic channel near the active site, whereas the *S. solfataricus* lactonase has a flexible loop leaving the hydrophobic channel unrestricted. As a result the acyl chains of substrates interact with surface residues of the α-helix in *V. moutnovskia* lactonase rather than sitting in the channel, so its activity is no longer restricted to substrates with long acyl chains.

A gluconolactonase encoded by a thermophilic *Planctomyces* genome was cloned and expressed at high levels in an *E. coli* host, purified and successfully crystallised. The crystals had a space group of P 3 2 1 and diffracted to a resolution of 2.41 Å. This enzyme was intended to be used by an industrial

partner for synthesis of metabolite standards for mass spectrometry and diagnostics. Attempts were made to clone, purify and express two other lactonases from thermophilic metagenomes obtained from terrestrial hot springs: an enol lactonase and a second quorum sensing lactonase. Homology modelling was used to create predicted structures for both of these enzymes. The quorum sensing lactonase showed a 2.5 Å difference in the position of a catalytic serine and a 3.1 Å difference in a catalytic histidine in comparison to a mesophilic homologue. The enol lactonase contained an aspartic acid in place of a catalytic serine found in a mesophilic homologue.

A DXP synthase from an anaerobic Gram-negative bacterium *Thermovibrio ammonificans* was successfully cloned, over-expressed and purified. Crystals were successfully produced although these diffracted only to low resolution. A DXP synthase from an anaerobic Gram-positive bacterium *Carboxydothermus hydroformans* was successfully cloned, however the protein was expressed primarily in the insoluble fraction. Homology models were made for these two enzymes. Both enzymes showed strong similarity with mesophilic DXP synthases in terms of tertiary structure and positions of active site residues. Visual analysis revealed an increase of 15-20 % in the number of hydrophobic interactions within the enzymes and a high proportion of charged residues at the dimer interface, which would confer increased thermostability. The hope was to obtain high resolution diffraction data to assist in understanding what allows these enzymes to utilise pyruvate as a substrate compared to transketolase, a related enzyme, which uses hydroxypyruvate.

Table of contents

3.   Quorum sensing lactonase

4.   Gluconolactonase

List of figures

List of tables

Abbreviations

Bis-Tris	Bis-(2-hydroxy-ethyl)-amino-tris(hydroxymethyl)-methane
ddH$_2$O	Double distilled water
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DXP	1-Deoxy-D-xylulose-5-phosphate
dNTP	Deoxyribonucleotide triphosphate
EDTA	Ethylene diamine tetra acetic acid
GF	Gel filtration
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulphonic acid
IDT	Integrated DNA Technologies
I-TASSER	Iterative Threading ASSEmbly Refinement
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	Kilodalton
MOPS	(3-(N-morpholino)propanesulphonic acid)
NiAC	Nickel affinity chromatography
nt	Nucleotide
PCR	Polymerase chain reaction
PDB	Protein data bank
Phyre2	Protein Homology/analogY Recognition Engine V 2.0
QMEAN	Qualitative Model Energy ANalysis
RCF	Relative centrifugal force
RMSD	Root-mean-square deviation
rpm	Revolutions per minute
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis

TAE buffer	Tris, acetic acid and EDTA buffer
TDP	Thiamine diphosphate
Tris	Trisaminomethane
YASARA	Yet Another Scientific Artificial Reality Application

<u>Acknowledgments</u>

1. Introduction

1.1 <u>Applications of biocatalysis</u>

Enzymes are currently being used as biocatalysts in many industrial processes since they are highly selective for their substrates and products, they are usually stereoselective, and the enzymes are potentially reusable (Schmid *et al.*, 2001). This is unlike traditional chemical synthetic methods, which can involve high temperatures, high pressures, toxic transition metals or a combination of all of these. This makes the process costly, requiring a lot of energy to create the high temperatures and pressures, as well as environmentally unfriendly with regards to the disposal of the toxic heavy metal catalysts. In comparison, using enzymes as catalysts in pharmaceutical and industrial processes requires much lower pressure, temperature and they can usually substitute for the reactions carried out using heavy metal catalysts.

A market research report published in 2016 from Grand View Research, Inc. (https://www.grandviewresearch.com/industry-analysis/enzymes-industry) showed that the global market for enzymes in 2015 was 1.18 billion USD. By 2024 it has been predicted that this will increase to an estimated 17.5 billion USD. The areas where the use of enzymes is expected to grow the fastest over the coming years are animal feed and speciality enzymes designed for diagnostics or industrial biocatalysis.

Enzymes are divided into six different classes, hydrolases, transferases, ligases, isomerases, oxidoreductases and lyases (Table 1). This was determined by the Enzyme Commission, part of the International Union of Biochemistry and Molecular Biology (https://iubmb.org).

Table 1. Enzyme classifications.

Enzyme type	Reactions catalysed
Hydrolase	Hydrolytic cleavage
Transferase	Transfer of a functional group between two molecules
Ligase	Formation of a new bond joining two large molecules together
Isomerase	Intramolecular rearrangement
Oxidoreductase	Oxidation and reduction
Lyase	The removal or addition of a molecule to a double bond

The first instance of biocatalysis being used in an industrial process was in the 1930's where *Acetobacter suboxydans* was used was used to convert D-sorbitol and into L-sorbose for the production of L-ascorbic acid (vitamin C). Also, yeast was used for the conversion of benzaldehyde into (R)-phenylacetyl carbinol for the production of D-ephedrine for treating low blood pressure (Holland 2000).

Examples of biocatalytic processes include the breakdown of amides by penicillin acylase, hydrolysis of esters to acids by esterases and lipases, the formation of malic acid via condensation by fumarase, conversion of glucose to fructose by glucose isomerase and the removal of a halogen by dehalogenases (Taylor, 1998). Many semi-synthetic antibiotics are now being made by utilising biocatalysis instead of chemical conversions based on stoichiometry. The α-amino acid ester hydrolase, α-acylamino-β-lactam acylhydrolase and ampicillin acylase are all enzymes that have been reported capable of producing semi-synthetic cephalosporins from 7-aminocephem or 6-aminopenam with α-amino acid esters (Ryu and Ryu, 1986).The immobilised enzyme penicillin acylase is being used in the manufacturing process of β-lactam antibiotics. Condensation of 6-aminopenicillanic acid with D(−)-4-hydroxyphenylglycine or D(−)-phenylglycine by this enzyme will make amoxicillin or ampicillin, respectively. Also cefadroxil and cephalexin can be made from 7-aminodesacetoxycephalosporanic acid in the same manner (Bruggink *et al.*, 1998).

Enantiomers can have radically different biological effects in biological systems. An example is thalidomide where the *(R)-*enantiomer has a sedative effect, previously used to treat morning sickness, while the *(S)-*enantiomer is extremely teratogenic. When first marketed, the drug was administered as a racemate of both enantiomers and caused severe birth defects. Another example is terodiline, which was originally marketed as a racemate for the treatment of urinary incontinence. It was later withdrawn due to one of the side effects being severe heart arrhythmia, which was discovered to be caused by the *(R)-*enantiomer. The racemate version of Benoxaprofen was also removed from the market as an anti-inflammatory drug due to rapid accumulation of the active *(S)-*enantiomer. This was due to the unforeseen effect of the *(R)-*enantiomer undergoing chiral-inversion within the body to the more potent *(S)-*enantiomer making the dosage of the active compound much higher than anticipated (Srinivas *et al.*, 2001). Of course not all chiral compounds have a negative effect, some have no observable biological effect. Methylphenidate is one such example, used in the treatment of children with ADHD and ADD, where the D-enantiomer is the medicinal compound and the L-enantiomer has no biological effect (Srinivas *et al.*, 2001). But for instances like those described above where the biological effects of one of the enantiomers was highly toxic, being able to select one enantiomer over the other is vital in drug development. As such there is far more scrutiny on racemates by regulatory agencies with new drug applications. When one isomer is identified as being a candidate for drug development, full characterisation of its optical isomer is required for scrutiny adding more time and cost to drug design.

Due to chirality having such significance in biological systems many enzymes have evolved to favour one enantiomer over another. This property has been of great interest to the pharmaceutical industry in producing chiral compounds. The asymmetric synthesis of chiral amines is important for a large number of pharmaceutical processes since they are important components of many drugs, such as delavirdine, an HIV reverse transcriptase inhibitor (Romero et al., 1993), however they remain a challenge to produce. To overcome this amine dehydrogenases have recently begun to be utilised. A reaction using three amine dehydrogenases: two L-phenylalanine dehydrogenases from a *Rhodococcus* species and *Bacillus badius;* and a leucine dehydrogenase from *Bacillus stearothermophilus*; along with a formate dehydrogenase to recycle the

nicotinamide cofactor, was developed that performed the amination of aromatic and aliphatic ketones and aldehydes with high efficiency and a high turnover of product. Furthermore, the enzymes produce amines with high stereoselectivity with more than 99 % of the product in the *(R)* conformation (Knaus *et al.*, 2017). The final step in the production of sitagliptin, an anti-diabetes drug, was an asymmetric rhodium-catalysed hydrogenation. However an alternative was developed using a transaminase enzyme which won the Merck Green Chemistry Award in 2010 and has since replaced the previous method. The transaminase itself does not naturally have activity to the intermediate dehydrositagliptin so was engineered using mutagenesis directed at the active site to accommodate the non-natural substrate, followed by rounds of directed evolution to improve yield, activity, enantioselectivity, and stability (Savile *et al.*, 2010).

Transaminases mediate the transfer of an amino group from the amino donor to the acceptor for chiral synthesis of amino acids or other amines, with the aid of a pyridoxal 5′-phosphate cofactor. These enzymes are currently of great interest for a number of factors (Guo and Berglund, 2017). These enzymes display a high level of enantioselectivity due to positioning of the cofactor and binding pockets, resulting in high amounts of chiral products (Fuchs, *et al.*, 2015). Also the cofactor is recycled after the reaction which is highly advantageous in industrial biocatalysis because additional cofactor regeneration is not required, unlike many other enzymes (Kara *et al.*, 2014). A transaminase from the crenarchaeon *Sulfolobus solfataricus* was found to catalyse the reaction of L-serine and pyruvate to 3-hydroxypyruvate and alanine. It also had activity towards other amino acids; methionine, asparagine, glutamine, phenylalanine, histidine, and tryptophan (Sayer *et al.*, 2012). This made it ideal to be used alongside a transketolase enzyme (Littlechild *et al.*, 1995) for the synthesis of a number of optically pure intermediates for drug production (Chen *et al.*, 2006).

Whole cells can be used in complex reaction pathways that require regeneration of co-factors for the enzymes as it is far easier and cheaper to allow a living cell to replenish these automatically as part of its natural processes to ensure the enzyme remains active (Schmid *et al.*, 2001). Due to their high specificity, there are few by-products of enzyme-based reactions; with increasing environmental awareness this becomes even more of a necessity as any waste must be dealt

with properly, so the less there is the cheaper the end product. The production of dorzolamide hydrochloride, a carbonic anhydrase inhibitor developed for the treatment of glaucoma, utilises whole microorganisms to produce an optically pure enantiomer. The fungus *Rhodotorula pilimanae* is used in the asymmetric production of the chiral hydroxy sulphone from ketosulphone (Blacker and Holt , 1997). A problem with whole cell reactions is that the desired product may be further metabolised by other enzymes present in the cell. Cells may also have to be lysed unless the product is excreted and there would be significant contamination from other products produced by the cells as well as nutrients in the growth medium that would need to be removed.

1.2 <u>Thermophiles</u>

An active protein is held together by a combination of forces: van der Waals, electrostatic, hydrophobic interactions and hydrogen bonds. When these forces are disrupted by things such as elevated temperatures, the proteins unfold into inactive structures and form aggregates where the hydrophobic residues of different unfolded proteins interact with each other. Once this occurs it is not usually possible to get back the original protein structure (Littlechild *et al.*, 2013). However, many organisms are found in environments where mesophilic proteins would immediately unfold due to the high temperatures. Organisms adapted to these environments, known as thermophiles and hyperthermophiles, possess proteins that are modified to resist the high temperatures.

Thermophilic enzymes are of great interest for industrial applications due to their resistance to high temperatures. This is because running reactions at higher temperatures give several advantages over those done at temperatures that mesophilic enzymes are stable. At higher temperatures there is an increase in the solubility of organic compounds allowing higher concentrations to be achieved and fewer rounds of product extraction to lower manufacturing costs. The higher temperatures also cause a drop in the viscosity of the solvent and an increase in the rate of diffusion of organic compounds. Finally, higher temperatures are expected to produce higher rates of reactions due to smaller boundary layers, the layers of stationary fluid surrounding an immersed object, which in this case would be the enzymes themselves (Krahe, *et al.*, 1996; Becker

et al., 1997). An example of a thermophilic enzyme currently being used in industrial applications is in the production of corn syrup. A xylose isomerase is used in the conversion of glucose to fructose and xylose to xylulose (Chanitnun and Pinphanichakarn, 2012).

Thermophilic enzymes have evolved a number of ways in order to prevent denaturation on exposure to the high temperature environments that the host organisms have to survive in. While there is a pronounced difference between mesophilic and thermophilic enzymes, where thermophilic enzymes possess fewer uncharged polar amino acid residues and more hydrophobic and proline residues. There is much less difference between thermophiles from different temperature ranges (Zhou *et al.*, 2008). It is often seen that glycine and lysine residues are replaced with alanine and arginine respectively. The reason for this is that a higher alanine content would help stabilise α-helices as its features best suit α-helices in that it possesses a small methyl group that will not cause steric hindrance while also providing hydrophobic interactions to stabilise the structure (Quellère, 2007). Arginine is hypothesised to be less susceptible to chemical attack at high temperatures compared to lysine (Vieille and Zeikus, 2001). Besides an increase in the number of ionic and hydrophobic interactions found in the proteins there is also often an absence of cysteine as it is particularly susceptible to oxidation at high temperatures. Thus disulphide bridges are not that common in mesophilic proteins. The exception to this are extracellular proteins where they are not exposed to the reducing environment of the cell and can contribute to stabilising the protein. This is not true for hyperthermophilic archaeal proteins such as those from *Aeropyrum pernix* where disulphide bonds are found in many cytoplasmic proteins (Littlechild *et al.*, 2013). There are also fewer asparagine and glutamine residues as they are easily deaminated (Hess *et al.*, 1995). There is an increase of aromatic and proline residues as these increase the rigidity of the protein (Fleming and Littlechild, 1997). Often there are also more hydrogen bonds present within the protein. Changes to the secondary and tertiary structure is also observed with fewer and shorter external loop regions being present or alternately they may be anchored though ion pairs, hydrogen bonds or hydrophobic interactions to the main structure (Vieille and Zeikus, 2001). Metal ions may also be bound to loop regions to stabilise these further (Mallick *et al.*, 2002). Similarly the N and C termini may be anchored to

the core region to increase stability (Macedo-Ribeiro *et al.*, 1996). There have also been changes to the quaternary structure that have been observed. Thermophilic proteins are more likely to form higher oligomeric states than their mesophilic counterparts which is thought to increase thermal stability. Studies on monomeric mutants of dimeric enzymes show significant decreases in the temperature at which the protein denatures (Thoma *et al.*, 2000). Ultimately, there is no single rule for how stability of enzymes in hot environments is achieved, rather the adaptations differ between species.

Alcohol dehydrogenase from the hyperthermophilic archaeon *Aeropyrum pernix* is an example of an enzyme designed to withstand high temperatures. The organism is heterotrophic and grows at temperatures of 90–95 °C and at a pH of 7.0. The enzyme is a tetramer and appears to have adaptations to maintain this despite the high temperatures. There are four ionic bonds between subunits for a total of sixteen inter-subunit ionic bonds. There are also extensive hydrophobic pockets located at the subunit interfaces. These features, along with a disulphide bond within each monomer, infer the ability of the enzyme to survive for one hour at 90 °C with only a loss of 50 % activity (Guy *et al.*, 2003). In two pyrrolidone carboxyl peptidases from the thermophiles *Pyrococcus furiosus* and *Thermococcus litoralis*, there is an insert of five amino acids in both that is not present in the mesophilic variant. These create a C-terminal loop that forms the core of a tetramer which is attributed as the main form of stabilising this protein as the mesophile enzyme forms a dimer (Littlechild *et al.*, 2007).

There are many cases where thermophilic enzymes are being utilised for their stability or other unique traits. The polymerase from the hyperthermophilic archaeal *Pyrococcus* species is currently being widely used in PCR reactions as it has demonstrated higher fidelity in comparison to its mesophilic counterparts. A γ-lactamase from the thermophile *Sulfolobus solfataricus* was purified and immobilised as a cross-linked polymerised enzyme in a microreactor (Hickey *et al.*, 2009). This was to catalyse the reaction of a γ-lactam to a single enantiomer of a γ-bicyclic lactam product which is important for the synthesis of the anti-HIV compound, Abacavir (Taylor *et al.*, 1993). A thermophilic lipase was also found in *S. solfataricus*, its optimal activity was towards esters of mid-chain length in temperatures around 70 °C. Lipases such as this one have potential use in the

production of drug intermediates (Littlechild *et al.*, 2007). A thermophilic dehalogenase enzyme was located in the crenarchaeon *Sulfolobus tokadaii* and was cloned and purified due to it being able to produce chiral halo-carboxylic acids which are needed for both fine chemical and pharmaceutical industries (Rye *et al.*, 2009).

Thermostable enzymes are also more stable to organic solvents and more resistant to proteolysis. This is advantageous as some industrial processes must been done in the presence of solvents since the non-natural substrates are not soluble in water. Reactions carried out in organic solvents also can have other advantages such as less effort required in recovering the product, change in the equilibrium and less risk of contamination in bioreactors (Valadez-Blanco *et al.*, 2008). In non-polar solvents, lipase and esterase enzymes generally retain the same catalytic activity as they do when in water, indicating water remains bound to the enzyme modulating activity. It was shown that water forms a shell around the enzyme when put into solvents such as hexane. With polar solvents such as alcohols, however, there is an observable difference in activity where the solvent partially replaces the role of bound water required for enzyme activity. The remaining water forms small clusters loosely bound to the surface of the enzyme (Micaêlo and Soares, 2007). Other classes of enzyme are less stable to high percentages of organic solvents, which can denature the enzyme and reduce activity.

To determine whether proteins retain their native conformation in organic solvents a study was done using solid state NMR. The structure of a serine protease cutinase from *Fusarium solani* was observed in hexane, 3-pentanone, diisopropyl ether, ethanol, and acetonitrile. Here it was shown that the overall structure of the enzyme differed depending on the hydration levels. The root mean squared deviation (RMSD) of the protein when exposed to a solvent formed a bell shaped curve, revealing an optimal level of hydration required for it to assume its native conformation (Micaêlo and Soares, 2007).

Work to impart thermal stability to mesophilic proteins has been going on for quite a while. A number of methods are employed including rational design using techniques such as site-directed mutagenesis and directed evolution (Song and

Rhee, 2000). A lactate oxidase, used in biosensors to measure the concentration of lactate in the blood, was put through directed evolution using error-prone PCR to increase its stability at high temperatures. The group managed using this method to increase the half-life of the enzyme at 70 °C to about 36 times of the wild-type enzyme (Minagawa *et al.*, 2007).

1.3 <u>Psychrophiles</u>

The earth has many regions that are colder than most life can survive in such as the Antarctic. Psychrophiles thrive in cold environments and even at temperatures below freezing and this group includes any organism adapted to temperatures below 7 °C. Examples include *Moritella profunda*, a psychropiezophilic bacteria adapted to 2 °C temperatures and the deep sea. *Ascomycotina*, *Basidiomycotina*, *Deuteromycotina*, and *Zygomycotina* are all families of fungi that thrive at low temperatures and are responsible for some food spoilage while in refrigeration (Maurer, 2003). Microbial communities have even been found in porous rocks in Antarctic dry valleys which are adapted to surviving temperatures as low as -60 °C (Cary, *et al*, 1982).

Psychrophiles produce antifreeze proteins which are a group of ice-binding proteins that prevent ice growing by binding of the protein to the surface of the ice crystal. This alters the geometry of the ice crystals forcing them to form in a convex pattern, decreasing the freezing point (Kristiansen and Zachariassen, 2005). Cold shock proteins are expressed when cells are exposed to temperatures below those they normally grow at. These bind to RNA and act as RNA-chaperones to regulate translation by destabilising the RNA to make it single stranded. These proteins are constantly being produced in psychrophiles as opposed to transiently in mesophiles and thermophiles (Lee *et al.*, 2013).

Psychrophilic proteins have been adapted to function at lower temperatures where water viscosity, substrate diffusion rates and enzyme kinetics all differ from those of mesophiles and thermophiles. Overall their structures are more flexible with longer loop regions. They have a reduced of proline residues due to folding and isomerisation of proline being a rate limiting step for psychrophilic proteins and increasing the flexibility of the protein backbone, as well as having prolyl

[23]

isomerases to aid in the process. The enzymes are less stable than their mesophilic counterparts and localised increases in flexibility at the active site are responsible for the high activity at low temperatures (Feller, 2013). They have more nonpolar residues on their surface which destabilises the water shell around the enzyme. They have a lower abundance of arginine residues, which increases backbone flexibility as well as fewer hydrogen bonds to reduce the stability within the enzyme core (Paredes *et al.*, 2011). Their active sites are larger and more accessible with smaller amino acid side chains being favoured in them (Aghajari *et al.*, 2003). As a consequence of this psychrophilic enzymes are often more promiscuous with substrate specificity. The k_{cat} of these enzymes is higher than their mesophilic counterparts over temperatures between 0-30 °C (Feller *et al.*, 1996). Substrates bind less firmly to the active site in cold conditions giving rise to higher K_m values. The majority of psychrophilic enzymes improve the k_{cat} at the expense of K_m, leading to suboptimal values of the k_{cat}/K_m ratio. The theory behind this is that a weaker substrate binding affinity means the enzyme-substrate complex is at a higher energy state. This means less energy is required to overcome ΔG to complete the conversion to the products (Struvay and Feller, 2012). Whereas thermophilic proteins maintain a low entropic state to resist high temperatures, psychrophilic proteins employ the opposite strategy to have a much higher entropic state to ensure there is sufficient flexibility within the protein for catalytic activity (Feller, 2013).

1.3 Lactonases

Lactonases are a class of enzyme that catalyse the hydrolysis of the ester bond within a lactone ring. Macrocyclic lactones are currently in use as broad spectrum drugs for treatment of parasitic infections. Testolactone, an inhibitor of aromatase, a key enzyme for oestrogen biosynthesis, is used for the treatment of advanced stage breast cancer (Cepa *et al.*, 2005). Lactones can also be used for the synthesis of polymers (Bains *et al.*, 2011). Studies to date have identified three structurally diverse classes of lactonase enzymes. These are the enol lactonases, gluconolactonases and the quorum sensing lactonases.

The primary role of enol lactonases is as part of a pathway of enzymes which make phenol a viable source of energy (Fig. 1.1). Phenol is degraded into 3-

[24]

oxoadipate-enol-lactone which is catalysed to 3-oxoadipate by 3-oxoadipate-enol-lactonase. Following this it is converted into 3-oxoadepyl CoA by a transferase, then into acyl CoA and succinyl CoA by an acyltransferase and a thiolase respectively (Basha *et al.*, 2010). Both of these go on to feed directly into the citric acid cycle to generate energy for the cell. Gluconolactonases are involved in pathways in animals for the synthesis of L-ascorbic acid, an important antioxidant, from D-glucuronate, and for synthesis of pentose phosphates which are used in synthesising nucleic acids (Fig. 1.2). Glucose is converted in these pathways eventually to an intermediate, L-gulonate by glucoronate reductase, and is then converted to L-gulonolactone by gulonolactonase catalysing the reverse reaction of what is normally seen by condensing it to a lactone ring in response to the equilibrium where the lactone is favoured over the L-gulonate (Fig 1.2).

Quorum sensing lactonases have been shown to disrupt quorum sensing involving the secretion of N-acyl homoserine lactones, primarily in Gram-negative bacteria (Fig 1.3). These lactonases are used either by species that emit lactones as quorum sensing molecules to recycle them, or by competing organisms. The genes under the control of these molecules usually are related to virulence and biofilm formation. The density of quorum sensing molecules are a way for cells to measure the local population, with more cells creating higher concentrations of signalling molecules in the surrounding environment. They are also a way for bacteria to sense the dimensions and how open its surroundings are through a method known as diffusion sensing, where diffusion of signal molecules away from the cell indicate an environment that is not restrictive in size. This allows them to probe the worthwhileness of producing costly molecules such as extracellular enzymes (Hense *et al.*, 2007). It has been suggested that disrupting quorum sensing would be less likely to give rise to new resistant strains, due to the fact that there is less selective pressure, which makes this a topic of great interest (Singh, 2009). A lactonase from *S. solfataricus* was immobilised on a nanoalumina membrane, and although it did not inhibit the growth of the *Pseudomonas aeruginosa* being grown there, it did inhibit expression of exoproteases and elastases, proteins associated with infection and resistance to the immune system (Kuang *et al.*, 2011), and reduced the concentration of bacterial cells by approximately five times (Ng *et al.*, 2011). A quorum sensing

lactonase from *S. solfataricus* was purified and found to inhibit the virulence of isolates of *P. aeruginosa*, a Gram negative pathogenic bacterium. This bacterium is responsible for a number of human infections including otitis, keratitis, pneumonia, and diabetic foot ulcers. It reduced the secretion of both proteases and pyocyanin, known virulence factors, as well as inhibiting biofilm formation. When compared to molecules known to inhibit quorum sensing such as brominated furanones which prevent bacteria from sensing any quorum signalling molecules in solution, the lactonase was found to be far more effective at disrupting the quorum sensing of the bacteria (Guendouze *et al.*, 2017). While virulence genes are not considered necessary for bacterial survival, meaning there is less selective pressure (Rasko and Sperandio, 2010), a recent paper suggested that there is a possibility in the rise of resistance to quorum quenching (Defoirdt *et al.*, 2010). They showed that within the bacterial population there were variances in the expression of quorum sensing genes, both those for producing the molecules and those for perceiving them. They argued that there is a risk of resistance if this variation induces differences in fitness under quorum quenching conditions.

It has been postulated that quorum sensing lactonase might be produced by some bacteria as a method of metabolising the lactone molecules secreted by other species of bacteria to use as a source of both nitrogen and energy. *Variovorax paradoxus*, a rod-shaped bacterium isolated from soil samples was found to be capable of growth on plates containing acyl-homoserine lactones as the sole source of energy and nitrogen. A hypothetical pathway for this indicates the lactonase to be responsible for cleaving the lactone ring and releasing nitrogen for use by the bacterium (Fig. 1.4) (Leadbetter and Greenberg, 2000).

The mechanism by which lactonases hydrolyse the ester ring was proposed by (Momb *et al.*, 2008) and is shown in Fig. 1.5. They investigated acyl-homoserine lactonase from *Bacillus thuringiensis* using X-ray crystallography and site directed mutagenesis. Two divalent zinc ions present in the active site bind to the lactone ring on the side less hindered by its side chain, with the carbonyl oxygen interacting with the first zinc ion and the leaving oxygen with the second zinc ion. This interaction between the zinc ions and the lactone ring is stabilised by the interaction of the phenol group of Tyr 194 with the carbonyl oxygen of the lactone.

Phe 107 forms a hydrogen bond with the carbonyl oxygen of the lactone's amide group facilitated by a water molecule between them. The zinc ions polarise the lactone bond giving the carbonyl carbon of the ring an increased positive charge and making it susceptible to a negatively charged hydroxide ion. The addition of this hydroxide group forms a tetrahedral intermediate and causes cleavage of the C–O bond. It was also proposed that the negative charge on the leaving group is stabilised by the zinc ions.

3-oxoadipate-enol-lactone
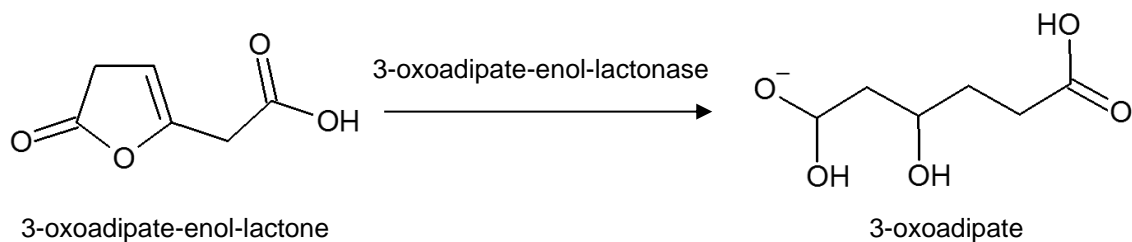
3-oxoadipate

Fig. 1.1. Reaction of 3-oxoadipate-enol-lactone to 3-oxoadipate catalysed by 3-oxoadipate-enol-lactonase. Image drawn using ChemSketch (Advanced Chemistry Development Inc., 2013)

Glucose-6-phosphate

Glucose-6-phosphate dehydrogenase

6-phosphogluconolactone

6-phosphogluconolactonase

6-phosphogluconate

6-phosphogluconate dehydrogennase

$NADP^+$

$NADPH + CO_2$

Ribulose-5-phosphate

Ribose-5-phosphate

AMP/GMP/UMP/CMP

Glucose

$NADP^+$

NADPH

UDP-glucose

UDP-glucose dehydrogenase

$2NAD^+ + H_2O$

$2NADH + 3H^+$

UDP-D-glucuronorate

H_2O

UDP

D-glucuronate

Glucuronate reductase

$NADPH + H^+$

$NADP^+$

L-gulonate

Gulonolactonase

H_2O

L-gulonolactone

Gulonolactone oxidase

$2H^+$

L-ascorbic acid

Detoxification and removal of organic compounds

[28]

Fig 1.2. Gluconolactonase catalysed reaction pathway. Gluconolactonases are highlighted in yellow.

Low cell population density

Signalling molecule
e.g. lactones

Transcriptional
regulator

High cell population density
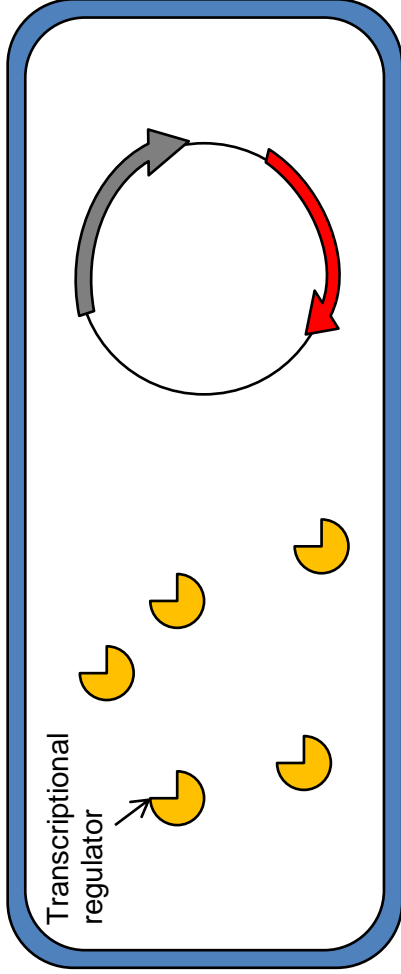
Transcription of
biofilm and virulence
genes
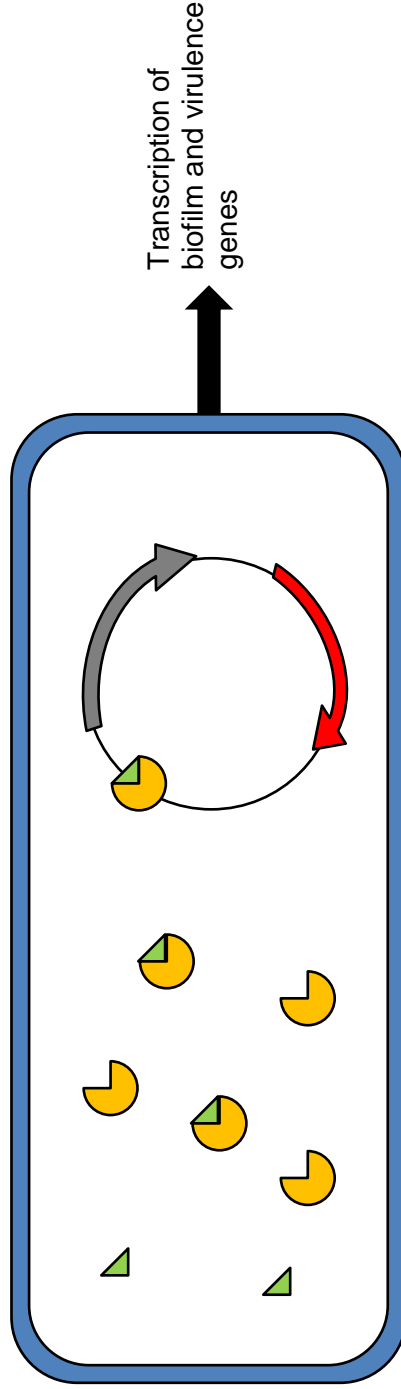
Fig. 1.3. Quorum sensing molecules regulate expression of population-dependent genes. When the population density rises, so too does the concentration of the signalling molecules. These bind transcriptional regulators that bind to promoter regions and induce expression of quorum sensing regulated genes.
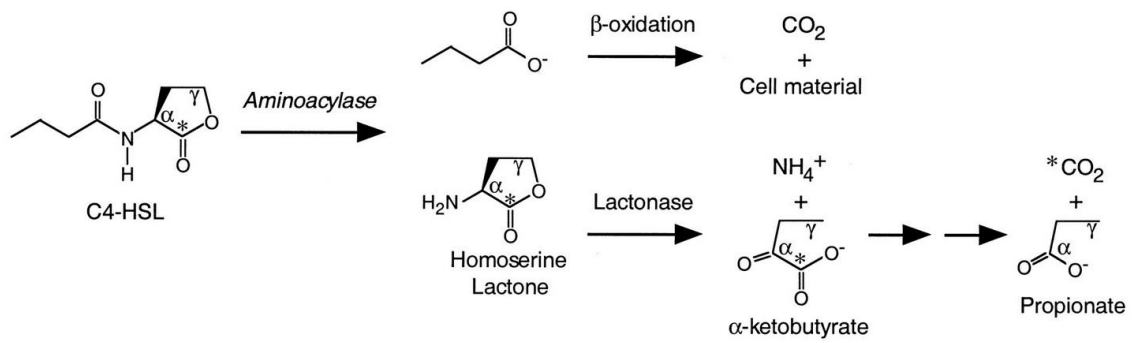
Fig. 1.4. Hypothetical degradation pathway used by *V. paradoxus* in utilising lactone signalling molecules as a source of energy. An aminoacylase releases a fatty acid molecule for β-oxidation and a lactonase hydrolyses the ester bond releasing an ammonium ion. Image taken from (Leadbetter and Greenberg, 2000).

.
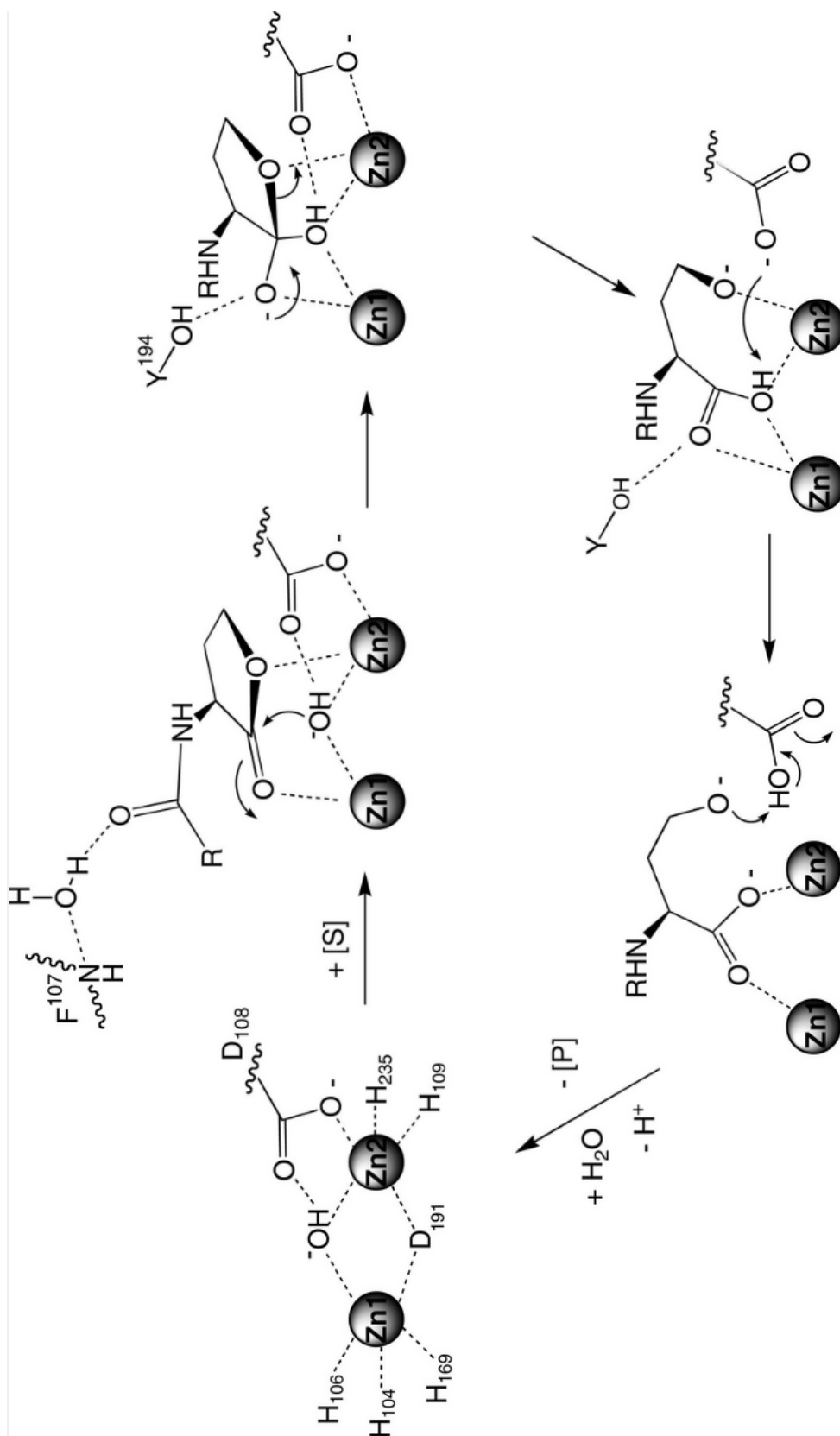
Fig. 1.5. Proposed mechanism of acyl-homoserine lactonase from *B. thuringiensis*. Image taken from (Momb *et al.*, 2008). Two zinc ions are bound closely in the active site and bind to the lactone ring with the OH group of Tyr 194 stabilising the complex. Polarisation of the ring by the two Zn^{2+} causes a reaction with an OH^- ion and results in cleavage of the lactone ring.

[31]

1.4 DXP Synthases

Isoprenoids are essential in all living organisms as they form the basis for the production of sterols, carotenoids, dolichols, vitamin A, vitamin E, vitamin K and the co-factors pyridoxal-5-phosphate and thiamine diphosphate (Brammer *et al.*, 2011; Morris *et al.*, 2013; Smith *et al.*, 2014). Most human pathogens, ranging from bacteria to the malarial parasite, and the chloroplasts of higher plants and green algae utilise the methyl erythritol phosphate pathway (MEP) (Brammer *et al.*, 2011; Xue and Ahring, 2011; Morris *et al.*, 2013). This pathway differs from that for isoprenoid synthesis in human cells which uses the mevalonate pathway, shared with archaea and fungi. Therefore it has generated a lot of interest as a potential target for new anti-microbial compounds since the potential side effects on the host would likely be minimal due to an alternative pathway being used.

1-Deoxy-D-xylulose 5-phosphate (DXP) is a key component and the first step in the MEP pathway. It is produced from pyruvate and D-glyceraldehyde-3-phosphate and the reaction is catalysed by DXP synthase. This enzyme uses thiamine pyrophosphate, one of the products of DXP synthase-catalysed reaction pathways, as a cofactor and has the potential to carry out a biosynthetic reaction resulting in 100 % conversion since one of the products is CO_2, meaning there is no equilibrium formed since one of the products is lost as a gas from the solution. This also makes DXP enzymes of interest for industrial reactions where yield is important in reducing cost and waste. Targeting this enzyme as an anti-microbial is hampered by identifying inhibitors of this enzyme that do not also interact with other thiamine pyrophosphate dependent enzymes found within mammalian cells, such as pyruvate dehydrogenase and transketolase enzymes (Samson *et al.*, 2017). The enzyme itself possesses low sequence identity to both transketolase and pyruvate dehydrogenase, with around 20 % sequence identity to each (Lois *et al.*, 1998).
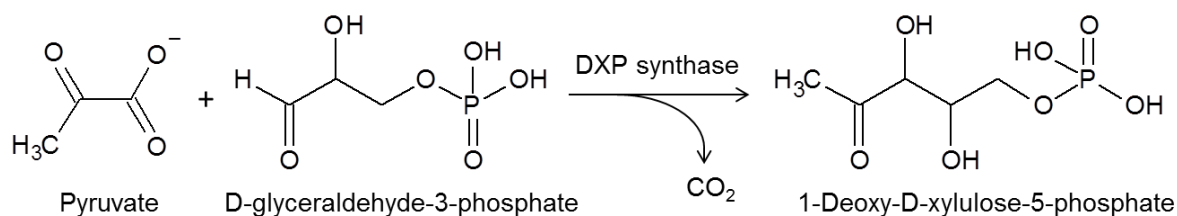
Fig. 1.6. Reaction pathway for 1-deoxy-D-xylulose-5-phosphate synthesis from pyruvate and D-glyceraldehyde-3-phosphate catalysed by DXP synthase as part of the isoprenoid biosynthesis pathway. The reaction is irreversible due to CO_2 product loss by gaseous diffusion. Image produced with ChemSketch.

To date there are two DXP synthase enzyme structures in the Protein Data Bank, which are those from *Escherichia coli* (PDB code 2O1S) and *Deinococcus radiodurans* (PDB code 2O1X) (Xiang *et al.*, 2013). These two DXP synthases were shown to be dimers with three domains, with the active site located on the interface between domain one and domain two of the monomer. This is very different from transketolase, where the active site is located at the interface between the dimer. The C2 of the thiamine pyrophosphate interacts covalently with pyruvate. A histidine in the active site, His 49 in *E. coli* DXP synthase, the equivalent of His 51 in *D. radiodurans* DXP synthase, is involved in binding with D-glyceraldehyde-3-phosphate and is thought to participate in proton transfer during the reaction (Nikkola et al., 1994). The results of a crystal soak with *D. radiodurans* DXP synthase in high concentrations of D-glyceraldehyde-3-phosphate indicated possible interactions between the side chains of His 51, His 304, Tyr 395, Arg 423, Asp 430 and Arg 480 and the substrate.
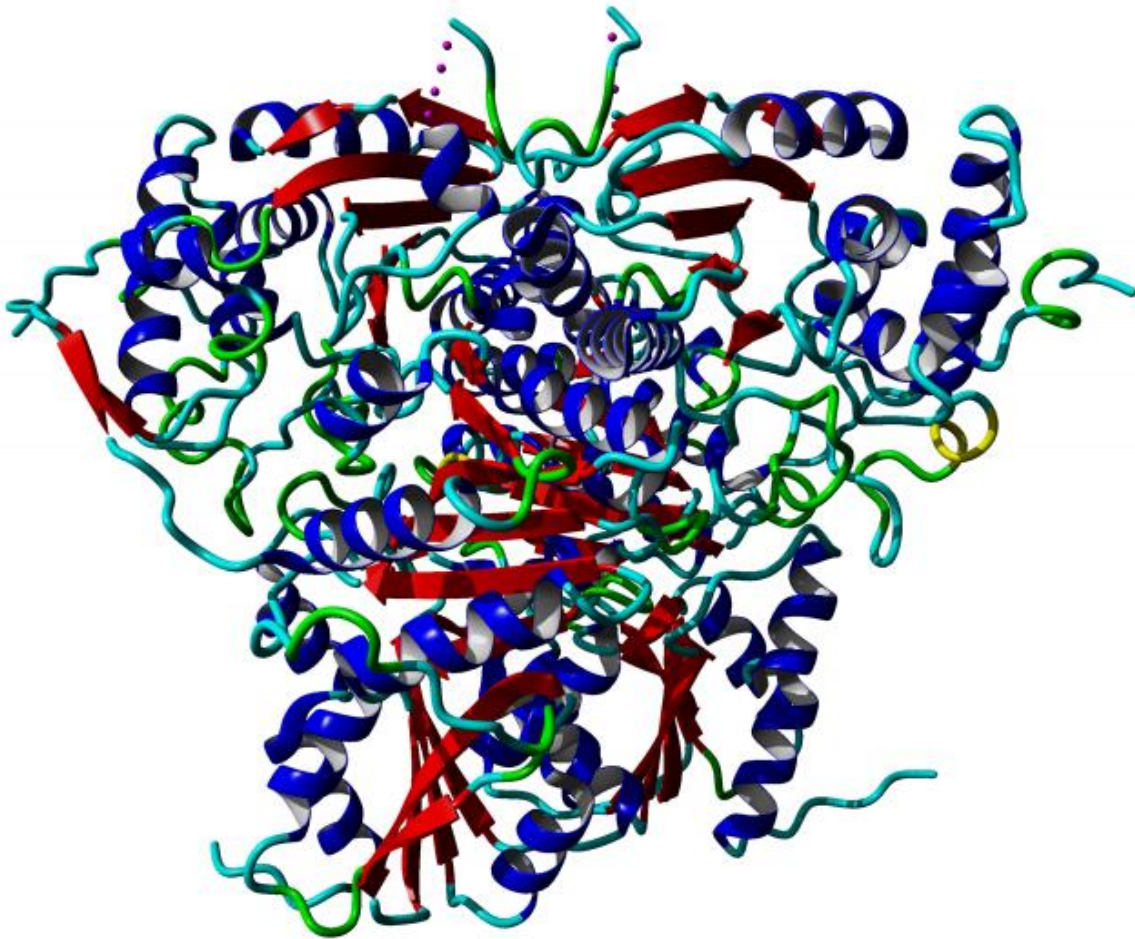
Fig. 1.7. Structure of DXP synthase from *D. radiodurans*. Crystal structure from the Protein Data Bank (code 2O1X). Made using YASARA (Krieger and Vriend, 2014), blue are α-helices, red are β-strands, green are turns and cyan are random coils. The enzyme forms a dimer with an α/β fold and a central parallel β-sheet in-between α-helices.

1.5 Principles of crystallisation

A supersaturated protein solution can be achieved with the addition of a precipitant. Using this it is possible to achieve a slow, controlled precipitation of the protein from solution without denaturing it. The most commonly used precipitants are Polyethylene glycols (PEG) due to being strong precipitant while also being a weak denaturant. Other precipitants used include ionic compounds such as salts and organic solvents, though the latter will interact with hydrophobic regions of proteins and can denature them.

To achieve well-ordered protein crystals of sufficient size it is necessary to maintain the time in which the solution is in the nucleation phase to a minimum (Fig. 1.8). This is because excess nuclei will deplete the protein concentration generating a large number of crystals too small to use or form crystal clusters where several nuclei formed together. As such it is ideal to obtain a concentration of precipitant and protein where a few nuclei form and decreasing the protein concentration to the metastable phase due to this precipitation. It is also necessary to decrease the volume of the droplet and thus increase the concentrations, without this there would be a considerable amount of protein still in solution.

Current crystallisation methods include protocols such as microbatch crystallisation (Fig. 1.9 A). A droplet containing a mixture of the protein and precipitant is dispensed into a well and covered with a layer of oil consisting of 50 % silicone oil and 50 % paraffin oil, known as Al's oil. This oil allows partial diffusion of water out of the droplet into the surrounding oil over time, decreasing the volume. The speed of diffusion can be altered by increasing or decreasing the concentration of silicone oil which is the part of the oil that allows water to diffuse out, so a higher concentration allows faster diffusion and a lower concentration slows the diffusion of water. Other methods include sitting drop and hanging drop vapour diffusion (Fig. 1.9 B). This is where the droplet is suspended on a surface exposed to the air in a sealed environment with a reservoir of the reagent that was added to the droplet to crystallise the protein solution. With a lower concentration of water in the reservoir, water slowly evaporates from the droplet into the surrounding atmosphere and into the solution in the reservoir until

an equilibrium is reached where the reagent concentration in the drop is approximately the same as the reservoir. Though the same in principle there can be observable differences between the results of both sitting and hanging drop vapour diffusion. It is proposed that the difference in the geometries in each method alter the kinetics of the processes occurring (Littlechild, 1991).

The main advantage of vapour diffusion over standard microbatch is diffusion is slower, giving more time for a single large crystal to grow before the supersaturation results in other sites of nucleation in the solution to form. However this can be remedied by adjusting the concentrations of Al's oil. Reducing the concentration of silicone oil can reduce the rate of diffusion to that seen in vapour diffusion. This versatility is the main reason microbatch was the method of choice for this project. Also vapour diffusion comes with the disadvantage of time and cost. While companies such as Douglas Instruments have made recent advances in automating the process of dispensing samples, traditionally it has been more labour intensive than microbatch which has been automated efficiently for some time. Since each well requires a reservoir of the crystallisation screen solution a lot more material is required increasing costs.

When crystallisation conditions have been established for a protein, the next step is to determine the best concentrations of additives that were present in the original by altering the concentration of the protein, salts, precipitants and the pH. Different temperatures may also be used to either slow or speed up the formation of crystals potentially giving better results. Also the addition substrates and inhibitors may also provide more optimal conditions for protein crystallisation.
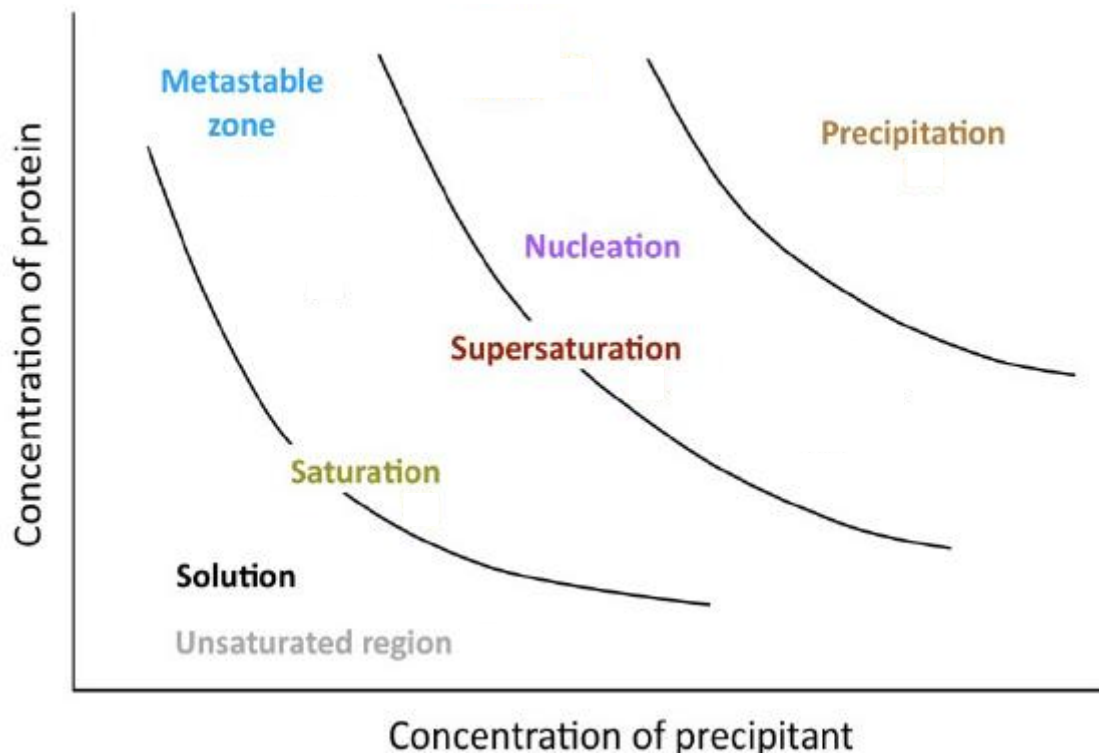
Fig. 1.8. Protein solubility curve. Modified from Alternative Protein Crystallization Technique (Nemoviov and Kut, 2012). The precipitation zone is where the protein molecules immediately precipitate into amorphous aggregates. The nucleation zone is the zone is where the protein molecules precipitates into a crystalline form. The metastable zone is where the protein nuclei grow with protein leaving solution until an equilibrium is reached.

A

B

Crystallisation droplet
(Sample and reagent)

Reservoir
(Crystallisation reagent)
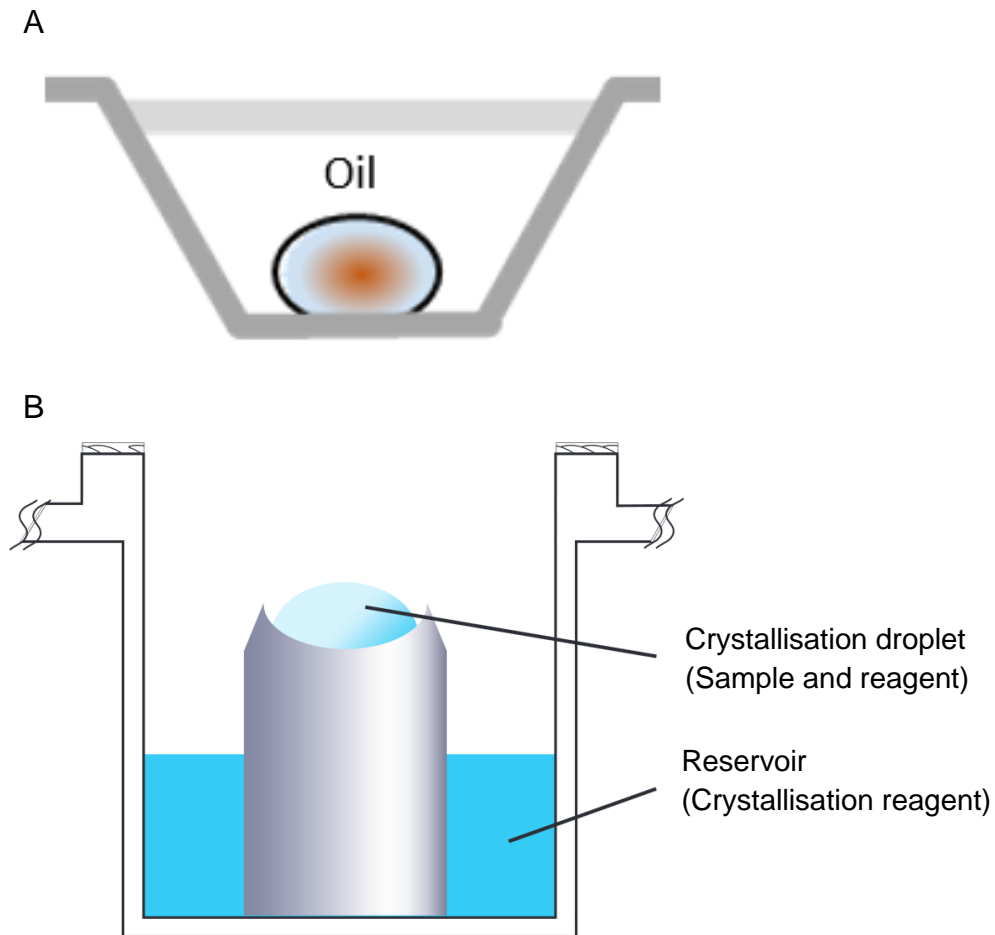
Fig. 1.9. Examples of methods for protein crystallisation. (A) Microbatch crystallisation (taken from Douglas Instruments April 2016 newsletter). (B) Sitting drop crystallisation (modified from www.hamptonresearch.com accessed on 11/7/18).

1.6 Principles of X-ray diffraction

Radiation that comes in contact with electrons surrounding atoms is scattered. Depending on the shape of the electrons field the waves of radiation are scattered differently. Using this it is possible to determine the shape and size of the electron field and produce a structure of atoms that would be present in such a field. But the scattering from single molecules does not produce a signal strong enough to detect. So a structure of molecules in an ordered and repeating crystal conformation is required to amplify the signal of the scattering by multiplying and combining the waves. For crystallography X-rays are the chosen waves as the wavelength is similar to that of the inter-atomic distances (Lytle, 1966). With this the intensity of the diffraction peaks can be used to determine the electron density of molecules in crystal structures.

Waves not only possess an amplitude but also a phase, the latter of which is lost when the wave is measured by a detector. As such the dataset cannot be completed without this vital element. To solve this, methods such as multiple isomorphous replacement are used to induce a phase shift between different data which can be compared to determine the phase. To do this the protein is either co-crystallised with a heavy metal or the crystal soaked in a heavy metal solution. The unit cell of the crystal must be the same as one analysed without the presence of the heavy metals. The location of the heavy metal is then determined within the structure as it gives a far stronger signal than other atoms present in the structure, giving not only the amplitude but also the phase. The easiest and most common method used for phase determination is molecular replacement where the phase from a similar or related protein is used to determine the phase of the protein whose structure is being analysed. With so many different protein structures currently solved, this is usually the method of choice. Discovery of a new protein fold is now unusual and a related protein structure is usually available.

1.7 The HotZyme project

This project was linked to the HotZyme grant, funded by the EU Framework 7 programme for research and technological development and the THERMOGENE ERA–IB grant. The aims of the project are to discover novel thermophilic enzymes of industrial interest from newly sequenced thermophilic bacterial and archaeal genomes, and metagenomes. Hot springs present a unique and abundant source of thermophilic species. The species that are usually found in these environments are bacteria, archaea and their viruses. Samples were collected by two groups within the HotZyme project, that of Professor Elizaveta Bonch-Osmolovskaya (Russian Academy of Sciences, Moscow), and the group of Dr Xu Peng (University of Copenhagen). It was noted that higher temperatures and more acidic environments corresponded to a decrease in biodiversity, with more of the population skewed towards archaea than at lower temperatures and higher pHs. The group speculated that this was due to differences between the composition of cell walls between bacteria and archaea. During sequencing of the samples, 58 % of the reads could not be assigned to any known phyla. This supports the idea that many species of microorganisms have not yet been identified and characterised, especially those in hot and other extreme environments (Menzel *et al.*, 2015). HotZyme aimed to clone and over-express high priority novel hydrolases from hot springs located in several different countries and to characterise them both biochemically and structurally and, in some cases, modify their substrate specificity by targeted mutagenesis experiments to optimise their use as industrial biocatalysts.

1.8 Aims of this project

This project focused on cloning, overexpressing, purifying and characterising biochemically and structurally novel thermophilic enzymes identified from genomic and metagenomic resources. Enol lactonases, gluconolactonases and quorum sensing lactonases that have applications to the pharmaceutical industry were investigated as they are of interest to Sigma-Aldrich, who will be providing newly synthesised lactone substrates for assays. DXP synthases identified from thermophilic genomes were investigated as part of the THERMOGENE project, funded by an ERA-IB grant, whose goal is to identify and characterise,

biochemically and structurally, novel thermostable transferases from newly sequenced thermostable genomes and metagenomes.

2. General Materials and Methods

2.1 Gene discovery from thermophilic genomic and metagenomic DNA

Hot springs were sampled and sequenced by the groups of Professor Elizaveta Bonch-Osmolovskaya (Russian Academy of Sciences, Moscow), and Dr Xu Peng (University of Copenhagen) as part of the HotZyme project. The hot springs that were sampled in the project were located in China, Iceland, Italy, Russia, and the USA. Temperatures in these hot springs varied between 61 °C and 92 °C, and a range of pH's between 1.8 and 7.0. DNA was extracted and cleaved into 170 nt fragments on average and the samples were Illumina sequenced, assembled and the metagenomes clustered through the Uniprot and Swissprot databases.

2.2 Gene cloning

2.2.1 Gene identification

Genomic data were obtained from the online genomic and metagenomic repository ANASTASIA (Automated Nucleotide Aminoacid Sequences Translational plAtform for Systemic Interpretation and Analysis) (Ladoukakis *et al.*, 2014) constructed by the group of Professor Fragiskos N. Kolisis (National Technical University of Athens, Greece) who were collaborators of the HotZyme project. Homologues in the Protein Data Bank (PDB) were identified using the BLAST search engine (Altschul *et al.*, 1990) to identify similar proteins within the PDB.

2.2.2 Amplification of genomic DNA

Genomic DNA was amplified using the GenomiPhi DNA amplification kit (GE Healthcare, USA). 1 ng of DNA in 10 µl of double distilled water (ddH$_2$O) was heated to 95 °C for 3 min, then cooled to 4 °C on ice. 9 µl of reaction buffer and 1 µl of the enzyme mix was then added and incubated at 37 °C for 1.5 hours. After this the sample was heated to 65 °C for 10 min to inactivate the enzyme.

2.2.3 PCR amplification

Genes were amplified by polymerase chain reaction (PCR) with Phusion polymerase (New England Biolabs, USA). A mix of 5 µl *Pfu* 10X buffer, 1 µl of 2000 u ml^{-1} *Pfu* polymerase, 6 µl MgSO$_4$ at 50 mM, 1 µl of 10 mM dNTPs, 2.5 µl of 100 % DMSO, 1 µl of 100 µM of each forward and reverse primer, 0.5-50 µg µl^{-1} DNA and made up to a total of 50 µl with nuclease free water. Samples were put into a Bio-Rad thermocycler T100 (Bio-Rad, USA) with the following protocol: 98 °C for 2 min, then 35 cycles of: 95 °C for 1 min, 55 °C for 30 s (or other temperatures appropriate for the T_m of the primers), 72 °C for 2 min kb^{-1}; then a final step of 72 °C for 5 min. Primers for PCR were ordered from Invitrogen; primer sequences are shown in Table 2.

2.2.4 Agarose gel electrophoresis

Gels were made with 0.8 % w/v agarose dissolved in TAE buffer (40 mM Tris-HCl, pH 8.3, 20 mM acetic acid, 1mM EDTA) containing 0.5 mg ml^{-1} Midori Green Advance (Nippon Genetics, Germany) to stain the DNA for visualisation under UV light (302 nm). Samples were loaded along with a 1 kb ladder (New England Biolabs, USA) for measurement. Gels were run in a Sub-Cell GT Cell (Bio-Rad, USA) at 100 V in TAE buffer for 45 min and imaged with a BioDoc-It imaging system (UVP, Canada).

2.2.5 Ligation independent cloning

Following genomic DNA amplification or PCR, a clean-up step was performed using SureClean Plus (Bioline, UK). An equal volume of SureClean Plus was added to the solution and incubated for 10 min. It was then centrifuged at RCF = 14000 for 10 min and the supernatant removed. A volume of 70 % ethanol equal to double the sample volume was then added, vortexed and centrifuged again for 10 min. The supernatant was removed and the pellet resuspended in nuclease free water. The PCR product was inserted into either pLATE31 or pLATE51 expression vectors (Thermo Scientific, USA) with the following protocol: a solution containing 2 µl LIC buffer, 2 µl PCR product, 5 µl nuclease free water and 1 µl of 5 U µl^{-1} T4 DNA polymerase was incubated for 5 min at room

temperature to create overhangs, after which 0.6 µl of 0.5 M EDTA pH 8.0 was added to stop the reaction. Then 1 µl of the pLATE vector at 60 ng (0.02 pmol) was added and the mix incubated for a further 5 min.

Table 2. Forward and reverse primers used in the cloning of the specified genes.

Primer name	5′ to 3′ sequence[1]
Lac11 forward	AGAAGGAGATATAACTATG**CGGAAGCTTCT**
Lac11 reverse	GTGGTGGTGATGGTGATGGCC**TCAGAAACCCAACC**
DXP Ta forward	AGAAGGAGATATAACTATG**ATTTTAGATAAAGTCAACAGCCCAGACG**
DXP Ta reverse	GTGGTGGTGATGGTGATGGCC**TTCACGAACCGATTCCTTAATTCCGTC**
DXP Ch forward	AGAAGGAGATATAACTATG**CCTATTCTTGAACGCATCTC**
DXP Ch reverse	GTGGTGGTGATGGTGATGGCC**TAGTTTACTCAGGATTTCACGAAACTT**

[1] In normal font are the vector overhang portions of the primers and in bold are the gene-specific overhangs.

2.2.6 Gene synthesis

Two genes from metagenomic DNA unavailable at the time for PCR were synthesised by ATUM (formerly DNA 2.0) (California, USA). Genes were codon optimised for expression in *E. coli* by the company's own software and inserted in expression plasmids ready for transformation into cell lines. The expression vector was the Electra plasmid pD441-CH containing a T5 promoter with strong ribosome binding affinity, under the control of isopropyl β-D-1-thiogalactopyranoside (IPTG), and a C-terminal histidine tag. Plasmids were supplied as a dry precipitate and were resuspended in TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA) to a final concentration of 10 ng µl^{-1} with incubation at 50 °C for 20 min.

[45]

2.2.7 Plasmid minipreps

Plasmids were purified from cells using GeneJET Plasmid Miniprep Kit (Thermo Scientific, USA). A single colony from an agar plate containing cells transformed with the recombinant DNA was grown in 10 ml of LB broth overnight at 37 °C with shaking at 200 rpm. The culture was centrifuged at 6800 rpm for 2 min and the pellet resuspended in 250 µl resuspension buffer containing 50 mM Tris-HCl pH 8.0, 10 mM EDTA and 100 µg ml^{-1} RNase A. 250 µl of lysis buffer (200 mM NaOH, 1 % SDS) was added and mixed. 350 µl neutralisation buffer (4.2 M guanidinium chloride-HCl pH 4.8, 0.9 M potassium acetate) was added. The sample was centrifuged for 5 min at 12100 rpm. The supernatant was loaded on to a GeneJET Spin Column and centrifuged at the same speed for 1 min. 500 µl of wash solution (10 mM Tris-HCl pH 7.5, 80 % ethanol) was added. The column was centrifuged for 1 min and a second wash step performed before centrifuging it dry for another min to remove any residue. 50 µl elution buffer (10 mM Tris-HCl, pH 8.5) was added to the membrane and incubated for 2 min before being centrifuged for 2 min.

2.2.8 Transformation of competent cell lines

E. coli cloning cell lines used were 5-alpha competent cells (New England Biolabs, USA). All DNA sequencing was carried out by Eurofins Genomics. The *E. coli* expression cell lines were BL21 (DE3) (New England Biolabs, USA), BL21 RIPL (Agilent, USA), BL21 pLysS (New England Biolabs, USA), Rosetta (Merck Millipore, USA) and Arctic Express (DE3) (Agilent, USA). Cells were transformed by heat shock at 42 °C for 30 s and selected for either ampicillin or kanamycin resistance on LB agar plates.

2.2.9 Colony PCR

Successful transformants were tested using colony PCR with DreamTaq DNA polymerase master mix (Fermentas, USA) and the relevant primers. The same protocol, as previously described for the PCR reaction, was used with the exception that, instead of adding DNA to the mix, a small amount of a single

colony was added by taking cells with the end of a pipette tip and thoroughly pipetting the tip in the master mix solution.

2.2.10 Takara chaperone plasmid transformations

Five Takara plasmids containing chaperone genes (pG-KJE8, pGro7, pKJE7, pG-TF2 and pTF16; Takara Bio, Japan) were each transformed into *E. coli* expression cell line BL21 (DE3). It was then necessary to make the transformed cells competent again. Cells were selected for on agar plates containing 20 µg ml^{-1} chloramphenicol and then grown overnight in 10 ml LB media at 37 °C 200 rpm. 1 ml of this was inoculated into 100 ml of LB medium and incubated for 1.5 – 3 hours at 37 °C and 200 rpm. The cells were put on ice for 10 min and then centrifuged at RCF = 6000 for 3 min. The supernatant was discarded, and the cells resuspended in ice cold 0.1 M $CaCl_2$ and left on ice for 20 min. Cells were spun down again as previously described and resuspended in 0.1M $CaCl_2$ and 15 % v/v glycerol before being flash frozen in liquid nitrogen and stored at −80 °C. Cells were then ready for transformation with expression plasmids.

2.3 Enzyme purification

2.3.1 Bacteria cultures and storage

Glycerol stocks were made by mixing 50 % sterile glycerol with an equal volume of overnight bacterial culture, which was then flash frozen in liquid nitrogen and stored at −80 °C.

2.3.2 Protein expression

Transformed *E. coli* expression cell lines were grown in 1 litre high salt lysogeny broth (LB) medium, inoculated with 20 ml of overnight culture. Flasks were put into shaking incubators at 37 °C, 200 rpm and expression induced with 1 mM IPTG at an OD at 600 nm of 0.6 corresponding to the log phase of bacterial growth as standard. For optimisation purposes these values were varied depending on the protein being expressed. Cells were harvested after 4 hours of expression by centrifugation at 4700 rpm.

2.3.3 Expression trials

Cell cultures were grown as stated above with one or more variables altered to achieve high levels of expression of soluble protein. Cells were grown at temperatures of 37° C until induction when the OD 600 nm was 0.6. After the induction the culture either remained at 37°C for 4 hours, or 30°C for 6 hours or 12°C and left overnight. The IPTG concentration was varied between 1 mM and 0.1 mM.

2.3.4 Autoinduction

500 ml of ZYM media was made consisting of 16 g l^{-1} tryptone, 10 g l^{-1} yeast extract and 5 g l^{-1} NaCl$_2$ and autoclaved. To this was added: 25 mM (final concentration) Na$_2$HPO$_4$; 25 mM KH$_2$PO$_4$; 50 mM NH$_4$Cl; 5 mM Na$_2$SO$_4$; 2 mM, MgSO$_4$; 0.5 % v/v glycerol; 0.05 % w/v glucose; 0.2 % w/v α-D-lactone. A solution of trace metals was also added to the following final concentrations: 50 mM FeCl$_3$; 20 mM CaCl$_2$; 10 mM MnCl$_2$; 10 mM ZnSO$_4$; 2 mM CoCl$_2$; 2 mM CuCl$_2$; 2 mM NiCl$_2$; 2 mM Na$_2$MoO$_4$; 2 mM Na$_2$SeO$_3$; 2 mM H$_3$BO$_3$. Medium was induced with 10 ml of overnight cultures and grown for 48 hours at 12 °C and 200 rpm. Cultures were then harvested by centrifugation at the same speed as stated before.

2.3.5 Takara plasmid co-expression

Cells containing both a Takara plasmid (Takara Bio Inc, Japan) and a vector containing the gene of interest were grown as standard at 37 °C and 200 rpm in 1 litre LB media, inoculated with 20 ml of overnight culture. Media contained 20 µg ml^{-1} chloramphenicol for Takara plasmid selection, and a second antibiotic to select for the gene-containing vector. Expression of the chaperones was induced at the beginning of growth: 0.5 – 4 mg ml^{-1} L-arabinose used for expression of pGro7, pKJE7 and pTF16; 1 – 10 ng ml^{-1} tetracycline was added for the induction of pG-Tf2; and both 0.5 – 4 mg ml^{-1} L-arabinose and 1 – 10 ng ml^{-1} tetracycline

was added for the induction of pG-KJE8. Cells were induced for expression of the gene of interest as stated above at OD of 0.6 - 0.8 at 600 nm.

2.3.6 SDS-PAGE gels

Protein samples were mixed with loading buffer containing 100 mM Tris-HCl pH 8.0, 2 % 2-mercaptoethanol v/v, 40 g l^{-1} SDS, 2 g l^{-1} bromophenol blue, 20 % glycerol v/v and heat denatured at 100 °C for 5 min. Samples were then loaded onto ExpressPlus PAGE gels (GenScript, USA) and a Spectra Multicolor Broad Range Protein Ladder (ThermoFisher, USA) was also loaded. The gel was run at 140 V for 60 min in running buffer containing: 20 mM Tris-HCl, 20 mM MOPS, 300 mM SDS, 1 M EDTA at pH 8.3. Proteins were detected with InstantBlue protein stain (Expedeon, UK) by leaving the gel shaking in the stain for 1 hour followed by another hour in deionised water to destain the gel to remove dye from the background.

2.3.7 Purification of enzymes

Pelleted *E. coli* cells harvested from an expression culture were resuspended in 50 mM Tris-HCl pH 7.5, with 10 ml of buffer per gram of cell paste and subjected to sonication at an amplitude of 15 μm for 20 s pulses, repeated 10 times with 40 s breaks between for cooling on ice. The samples were then centrifuged at RCF = 24000 for 30 min to remove the cell debris. For nickel affinity chromatography, samples were passed through a 1 ml HisTrap HP column (GE Healthcare, USA) and eluted with Tris-HCl pH 7.5 buffer containing 1 M imidazole. Samples were then concentrated down to less than 2 ml with a Vivaspin 20 concentrator with a 10 kilodalton (kDa) cut off (Sartorius, Germany) and then put through a gel filtration (GF) GF200 column (Superdex 200 Hiload 16/60, GE Healthcare, USA) with 50 mM HEPES buffer pH 7.5 containing 100 mM NaCl. 1.5 ml fraction sizes were collected with both purification steps. An ÄKTApurifier (GE healthcare, USA) was used to run each column.

2.3.8 Western blotting

SDS-PAGE gels were run as duplicates, with one being stained as described above and the other left unstained. The proteins, separated in the gel, were transferred to a nitrocellulose membrane (Sartorius, Germany) using a Pierce G2 Fast Blotter (Thermo Fisher Scientific, USA). Transfer was at 25 V, 1.3 A for 15 min. The membrane was then put into the iBind Western System (Thermo Fisher Scientific) soaked in 6 ml of iBind mix made from: 6 ml iBind buffer, 300 µl additive and 23.7 ml ddH$_2$O. This was then placed on an iBind card also soaked in 6 ml of the iBind mix. Anti-his tag antibodies and anti-mouse antibodies were diluted accordingly to the manufacturer's recommendations. The membrane and card were left for either 2.5 hours or overnight for better resolution. Using lateral flow capillary diffusion, the membrane was exposed to the primary antibody, then washed with iBind mix, then the secondary antibody, with a final wash step at the end. The results were viewed using an Odyssey CLc Imaging System (LI-COR, USA).

2.4 Crystallisation trials

2.4.1 Crystal screening

Protein was concentrated using a Vivaspin 20 concentrator with a 10 kDa cut-off (Sartorius, Germany) and its concentration measured using a NanoDrop 2000c spectrophotometer (Thermo Scientific, USA). An Oryx 8 crystallisation robot (Douglas Instruments, UK) was used to set up microbatch crystallisation trials in Hampton 96 well microbatch plates (Greiner, Austria). 1 µl droplets were made with a 50:50 mix of protein and crystallisation screen. Plates were incubated at 19 °C. The screens used were Sigma 82009 and 70437 (Sigma Aldrich, USA), MDL-01 and MDL-02, JCSG plus, MIDAS, Morpheus, Clear Strategy, Stura Footprint and Macrosol (Molecular Dimensions Limited, UK). The lac11 co-crystallisation trials were performed by incubating lac11 at room temperature with either 100 mM glucose or 1 mM of either ZnCl$_2$ or CoCl$_2$ prior to being dispensed into crystallisation dishes. Crystals were frozen and stored in liquid nitrogen.

2.4.2 Crystal optimisation

The X-step optimisation software (Douglas Instruments, UK) was used to calculate dispensing volumes for each well for crystallisation optimisation. Concentrations of the lac11 protein were varied between 3.2 mg ml^{-1} and 5 mg ml^{-1}, and 2-methyl-2,4-pentanediol (MPD) between 10 % to 30 % v/v. The concentration of PEG 8000 was 25 g l^{-1} and Bis-Tris-HCl pH 6.5 was 0.1 M.

2.5 Assays

2.5.1 Lactonase pH-based colorimetric assay

Reactions were performed in a solution of 2.5 mM BICINE, pH 8.0-8.3 containing 0.2 M NaCl, 0.2 mM cresol purple, and 10 mM of the lactone substrate. 195 µl of this solution was loaded into each well. 5 µl protein at 2 ng ml^{-1} was added immediately prior to monitoring absorbance and mixed thoroughly. The decrease in absorbance at 577 nm was monitored in a Greiner 96 well plate and analysed in an Infinite 200 Pro (Tecan) plate reader for up to 30 min or until a complete colour change was observed. All assays were carried out in triplicate.

2.5.2 Differential scanning fluorimetry

Protein unfolding in various conditions was measured using Sypro Orange dye diluted from a 5000 X commercial stock solution by adding 1 µl to 124 µl ddH$_2$O. Within a 96 well white PCR plate (Fisher Scientific, USA) the following was added: 5 µl protein to a final concentration of 0.1 mg ml^{-1}; 2.5 µl Sypro Orange from diluted stock; 11.5 µl 50 mM HEPES pH 7.5, 100 mM NaCl and was covered with a qPCR seal. Then the plate was centrifuged at RCF = 500 for 2 min to remove air bubbles. It was then measured in a Mx3005P quantitative PCR machine (Stratagen, UK) at 580 nm between 25°C and 99°C over 40 min.

2.6 Bioinformatics

All programs used are part of the CCP4 integrated suite of crystallography programs unless stated otherwise (Winn *et al.*, 2011). The following software was used:

- Molrep version 11.0 /22.07.2010/ was used for molecular replacement.
- BALBES from the York Research Database (Long *et al.*, 2007) was also used for molecular replacement (accessed March 2014).
- REFMAC version 5.3 was used for structural refinement.
- COOT version 0.8.5 was also used for structural refinement.
- CCP4I version 7.0.060 was used in structural analysis.
- PyMOL version 2.3 was used for production of protein figures.

2.6.1 Multiple sequence alignment

Clustal Omega (Sievers *et al.*, 2011) from EBI Database was used to perform multiple sequence alignments. Parameters used were the default.

2.6.2 X-ray data collection and analysis

Data was collected using the Diamond Synchrotron Light Source (Harwell Science and Innovation Campus, Oxfordshire) and the data was processed by the programs Xia2 (Winter, 2010) and iMOSFLM (Leslie and Powell, 2007). Molecular replacement was performed by MOLREP (Vagin and Teplyakov, 2010) with the structure of *Xanthomonas campetris* lactonase (PDB code 3DR2) which has 35 % sequence identity to lac11. The program COOT (Emsley and Cowtan, 2004) was used to perform structural visualisation and refinement along with REFMAC (Murshudov *et al.*, 1997).

2.6.3 Homology modelling

Three different servers were used to generate homology models: SWISS-MODEL (Biasini *et al.*, 2014), Phyre2 (Kelly *et al.*, 2015) and I-TASSER (Zhang, 2008). Models were generated with all the servers for each protein and compared

to determine the most accurate one using Rampage (Lovell *et al.*, 2003). Models were viewed using YASARA (Krieger and Vriend, 2014), CCP4mg (McNicholas *et al.*, 2011) and superposed with Chimera (Pettersen *et al.*, 2004) due to incompatibility between the PDB files of the models generated by Phyre2, and the YASARA software which prevented superposition with files from the Protein Data Bank.

2.6.4 Analysis of homology modelling

The Rampage server, from the University of Cambridge (Lovell *et al.*, 2003) was used to generate Ramachandran plots and list the amino acids with irregular conformations. The QMEAN (Qualitative Model Energy ANalysis) server from SWISS-MODEL was used to generate and give an estimation of the quality of the homology models. The models were scored with the more negative the value the worse the structure, with the score being ideally as close to 0 as possible. The score was based on: the local geometry, where the torsion angle potential of three consecutive amino acids was analysed; long range interactions between secondary structures within the model; the solvation potential to analyse the buried status of amino acids such as hydrophobic ones; and analysis of observed and predicted secondary structure.

2.6.5 Oligomeric state analysis

PISA, part of the suite of programs in CCP4I, was used to test structures and calculate the macromolecular surfaces and interfaces, assess their properties and determine probable oligomeric states. The online server FRODOCK 2.0 was also used for protein-protein docking to predict dimeric models.

2.6.6 Sequence and secondary structure alignment

ESPript 3.0, developed by Lyon University (Lyon, France), was used to obtain sequence and secondary structure similarities from aligned sequences. Sequence alignments were created in Clustal Omega and uploaded with the PDB files for the structure of interest and a reference structure. Default parameters were used for analysis (Robert and Gouet, 2014).

3. Quorum sensing lactonase

3.1 Introduction

Quorum sensing lactonases were investigated due to the high interest by industrial partners for their potential to disrupt biofilm formation in bioreactors and water filtration systems. A quorum sensing lactonase was identified within the newly sequenced genome from the thermoacidophilic anaerobic crenarchaeon *Vulcanisaeta moutnovskia*. The genome was sequenced by Prof. Elizaveta Bonch-Osmolovskaya from a hot spring in Kamchatka, Russia (Gumerov *et al.*, 2011). The lactonase was cloned into an over-expression *E. coli* cell line by the group of Prof. Bettina Siebers, University of Essen, Germany. The lactonase was cloned with a C-terminal His-tag into a pLATE31 vector and into the *E. coli* BL21-CodonPlus (DE3) RIPL cell line (Agilent Technologies, USA). A second possible quorum sensing lactonase was identified and investigated from a metagenome using the ANASTASIA platform generated within the HotZyme project. Our goal was to structurally characterise these enzymes and compare their structures to known mesophilic homologues and determine differences that could be attributed to their adaptation to hot environments, as well as possible differences in substrate specificity through comparing their active sites.

3.2 Materials and methods

3.2.1 Co-crystallisation

The substrate γ-butyrolactone was added to the protein sample to a final concentration of 10 mM and to a sample of 11 mg ml^{-1} of protein. The solution was incubated at room temperature for 20 min before being dispensed into the microbatch crystallisation plates.

3.3 Quorum sensing lactonase from *V. moutnovskia*

E. coli BL21 (DE3) pLysS cells were provided already containing a pLATE31 vector with the gene for the *V. moutnovskia* lactonase inserted. The cells were grown in LB media at 37 °C and 200 rpm and the lactonase expression induced with 1 mM IPTG. The lactonase was successfully purified with a combination of nickel affinity chromatography and size exclusion chromatography using

standard Tris-HCl buffer as described in material and methods (2.3.6). Fig. 3.1 shows the 280 nm trace from the gel filtration and the SDS-PAGE gel showing the pure protein at its molecular weight of 35 kDa.

The activity of the lactonase was determined using the substrates shown in Fig. 3.2 using the colourimetric assay described in the materials and methods. The lactone substrates were provided by one of our HotZyme collaborators, Dr Roland Wohlgemuth, Sigma Aldrich, Merck. The colour change for the initial assays was very slow, taking more than 24 hours. The addition of 1 mM $MnCl_2$, which was selected due to a related enzyme from *Sulfolobus islandicus* having Mn^{2+} as a cofactor (Hiblot et al., 2012), increased the rate of reaction so results were obtained within an hour. This suggests that the cofactor for this enzyme is Mn^{2+}. The results of assays done with the addition of $MnCl_2$ are shown in Table 3. It is shown that the enzyme has activity with a wide variety of different lactone substrates with various lengths of acyl chains. The substrates where activity was observed are γ-butyrolactone, γ-valerolactone, γ-caprolactone, whiskey lactone and δ-dodecalactone. No measurable activity was seen for mevalonolactone or δ-decalactone. Activity towards D-galactono-γ-lactone could neither be confirmed nor ruled out as the substrate was very unstable in solution and auto-hydrolysed rapidly due to the control producing a colour change as fast as samples containing the lactonase. For two of the substrates where activity was known, γ-valerolactone and γ-caprolactone, the optical isomeric forms of these were tested to determine stereoselectivity of the enzyme. Results in Table 3 indicated that while activity is seen with both isomers, it seems to favour the D form of these substrates as there was a greater pH change observed compared to the L form.
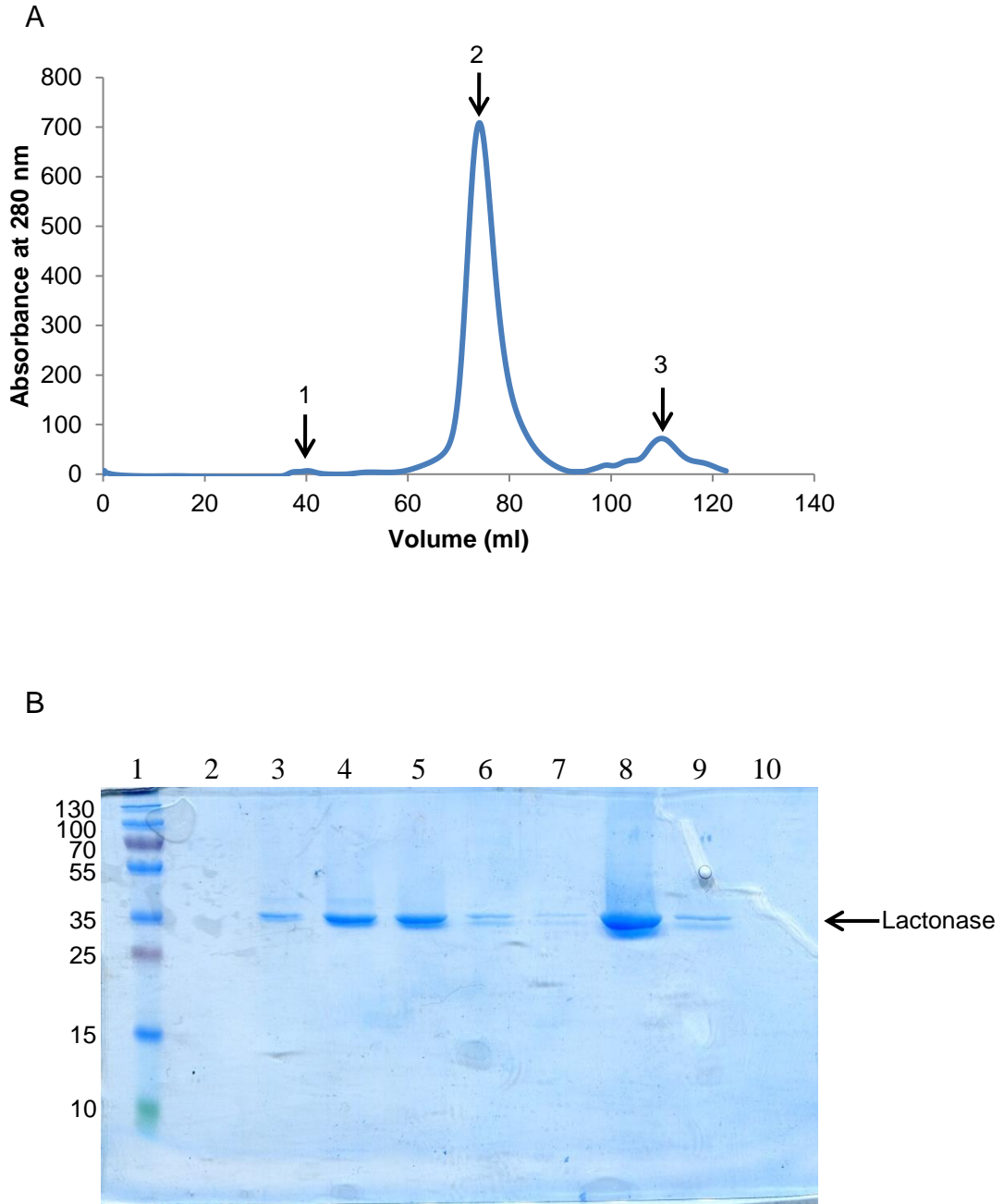
Fig 3.1. Purification of *V. moutnovskia* lactonase. A) Gel filtration trace measuring absorbance at 280 nm. Arrows show peaks analysed by SDS-PAGE. B) Analysis of gel filtration by SDS-PAGE: Lane 1. Marker (sizes in kDa); 2. Peak at 1l; 3-9. Peak 2; 10. Peak 3.
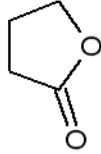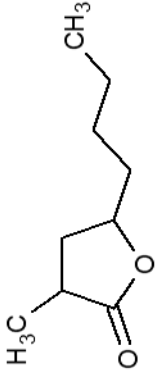
Fig 3.2. Illustration of lactone substrates used in assay.

Table 3. Relative enzyme activity of *V. moutnovskia* lactonase for different substrates
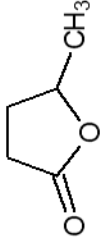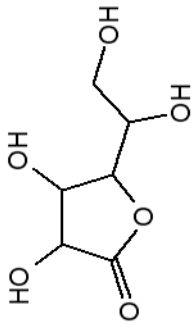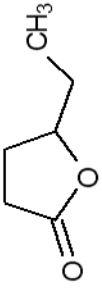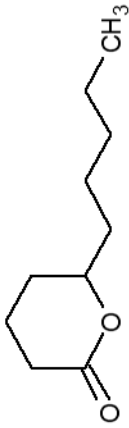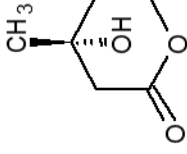
Lactone	Activity[1]
γ-butyrolactone	++
γ-valerolactone	++
L-γ-valerolactone	+
D-γ-valerolactone	++
mevalonolactone	−
γ-caprolactone	++
L-γ-caprolactone	+
D-γ-caprolactone	++
whiskey lactone	+
D-galactono-γ-lactone	−
δ-decalactone	−
δ-dodecalactone	++

[1] ++ represents ≥0.15 nmoles of H^+ ions produced above the value obtained for the blank. + represents 0.05-0.15 nmoles H^+ ions produced above the blank. − represents no activity measured.

The contribution of various buffers to the stability of the enzyme was tested using differential scanning fluorimetry (thermal shift assay) to determine which buffer would be most appropriate for storage (Fig. 3.3). This would be of benefit when attempting to crystallise the enzyme as a stable enzyme would have reduced flexibility enabling its molecules to pack together to form a well-ordered crystal. The concentration of protein used was 0.9 mM. Overall the enzyme exhibited good thermal stability with unfolding not seen below 72.0 °C, although this may not represent the activity retention of the enzyme. The enzyme stability was lowest with Bis-Tris propane where the melting temperature was measured at 74.3 °C. The MOPS and HEPES buffers showed similar melting temperatures for the protein 80.2 °C and 79.8 °C, respectively. The Tris-HCl buffer was shown to be the most stabilising of the buffers tested with an enzyme melting temperature of 83.5 °C, suggesting that it is optimal for the crystallisation trials.
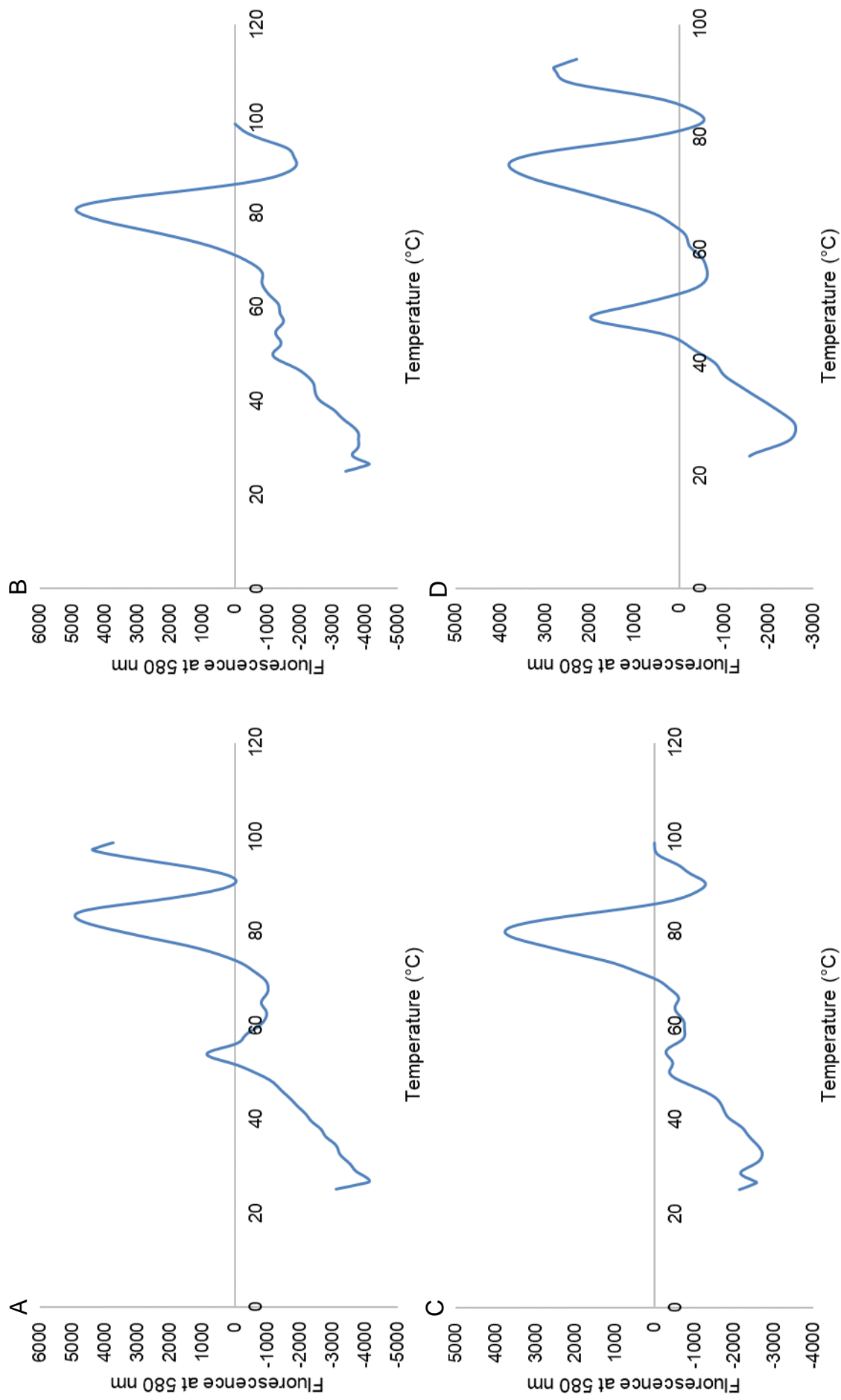
Fig. 3.3. Differential scanning fluorimetry for *V. moutnovskia* lactonase in different buffers: A is Tris-HCl; B is HEPES; C is MOPS; D is Bis-tris propane

The *V. moutnovskia* lactonase was purified and concentrated using a centrifugal concentrator to between 11 and 14 mg ml^{-1} for initial crystallisation trials as concentrations of 10 mg ml^{-1} and above are a common range for protein crystallisation in screens from such as JCSG-plus and good starting point to test (Schulz, 2007). Crystallisation trials were carried out using sitting drop microbatch screening and sitting drop vapour diffusion screening. All available screens were tested and crystals were obtained from 10 conditions, with the best crystals from JCSG-plus as seen in Fig. 3.4. The conditions were: 0.1 M HEPES pH 7.0 and 10 mg ml^{-1} PEG 6000. A pre-made cryo-protectant with similar conditions was used when harvesting crystals, containing 0.1 M Tris-HCl pH 7.5, 10 % w/v PEG 6000 and 30 % v/v PEG 400. Data were collected from these crystals at the Diamond Synchrotron and analysed using the CCP4 suite of programs. Unfortunately, none of the crystals diffracted to high enough resolution for the structure to be determined and the data could not be processed by Xia2 (Winter, 2010) or iMosflm (Battye *et al.*, 2011). To try and improve the diffraction quality, crystals were grown in the presence of one of the confirmed substrates, γ-butyrolactone at a concentration of 10 mM. No crystals were obtained using this method.
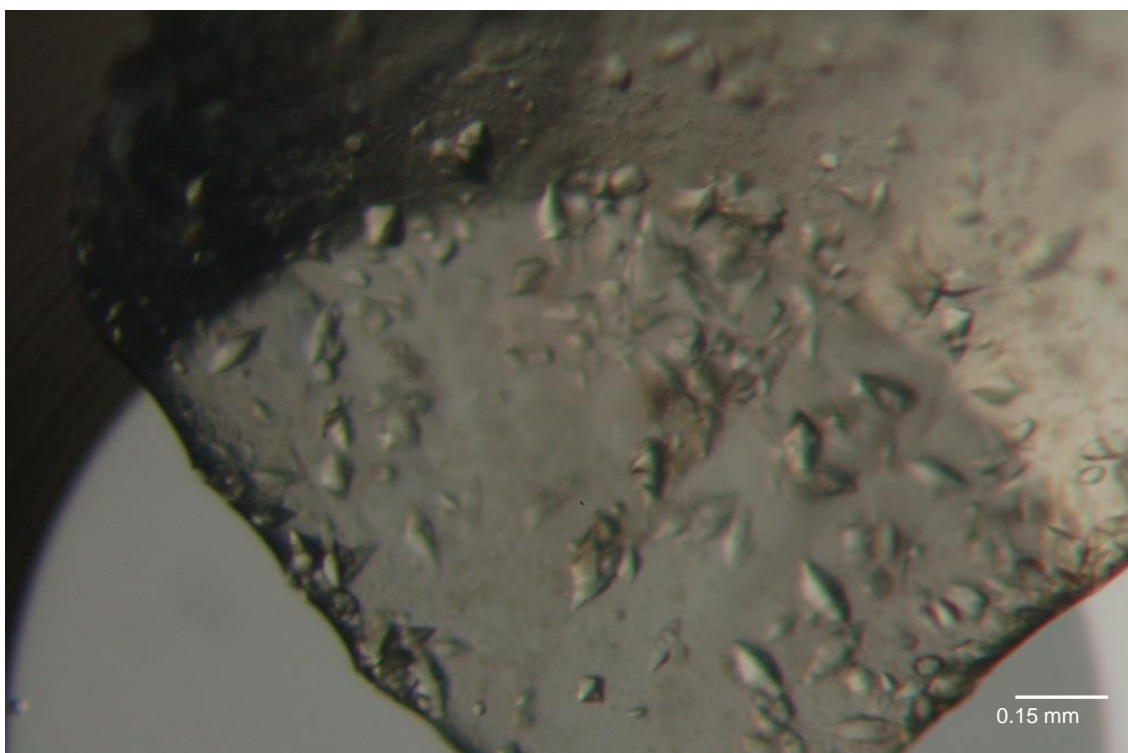
Fig. 3.4. Crystals grown of the *V. moutnovskia* lactonase using the JCSG-plus screen. The conditions were: 0.1 M HEPES pH 7.0 and 10 mg ml^{-1} PEG 6000.

Before our studies could be completed another group reported in a paper (Hiblot *et al.*, 2013) that the structure of this enzyme had been determined to 2.4 Å resolution with a space group of P 6 2 2 (PDB code 4RE0). This was later improved upon with another crystal structure by the same author to a higher resolution of 1.8 Å and a space group of P 6 4 (PDB code 4RDZ) (Hiblot *et al.*, 2015).

The structure from the other group revealed the enzyme to be a homodimer with a (β/α)$_8$ barrel topology with two metal cations located in the active site (Fig. 3.5 A); this is similar to other known lactonase structures of this class (Draganov, 2010). The cobalt ions as seen in Fig. 3.5 B are bridged by a catalytic water molecule and are coordinated by His 23, His 25, His 171, His 200, Asp 257 and a carboxylated Lys 238. The most significant difference when compared to *S. solfataricus* lactonase (PDB code 4G2D) is loop 8 in the active site, which is shorter in the *V. moutnovskia* lactonase due to a different distribution of proline residues to disrupt the α-helix making it more compact in comparison. Unlike the *S. solfataricus* lactonase where there are 3 proline residues evenly distributed

along the loop, in the *V. moutnovskia* lactonase the prolines are located at the beginning and end of the α-helix. Most of the charged residues are located on the surface of the protein and form a total of 40 salt bridges within each of the monomers. This is a higher amount compared to the *S. solfataricus* lactonase which possesses 36 salt bridges. The side of the protein where the active site is located is negatively charged whereas the opposite side of the protein is more positively charged.
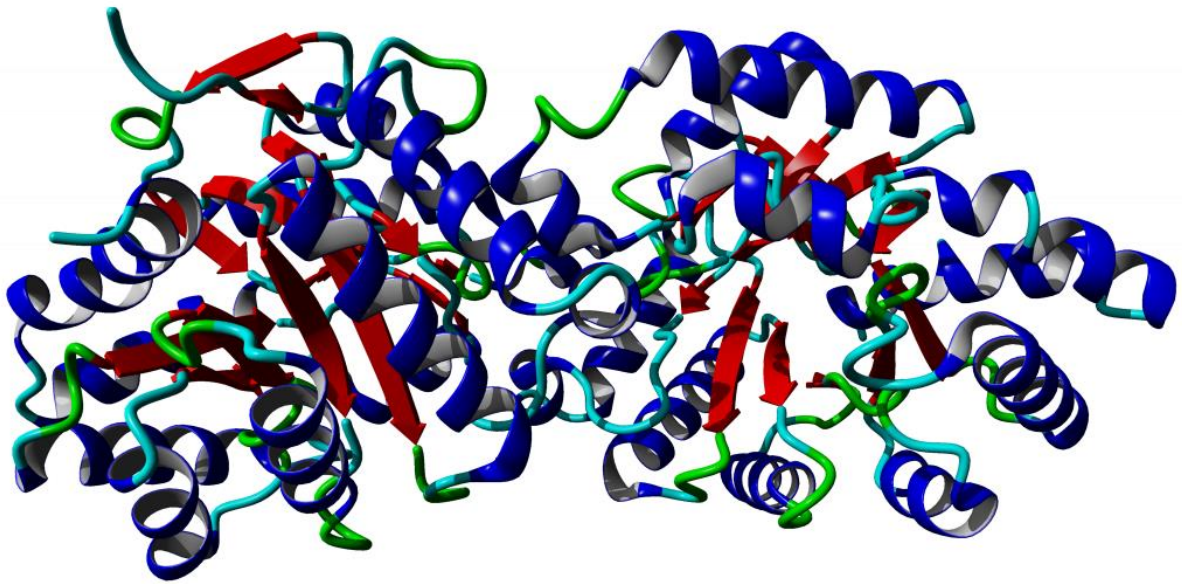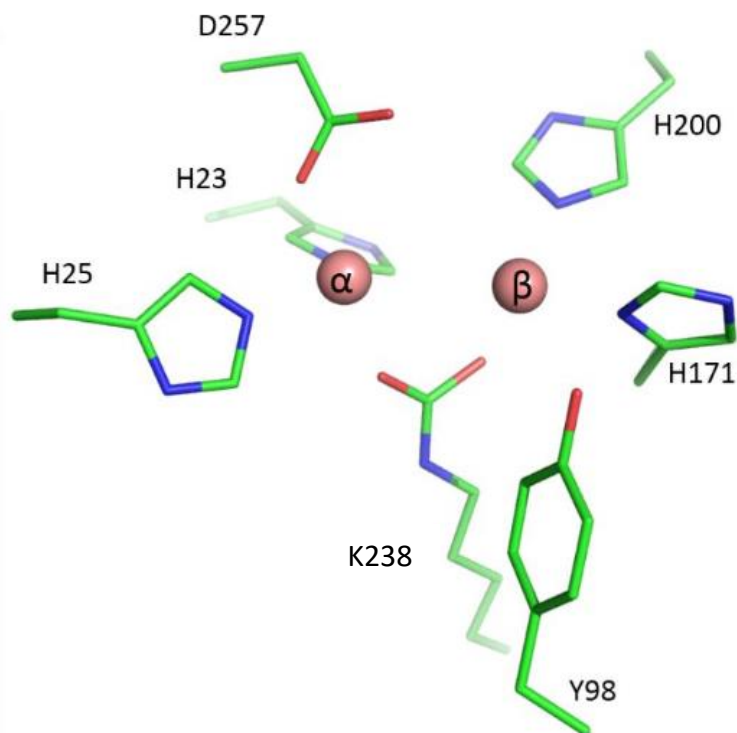
Fig. 3.5. (A) Structure of lactonase from *V. moutnovskia* resolved to 1.8 Å (PDB 4RDZ) viewed using Yasara. Blue are α-helices, red are β-strands, green are turns and cyan are random coils. (B) the active site, modified from (Hiblot *et al.*, 2015). In the active site are 2 catalytic waters (labled α and β) and a methylated lysine. The binding of the different lactones to the active site of the *V. moutnovskia* lactonase was assumed to be by a similar mechanism to that observed for the lactonase from *S. solfataricus*.

Our results on substrate specificity in Table 3 did indicate that the lactonase did have activity towards a broad range of substrates with little regard towards the length of the side chain and that it favoured the D isomers. These results differ from the current proposed models for the *S. solfataricus* enzyme (Ng *et al.*, 2011) where binding to the hydrophobic pocket was required for activation of the enzyme (Elias *et al.*, 2008). Looking at the structures of the enzymes for a possible explanation to this, the main cause is that the *V. moutnovskia* and *S. solfataricus* lactonases have structural differences between their loop 8 regions (Fig. 3.6 B). In the *V. moutnovskia* lactonase this region forms part of an α-helix and is not as flexible as its counterpart in *S. solfataricus* lactonase resulting in the side chains of Val 270 and Val 274 sterically preventing a hydrophobic channel from forming (Fig. 3.6 C). This allows the acyl chains to interact with the surface residues of loop 8 which results in an enzyme with activity to substrates that is independent of the size of the acyl chains.

A

B

Loop 8

C

D
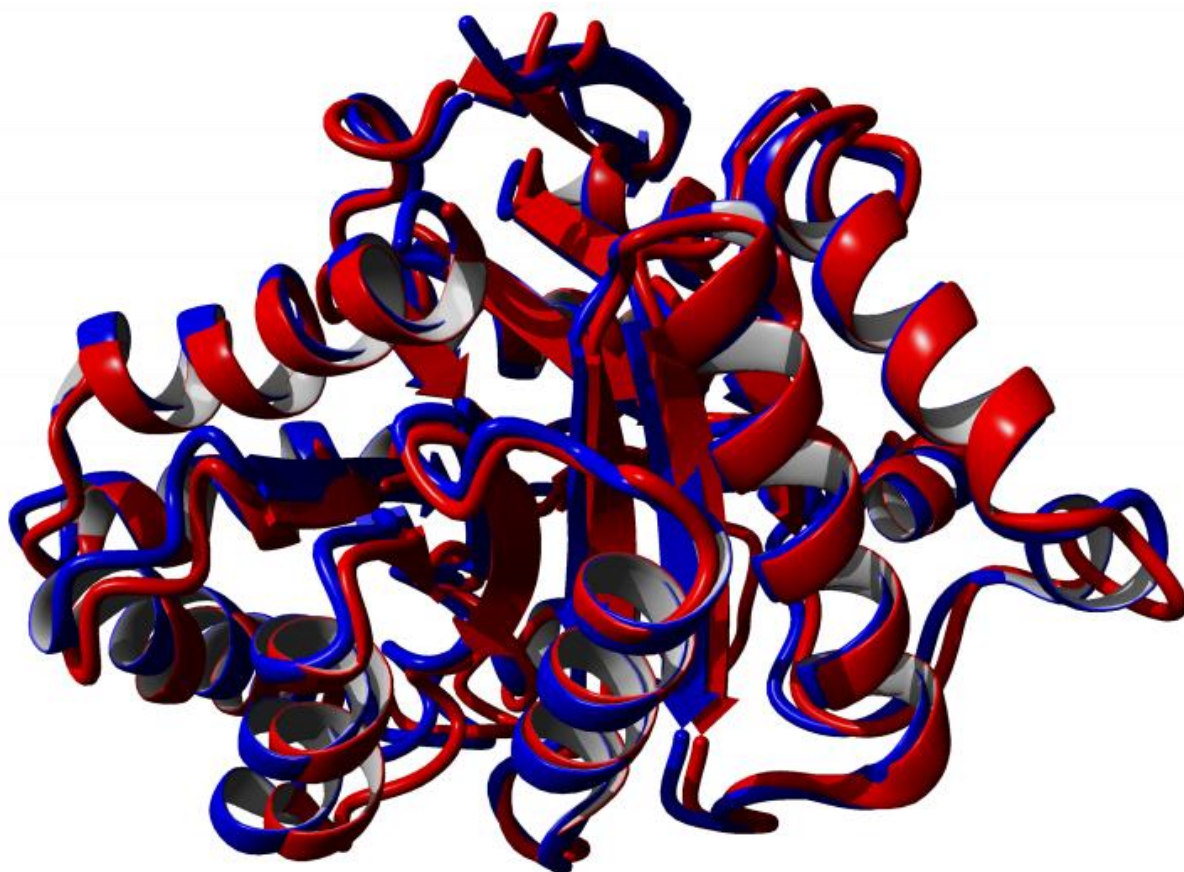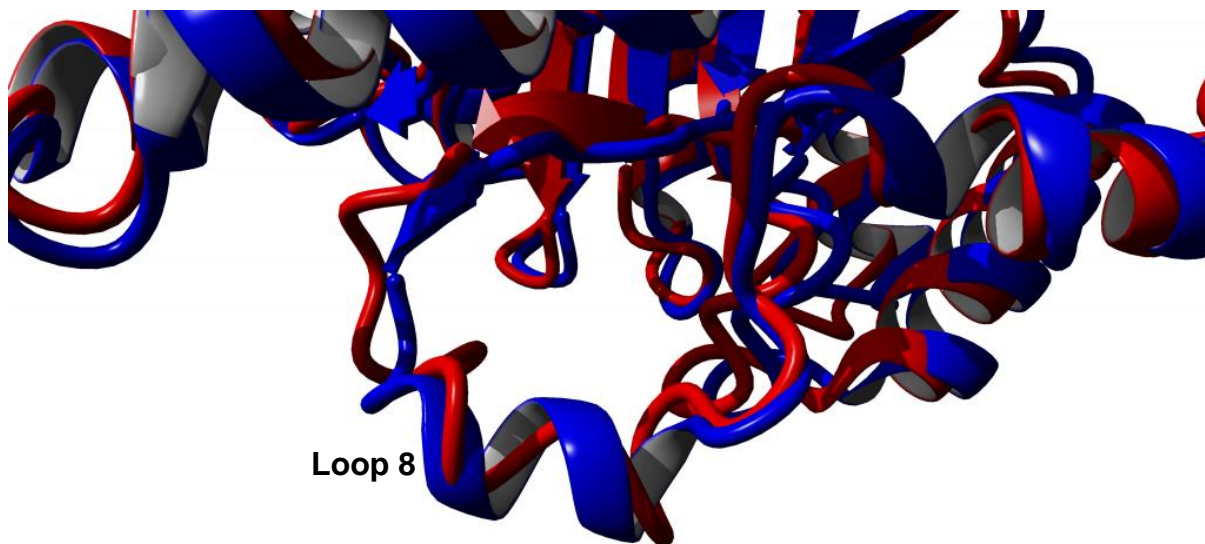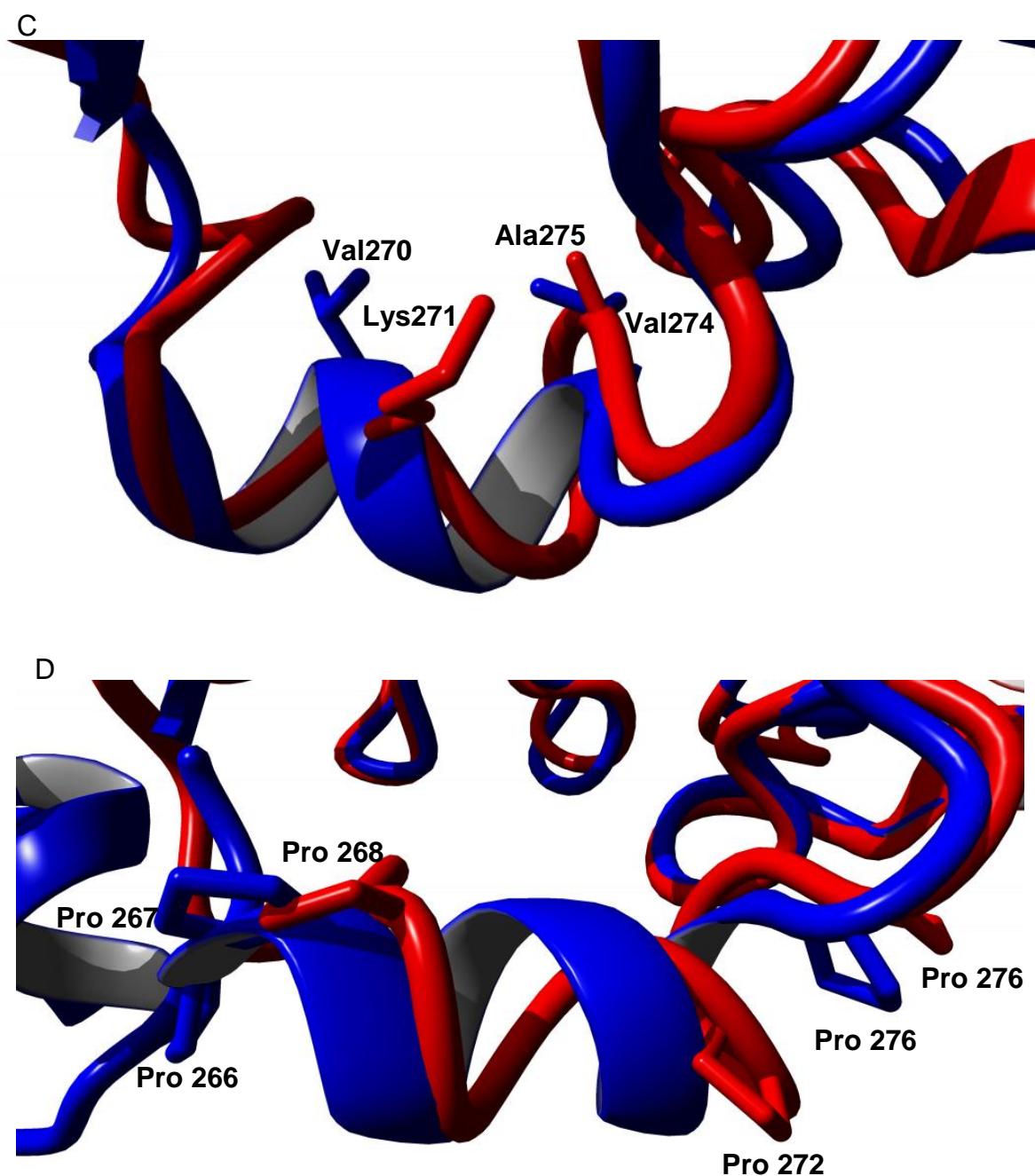
Fig. 3.6. Comparison of the *V. moutnovskia* lactonase (blue) and the *S. solfataricus* lactonase (PDB code 4G2D) (red). (A) Superposition of both lactonases showed an overall mean RMSD calculated by YASARA as 5.2 Å. (B) Loop 8 region, where the *S. solfataricus* lactonase forms a flexible helical loop and the *V. moutnovskia* lactonase forms a more rigid α-helix. (C) Valine residues on the loop 8 region of the *V. moutnovskia* lactonase cause far more steric hindrance to the hydrophobic binding pocket than the *S. solfataricus* lactonase where the loop is more flexible and the side chains of the equivalent amino acids, Lys 271 and Ala 275, either do not point to the pocket as is the case of Lys 271 or are smaller like Ala 275. (D) Proline residues located on the loops, both contain three, in *S. solfataricus* lactonase Pro 268, Pro 272 and Pro 276 are evenly distributed along the loop unlike the *V. moutnovskia* where Pro 266, Pro 267 and Pro 276 are located at the ends of the loop, allowing for a more rigid structure without them disrupting the overall secondary structure of the enzyme.

3.4 Quorum sensing lactonase from the Tomsk metagenome

A potential quorum sensing lactonase was identified within the metagenome "Tomsk", collected by Prof. Elizaveta A. Bonch-Osmolovskaya from hot springs in Russia. The temperature of the hot spring sampled was 46 °C and was around pH 7 (Ferrandi *et al.,* 2015). A BLAST search showed that the amino acid sequence had 25 % identity to a known quorum sensing lactonase in the Protein Data Bank from an uncultured bacterium (PDB code 5EGN). The gene sequence was synthesised and codon optimised for expression in *E. coli* by ATUM (California, USA). The gene was supplied in a pD441-CH vector, part of the Electra expression system developed by ATUM. Expression cell lines were successfully transformed with the recombinant DNA and expression trials conducted. The only conditions the protein was successfully expressed within the soluble fraction were using Arctic Express (DE3) cells grown at 30 °C and then transferred to 12 °C after induction with IPTG. However the amount of soluble protein obtained from this was still very low, with bands on SDS-PAGE gels being barely visible so the quantity of soluble protein needed to be increased to obtain sufficient quantities for crystallisation studies. The reason that this cell line produced better results than others was thought to be attributed to the expression in these cells of additional chaperones, Cpn10 and Cpn60 from the psychrophilic bacterium, *Oleispira antarctica*. So, five BL21 (DE3) cell lines containing different TAKARA plasmids were made and each transformed again with the Tomsk lactonase gene in the pD441-CH vector. The TAKARA plasmids encode for different combinations of chaperones to aid in the correct folding of proteins which should increase the amount of soluble protein produced. These chaperones are: dnaK, dnaJ, grpE, groES, groEL and tig. Out of the 5 TAKARA vectors tested the one that showed the highest amount soluble protein expressed was the pKJE7 plasmid, which contained the chaperones dnaK, dnaJ and grpE.

The Tomsk lactonase was successfully purified using a combination of nickel chromatography to bind the histidine tag and gel filtration chromatography to separate any remaining proteins based on size (Fig 3.7). However the trace from the gel filtration chromatography had three peaks, one occurred within the range of a protein of this size would be expected to elute. The other two were between

100 ml and 125 ml, a range low molecular weight molecules would elute in. Analysis by SDS-PAGE and Western blotting in Fig. 3.7 showed that the first peak contained a band that corresponded to the size of the Tomsk lactonase (29 kDa) and also had a histidine tag. The second peak contained a smaller molecular weight protein that also contained a histidine tag, this was most likely a cleaved product of the lactonase protein, which was possibly due to proteolytic cleavage of a flexible region of the protein in the lysate after sonication. The third peak did not show any band on the SDS gel and was most likely imidazole from the nickel affinity chromatography. This is because the wavelength of 280 nm is used to detect the aromatic amino acids present in proteins, tryptophan, phenylalanine, histidine and tyrosine. Since imidazole is also an aromatic compound it too would show an absorbance in that spectrum and would elute last from the column due to it being far smaller than any protein.

A

B

Lactonase

[71]
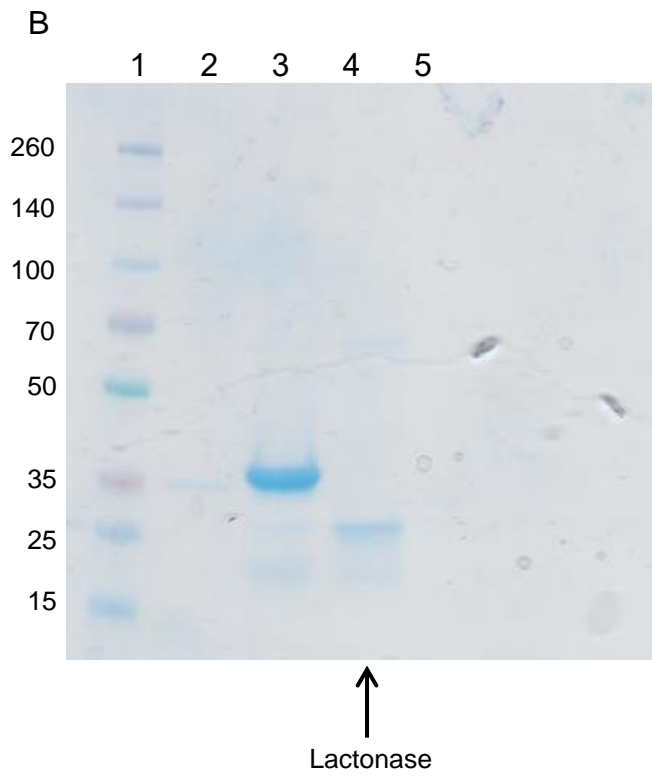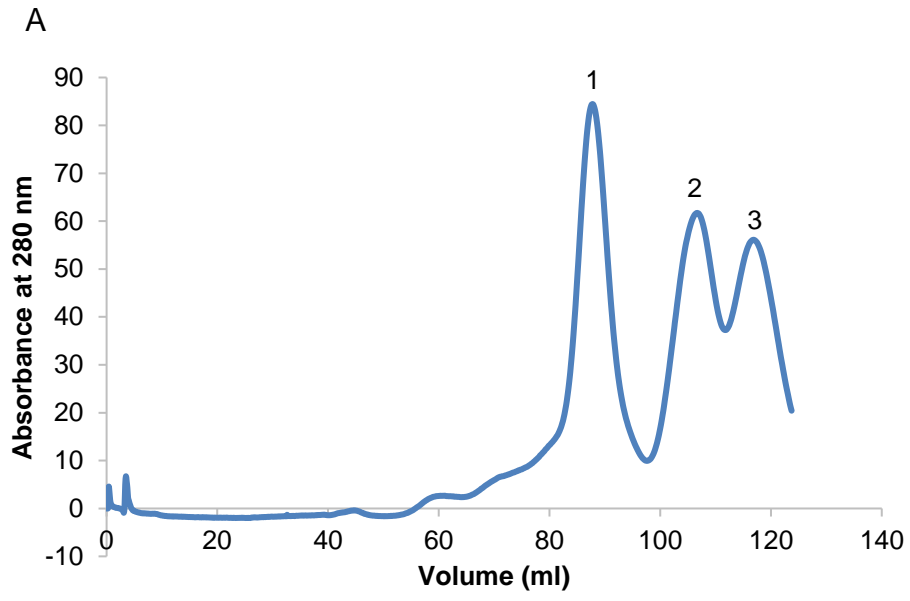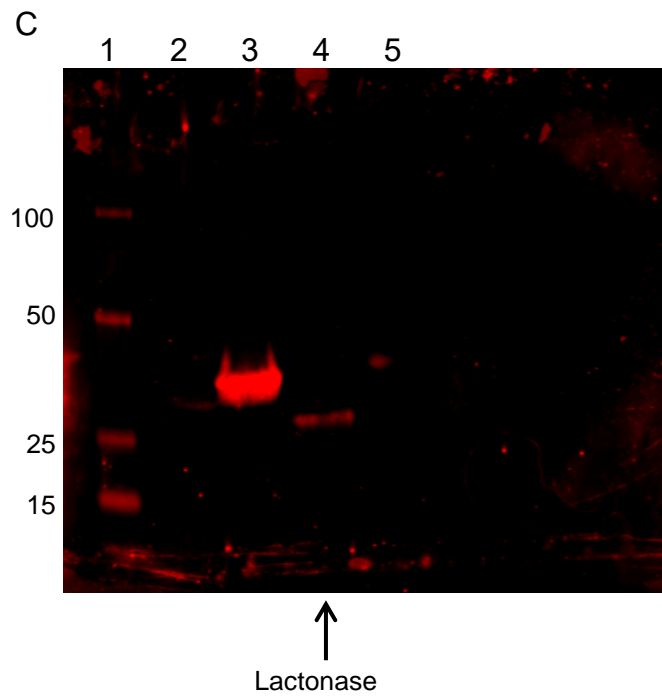
Fig. 3.7. Purification of lactonase from the Tomsk metagenome. A) Gel filtration trace measuring absorbance at 280 nm. B) Duplicate SDS-PAGE of Tomsk lactonase before Western blot: Lane 1. Marker (sizes in kDa); 2. Peak 1 fraction from gel filtration chromatography; 3. Protein with a histidine tag to act as a positive control to ensure the antibodies worked; 4. Peak 2; 5. Peak 3. C) Western blot of Tomsk lactonase, only 4 marker proteins in the ladder are visible as they are the only ones also possessing a histidine tag. A band corresponding to the correct molecular weight of the protein at around 29 kDa was visible indicating the correct protein had been purified.

The total yield of protein was very low and even cell paste from 6 litres of LB media only produced enough protein to make 2 microbatch crystallisation plates at a time at 12 mg ml^{-1} protein concentration, which is 0.2 µg of protein per 1 litre flask, restricting the available options for trials due to the cost in materials and time. Out of the screens tested, some crystals were found in the Stura Footprint screen in 0.2 M MgCl$_2$, 0.1 M HEPES pH 7.5 and 15 % 2-propanol (Fig. 3.8). While it did appear the crystals were protein they diffracted so poorly no data were able to be collected. Optimisation trials of this condition where the concentrations of the protein and the 2-propanol were varied on either the Y or X axis in the microbatch plate did not yield any further crystals.
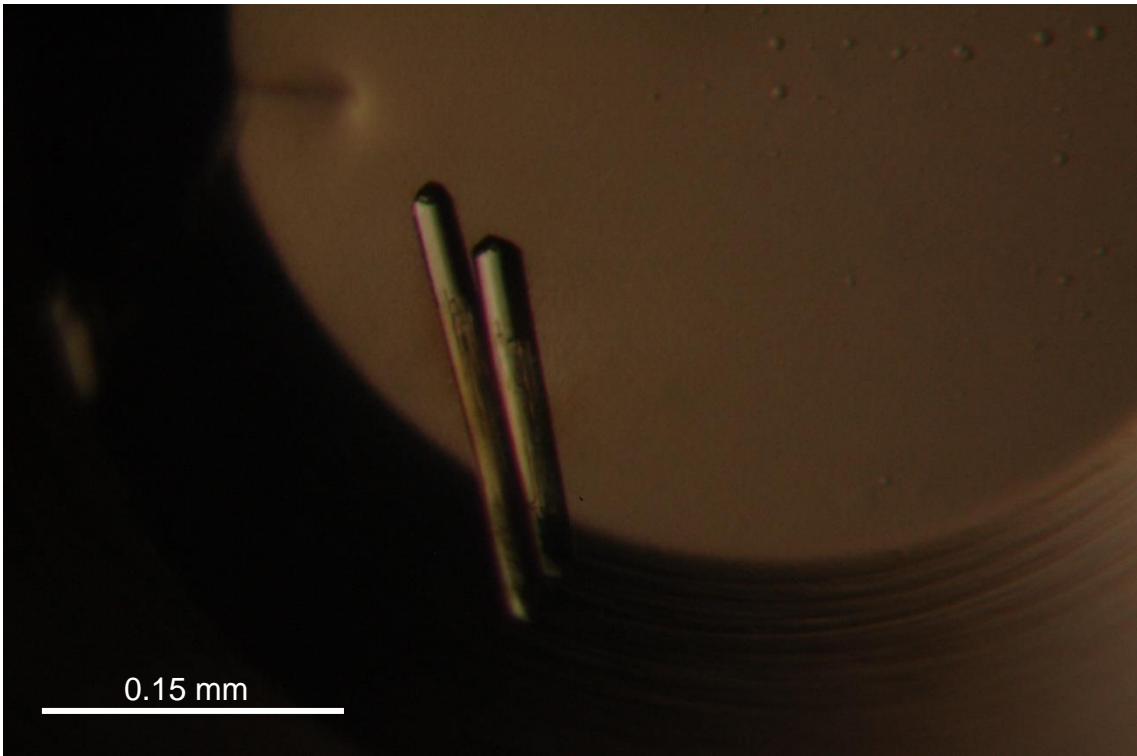
Fig. 3.8. Crystals formed by Tomsk lactonase at, 0.2 M $MgCl_2$, 0.1 M HEPES pH 7.5 and 15 % 2-propanol.

Attempts to obtain further crystals for X-ray diffraction were halted when there were no further crystals obtained from the screens and attempts to improve the yield of soluble protein to enable a broader range of trials to be conducted failed. So homology modelling was done with the enzyme as described in the material and methods. The best structure obtained from this method appeared to be one of the 5 models generated by I-TASSER. The analysis of the structure gave a QMEAN of −4.74 and superposition gave a RMSD of 2.6 Å with the closest homologue of the same class of enzyme, an N-Acyl homoserine lactonase from an unknown bacterium (PDB code 5EGN) identified from a BLAST search. This value for the RMSD is relatively low indicating the structures are very similar and a good value to obtain considering there is only 25 % sequence identity between the two structures. The Ramachandran plot (Fig. 3.9 A) showed 81.4 % of all amino acids in a favoured conformation, 12.8 % in an allowed conformation, and the remaining 5.8 % as outliers. These outliers were identified as: Ala 8, in the middle of a loop region; Pro 9, on the same loop; Ser 79, the last amino acid of a loop before an α-helix, which is not unusual to be an outlier due to the transition in secondary structure; Leu 130, most likely because the possible alternative conformation would cause a steric clash with an adjacent proline on the same loop; Pro 131, in the middle of a loop; Asp 137, which is the last amino acid on a loop before an α-helix and an alternative conformation would cause a steric clash with the helix; Ala 192, located on a helical loop; Pro 197, on the same loop as Ala 192; Pro 200, at the end of the helical loop; Arg 223, on a loop facing the solvent.

The structure was a standard α/β hydrolase fold, with a β-sheet consisting of eight β-strands connected by six α-helices (Fig. 3.8 B). Two active site residues important for activity were identified in the structure of the closest homologue, PDB 5EGN, a quorum sensing lactonase from an uncultured bacterium. These residues were Ser 93 and His 242 (Liu *et al.*, 2016) and the equivalent residues in the Tomsk lactonase model were identified by sequence alignment with Clustal Omega (Sievers *et al.*, 2011) to be Ser 203 and His 247. Comparison of the active site with the closest homologue (Fig. 3.9 C) revealed the positions of the Tomsk lactonase model did not align exactly and the position differed by 2.5 Å for Ser 203 and 3.1 Å for His 247. This could infer a difference in substrate specificity between the Tomsk lactonase and the 5EGN lactonase, however because it is

just a homology model this is hard to say with any certainty for such a small difference in the positions. Since there is no literature for the other structure we do not know what its substrate specificity is and as such making a comparison would not be possible unless the 5EGN lactonase was obtained, purified and had activity assays conducted in parallel with the Tomsk lactonase. The side chains of the arginines and lysines were examined for the presence of any glutamic acid or aspartic acid side chains within 3.2 Å of each other using COOT (Emsley and Cowtan, 2004) to identify any possible salt bridges that may be formed. A total of 15 salt bridges were identified within the structure of the Tomsk model. This is a significant increase over the mesophilic lactonase which had 7. This indicates a high probability that salt bridges provide an increase in the stability of the Tomsk lactonase to allow the enzyme to remain active at high temperatures.

Fig. 3.9 D shows the surface of the model represented by white for hydrophobic regions, red for negatively charged and blue for positively charged regions. The active site is observed to be open and exposed to the solvent, as seen by the cavity indicated in the image. There is no indication of a hydrophobic binding channel as seen in the *S. islandicus* lactonase. Most of its active site consists of α-helices rather than loops as seen in the *S. islandicus* lactonase, giving a more compact and rigid structure. This could suggest that like the *V. moutnovskia* lactonase its activity is not restricted to lactones with long acyl chains.

A

B

C

D

Active site
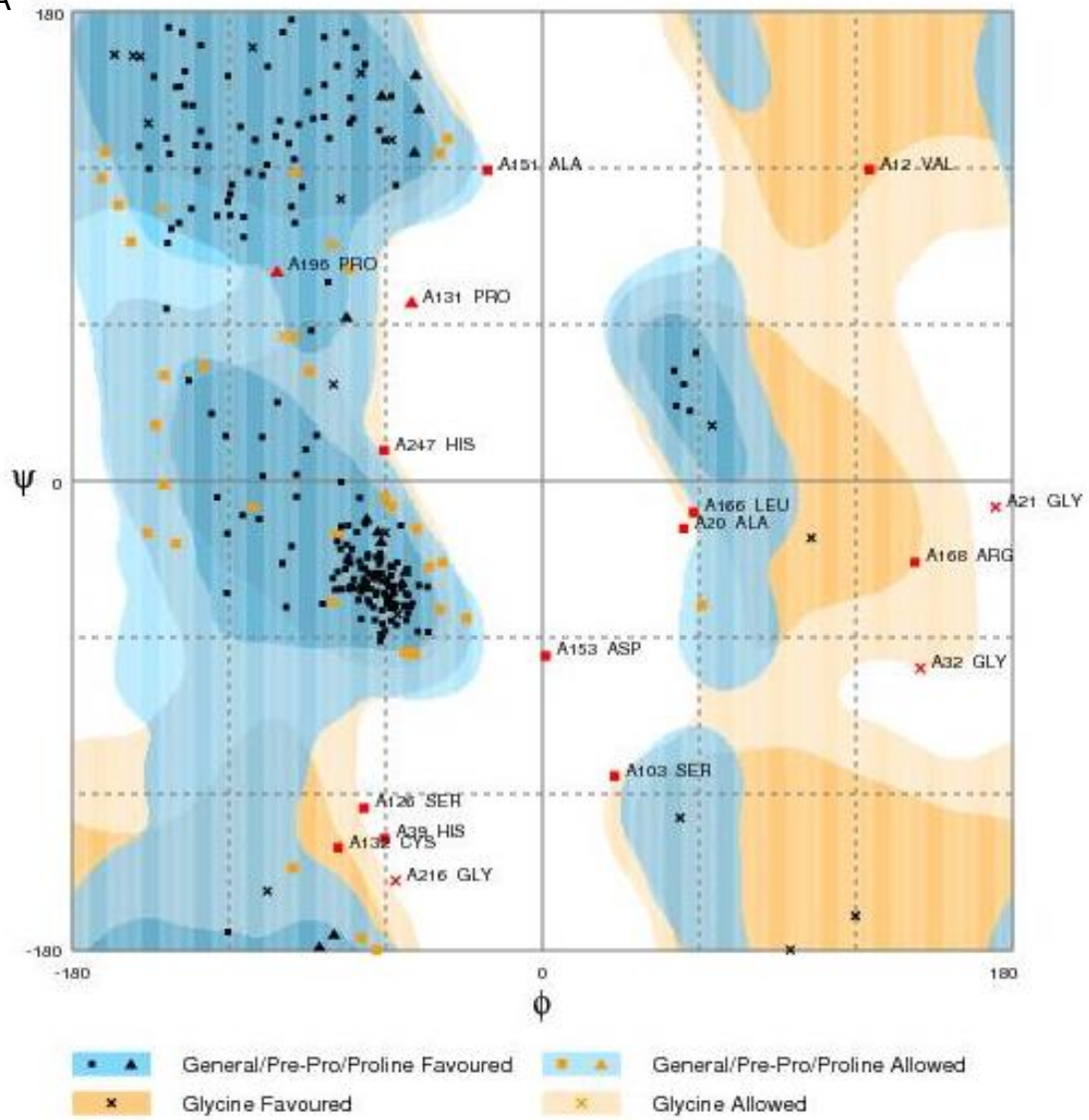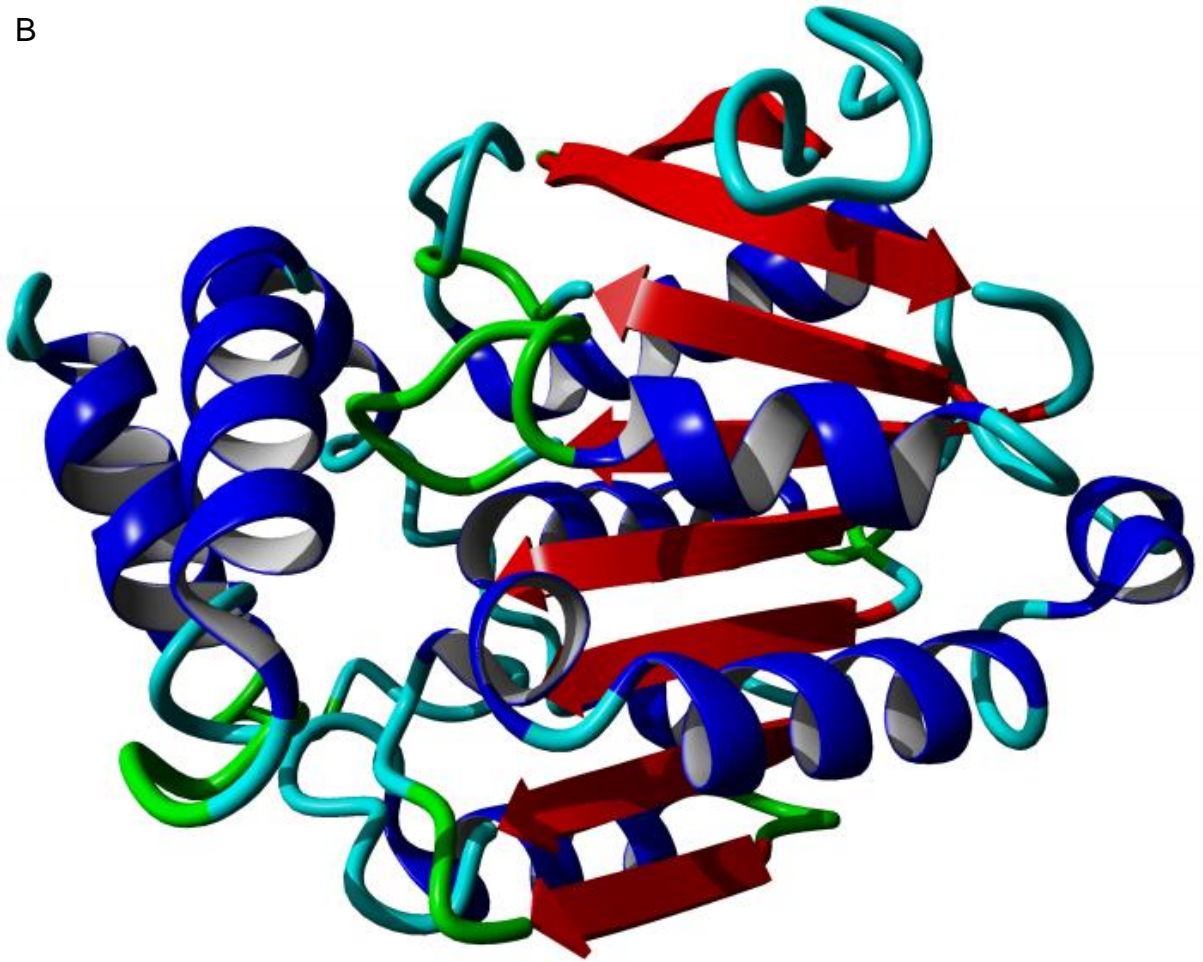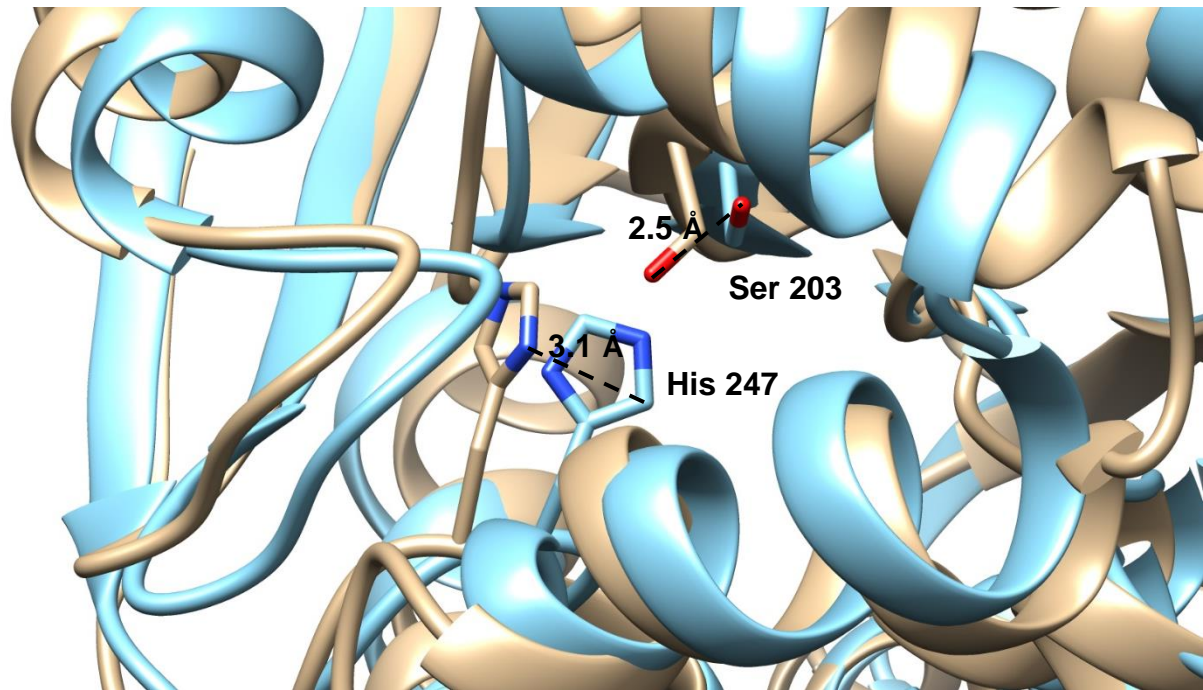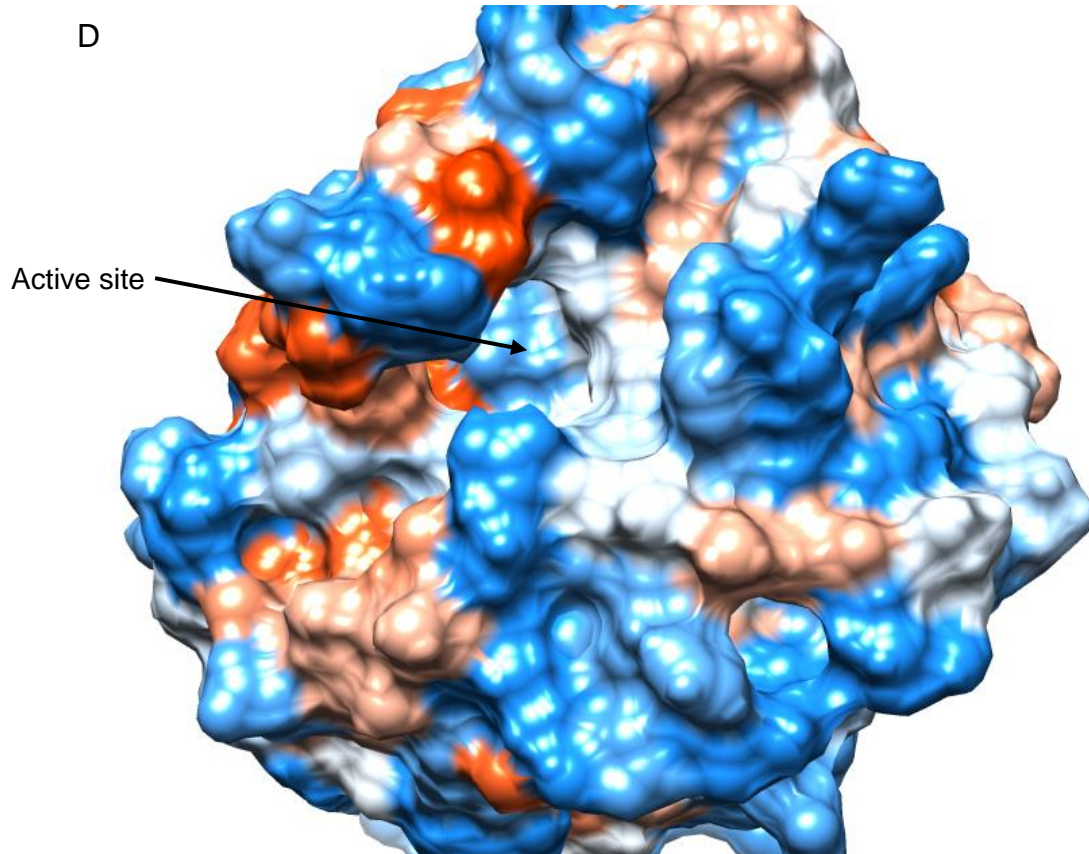
Fig. 3.9. Homology modelling of the Tomsk lactonase. (A) Ramachandran plot of the model showing 81.4 % of residues in a favoured conformation, 12.8 % in an allowed conformation, and 5.8 % as outliers highlighted in red. (B) Structure of the model generated by I-TASSER (Zhang, 2008)Maduro, 2014). Blue are α-helices, red are β-strands, green are turns and cyan are random coils. The structure has a standard α/β hydrolase fold with a core of β-sheet consisting of eight β-strands connected by 6 α-helices. (C) Active site histidine and serine residues compared between the two superposed structures, Tomsk lactonase model in blue, 5EGN structure in brown. A shift of 3.1 Å was seen in the His 247 and a shift of 2.5 Å in the Ser 103 from the 5EGN structures residues. (D) Surface representation of the Tomsk lactonase, white is uncharged, red is negatively charged and blue is positively charged. The arrow indicates the location of the active site.

Looking at the composition of amino acids with another quorum sensing lactonase from the mesophile *Bacillus thuringiensis* (PDB code 2A7M) in Table 4, there is a noticeable difference between the percentages of certain amino acids. Most notably the Tomsk lactonase consists of 15.9 % alanine, compared to the 2A7M of only 4.0 % this means there is a difference of 11.9 % overall. Although the weakest hydrophobic amino acid, such a greater amount in a thermophilic protein would implicate that there is an increased amount of hydrophobic interactions stabilising the enzyme. Most of the substitutions for alanine occur in α-helices (Fig. 3.10) as these are the most suitable amino acids for stabilising helices due to having small hydrophobic side chains. There are also 0.8 % more proline residues present, slightly more but not enough to be a major contributing factor. There was 7.7 % more arginine residues indicating there to be more ionic interactions as well, most likely in regions on the protein surface. Most of the arginine are present in α-helices and the loop regions which are very exposed to the solvent in the structure so arginine would help stabilise them through interacting with water molecules. There were 4.9 % fewer lysine residues, several were seen to be substituted for arginine which is common for thermophilic enzymes as arginine is able to form more interactions due to having two charged groups on its side chain (Broadwater *et al.*, 1994). Being capable of forming electrostatic interactions in three directions means arginine can improve the overall stability of the enzyme to a greater extent than lysine. Also it has been theorised that the higher pKa of arginine may also generate more stable ionic interactions than lysine (Sokalingam *et al.*, 2012). The rest of the lysines were mostly substituted for hydrophobic residues in loop regions.

Visual inspection of the structure with COOT revealed that there was a slight increase in the number of hydrophobic interactions seen in the structure in comparison to the 5EGN structure it was modelled after. Overall it appeared that any extra thermal stability the enzyme possessed would be from the additional salt bridges rather than hydrophobic interactions. As the enzyme was acquired from a hot spring with a temperature of 46 °C it would not display many obvious adaptations over mesophilic proteins as there is a less than 10 °C difference between the environments.

Table 4. Percentage composition of amino acids in the Tomsk lactonase and 2A7M.

Amino acids	5EGN	Tomsk
Ala	4.0 %	15.9 %
Arg	2.8 %	10.5 %
Asn	4.0 %	1.4 %
Asp	4.4 %	8.3 %
Cys	1.6 %	2.2 %
Gln	2.4 %	1.4 %
Glu	10.8 %	3.3 %
Gly	7.2 %	9.1 %
His	3.6 %	3.3 %
Ile	7.6 %	2.2 %
Leu	10.4 %	12.3 %
Lys	5.6 %	0.7 %
Met	2.0 %	2.9 %
Phe	4.8 %	2.9 %
Pro	6.4 %	7.2 %
Ser	5.6 %	4.3 %
Thr	5.2 %	2.9 %
Trp	0.4 %	1.8 %
Tyr	4.4 %	0.7 %
Val	6.8 %	6.5 %

```
2A7M

                1        10        20        30        40            50
2A7M    ...........MTVKKLYFIPAGRCMLDHSSVNSALTPGKLINLPVWCYLLETEEGPILV............DTG
tomsk   MLAARPSAPDGVWHGEVSVAGGKVSFAMAGDG......LPLILLHGWTLDHRMWRPQIGQLSRDFLLVMPDRRGCASTA
tomsk


2A7M
                                           60        70        80        90
2A7M    MPES..................................AVNNEGLFNGTFVEGQILPKMTEEDRIVNILKRVGYEPDDLL
tomsk   PPDLSREAEDVIAIADFLGFERFGLLGLSQGAVVALDVARKFSSRLTGLVVSGAPLPCLVERDEAIPLDRYRAMV.....
tomsk


2A7M
        100       110       120       130       140       150       160       170
2A7M    YIISSHLHFDHAGGNGAFTNTPIIVQRTEYEAALHREEYMKECILPHLNYKIIEGDYEVVPGVQLLYTPGHSPGHQSLFI
tomsk   ......AAGDVAGMRRDWARHPLMRT.......................................HDPDARTL..
tomsk


2A7M
        180       190       200       210       220       230       240
2A7M    ETEQSGSVLLTIDASYTKENFEDEVPFAGFDPELALSSIKRLKEVVKKEKPIIFF..GHDIEQEKSCRV...........
tomsk   ........AAAMLADYDGRDLAAVSEPPGLPRE....VLSH......LAVPVLALAGEHDTPWRRACAAALADCAPRGRH
tomsk


2A7M
                          250
2A7M    ................F......PEYI............
tomsk   ALIGRAGHLANCDNPQDFNALAGFLRTCADPRTRANR
tomsk
```

Fig. 3.10. Secondary structure and amino acid sequence alignment between the Tomsk lactonase and the 2A7M lactonase made using the ESPript 3 sequence and structure alignment server (Robert and Gouet, 2014). Residues highlighted in red are conserved, residues in blue boxes are conserved substitutions.

3.5 Summary

Two quorum-sensing lactonases from thermophilic hot spring microorganisms were successfully overexpressed and purified. The activity of a lactonase from the thermoacidophilic anaerobic crenarchaeon *V. moutnovskia* was tested against a range of lactone substrates and it shows a preference for the D stereoisomer of γ-valerolactone and γ-caprolactone. Crystallisation trials produced crystals for both of the enzymes. The crystal structure of the *V. moutnovskia* enzyme was obtained by another group and has allowed rationalisation of the substrate specificity that we have observed. The lactonase from the *V. moutnovskia* enzyme showed the absence of a hydrophobic binding pocket seen in other quorum-sensing lactonases. This was caused by substitution of proline residues in the nearby loop region that resulted in the formation of an α-helix that blocked the binding pocket. This would explain why the enzyme is active with short acyl-chain lactones if binding of a sufficiently long acyl chain to a hydrophobic pocket is no longer necessary for activity.

Homology models were obtained for the quorum-sensing lactonase from the Tomsk metagenome which showed the enzyme did not contain many more hydrophobic interactions compared to a mesophilic homologue. Instead there appeared to be more salt bridges in the enzyme which likely provide the extra thermal stability. There was also an absence of a hydrophobic binding pocket anywhere near the active site which could indicate similar activity to the *V. moutnovskia* lactonase where the enzyme can use lactones with short acyl chains as substrates.

The next step in the investigation of lactonases was to study another class of this enzyme, specifically a gluconolactonase.

4. Gluconolactonase

4.1 Introduction

A gluconolactonase (lac11) was identified from a *Planctomyces* genome isolated by the group of Prof. Elizaveta A. Bonch-Osmolovskaya, Russian Academy of Science.

4.2 Materials and methods

All methods were the same as described in the general materials and methods except for the following:

4.2.1. Sample concentration

Samples were found to bind to the membranes of the standard Vivaspin 20 concentrators causing them to precipitate so concentrators containing different membranes were trialled to see which would avoid this issue. The one selected was the Vivaspin Turbo 15 with a 10 kDa cut off (Sartorius, Germany).

4.2.2. Analytical size exclusion chromatography

A Superose 12 10/300 GL column (GE Healthcare, USA) was used with the standard buffer for size exclusion chromatography. The sample was added at a volume of 500 µl at 2 mg ml^{-1} and run at 1 ml min^{-1}.

4.2.3 Crystal seeding

Seed stocks were made with known or suspected protein crystals. Crystals grown from 90 mM NaNO$_3$; 100 mM Tris-HCl pH 8.5; and 30 % PEG 550 MME, were crushed with a glass rod in the well and transferred to an Eppendorf tube on ice with 50 µl of the stock solution which were then subjected to vortex with a plastic bead in the tube and frozen at −80 °C. Each well in the microbatch plate contained 0.3 µl of the protein solution, 0.2 µl of the screen solution, and 0.1 µl of the seed stock.

4.2.4 Microscale thermophoresis

Binding affinity between the lac11 enzyme and glucose was calculated using a NanoTemper (NanoTemper Technologies, Germany). 10 µl of 10 nmol of the protein, diluted with microscale thermophoresis optimized buffer containing: 50 mM Tris-HCl pH 7.6, 150 mM NaCl, 10 mM $MgCl_2$, 0.05 % Tween-20, was incubated with 10 µl 20 nM of the fluorescent tag NT-647 and a range of glucose concentrations and loaded into hydrophilic glass capillaries. The NT-647 tag contains 3 Ni^{2+} metal ions which bind to the histidine of His-tagged proteins and fluoresces at 633 nm. These samples were then placed in the Nano Temper machine and the change in emission from the fluorescent tag was monitored while a laser heated a single point in the capillary to monitor diffusion of the tagged protein.

4.2.5 Enzyme solvent stability

The enzyme was incubated at 0.4 mg ml^{-1} in solvent concentrations between 10 % and 60 % v/v at room temperature in HEPES gel filtration buffer for 20 min before being tested in the pH based assay as described in the general materials and methods to see if activity had been retained. The solvents used were acetonitrile, methanol, ethanol, DMSO and isopropanol. The solvent/protein mixture was then diluted 1:200 after the incubation process to minimise the chances of it interfering with the assay.

4.2.6 Enzyme activity retention after heat exposure

The enzyme was incubated at 0.4 mg ml^{-1} at temperatures up to 90 °C for 20 min. The enzyme was cooled on ice for 5 min before assaying at room temperature. A control was carried out using the enzyme left in standard GF buffer (50 mM HEPES pH 7.5, 100 mM $NaCl_2$) at room temperature for the same time which was used as a 100 % activity.

4.2.7 Differential scanning fluorimetry

Experiments were carried out as described in the general materials and methods section 2.5.2. Three different buffers were tested with lac11, all at pH 7.5: Tris-HCl, HEPES and MOPS all at 50 mM. Bis-Tris propane, at the same concentration and pH as the previous buffers, was also tested with the *V. moutnovskia* lactonase. The addition of potential stabilisers, citrate, malonate and malic acid, all at 1 mM, were tested with lac11 with Tris-HCl buffer. Divalent ions that could be potential cofactors were tested with lac11, these were; calcium chloride, magnesium chloride, zinc chloride, manganese chloride and cobalt chloride at concentrations of 1 mM.

4.2.8 Kinetics assay

A titration using acetic acid between 0 and 0.35 mM with a total volume of 200 µl was carried out to determine the linear part of the standard curve. A 5 mM stock of acetic acid was added at 1 µl increments, increasing the total concentration of acid by 25 µM each time. Assays were performed as described in the general materials and methods with the addition of acetic acid to a final concentration of 0.26 mM with readings taken every minute for 20 min.

4.2.9 Lysine methylation

The methylation of surface lysine residues has been shown to improve protein crystallisation by altering the interaction of protein molecules in the potential crystal packing (Walter *et al.*, 2006). 40 µl of 1 M formaldehyde and 20 µl of borane dimethylamine complex (97 % v/v) per 1 ml were added to the solution containing the protein which was kept at concentrations lower than 1 mg ml^{-1}. This was gently mixed at 4 °C for 2 hours. This procedure was done a second time to the solution and after the second 2 hour mixing step a final addition of 10 µl borane dimethylamine complex (97 % v/v) per 1ml of solution was added and incubated overnight. Samples were subjected to gel filtration chromatography to remove any denatured or aggregated protein and to remove the formaldehyde and borane dimethylamine complex from the solution which would influence protein crystallisation.

4.3 Results and discussion

Genomic DNA from the *Planctomyces* genome that was sequenced was provided by the group of Prof. Elizaveta A. Bonch-Osmolovskaya at the Russian Academy of Science, Moscow, Russia. The gene encoding the gluconolactonase was amplified through PCR and inserted into a pLATE31 vector. The recombinant DNA was then successfully inserted into and amplified in *E. coli* XL10-Gold ultracompetent cells. Expression trials were conducted with the lac11 to determine which expression cell line and conditions were required to optimally over-express the protein within the soluble fraction. Through analysis by SDS-PAGE the protein was identified in the soluble fraction of *E.coli* RIPL cells grown at 37 °C after induction with 1 mM IPTG. The enzyme was successfully purified and the trace from gel filtration chromatography is shown in Fig. 4.1 and the corresponding SDS-PAGE in Fig. 4.2.
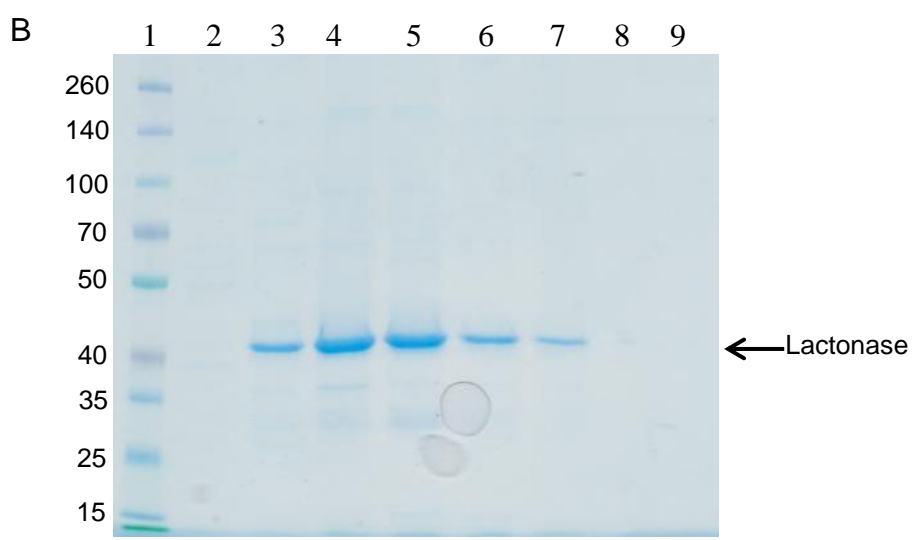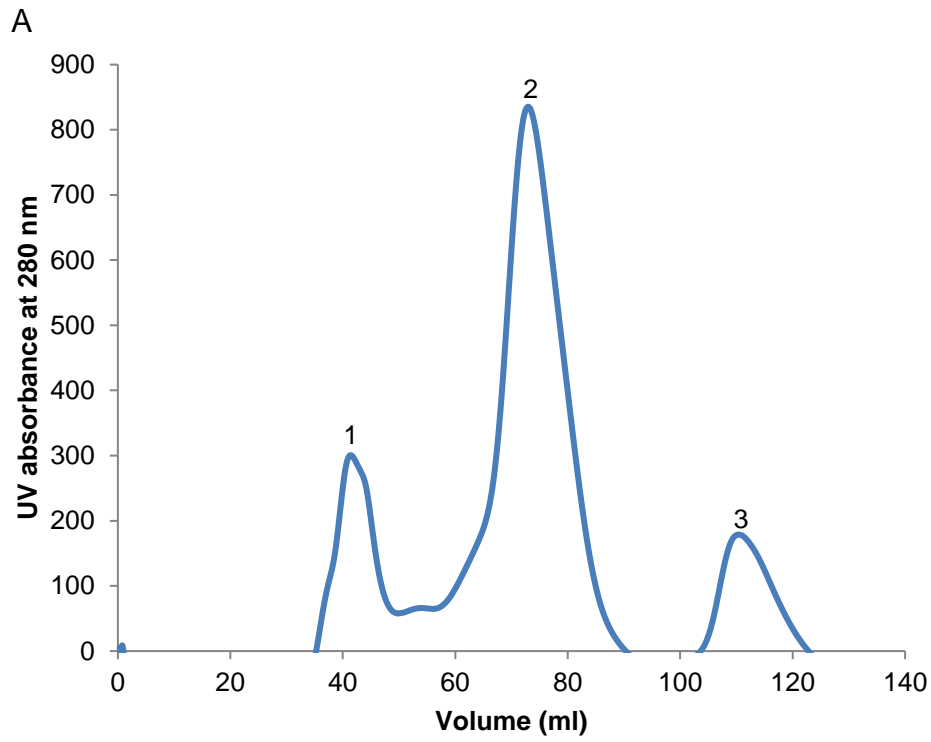
Fig 4.1. Gel filtration trace (A) and analysis by SDS-PAGE of gel filtration (B) of lac11:
Lane 1. Marker (sizes in kDa); 2. Peak 1; lane 3-8. Peak 2; lane 9. Peak 3.

A selection of newly synthesised sugar lactone substrates, as shown in Fig. 4.2, were provided by our industrial collaborator, Dr Roland Wohlgemuth (Sigma Aldrich/Merck). Since the co-factor for this enzyme was unknown, three divalent ions were tested to see if their presence resulted in a change of activity as divalent ions commonly seen in other gluconolactonases are integral to the proposed mechanism of hydrolysis by these lactonases (Momb *et al.*, 2008). Examples are: a mouse gluconolactonase PDB 4GN7, which contains a calcium ion; a human gluconolactonase PDB 4GNB which contains a calcium ion; and a gluconolactonase from *Trypanosoma brucei* PDB 3EB9 which contains a zinc ion. Each assay was duplicated with 2 mM of one of the following added to try and determine the metal cofactor: $MgCl_2$, $ZnCl_2$ and $MnCl_2$. The results (Table 5) show activity towards all three of the substrates tested regardless of which divalent ion was present and in the absence of the metal ions. This confirmed the enzyme to be a gluconolactonase but did not provide information of any required co-factor. Most likely the divalent cofactor is strongly associated with the active site of the enzyme and not easily displaced during the purification process. Overall the enzyme displayed little preference towards any particular substrate as all but one of these substrates was shown to be hydrolysed by the lac11. The only substrate where there is little or no activity was L-rhamnono-1,4-lactone. Interestingly the enzyme shows activity towards L-rhamnono-1,5-lactone as well as the other lactones which include members with a 1,4 structure as well. D-altrono-1,4-lactone is a chiral isomer of L-rhamnono-1,4-lactone where the only difference is the chirality of the R group of the side chain but has activity. L-rhamnono-1,4-lactone is also the only substrate tested with a different chirality of the R group among all the 1,4 lactones. The chirality of the hydroxides on the ring has no apparent effect on activity. So the hydrolysis of 1,4 lactones appears to be dependent on the chirality of the R group bonded to the C-4 in the lactone ring.
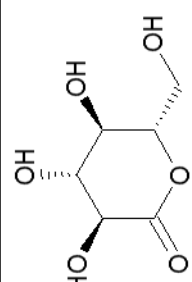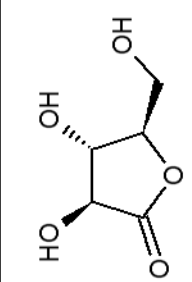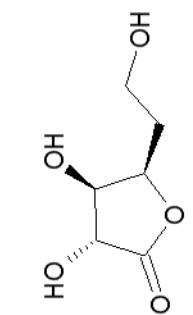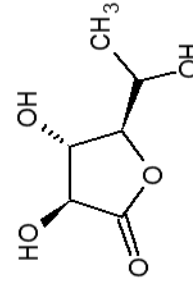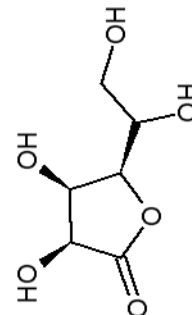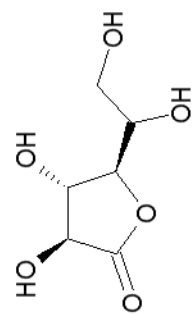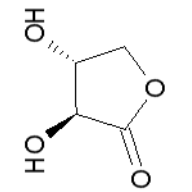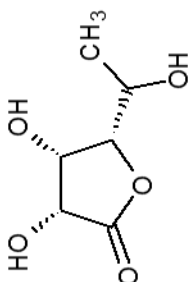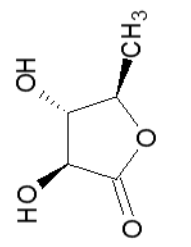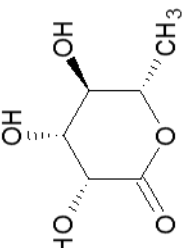
Fig. 4.2. Illustration of gluconolactone substrates used in assay

D-xylono-1,4-lactone

L-fucono-1,4-lactone

D-altrono-1,4-lactone

L-rhamnono-1,4-lactone

D-rhamnono-1,5-lactone

D-glucono-1,5-lactone

D-glucuronic acid-1,4-lactone

L-gulonic acid-1,4-lactone

D-threono-1,4-lactone

D-arabino-1,4-lactone

Table 5. The activity of lac11 towards the different substrates.

Gluconolactone	Activity[1]
D-glucono-1,5-lactone	+
D-glucuronic acid-1,4-lactone	++
L-gulonic acid-1,4-lactone	++
D-threono-1,4-lactone	+
D-arabino-1,4-lactone	++
L-arabino-1,4-lactone	++
D-xylano-1,4-lactone	++
L-fucono-1,4-lactone	++
D-altrono-1,4-lactone	++
L-rhamnono-1,4-lactone	−
D-rhamnono-1,5-lactone	+

(1) ++ represents ≥0.15 nmoles of H^+ ions produced above the value obtained for the blank. + represents 0.05-0.15 nmoles H^+ ions produced above the blank. – represents no activity.

To examine the kinetics of the enzyme, pH based assays were carried out in triplicate and the results are shown in Fig. 4.3 after determining the titration curve to bring the assays as close to the point of colour change as possible. Analysis of the results showed that the kinetics were more likely allosteric-sigmoidal rather than standard Michaelis-Menten. For the kinetics with glucuronic acid-γ-lactone the data had an R^2 of 0.9621 and a Sy.x of 0.0554 when a Michaelis-Menten curve was applied. With an allosteric-sigmoidal curve it had an R^2 of 0.9897 and a Sy.x of 0.02961, indicating that this would be a better fit to the data. The same trend was seen with the other kinetic studies performed, with higher R^2 and Sy.x values seen for allosteric-sigmoidal curves indicating that this is the most probable model for this enzyme. The D-glucuronic acid-γ-lactone had a V_{max} of 0.8092 μmole ml^{-1} min^{-1} and a K_{prime} of 2.009 mM. This was the lowest K_{prime} of those tested indicating this to be the substrate the enzyme had the highest activity of the three. The gulonic acid-γ-lactone had a V_{max} of 0.5263 μmole ml^{-1} min^{-1} and a K_{prime} of 3.622 mM. The rhamnono-1,5-lactone had a V_{max} of 1.183 μmole/ml/min and a K_{prime} of 4.463 mM.
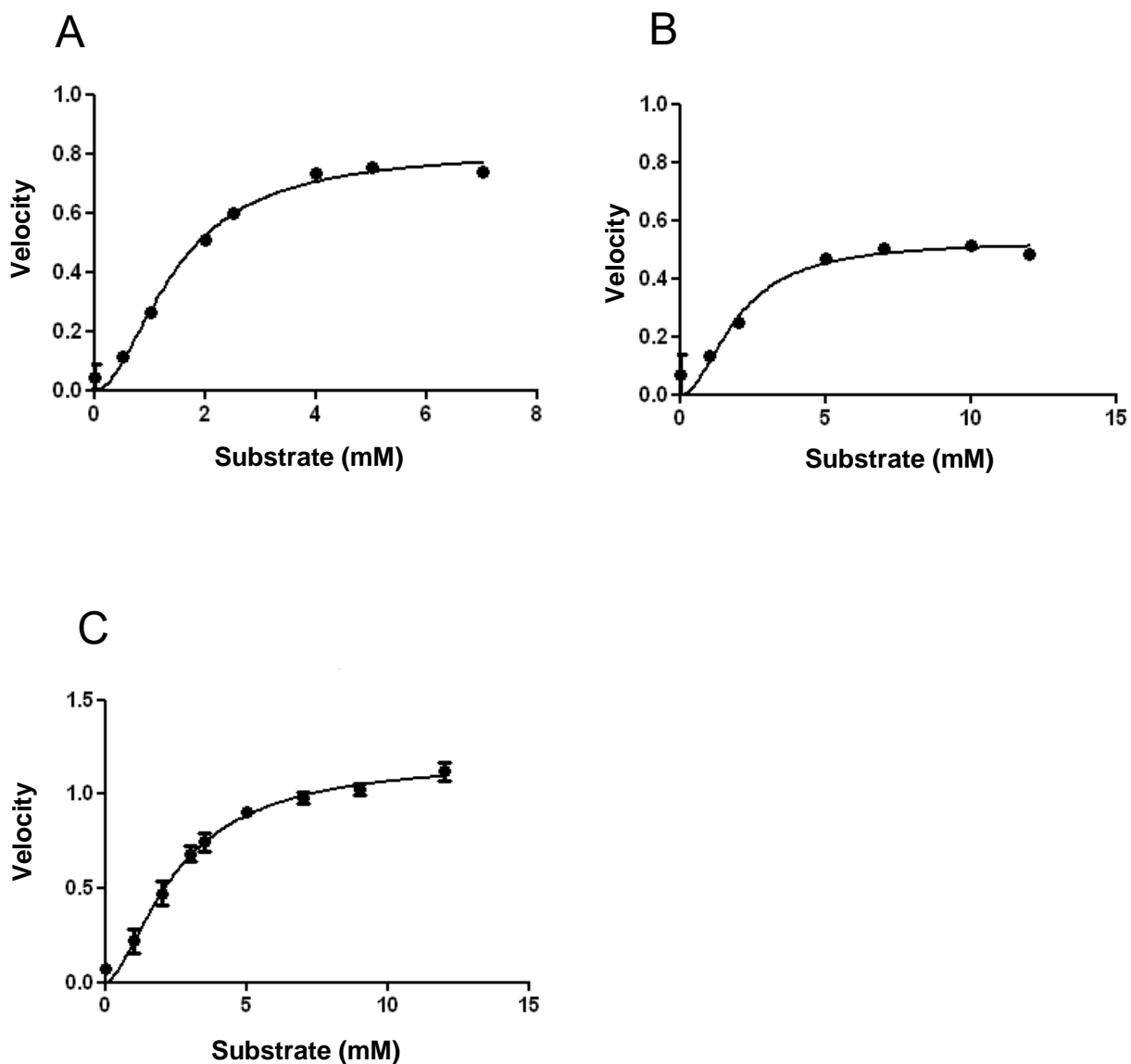
Fig. 4.3. Kinetics graphs of lac11. Kinetics for: (A) D-glucuronic acid-γ-lactone, with a V_{max} of 0.8092 μmole ml^{-1} min^{-1} and a K_{prime} of 2.009 mM; (B) L-gulonic acid-γ-lactone with a V_{max} of 0.5263 μmole ml^{-1} min^{-1} and a K_{prime} of 3.622 mM; and (C) D-rhamnono-1,5-lactone

To try and determine the optimal conditions to stabilise the protein and provide the highest chance of producing ordered crystals, differential scanning fluorimetry was carried out using a number of different conditions (Table 6). The results showed no change in the melting temperature of the lac11 with any of the small molecules tested in comparison to the Tris-HCl only control. This means none of the conditions tested increased the stability of the enzyme which could have improved the chances of obtaining an ordered crystal as usually an increase in stability also means reduced flexibility, which can disrupt formation of crystals. Out of the buffers tested Tris-HCl was the worst with the lowest temperature melting peak out of those tested, with a T_m of 56.5 °C. The melting temperature for both HEPES and MOPS was shown to be a higher with the lac11, a T_m of 96.0 °C for the former and 96.3 °C for the latter.

As the activity assays done earlier did not reveal which divalent ion is the co-factor for this enzyme, the thermal stability of the enzyme divalent ions added was measured. The measured T_m for the blank, $CaCl_2$, $MgCl_2$ and $MnCl_2$ all showed similar values, with $MnCl_2$ showing the lowest T_m at 50.6 °C and the highest of these four being the $MgCl_2$ with a T_m of 58.2 °C. The two exceptions were $ZnCl_2$ and $CoCl_2$ both of which showed a significant increase in the melting temperature. $ZnCl_2$ had a T_m of 87.4 °C and $CoCl_2$ had the highest T_m with 95.1 °C. This would suggest that one of these two would be the potential candidate for the cofactor of this enzyme, though it should be noted that with many Zn^{2+} containing enzymes, Co^{2+} will bind with greater affinity (Kremer-Aach et al., 1997). Thus while there may be a greater increase in thermal stability with $CoCl_2$, it is more likely that Zn^{2+} is the native co-factor as cobalt is a rare element in nature.

Table 6. T_m determined by differential scanning fluorimetry of lac11 with various buffers and additives.

Conditions	T_m ± SD °C
Tris-HCl	56.5 ± 0.1
Tris-HCl + sodium citrate	56.0 ± 0.1
Tris-HCl + malonic acid	56.8 ± 0.03
Tris-HCl + malic acid	56.3 ± 1.2
Tris-HCl + CaCl$_2$	54.0 ± 4.5
Tris-HCl + MgCl$_2$	58.2 ± 1.7
Tris-HCl + ZnCl$_2$	87.4 ± 6.3
Tris-HCl + MnCl$_2$	50.6 ± 0.5
Tris-HCl + CoCl$_2$	95.1 ± 0.5
HEPES	96.0 ± 0.2
MOPS	96.3 ± 0.07

As thermophilic enzymes are generally more resilient to the exposure of organic solvents that would normal denature a protein, the ability of this enzyme to resist denaturation was examined. The stability of the lac11 was tested in five different solvents: ethanol, methanol, acetonitrile, isopropanol and DMSO, with the results shown in Fig. 4.4. Despite a wide range of solvents being used as well as concentrations up to 60 % for all the solvents, full, or very close to 100 %, activity was seen for all activity assays with the enzyme. It is entirely possible that the enzyme did denature in the solvents as for concentrations 60 % or higher, a small amount of precipitation was seen when the enzyme was added to the solvent. Either the enzyme was stable enough to endure the solvents tested at these concentrations, or it was highly efficient at refolding itself to return to a functional form after partial denaturation on exposure to a solvent.
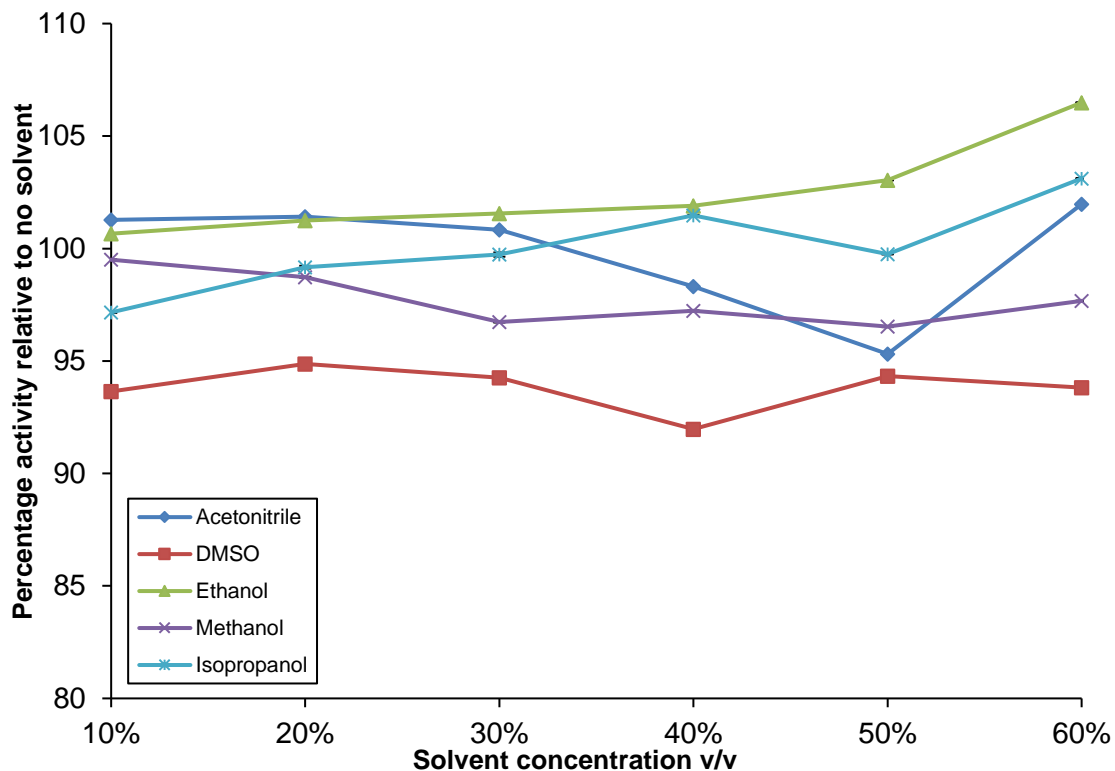
Fig. 4.4. Activity of lac11 towards D-glucuronic acid-1,4-lactone relative to the rate seen in standard HEPES buffer with no solvent.

In addition to the unfolding temperature, the activity retention of the enzyme was also analysed at higher temperatures. This method is less informative than differential scanning fluorimetry as there is no way to perform the assay at these high temperatures due to the unstable nature of the sugar lactones. It is highly likely that at high temperatures the substrates would degrade rapidly considering that at room temperature most substrates have degraded sufficiently to cause a colour change in the pH-based assay within 15-30 min, so gaining an accurate measurement of the enzymatic activity at elevated temperatures would not be possible. Overall the enzyme retains all of its activity after incubation at 80 °C for 20 min. At 90 °C there is a small decline in the activity to 96.5% of that seen for the control done at room temperature. The solvents studies showed a similar result with no significant difference in activity after exposure. So while there does appear to be a very small amount of activity lost, the majority of its function remains at this high temperature, or it is highly efficient at refolding as was suggested by the solvent stability assays.

Microscale thermophoresis is a method to identify interactions between molecules using a microscopic temperature gradient and measuring the rate of diffusion of labelled molecules. Thus changes to the molecule in terms of size or conformation will alter the rate of diffusion and enables identification of potential inhibitors and substrate analogues due to binding by these molecules to enzymes. An assay was carried out to try and identify potential substrate analogues for lac11 due to the high instability of the native substrates making them unsuitable for co-crystallisation trials. D-glucose was identified in the PDB database as a substrate analogue for another solved gluconolactonase and was visible within the active site of the structure (PDB code 4GN9). The binding affinity between glucose and lac11 was analysed to determine whether or not it would be a suitable analogue for co-crystallisation. The results from this are shown in Fig. 4.5. Overall it was determined that D-glucose did indeed bind to the lac11 as the fluorescence increased with the concentration of glucose meaning the protein was diffusing slower from the focal point due to binding of the ligand and possessed a binding affinity of 520 µM ±0.06 nM, indicating its high potential for stabilising the protein complex to try and obtain more ordered crystals.
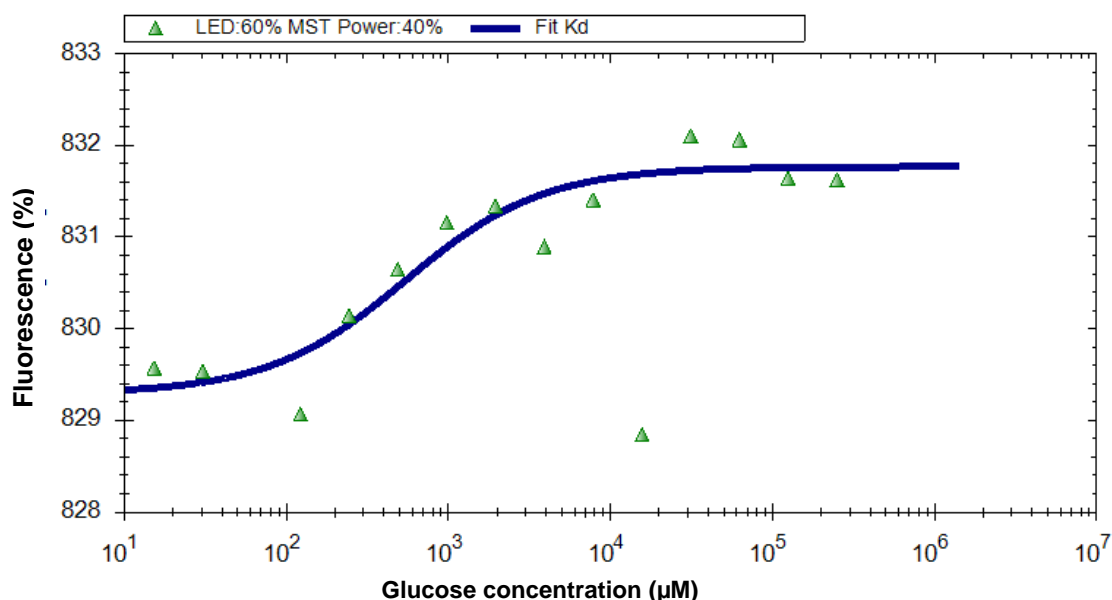
Fig. 4.5. Analysis of data from the microscale thermophoresis experiment for lac11. The dissociation constant (K_d) was calculated with an exponential increase in glucose concentration. The y axis shows the fluorescence of the NT-647 tagged lac11 and the x axis shows the concentration of D-glucose in µM.

The lac11 was concentrated to 9.6 mg ml^{-1} and screened for crystal growth. A large number of crystals were seen from the Morpheus and JCSG plus screens, with the typical morphology being long, thin needle-like crystals. When data was collected at the Diamond Synchrotron these were shown to be highly disordered and completely unsuitable for data collection. The only crystals to diffract well were from the JCSG plus screen and from the following conditions; 40 mM potassium dihydrogen phosphate, 16 % w/v PEG 8000 and 20 % v/v glycerol. The cryoprotectant that was made had the same composition of buffer, pH and salt as the screen with the exception of the glycerol concentration being raised to 30 % v/v. The best resolution that these crystals diffracted to was 2.7 Å. Additional electron density could be seen in the active site of the enzyme which could suggest the presence of a Zn^{2+} ion. Subsequent trials were performed again, this time with the addition of a seed stock. Crystal seeding, apart from stimulating crystal growth in conditions where there usually would not be any, has been known to provide higher resolution crystals (Bunker *et al.*, 2012). Crystals from the JCSG plus screen were selected as these had been confirmed as disordered protein crystals. These were used in the JCSG plus screen again as well as MDL-01 and 02. However this also failed to provide crystals that would diffract to an improved resolution.

From the results obtained from differential scanning fluorimetry the buffer was changed to HEPES from Tris-HCl as this was identified as more stabilising for the enzyme than the original buffer. In addition to this, co-crystallisation was performed in addition to standard screening with the substrate analogue glucose, which had been identified through microscale thermophoresis, at a concentration of 100 mM as glucose can act as a cryoprotectant and thus reduce the risk of ice crystals forming (Fig. 4.6 A). Co-crystallisation was also carried out with 1 mM CoCl$_2$ and ZnCl$_2$ as these metal ions also increased the stability of the enzyme and thus could provide more ordered crystals. The new crystals diffracted to 2.4 Å from the co-crystallisation trials with glucose (Fig 4.6 B and Table 7). The conditions within the well were 0.1 M sodium cacodylate pH 6.5, 5 % w/v PEG 8000 and 30 % MPD. Crystals from this condition were frozen directly without the use of a cryoprotectant due to the high concentration of MPD. Attempts to improve this resolution were made using crystal optimisation. Unfortunately, these trials did not yield any more crystals.
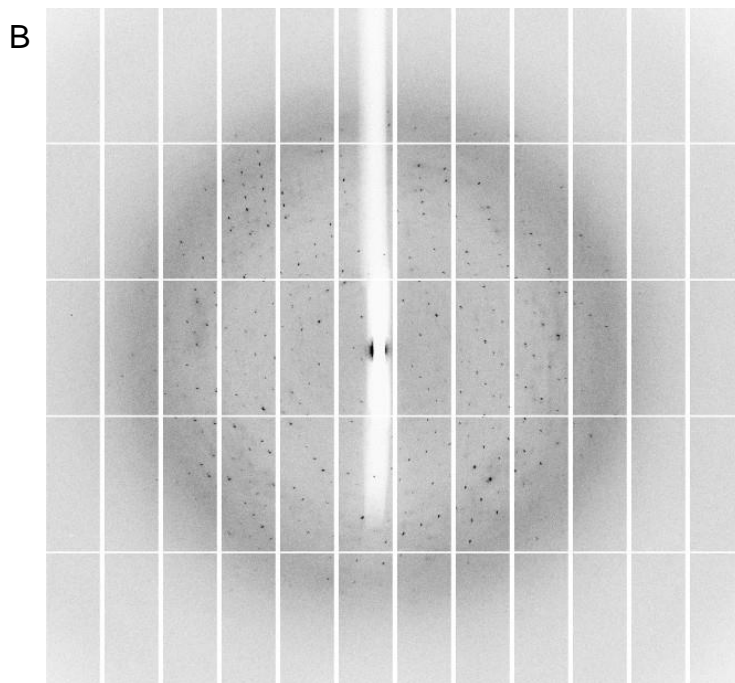
Fig. 4.6 Crystals and diffraction pattern of lac11. (A) Crystals obtained from the screen JCSG with the following conditions: 0.1 M sodium cacodylate pH 6.5, 5 % w/v PEG 8000 and 30% MPD. The protein concentration was at 9.4 mg ml^{-1} and had been incubated with 100 mM glucose at room temperature for 20 min. The final protein concentration in the well was 4.7 mg ml^{-1}. (B) Diffraction pattern obtained from Diamond Synchrotron showing resolution to 2.41 Å.

Table 7. Statistics from X-ray data collected at the Diamond Synchrotron for lac11 crystal diffraction

	Overall	Highest resolution
High resolution limit	2.41 Å	(5.90 Å - 2.41 Å)
Low resolution limit	30.20 Å	(30.20 Å - 2.64 Å)
Completeness	97.5%	(94.5% - 98.3 %)
Multiplicity	3.4	(3.4 - 3.3)
I/sigma	6.8	(11.6 - 2.2)
Rmerge	0.128	(0.057 - 0.700)
Unit cell dimensions:	151.995 Å 151.995 Å 90.401 Å 90.0° 90.0° 120.0°	
Spacegroup	P 3 2 1	

The phase was solved using the phase of a gluconolactonase from *Xanthomonas campestris* (PDB code 3DR2), a mesophilic bacterium associated with black rot in plants with which there was 36 % sequence identity with the lac11. To improve the accuracy of refinement the previously collected data to 2.7 Å was used to improve the phase through multi-crystal averaging. An R factor of 0.28 and a free R factor of 0.32 were achieved through structural refinement with COOT and REFMAC (Murshudov *et al.*, 1997). Improving these values proved difficult and after many attempts were made it was determined that for the time being these are the best that can be achieved with the current dataset. One of the things potentially keeping the R value and the free R value high was the presence of a large area of density to which none of the programs could fit the protein structure. This can be seen in Fig. 4.7 A, where there is an abundance of density present in a void between the protein units. Unlike the previous 2.7 Å data collected, there was no strong electron density within the active site. At the location of the divalent ion, seen in Fig. 4.7 B, there was expected to be a stronger diffraction caused by a heavy Zn^{2+}. However, the strength of the peak in the difference map would be indicative of something lighter divalent, probably Ca^{2+} ion at 3.0 Å root mean squared deviation. Whatever this ion may be, it was coordinated by 4 residues, Glu 79, Asn 175, Asn 230 and Asp 286. In addition to this, there was no obvious sign of glucose within the active site despite its presence within the well at a final concentration of 50 mM. The overall structure was a six bladed β-propeller with the active site present in the centre of all six blades (Fig. 4.7 C and D). Each propeller consists of at least four antiparallel β-strands apart from two, the top

two strands in the image which both contain an extra β-strand to a total of five. Superposition with 3DR2 (Fig. 4.7 E) showed structures to be similar with a RMSD of 3.1 Å. However a significant difference was seen in the side chains of the substrate binding pocket (Fig. 4.7 F). The four residues identified in the 3DR2 structure involved in substrate binding were Phe 151, His 104, Leu 64 and Trp 46 (Chen *et al.*, 2008). Equivalent residues were located in the lac11 structure. Both the histidine and tryptophan were conserved and located at 146 and 77 respectively. There is seen a conserved substitution of phenylalanine to tyrosine, both are of similar size and both are hydrophobic though with an addition of a hydroxide group tyrosine is less hydrophobic. Another conserved substitution was observed with leucine for valine, the α-carbon of both side chains were 2.0 Å apart. Additionally, the tryptophan side chains were in significantly different conformations, where the distance between the most distant carbons in the side chains was 4.9 Å. It was reported in the paper that the only substrate the enzyme had activity for out of those tested was glucono-1,4-lactone. The other substrates tested were D-gulono-1,4-lactone, L-gulono-1,4-lactone, and D-ribono-1,5-lactone. This is interesting as lac11 was shown by enzyme assay to have activity towards all but one of the substrates that were available (Table 5) suggesting it possesses greater promiscuity towards substrates than the lactonase 3DR2. Looking at the binding pocket, it is possible that the 4.9 Å shift in the tryptophan away from the other side chain residues, as well as the shorter side chain of the valine in comparison to the leucine, increases the size of the binding pocket. Lac11 not only possesses activity to smaller substrates than glucono-1,4-lactone, such a D-threono-1,4-lactone, but also to larger ones such as D-glucuronic acid-1,4-lactone. Both could be attributed to a larger binding pocket, crystallisation with a substrate or substrate analogue would need to be done to confirm the binding mechanism which would allow more accurate rationalisation for the differences observed. Fig. 4.7 G shows the surface of the structure represented by white for hydrophobic regions, red for negatively charged and blue for positively charged regions. The active site is observed to be open and exposed to the solvent on one side of the structure, as seen by the cavity indicated in the image. iMosflm (Battye *et al.*, 2011) was also used to for data integration to see if it could improve upon the quality of the results. There was no improvement over the previous programs used.

D

Asn175

Glu79

Asn230

Asp286

E

F

Trp 77

4.9 Å

Trp 46

Tyr 192

Phe 151

His 146

His 104

Val 95

2.0 Å

Leu 64

G

Active site

[105]

H
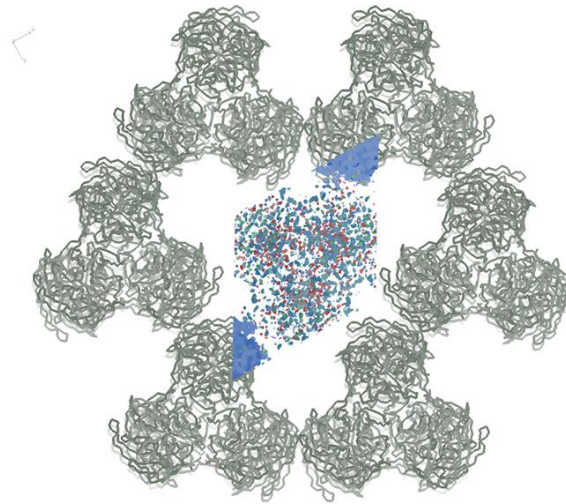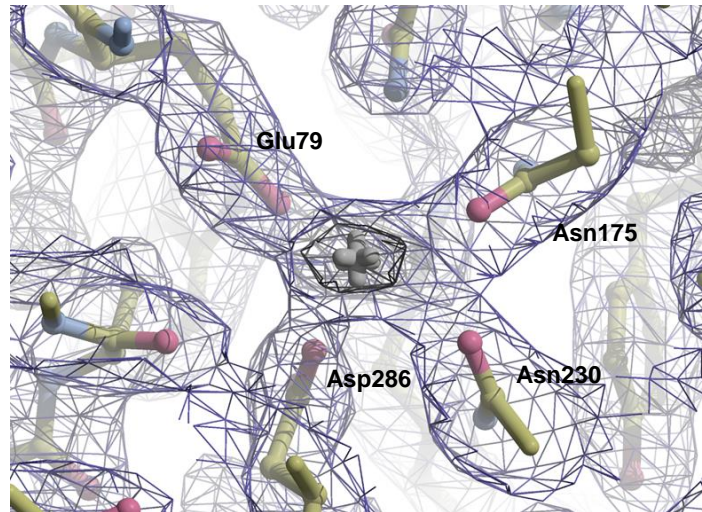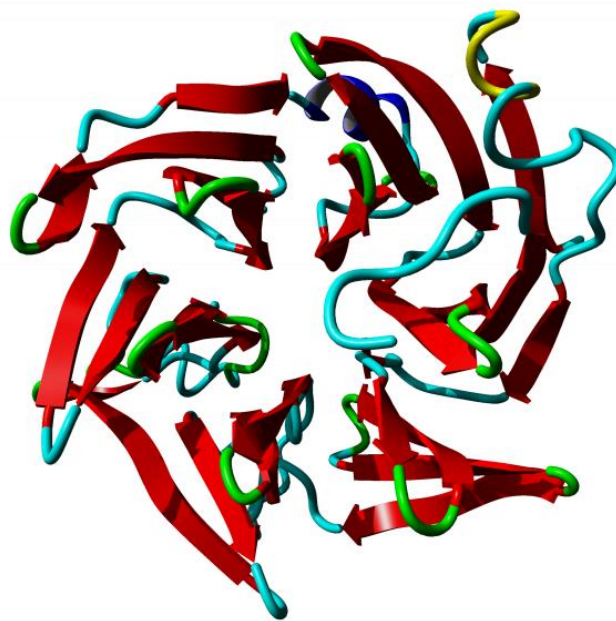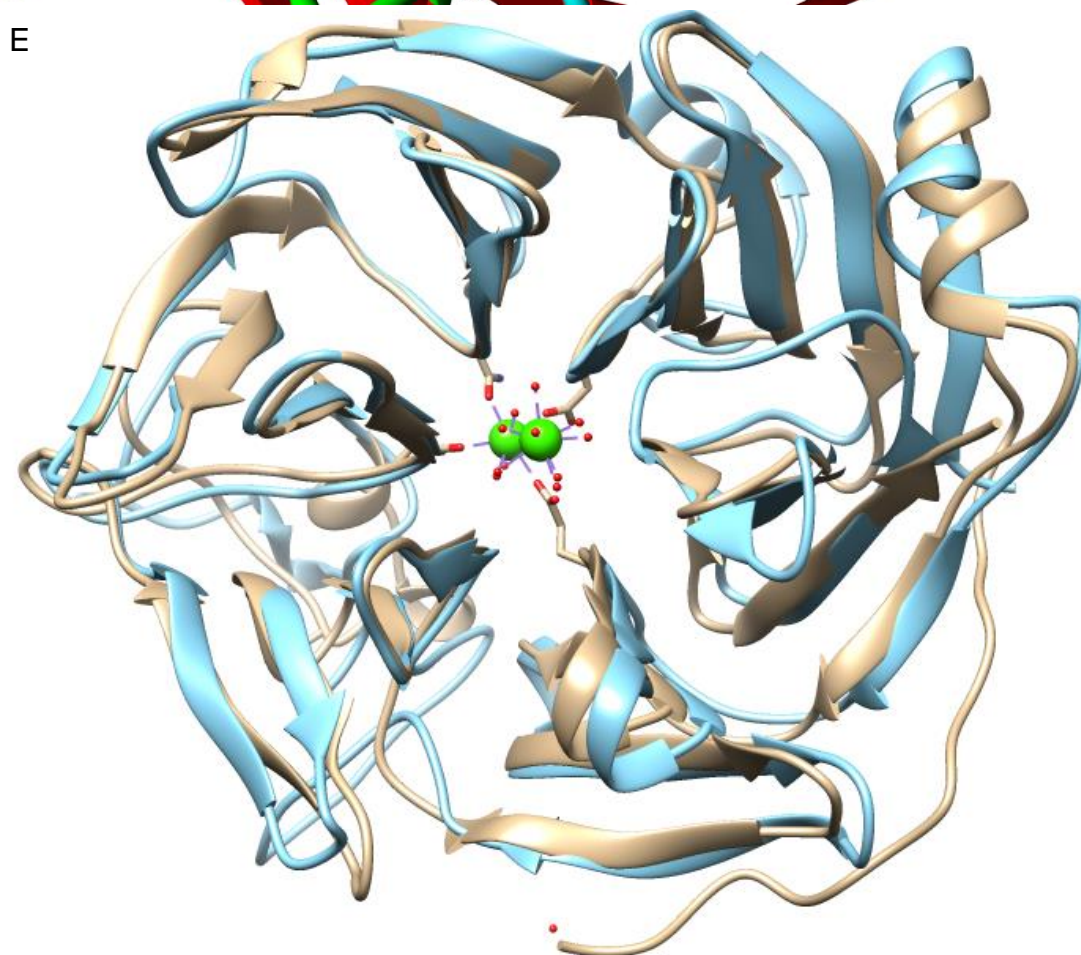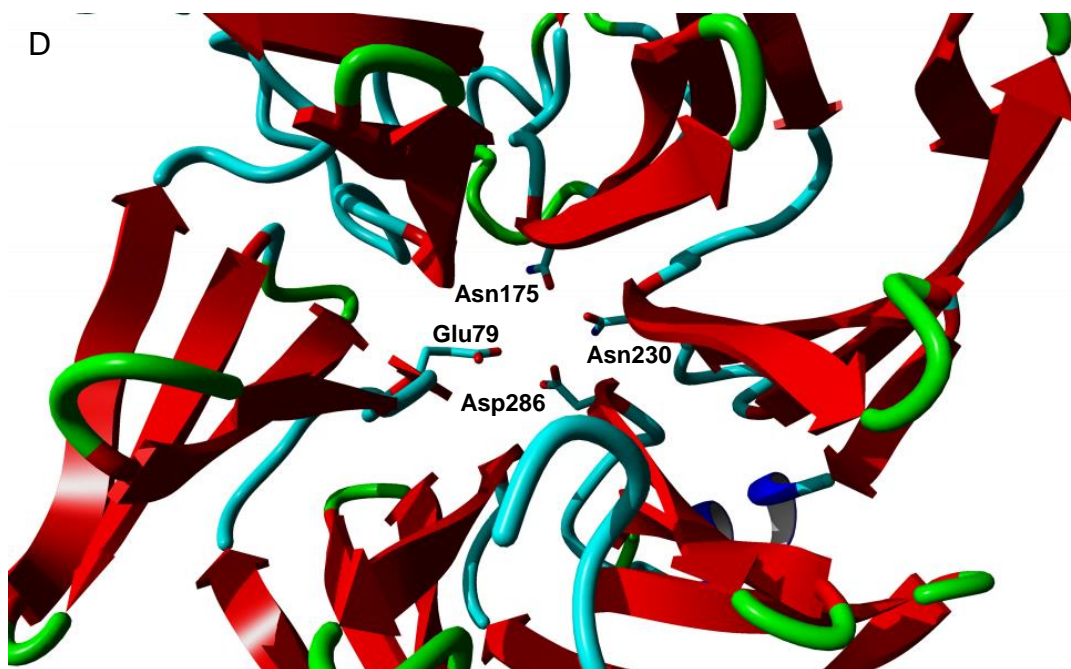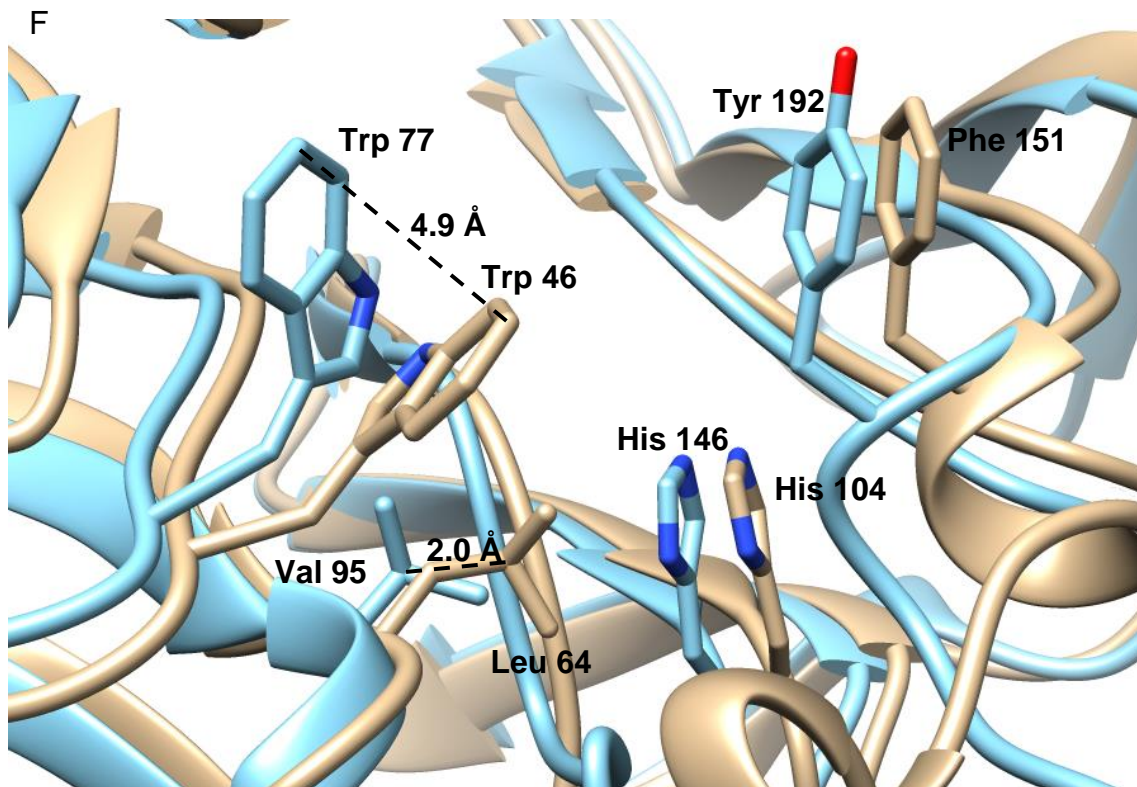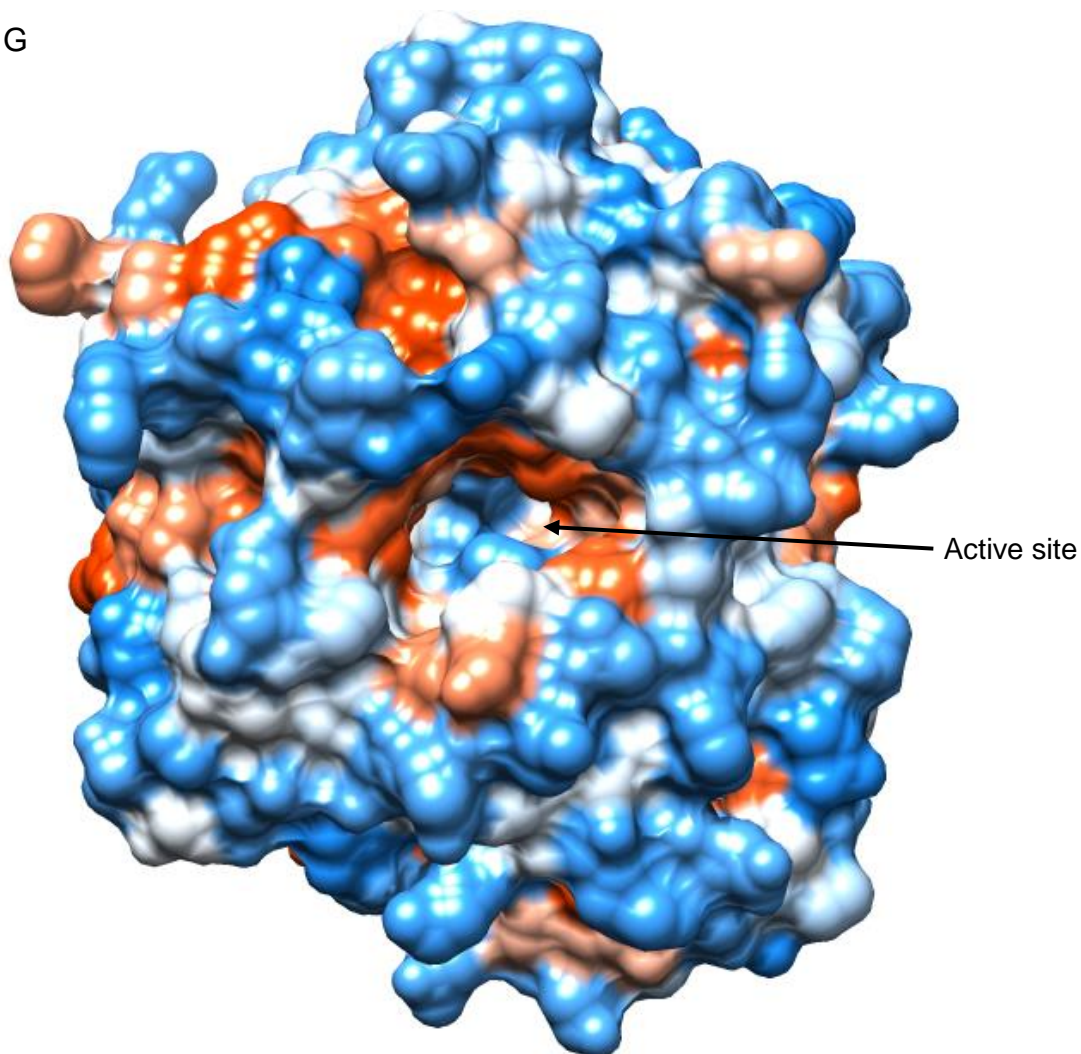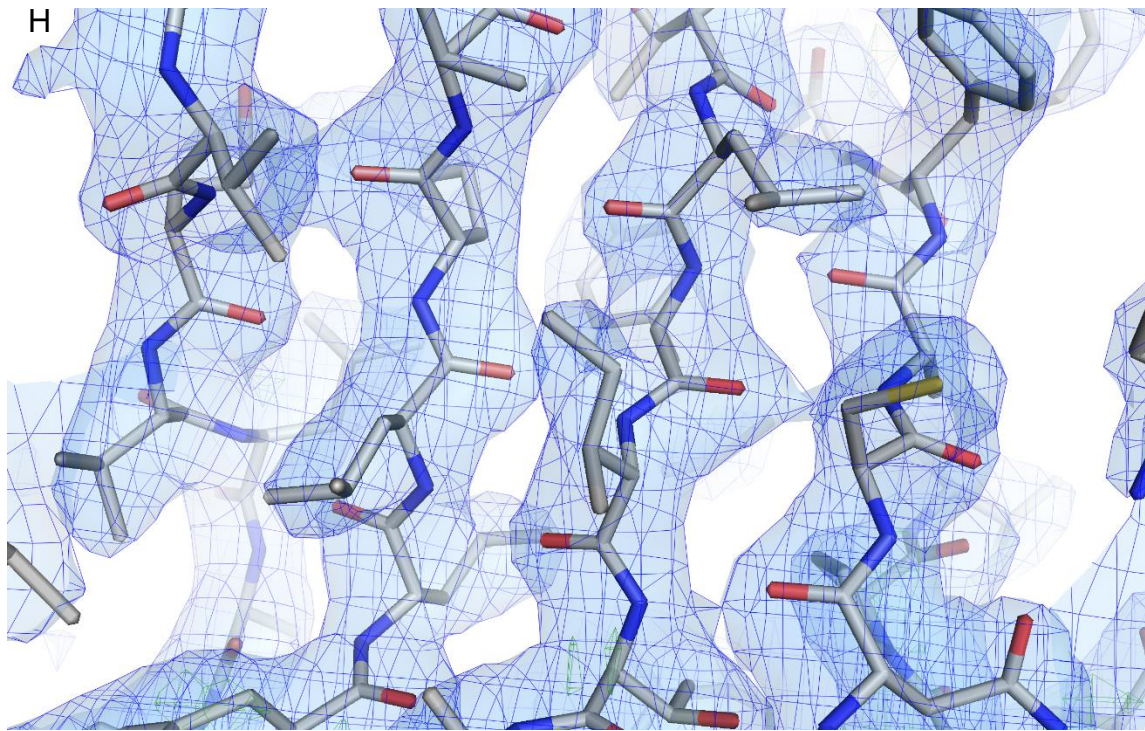
Fig. 4.7. Crystal packing and diffraction observed from the lac11 crystals. (A) The unit cell of the collected data. Within the crystal packing seen from the X-ray diffraction there is an abundance of high resolution density to which the protein (seen in grey) cannot be fitted using any available program. (B) Electron density map centred on the location of the divalent ion binding site, coordinated by a glutamine residue, an aspartic acid residue and two asparagine residues. (C) Structure of protein was found to be a six bladed β-propeller. (D) Active site residues were located in the core of the propeller structure. (E) Superposition of lac11 (blue) and lactonase 3DR2 (brown), RMSD was 3.1 Å. (F) Superposition of their respective substrate binding pocket residues. (G) Surface representation of the lac11 lactonase, white is uncharged, blue is positively charged, and red is negatively charged. The arrow indicates the location of active site. (H) Density for the native structure in the β -sheet region. The 2Fo-Fc map is shown in blue and is contoured at 1 sigma, the Fo-Fc map is contoured at 3 sigma (green). Image created with PyMOL (Delano, 2002).

It was unclear whether the lac11 was present as a dimer or trimer, as indicated in Fig. 4.7 A, or not owing to the fact that the predicted dimer interface seen during analysis in COOT was completely different to that found in the lactonase 3DR2 structure (Fig. 4.8 A), whose phase was used to solve the structure.

To determine whether or not the lac11 was a dimer, the program PISA (Krissinel and Henrick, 2007), part of the suite of programs from CCP4I, was used to analyse the structure to determine the likely oligomeric state. It determined that the interfaces present in the protein did not suggest it to form a oligomer in solution and it was most likely present as a monomer. To confirm this a sample of lac11 was passed through an analytical gel filtration column after purification by both nickel chromatography and standard size exclusion chromatography (Fig 4.8 B). Four peaks were seen, the first was aggregates and the fourth was in the void volume. The third peak corresponded to a size of 15 kDa, this would most likely be degraded protein as the sample used was an old sample stored frozen at −80 °C with 25 % glycerol so some damage would have occurred. The second peak corresponded to a size of 30 to 40 kDa, the size range the monomeric form of the protein. No peak in the range that the dimer would be present was seen, this would indicate that the protein was present as a monomer, unlike the lactonase 3DR2 structure previously solved. This makes it unique in comparison and makes it more like other six-bladed β-propeller proteins which usually exist as monomers, only a few of them form homodimers or oligomers (Paoli, 2001).

Interestingly, a stark difference observed between the two gluconolactonases is that while lactonase 3DR2 is stabilised by disulphide bonds (Chen *et al.*, 2008), the lac11 did not possess any intramolecular disulphides. Since the main role of gluconolactonases is in metabolic pathways, the lac11 is most likely adapted for conditions within the cell. A high number of disulphide bonds in a protein are more commonly seen in those secreted by the cell to make them more stable in the extracellular environment, meaning that lac11 has little need of them. However, it appears the lactonase 3DR2 is not exported either, a scan of its sequence, using the SignalP 4.1 Server from DTU Bioinformatics (Petersen *et al.*, 2011), revealed no export signal peptide present in the sequence, meaning both should be present intracellularly. A total of 9 salt bridges were identified within the structure of the lac11. Of particular note though is the salt bridge cluster formed

between Arg 316, Arg 318 and Glu 52. This increase of 2 salt bridges is not a significant increase over the mesophilic gluconolactonase from *E. coli* which had 7. This indicates that salt bridges do not provide a great deal of extra thermostability and that this must be attributed to other features such as hydrophobic interactions.
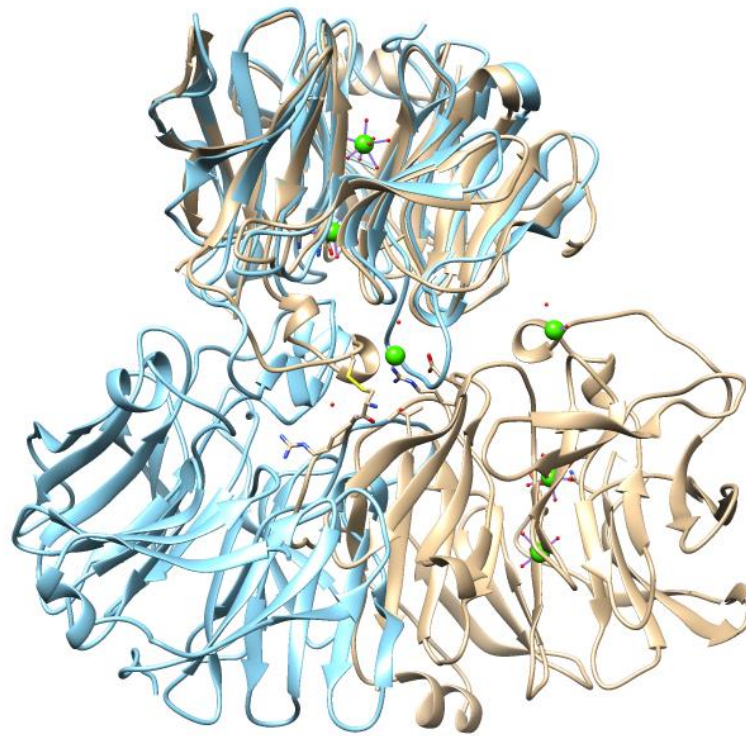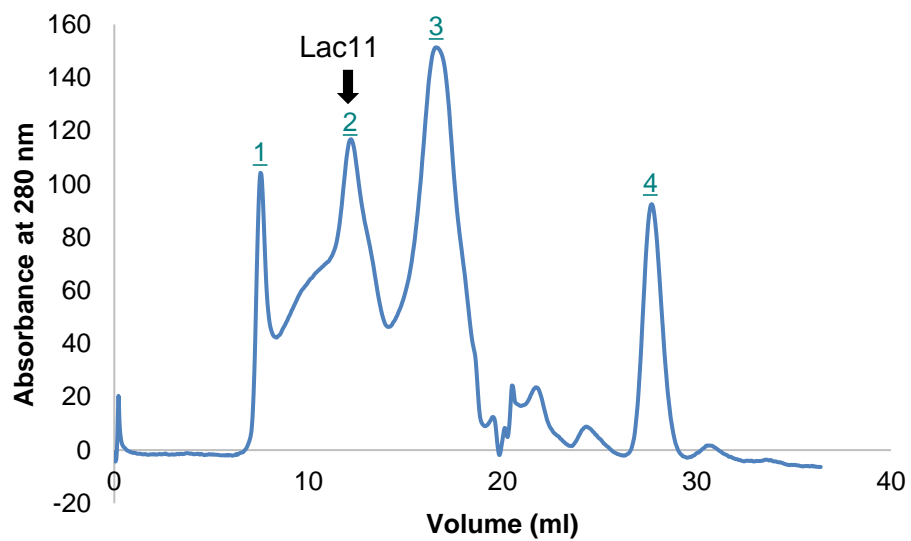
Fig. 4.8. (A) Superposition of lac11 (blue) as an artificial dimer and lactonase 3DR2 (brown) in Chimera (Pettersen *et al.*, 2004). Green balls are Ca^{2+} ions from 3DR2 and side chains shown are Mg^{2+} coordinating residues. Dimers form completely differently suggesting that lac11 might not be a dimer due to such a drastic difference. (B) Analytical size exclusion chromatography of lac11 sample previously frozen at -80 °C. Peak 1 was aggregate protein, peak 2 corresponded to a size of 30-40 kDa, peak 3 corresponded to a size of 15 kDa, peak 4 was at a total column volume.

The primary structure of lac11 was investigated and the percentages of each amino acid were compared to the gluconolactonase from the mesophile *X. campestris* (PDB code 3DR2) (Table 8). There was a reduction in the number of cysteine residues from 2.6 % to 1.1 % in the enzyme, this is standard as they are highly susceptible to oxidation at high temperatures. The biggest difference was a decrease in the number of alanine residues in lac11, with 6.5 % of the total amino acids consisting of these in lac11 in comparison to 10.2 % in the lactonase 3DR2. However most of the other hydrophobic amino acids in lac11 showed small increases: a greater amount of isoleucine with 1.9 % more, from 2.3 % to 4.2 %; a greater amount of methionine with 1.0 % more from 1.0 % to 2.0 %; a greater amount of phenylalanine more from 1.8 % from 3.3 % to 5.1 %; a greater amount of tyrosine with 2.1 % more from 1.3 % to 3.4 %. This trend suggests a substitution of weakly hydrophobic residues for stronger hydrophobic ones, and thus an increase in the hydrophobic forces present within the thermophilic enzyme in comparison to the mesophilic form. This would aid in stabilising the enzyme at high temperatures. The alanine residues present in the β-strands do appear to be substituted for more hydrophobic amino acids (Fig. 4.9). While those present in loop regions are substituted for charged and polar amino acids indicating that either interactions with water or the formation of ionic bonds with the main structure stabilise them.

Visual inspection of the structure with COOT revealed that there was approximately a 5 % increase in the number of hydrophobic interactions seen in the structure in comparison to the lactonase 3DR2 structure which was used as a model for molecular replacement. There were seen to be more prolines in loop regions as well as the loops themselves generally being shorter which would help increase the thermal stability of the lac11 protein.

Table 8. Percentage composition of amino acids in the lac11 and 3DR2.

Amino acids	3DR2	Lac11
Ala	10.2 %	6.5 %
Arg	7.9 %	6.2 %
Asn	2.0 %	3.1 %
Asp	9.2 %	7.0 %
Cys	2.6 %	1.1 %
Gln	4.6 %	3.7 %
Glu	2.6 %	4.8 %
Gly	9.8 %	12.4 %
His	3.9 %	0.6 %
Ile	2.3 %	4.2 %
Leu	10.2 %	9.9 %
Lys	0.7 %	5.1 %
Met	1.0 %	2.0 %
Phe	3.3 %	5.1 %
Pro	8.2 %	7.0 %
Ser	5.6 %	4.8 %
Thr	5.6 %	5.1 %
Trp	3.3 %	2.0 %
Tyr	1.3 %	3.4 %
Val	5.9 %	6.2 %

3DR2

```
                                          1         10        20        30        40
3DR2   ..................................MDSHCRVRPAGPAVPADCDPPRITHAALAARLGDARLLTLYDQAT
lac11  MRKLLGSSAIHVLSRRTGYGMYFLILALGVATQVVFGAEPSQNPSEKVAWIERADPGL....DDLIAPDAQVEVLAEGFE
lac11
```

1

3DR2

```
           50        60        70        80             90        100       110
3DR2   WSEGPAWWEAQRTLVWSDLVGRRVLGWREDGTVDVLLDATAFT............NGNAVDAQQRLVHCEHGRRAITRSDA
lac11  WSEGPVWIPEGGYLLFSDVPKNTIYRWKEGQGIDIFLKPSGYTGFRERGGESGSNGLALDRQGRLLLCQHGDRRVARWEK
lac11
```

3DR2

```
           120       130       140       150       160       170       180       190
3DR2   DGQAHLLVGRYAGKRLNSPNDLIVARDGAIWFTDPPFGLRKPSQGCPADPELAHHSVYRLPPDGSPLQRMADLDHPNGLA
lac11  .GCFITLADQYEGKPLNSPNDLVVKSNGDIYFTDPPYGMTPEAQRDPN..ALGFCGVYRISADGKLTLLVRDMTRPNGIA
lac11
```

2

3DR2

```
           200       210       220       230             240       250       260
3DR2   FSPDEQTLYVSQTPEQGHGSVEITAFAWRDGALHDRRH.......FASVPDGLPDGFCVDRGGWLWSSSGTGVCVFDSDG
lac11  FSPDEKTLYVAQSDPQRPLWMAF..PVREDGTLGEGKVFFDAKPWQQSGLPGLPDGMKVDQKGNLFATGPGGVNIFRPDG
lac11
```

3DR2

```
           270       280       290       300
3DR2   QLLGHIPTPGTASNCTFDQAQQRLFITGGPCLWMLPLP......
lac11  TFLGRIRVKVPTANCAFGDDGSTLYITADMYLLRVKTKTKGLGF
lac11
```

[112]

Fig. 4.9. Secondary structure and amino acid sequence alignment between the lac11 lactonase and the lactonase 3DR2 made using the ESPript 3 sequence and structure alignment server (Robert and Gouet, 2014). Residues highlighted in red are conserved, residues in blue boxes are conserved substitutions.

Fig. 4.10 shows the B-factor analysis of lac11 and the 3DR2 structure. This was done using the temperature factor analysis software in the CCP4i suite of programs (Winn *et al.,* 2011) to look at the displacement of atoms within the structures. Overall both structures had low B-factors with most of the residues below 40 Å^2. This indicates well-ordered structures with low flexibility. The residues in and around the active site are more ordered with lower B-factors than the rest of the structure. A similar trend is seen in the 3DR2 structure though the differences between the active site residues and the rest of the structure if not as pronounced as the lac11 structure, possibly due to its lower overall B-factor.
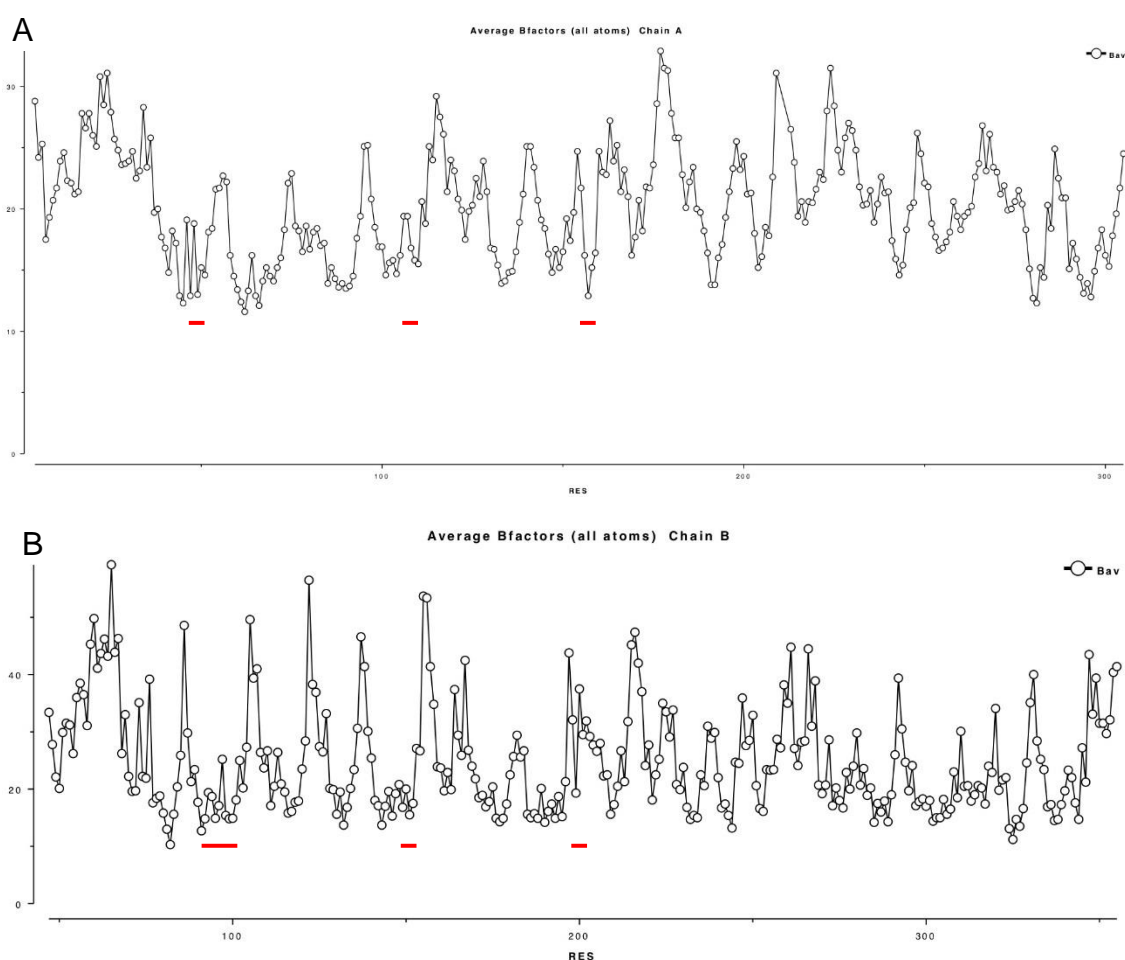
Fig. 4.10. B-factor analysis of the 3DR2 structure (A) and lac11 (B). The average B-factor, measure in Å^2, was plotted on the Y-axis against the residue number on the X-axis. Red lines highlight the residues located in the active sites of the enzymes.

Further crystallisation attempts were done to try and obtain a better structure with lower R values and higher resolution as ideally, they should be below 0.25 R $_{Free}$, instead the lowest value obtained was 0.32 R $_{Free}$. Optimisation of the crystallisation screen conditions that previously provided crystals which diffracted to 2.4 Å did not yield anymore crystals. The next method that was used to try and improve crystal quality was lysine methylation. The process itself required a larger quantity of protein to be purified beforehand as during the methylation step roughly 60 % of the protein was lost due to denaturation by the methylation conditions. When the protein concentration was measured using with a Nanodrop spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA) before methylation the concentration was 0.5 mg ml^{-1}, and after methylation it was 0.2 mg ml^{-1}. After methylation, crystallisation trials produced crystals in a new condition with a different morphology which appeared to be of a higher quality than the previously obtained crystals and due to their fragile nature when physically manipulated were highly likely to be protein (Fig. 4.11). The crystals grew in 0.1 M HEPES pH 8.5 and 25 % Sokalan CP7 from the MIDAS screen with the protein concentration of 7.2 mg ml^{-1}. First attempts to freeze these crystals proved problematic when the cryo-protectants contained 0.1 M HEPES pH 8.5 and either 40 % v/v PEG 400 or 40 % w/v glucose. Both of these conditions dissolved the crystals so one last crystal was frozen without any cryo-protectant but X-ray diffraction data collection was prevented by formation of ice. A cryo-protectant containing 0.1 M HEPES pH 8.5 and 40 % v/v Sokalan was made and fresh crystals produced. However before they could be frozen the crystals disintegrated overnight for some unknown reason which while very rare can happen with protein crystals (Rupp, 2009). Attempts to replicate the crystals produced in this condition failed, possibly due to the environmental factors in the room the crystals were growing in or modifications and repairs done to the equipment between crystallisation trials.
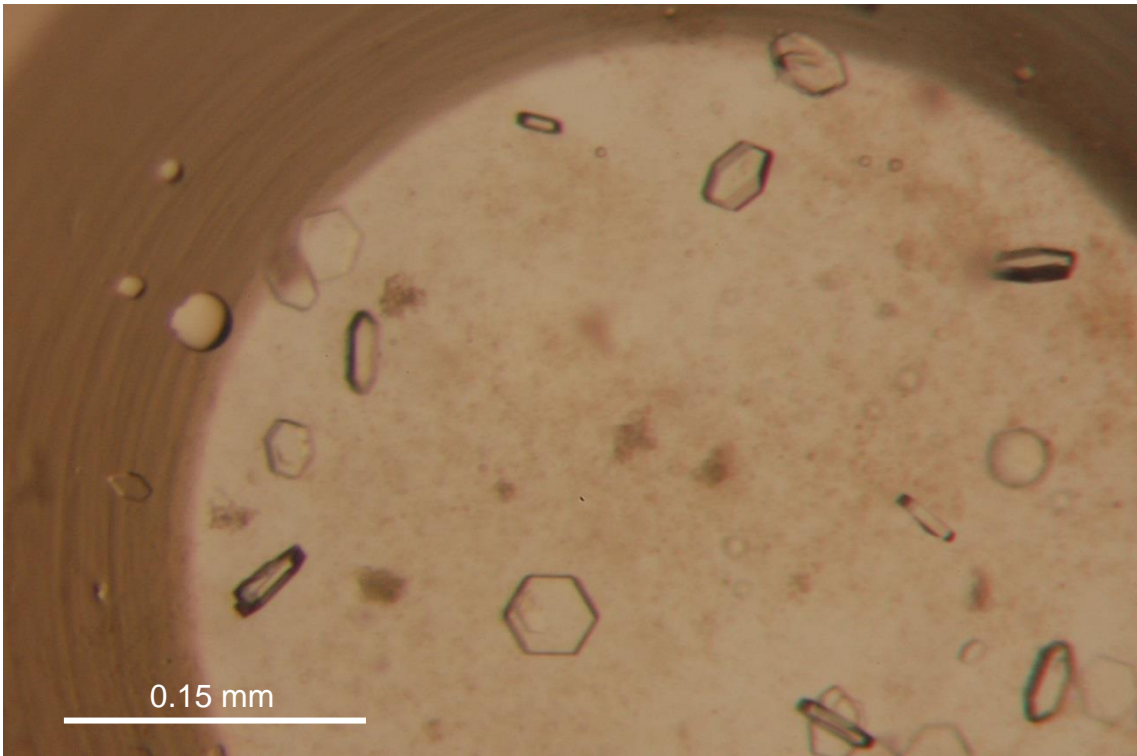
Fig. 4.11. Crystals formed by lac11 after lysine methylation in the MIDASplus screen in 0.1 M HEPES pH 8.5 and 25 % Sokalan CP7

4.4 Conclusion

The electron density map of the enzyme is high enough to see the general structure and has allowed relatively accurate placement of the side chains. Although the high R factor and free R factor were not fit for publishing, ideally both values should be below 0.25. Here both were above this value with the R factor at 0.28 and a free R factor at 0.32, as such a better crystal diffracting to higher resolution or one in a different space group would be required to try and obtain a structure with lower R values. Trials with crystal optimisation have yielded a large number of crystals, though so far not one has produced diffraction data anywhere near the 2.4 Å resolution already obtained. Understanding the structure and binding of the enzyme to a substrate or substrate-analogue would be of great benefit as very little is known about this class of lactonases. Currently there are only six different structures of gluconolactonases deposited in the Protein Data Bank: human, mouse, *T. brucei*, *X. campestris*, *Leishmania guyanensis* and *Leishmania braziliensis*.

Assays have shown the lac11 enzyme to possess activity for a broad range of substrates. Also lac11 does not lose any significant activity after exposure to high temperatures and high solvent concentrations making it of great interest for industrial applications. Also since gluconolactonase is essential in metabolic pathways leading to the synthesis of pentose and vitamin C, determining this structure when there are currently few structures available will aid in our understanding how these processes occur.

4.5 Summary

The crystal structure of the lac11 gluconolactonase was successfully solved to a resolution of 2.4 Å with a space group of P 3 2 1. The structure was revealed to be a six-bladed β-propeller and was confirmed to be most likely present as a monomer by PISA and analytical size exclusion chromatography. The enzyme showed activity towards a broad range of substrates including 1,4 and 1,5 sugar lactones and was stable at very high temperatures with the highest melting point measured at 96.3 °C. This stability appeared to be through substitution of alanine within the structure for residues capable of stronger hydrophobic interactions.

The next step in the investigation of lactonases was to study the last remaining class of this enzyme, enol lactonase.

5. Enol lactonase

5.1 Introduction

An enol lactonase, named B1-7Tnr1 (shortened to "B1 lactonase"), was identified using a BLAST search of a metagenome from a hot spring in Pozzuoli, Italy, collected by the group of Dr Xu Peng (University of Copenhagen). The conditions of the hot spring were 76 °C and at pH 3. The amino acid sequence had 29 % identity to a known enol lactonase from *Burkholderia xenovorans* in the Protein Data Bank (PDB code 2XUA), as revealed using a BLAST search. The metagenomic DNA was not available at the time so the sequence was synthesised and codon optimised for expression in *E. coli* by ATUM (California, USA). The gene was supplied in an pD441-CH vector, part of the Electra expression system

5.2 Results and discussion

The recombinant DNA containing the B1 lactonase gene was successfully transformed into *E. coli* XL10-gold ultracompetent cells with kanamycin resistance. The transformed cells were grown, recombinant DNA was purified by mini-preps and used to transform *E. coli* Bl21 (DE3) expression cells. Expression conditions were standard, as described in the materials and methods. SDS-PAGE revealed that none of the protein was detectable in the soluble fraction; all of it was located in the insoluble fraction. Expression conditions were altered, with lower concentrations of IPTG, lowering expression temperatures to 30 °C, and auto-induction all being tested; none showed any improvement to increasing the amount of soluble protein produced. Different expression cell lines, Bl21 (DE3) pLysS, BL21-CodonPlus (DE3) RIPL and Arctic Express (DE3) were also tested, using both the standard protocol and the variations mentioned above. Again, none showed any protein in the soluble fractions by SDS-PAGE for any of the cell lines or expression conditions.

Eventually, attempts to express and purify the B1 lactonase protein were halted in favour of homology modelling to solve the structure. The sequence was submitted as described in the materials and methods to I-TASSER (Zhang, 2008), Phyre2 (Kelly *et al.*, 2015) and SWISS-MODEL (Biasini *et al.*, 2014). Out

of all the models generated, the best one, determined by Ramachandran plots generated by Rampage, was from SWISS-MODEL.

The Ramachandran plot for the B1 enol lactonase (Fig. 5.1 A) showed 91.4 % of all amino acids in a favoured conformation, 7.5 % in an allowed conformation, and the remaining 1.1 % as outliers. These outliers were identified as: Arg 65, the first amino acid of a β-strand, which is not uncommon as amino acids that border transition in secondary structure can be found in an unfavourable conformation; Pro 138, located in the middle of a loop; and Arg 193, the first amino acid of an α-helix which, as with Arg 65, is a normal place to find an amino acid in an unusual conformation. The model was a standard α/β hydrolase fold with a β-sheet consisting of eight β-strands connected by six α-helices (Fig. 5.1 B). The analysis of the model gave a QMEAN of −3.02 and superposition gave a mean RMSD of 2.8 Å over all residues with the structure of the closest related enol lactonase identified from a BLAST search of the Protein Data Bank from *B. xenovorans* (PDB code 2XUA). Although this belongs to a different class of enzyme it was still logical for the program to select it as lactonases and epoxide hydrolases share common ancestry. Beyond both being α/β hydrolases, both proteins were found to be located within the same monophyletic group in the α/β hydrolase superfamily (Popiel *et al.*, 2014) and possessed homologous conserved residues. A total of seven salt bridges were identified within the structure of the B1 lactonase model. Of particular note is the salt bridge cluster formed between Arg 16, Arg 65 and Asp 29 as this double salt bridge to a single aspartic acid would confer stability to this region. The number of salt bridges is the same as in the mesophilic protein. This indicates that salt bridges do not provide any extra thermostability and that this must be attributed to other features such as hydrophobic interactions. The cap region as indicated in Fig. 5.1 B contains mostly loops and fewer α-helices than the 2XUA structure making it much more flexible in comparison. This would indicate that the substrate binding pocket likely exists in multiple conformations. This would mean the active site is affected by solvent dynamics more and is less restrictive with respect to substrate specificity.

Comparison of the active site (Fig. 5.1 C) with the epoxide hydrolase showed a similar catalytic triad (Bains *et al.*, 2011) with one small variation between the

residues, but the residues are in the same positions showing the catalytic triad is highly conserved between the two different enzymes. The triad in the epoxide hydrolase consists of two aspartic acid and a histidine, whereas in the lactonase one of the aspartic acids is substituted by a serine in the same position. This difference between a negatively charged residue and a polar residue would most likely be expected to cause a difference in substrate specificity or activity between the two enzymes. Fig. 5.1 D shows the surface of the model represented by white for hydrophobic regions, red for negatively charged and blue for positively charged regions. The active site is observed to be open and exposed to the solvent through a narrow channel, as seen by the cavity indicated in Fig. 5.1 D. This would potentially alter depending on the position of the cap region.

A

General/Pre-Pro/Proline Favoured   General/Pre-Pro/Proline Allowed
Glycine Favoured   Glycine Allowed

B

Cap domain

C

Ser 111

His 270

Asp 242

D

Active site
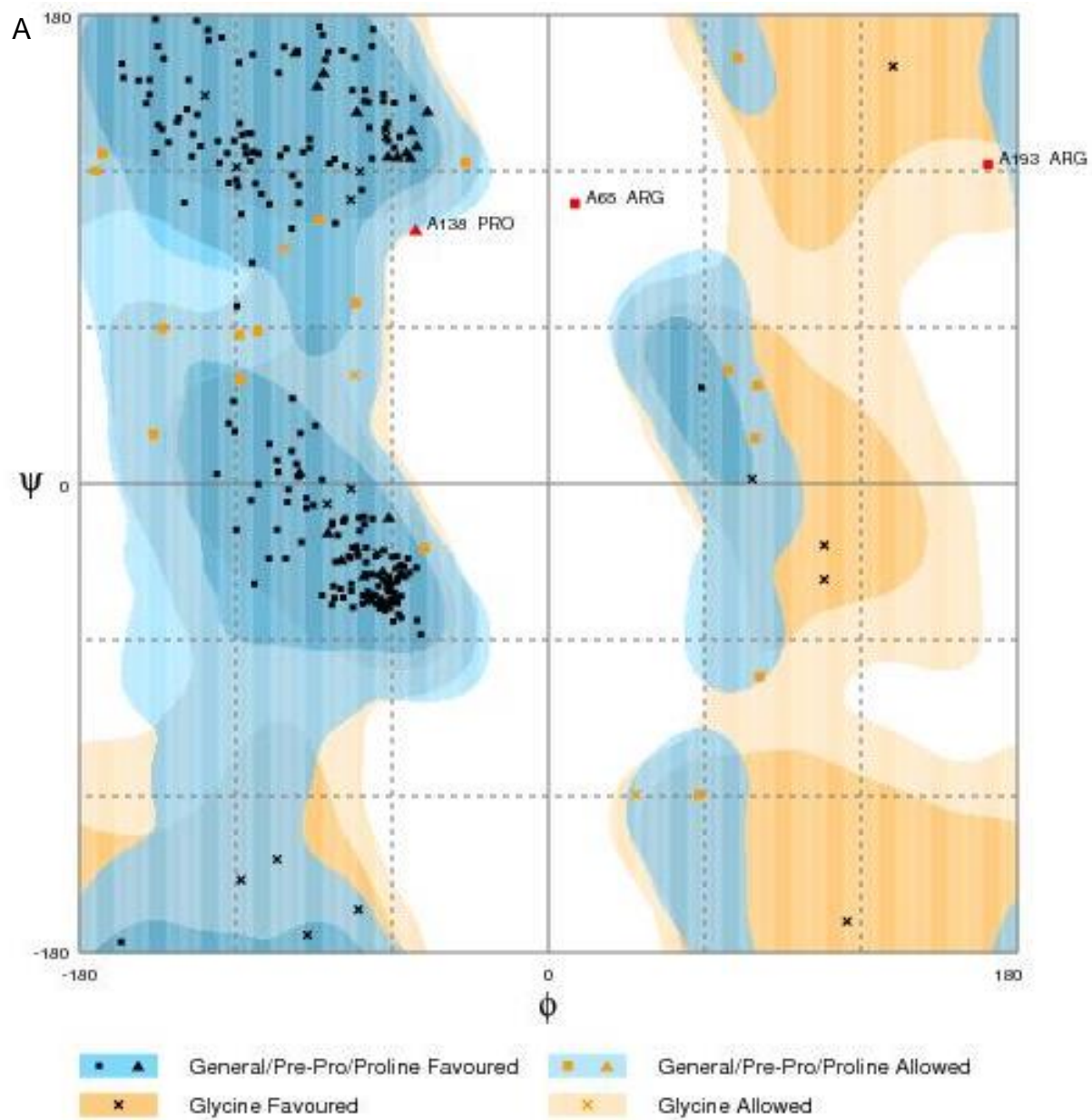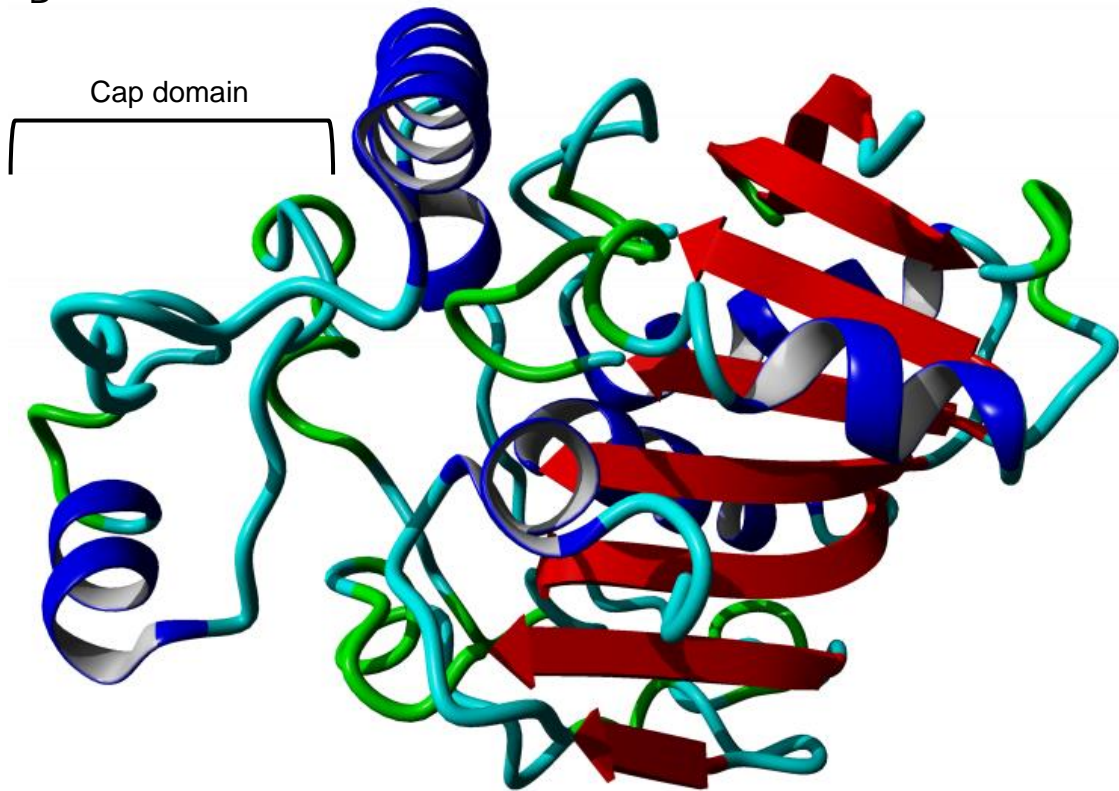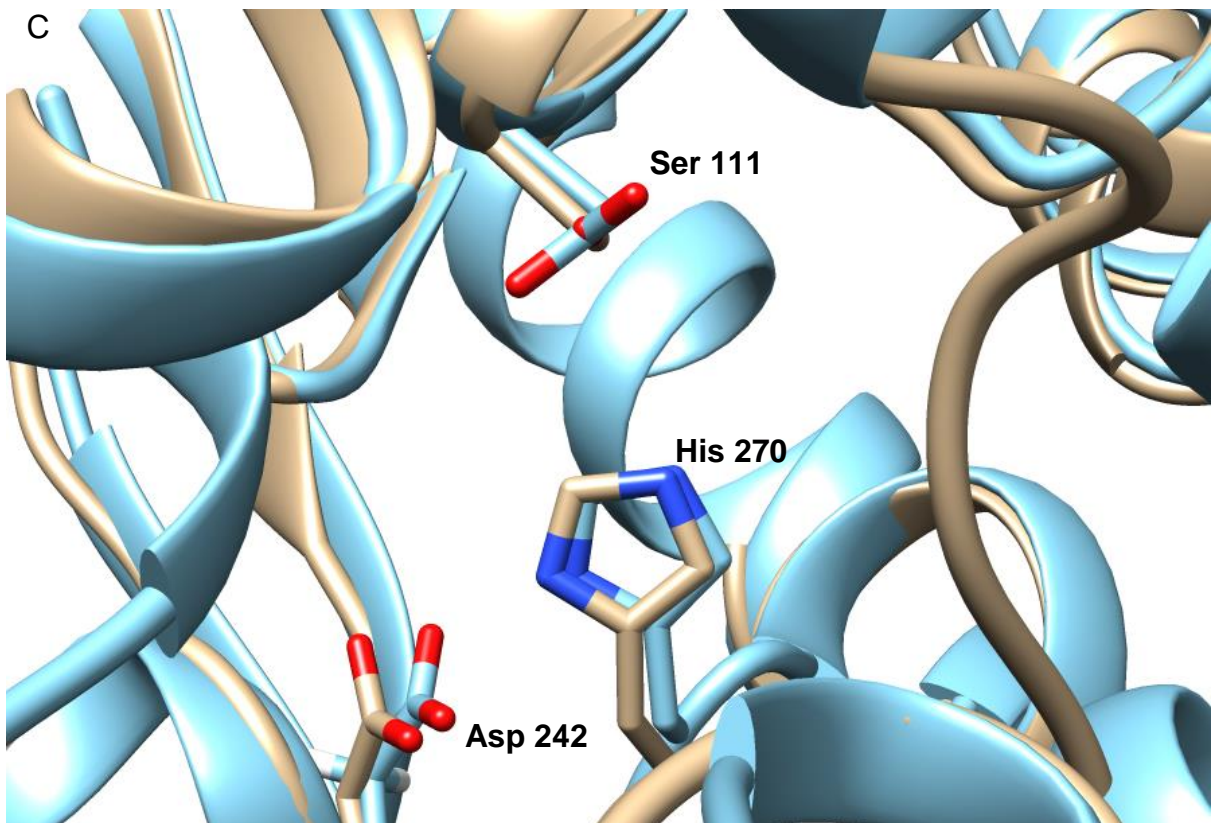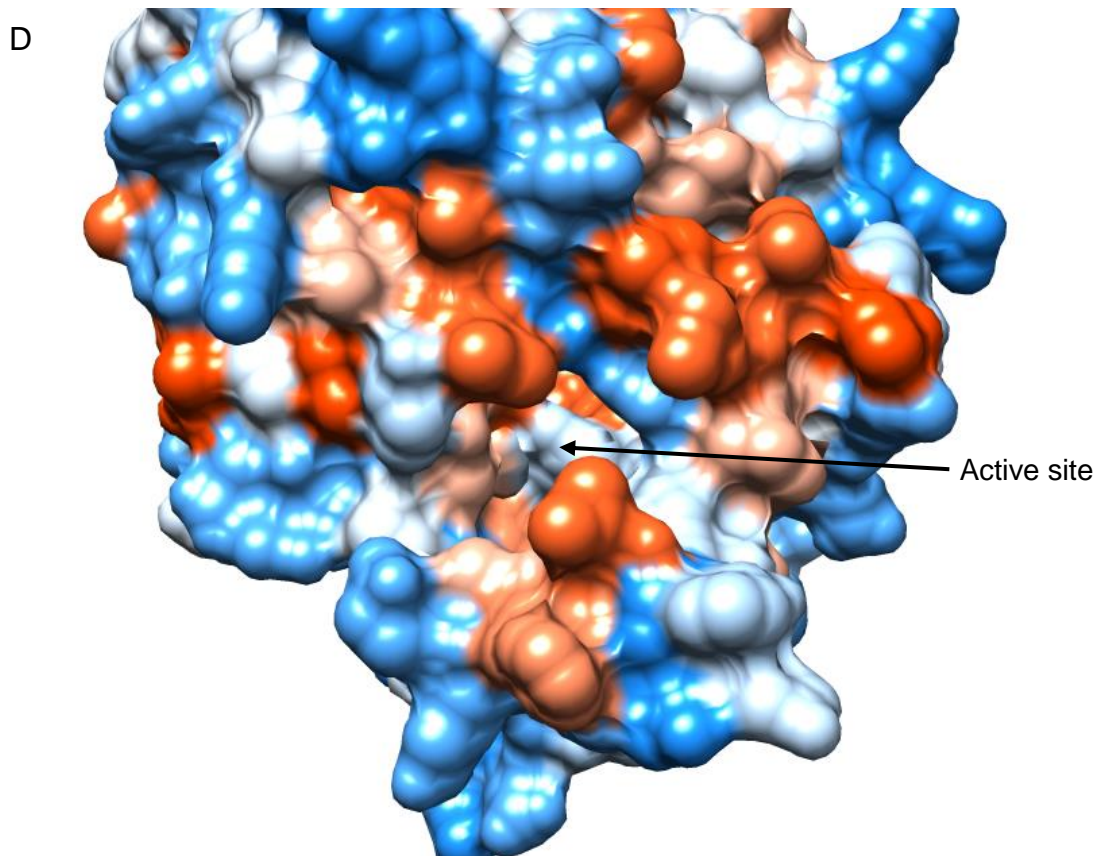
Fig. 5.1. Homology modelling of the B1 lactonase. (A) Ramachandran plot of the model. 91.4 % of the residues are in a in a favoured conformation, 7.5 % are in an allowed conformation, and 1.1 % are outliers which are highlighted in red. (B) Cartoon representation of the model generated by SWISS-PROT. The structure is a standard α/β hydrolase fold with a core of β-sheet consisting of eight β-strands (red) connected by six α-helices (dark blue). A cap domain can be seen as indicated (C) Active site histidine, aspartic acid and serine/aspartic acid residues compared between the two superposed structures, the B1 lactonase model in blue, and the 2XUA structure in brown. There were small variations in the positions of the side chains, but not enough to likely affect substrate specificity and activity from these alone. (D) Surface representation of the B1 lactonase, white is uncharged, red is negatively charged and blue is positively charged. Arrow indicates location of the active site.

Looking at the composition of amino acids with another enol lactonase from the mesophile *B. xenovorans* (PDB code 2XUA) there is a noticeable difference between the percentages of certain amino acids (Table 9). Most notably the B1 lactonase consists of 9.1 % alanine, compared to the 2XUA of 14.4 %, a difference of 5.3 % overall. With this there was also seen more leucine residues, with the B1 lactonase consisting of 14.3 % leucine and lactonase 2XUA consisting of 8.3 %, a difference of 6.0 %. This suggests that there is substitution of alanines for leucines, which would provide higher levels of hydrophobic interactions within the protein as leucine is far more hydrophobic than alanine. When the structure is examined (Fig. 5.2) this appears to be case for half of the alanine that are lost, mostly on the β-strands. But the others on α-helices are either substituted for residues that are more hydrophobic or charged residues. This is a typical adaptation with thermophilic enzymes to resist hot environments by creating more ionic interactions with the solvent facing side of the α-helices and stronger hydrophobic interactions with the core facing side. There is also 2.8 % more proline residues present in the B1 lactonase which would make the structure more rigid and stabilising it further. These extra prolines are mostly located on loop regions, stabilising them by reducing their flexibility. Cysteines are nearly absent (0.3 %) in the B1 lactonase, compared with 1.4 % in lactonase 2XUA. This is commonly observed in thermophilic proteins as cysteines are highly susceptible to oxidation at high temperatures. And while there is no indication that lysines are substituted for arginine as there is very little difference in the number of lysine present, there were more arginine seen with 9.4 % compared to 6.9 % indicating they are necessary for stabilising the enzyme. This is most likely achieved through its ability to form electrostatic interactions in three directions (Sokalingam *et al.*, 2012).These are primarily located in the loops and α-helices of the B1 lactonase.

Visual inspection of the structure with COOT revealed that there was a slight increase in the number of hydrophobic interactions seen in the structure in comparison to the epoxide hydrolase it was modelled after from *Mycobacterium thermoresistibile* (PDB code 5CW2). Approximately 10 % more hydrophobic interactions were seen in the B1 lactonase. Loop regions were seen to be shorter in B1 and had fewer glycines and more prolines in them that would increase their

stability. There was low structural homology between the two which meant there were few superposed secondary structure features to compare so it was difficult to be conclusive about the differences.

Table 9. Percentage composition of amino acids in the B1 lactonase and 2XUA.

Amino acids	2XUA	B1
Ala	14.4 %	9.1 %
Arg	6.9 %	9.4 %
Asn	2.9 %	1.4 %
Asp	6.1 %	5.2 %
Cys	1.4 %	0.3 %
Gln	2.0 %	5.6 %
Glu	6.3 %	4.5 %
Gly	7.6 %	7.0 %
His	4.3 %	2.1 %
Ile	5.1 %	3.1 %
Leu	8.3 %	14.3 %
Lys	2.9 %	2.8 %
Met	2.5 %	1.7 %
Phe	2.2 %	4.5 %
Pro	4.9 %	7.7 %
Ser	4.5 %	6.3 %
Thr	5.8 %	3.5 %
Trp	1.4 %	2.1 %
Tyr	2.5 %	2.1 %
Val	7.2 %	7.0 %
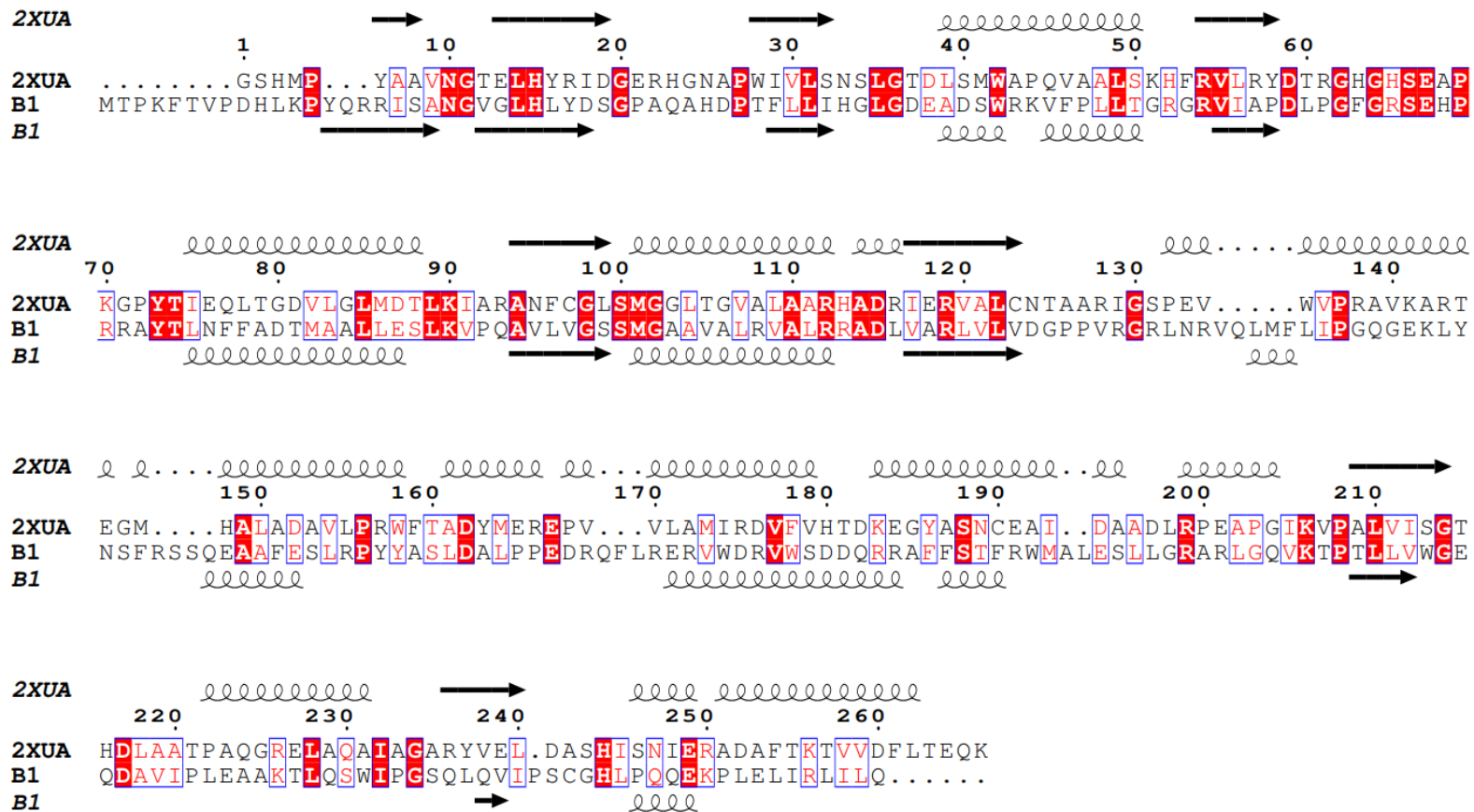
Fig. 5.2. Secondary structure and amino acid sequence alignment between the B1 lactonase and the 2XUA lactonase made using the ESPript 3 sequence and structure alignment server (Robert and Gouet, 2014). Residues highlighted in red are conserved, residues in blue boxes are conserved substitutions.

[129]

Lactonase and epoxide hydrolases are evolutionarily related and share similar structures and active site residues (Popiel *et al.*, 2014). To look into their similarities the homology model generated for the B1 enol lactonase was superposed to an epoxide hydrolase whose active site residues are known, from *Mus musculus* (PDB code 1CQZ), as the *M. thermoresistibile* epoxide hydrolase has no structure published to date. Both enzymes share 26 % sequence identity and when superposed in Chimera had a mean RMSD of 3.8 Å over all residues, showing that the overall structures of the enzymes are very similar. When the active sites were examined (Fig. 5.3) it was observed that the positions of the catalytic residues of the enol lactonase are mostly conserved in the epoxide hydrolase. Both the active site histidine and aspartic acid of the enol lactonase (His 270 and Asp 242) are conserved in the epoxide hydrolase, although the aspartic acid (Asp 495) has not been identified as being involved in catalytic activity for the epoxide hydrolase enzyme. The active site serine, Ser 111, in the lactonase has been substituted by an aspartic acid which would mean an additional negative charge in the active site, the same as for the 2XUA lactonase so it is a conserved substitution. The majority of the differences in the active site were observed when the residues identified as being involved in catalysis in the epoxide hydrolase but not the enol lactonase were compared. The tryptophan (Trp 334) in the epoxide hydrolases active site has been substituted for a methionine in the lactonase, since methionine is also a hydrophobic residue this was a conserved substitution. One of the tyrosines, Tyr 381, is located on an α-helix in the epoxide hydrolase and has been replaced by a loop instead in the lactonase that is not present in the same location as the α-helix, instead the loop moves away from the active site. The other tyrosine in the epoxide hydrolase important for activity, Tyr 465, is located on another α-helix that has also been replaced by a loop region in the lactonase. The loop is in a similar position to the α-alpha helix the Tyr 465 is located on but the residue in the equivalent location is a methionine which has its side chain pointing away from the active site, leaving the area where the tyrosine would have been located empty. Either these residues were unnecessary for activity and have been lost over time or have been lost to accommodate different substrate binding.
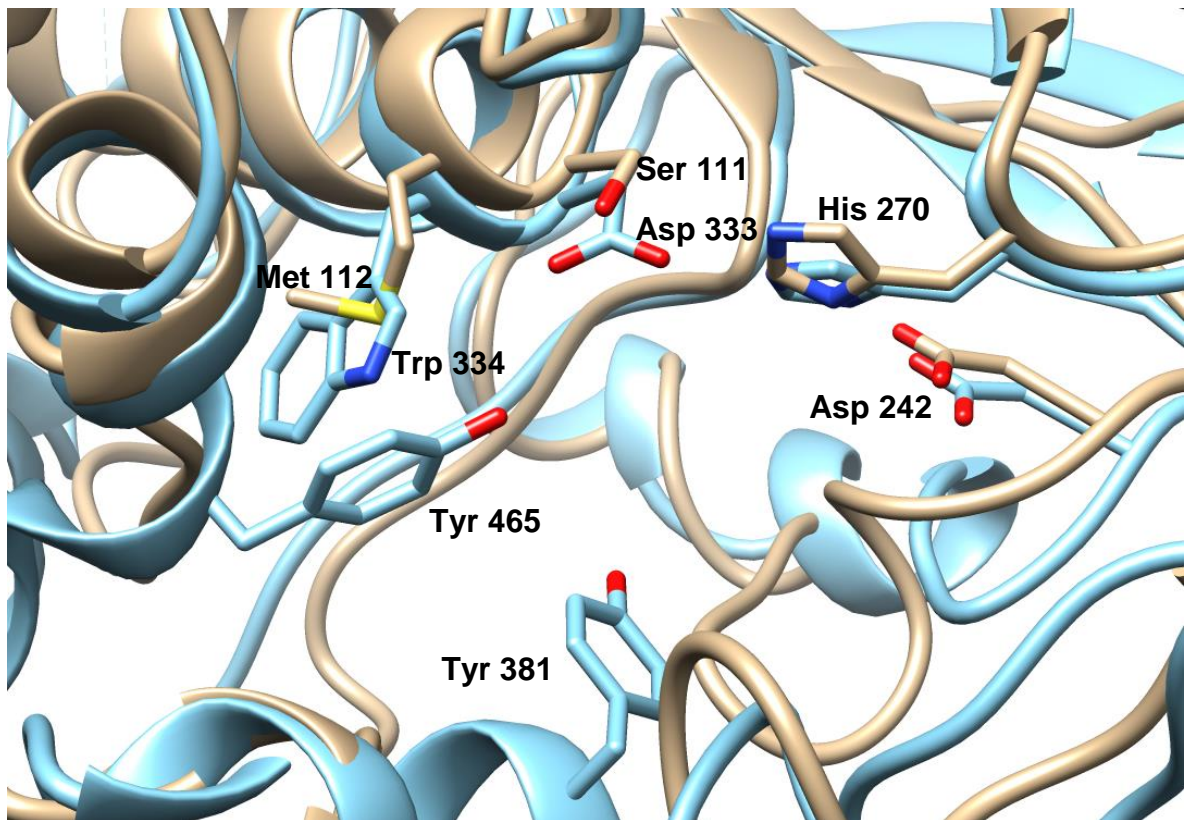
Fig. 5.3. Superposition of the active site residues of the B1 enol lactonase homology model (brown) and an epoxide hydrolase from *M. musculus* PDB code 1CQZ (blue). The catalytic histidine and aspartic acid residues from the lactonase were fully conserved in the epoxide hydrolase and there was a weakly conserved substitution of the polar serine residue for a charged aspartic acid. The active site tyrosines from the epoxide hydrolase were completely missing in the lactonase with nothing occupying the space they had been in and the epoxide hydrolase tryptophan was seen to have a conserved substitution for a methionine.

While unfortunate that the protein was not able to be purified in the soluble fraction, this is not an uncommon result when expressing recombinant protein in a foreign host. Currently it has been approximated that around 70 % of all recombinant proteins that are expressed form inclusion bodies (Yang *et al.*, 2011). The homology model itself gives some insight into the structure of the protein, with the most interesting aspect being the substitution of one residue of the catalytic triad. Without any soluble protein it will be impossible to test if this gives a difference in activity. A possibility would be to unfold and refold the insoluble protein, to attempt to get a properly folded and soluble protein. This was not done during this project as it is a time intensive experiment.

5.3 Summary

With the difficulties encountered trying to purify the protein in the soluble fraction a homology model was generated for the B1 enol lactonase. The model showed strong similarities to the enol lactonase from *B. xenovorans* (PDB code 2XUA) in terms of overall structure and active site residues. It was also shown to be similar to an epoxide hydrolase from *M. musculus* where although half of the active site residues from the epoxide hydrolase were not present in the enol lactonase, there was conserved substitution of a tryptophan and an aspartic acid for a methionine and a serine respectively in the enol lactonase and conservation of a histidine present in both enzymes.

With the last class of lactonase investigated, the next step of the project was to look at another class of thermophilic enzymes of great industrial interest to our collaborators, DXP synthases.

6. DXP synthase

6.1 Introduction

1-deoxy-D-xylulose-5-phosphate (DXP) synthases were investigated because of the close relationship with transketolase enzymes. Transketolases are found in the pentose phosphate pathway and biologically are involved with the rearrangement of sugar phosphates by transferring a two carbon unit from one sugar and transferring it to another using the cofactor thiamine pyrophosphate. The transketolases are used synthetically for industrial processes where they able to use hydroxypyruvate as a ketol donor with a variety of aldehydes to synthesise sugar products such as erythrulose. The release of CO_2 makes the reaction irreversible (Hobbs *et al.*, 1996). The DXP reaction catalyses the conversion of D-glyceraldehyde 3-phosphate and pyruvate to 1-deoxy-D-xylulose-5-phosphate and CO_2 (Demuynck *et al.*, 1991) with a very narrow specificity for sugar substrates. The industrial interest is generated from the fact that DXP synthases use pyruvate as a substrate in the catalysed reaction as opposed to hydroxypyruvate for transketolase which is a more expensive substrate. The goal was to solve the structures of the native DXP synthases, as well as substrates bound to the active site and a transition state, to determine what features allow these enzymes to use the substrate pyruvate and not hydroxypyruvate. This would further the goal of designing a transketolase enzyme that would not only have broader substrate specificity but could use the cheaper pyruvate as a substrate rather than hydroxypyruvate for the reaction. The reason thermophilic DXP synthases were investigated is because they would be more stable for other industrial applications. These enzymes should be also be less flexible so could hopefully crystallise more easily. The two structures currently available for DXP synthases are of limited resolution, one structure at 2.9 Å and the other at 2.4 Å.

6.2 Materials and methods

6.2.1 Resuspension of gBlock gene fragments

Synthetic genes ordered from Integrated DNA Technologies (IDT) (Iowa, United States) were provided as dry pellets and were resuspended in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) to a final concentration of 10 ng μl^{-1} of DNA. The solution was vortexed, followed by incubation at 50 °C for 20 min and then vortexed a second time to fully resuspend the pellets.

6.2.2 Buffers

Cell paste resuspension buffer was 50 mM HEPES pH 7.5 and 500 mM NaCl, elution buffer for nickel chromatography was 50 mM HEPES pH 7.5 and 1 M imidazole, and gel filtration buffer was 50 mM HEPES pH 7.5, 100 mM NaCl, 2 mM $CaCl_2$ and 50 µM thiamine pyrophosphate.

6.2.3 Protein sample preparation

Protein samples were concentrated to a range of concentration between 2 mg ml^{-1} and 22 mg ml^{-1} in a Merck Millipore centrifuge concentrator with a 50 kDa molecular weight cut off at 4000 rpm.

6.3 Results and discussion

Two DXP synthases were identified from genomes of two thermophilic organisms, one from an anaerobic Gram-negative bacterium, *Thermovibrio ammonificans* (TaDXP) and the other from an anaerobic Gram-positive bacterium, *Carboxydothermus hydroformans* (ChDXP). These were identified through a BLAST search (Altschul *et al.*, 1990), TaDXP was found to have 50 % amino acid sequence identity to a DXP synthase from *E. coli* (accession code 2O1S_A) and ChDXP was found to have 48 % sequence identity to a DXP synthase from *Pseudomonas aeruginosa* (accession code 2O1X_A) and they were shown to possess 57 % sequence identity to each other using Clustal Omega (Sievers *et al.*, 2011).

Genomic DNA from both organisms was amplified through use of a GenomiPhi DNA amplification kit, then the DXP synthase genes were amplified by PCR and inserted into both pLATE31 and pLATE51 vectors, the former of which has a C-terminal histidine tag sequence and the latter an N-terminal, to increase the chances of having the tag exposed and not buried within the structure for binding to a nickel affinity resin. These were then successfully cloned in *E. coli* XL10 gold cells and the amplified plasmid constructs were transformed for protein expression into *E. coli* BL21 (DE3) pLysS cells. Expression studies showed no protein in either the soluble or insoluble fraction. Analysis of the coding regions of the genes revealed many rare codons; 24 % of *T. ammonificans* and 17 % *of C. hydroformans* DXP synthase codons were below the 20 % codon usage threshold for *E. coli* (Maduro, 2014) (*E. coli* Codon Usage Analyzer 2) with several instances of consecutive rare codons in each gene which could cause translation to arrest. This is always a possibility when expressing a gene from a foreign organism in *E. coli*, particularly as distinguishable patterns of codon usage have been reported between thermophiles and mesophiles; it has been hypothesised that these differences are due to selective pressure on thermophilic species towards more stable mRNA molecules and thus better translation efficiency under high temperatures (Lynn, 2002). So, both genes were cloned in BL21-CodonPlus (DE3) cell lines which have been designed to express proteins containing more rare codons. However even this proved ineffectual to express either protein, so the coding sequences were optimised for expression in *E. coli*

and the codon-optimised genes were synthesised by Integrated DNA Technologies (IDT) (California, USA). The recombinant DNA was successfully cloned into XL10-Gold cell lines (Fig. 6.1).

A variety of expression trials were conducted for both N- and C-tagged constructs to determine the optimal conditions for protein expression, including expression in *E. coli* cell lines BL21 (DE3) pLysS, Rosetta, RIPL and Arctic Express. The concentration of IPTG used to induce expression was varied in different expression trials between 0.2 mM and 1mM. The expression with autoinduction media was also investigated, as well as altering the temperature after induction between 12 °C and 37 °C. For most trials the majority of the protein was present as inclusion bodies in the insoluble fraction. Overall it was not possible to obtain a significant amount of the *C. hydroformans* DXP synthase in the soluble fraction, with only a faint band visible at the corresponding molecular weight on SDS-PAGE gels (Fig 6.2). The DXP synthase from *T. ammonificans* was however successfully expressed in soluble form by induction with 1 mM IPTG and incubation for 4-5 hours at 37 °C before being cooled to 4 °C and left overnight before harvesting the cells. Yield was further improved by using a high salt resuspension buffer containing 500 mM NaCl; high salt concentration has been shown to improve the stability of thermophilic proteins (Mao *et al.*, 2007). The focus of study was accordingly shifted to purification and crystallisation of the DXP synthase from *T. ammonificans*.

TaDXP was successfully purified by nickel affinity chromatography and gel filtration (Fig 6.3). Confirmation that the protein purified was indeed the DXP synthase was confirmed using Western blotting which showed it possessed a histidine tag. It was found that the protein with a C-terminal histidine tag bound to the nickel column whereas the N-terminal histidine tagged protein did not indicating that the N-terminal tag was not exposed and most likely buried within the structure.
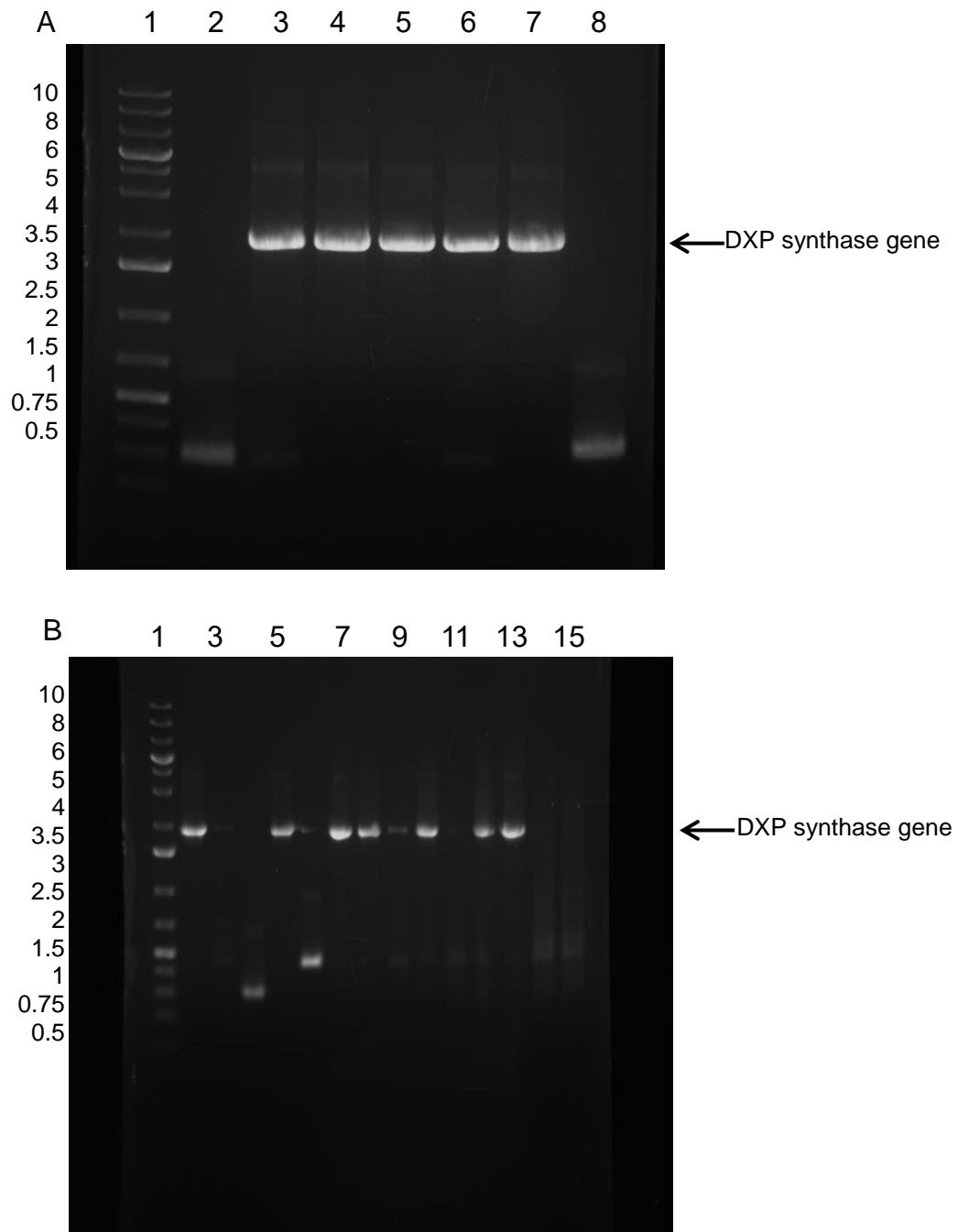
Fig. 6.1. DNA agarose gels of ChDXP gene amplification. A: ChDXP colony PCR. Lanes: 1 ladder (in kb); 2-8 colonies for gene in pLATE 51. B: TaDXP colony PCR. 1; ladder (in kb), 2-7; colonies for gene in pLATE 31, 8-16; colonies for gene in pLATE 51.
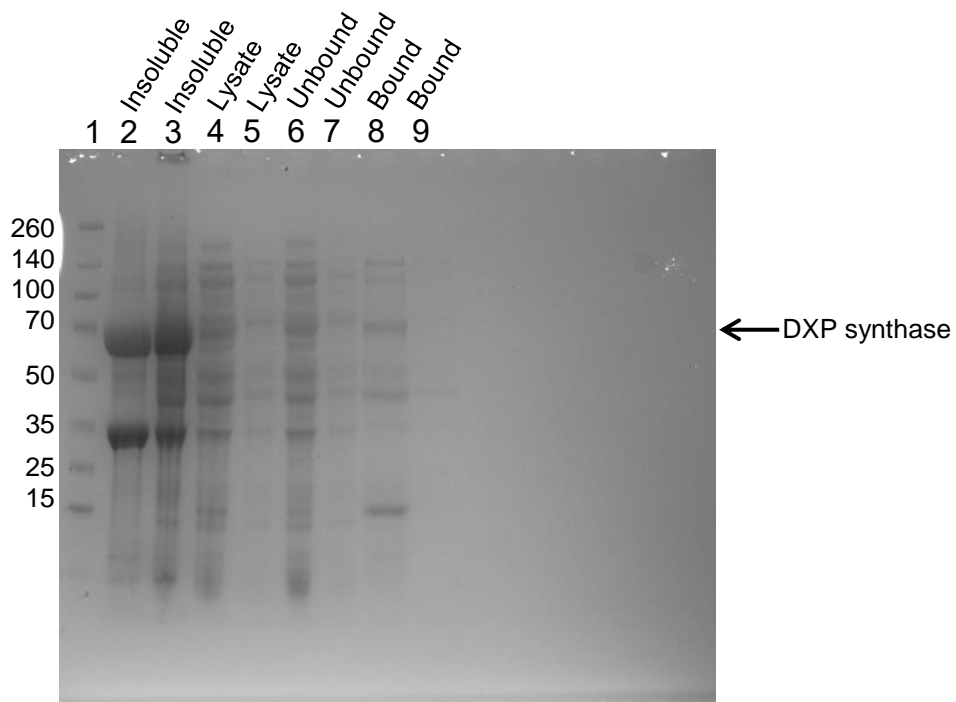
Fig. 6.2. A SDS-PAGE of ChDXP from nickel affinity chromatography. Lanes: 1. Ladder (in kDa); 2-3. Insoluble fraction; 4-5. Crude lysate; 6-7. Unbound fraction; 8-9. Bound fraction. Band intensities reveal the majority of the protein is present within the insoluble fraction.

[139]

A

B
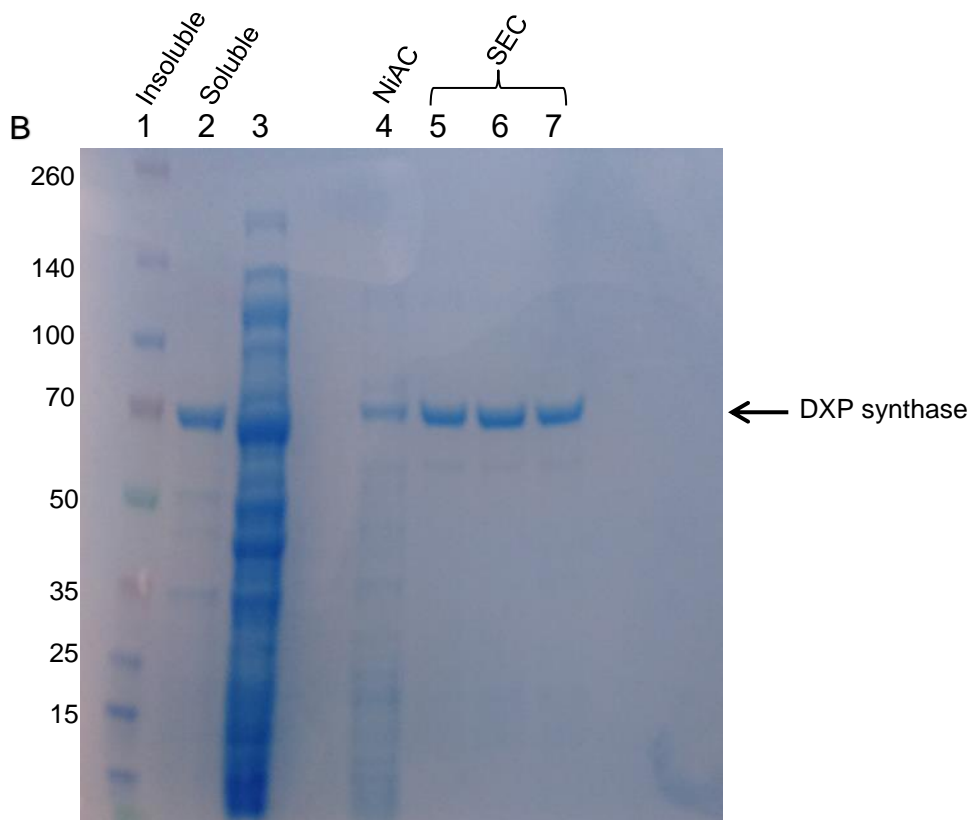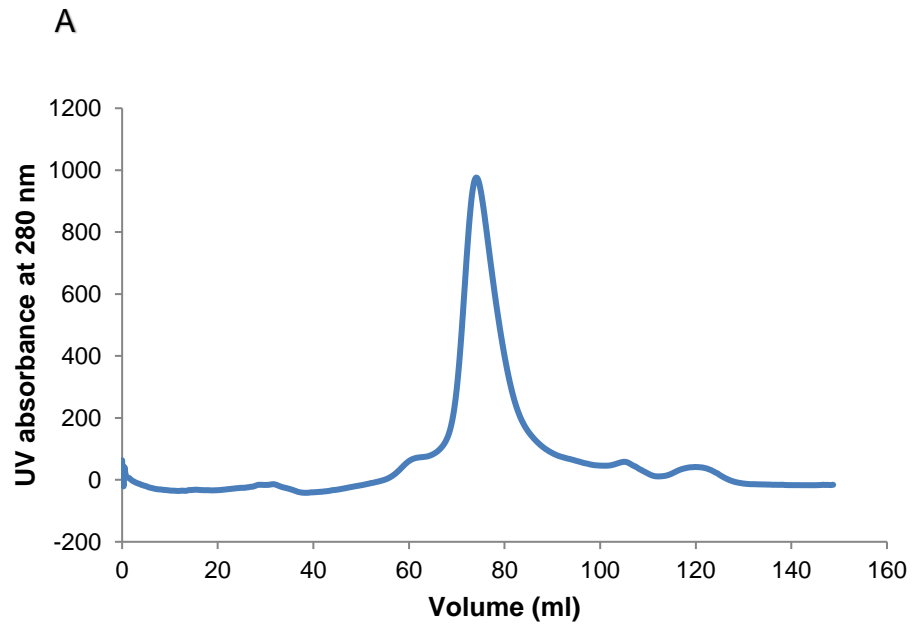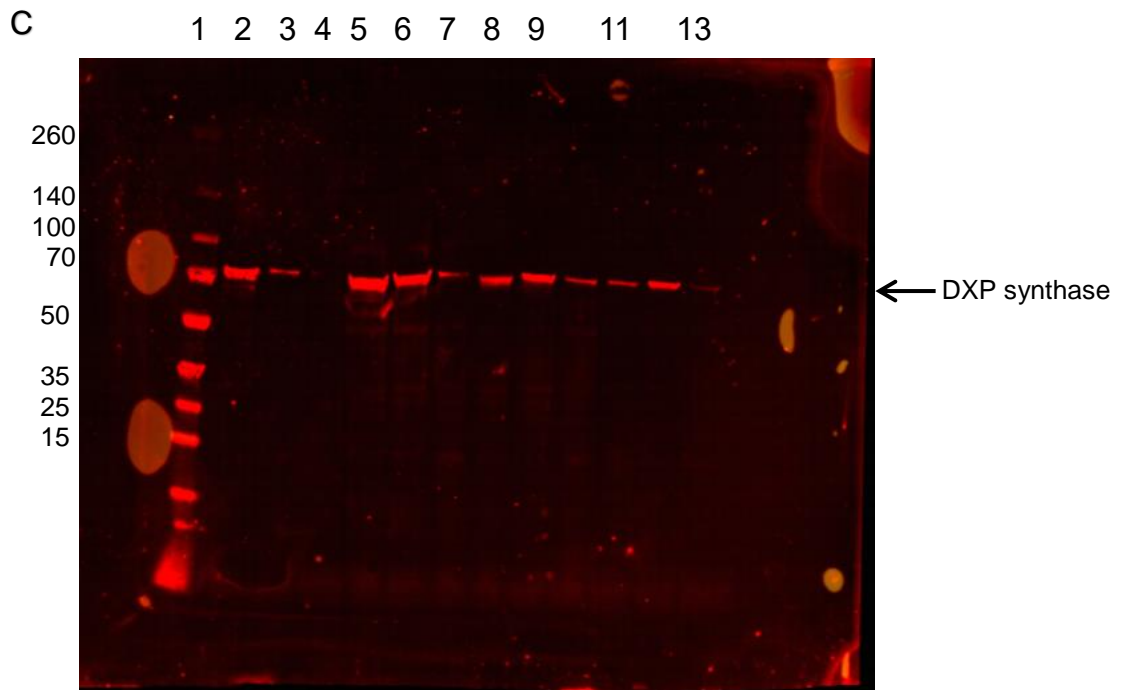
Fig. 6.3. (A) Size exclusion chromatography trace for the purification of TaDXP protein. (B) SDS-PAGE of TaDXP from nickel affinity and gel filtration chromatography. Lanes: 1. Ladder (in kDa); 2. Insoluble fraction; 3. Supernatant; 4. Elution fraction from nickel affinity chromatography (NiAC); 5-7. Peak fraction from gel filtration chromatography (C). Western blot of SDS-PAGE analysis of gel filtration chromatography peak. Bands represent proteins possessing a histidine tag confirming purification of the tagged DXP synthase protein. An identical protein ladder to the SDS-PAGE gel was used and all marker proteins possessed a histidine tag.

There was no straightforward way to perform an activity assay on the purified enzyme. Currently the only known assay available is a coupled-enzyme assay (Altincicek *et al.*, 2000). This current method uses pyruvate and D-glyceraldehyde 3-phosphate, which are converted to 1-deoxy-D-xylulose 5-phosphate. It is here that a second enzyme is required, DXP reductoisomerase, to convert the product to 2-C-methyl-D-erythritol 4-phosphate in an NADPH-dependent reaction. The conversion of NADPH to NADP$^+$ can be monitored spectrophotometrically at 365 nm. The main issue preventing the same assay being performed here is that DXP reductoisomerase is not commercially available. To acquire it would require synthesising the gene synthetically, cloning, expression and purification of the enzyme to carry out the assay. It was not considered cost effective nor time efficient to do this at this time for this project.

Differential scanning fluorimetry was performed on the purified protein to determine its thermostability at a range of different temperatures to determine the temperature where it denatures (Fig. 6.4). A standard buffer of 50 mM HEPES pH 7.5 and 100 mM NaCl was used. Two clear fluorescence peaks were observed giving T_m values at 57.0 °C and another at and 84.5 °C. DXP synthase and transketolase monomers contain three domains (Nikkola *et al.*, 1994; Xiang *et al.*, 2013). The peaks are most likely to correspond to the unfolding of the different domains and simultaneous dissociation of the dimer into monomers.
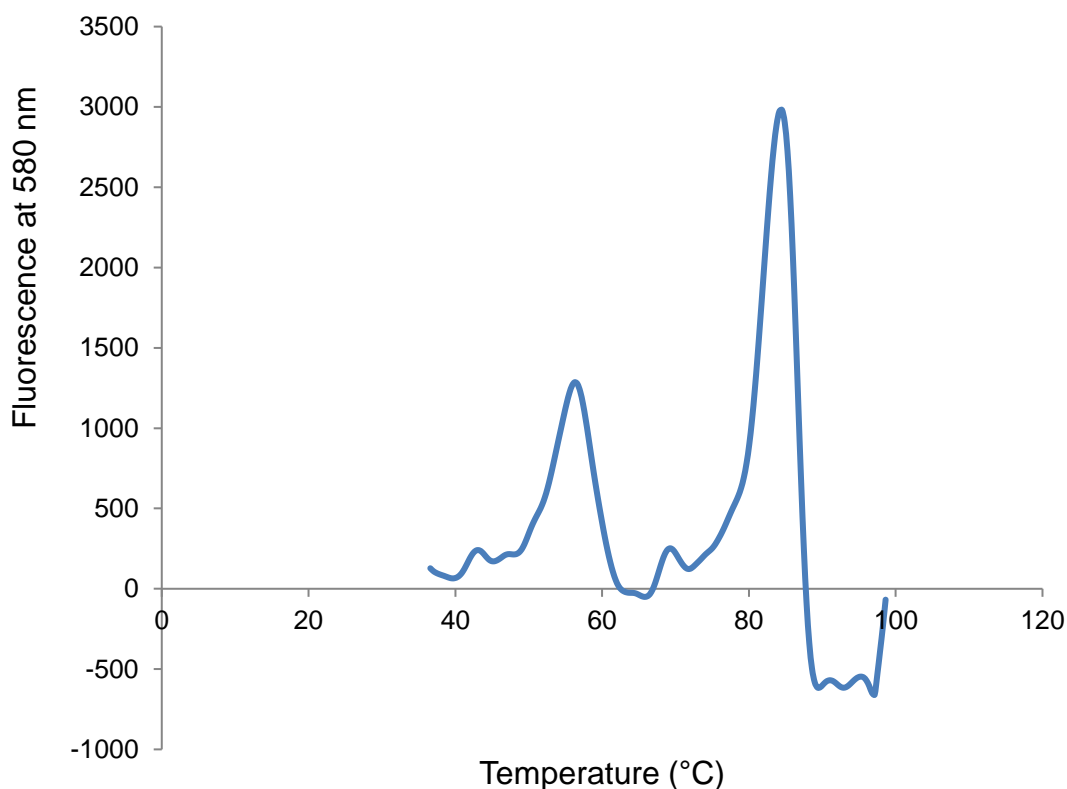
Fig. 6.4. Differential scanning fluorimetry for TaDXP using SYPRO Orange. Peaks indicate exposure of hydrophobic regions due to denaturation. Peaks are located at 57.0 °C and 84.5 °C. The peak at 57.0 °C most likely represents unfolding of the protein tertiary structure. The second peak is most likely unfolding of the secondary structure.

After successful purification the protein was concentrated in preparation for crystallisation trials. A 50 kDa cut off concentrator was used due to the large size of the protein for shorter centrifugation and to further reduce the chances of contaminants. The highest concentration of TaDXP protein that could be achieved before precipitation occurred was around 25 mg ml^{-1} so initial crystallisation trials by microbatch screening were conducted with 22 mg ml^{-1} protein concentration with JCSG-plus. Immediate precipitation was seen in most of the wells so concentrations were lowered and tested at 20, 16, 12, 8, 6, 4 and 2 mg ml^{-1} to try and find a concentration that produced cleared wells instead of precipitate to give an indication of what concentration to use for other crystal screens. Unfortunately the majority of wells in the microbatch screens, even those at the lowest concentration of 2 mg ml^{-1} showed precipitate. This was unusual when taking into consideration the maximum concentration this protein could achieve indicating the protein to be reasonably soluble. So crystallisation trials were repeated with the other screens in our possession: MIDASplus, Morpheus

II, SG1, MultiXtal and Stura FootPrint, Structure Screen 1 + 2, The PGA Screen and Proplex. The majority of wells from these conditions had precipitated. Crystals that had the appearance and properties of protein crystals were found in the MultiXtal screen produced with protein at 12 mg ml^{-1} in the following conditions: 0.5 M sodium chloride, 0.1 M sodium phosphate pH 7 and 30 % v/v PEG 300. Three crystals were frozen straight from the droplet as the high concentration of PEG 300 meant they were already cryo-protected. The remaining crystals in the well were used for seeding. It was not possible to confirm whether the crystals frozen were protein before creating seeds due to a three month wait for the next beam time at the Diamond Light Source Synchrotron. So, rather than wait, the seed stocks were used immediately to propagate more crystals and microbatch screening with seeding was conducted. The first screens tested were MultiXtal from which the protein originated from and JCSG-Plus. The official protocol from Douglas Instruments, who supplied the seed-making kit, called for 50 µl of stock solution to be used, however this caused small clusters of many crystals to form in the wells indicating that there was far too high a concentration of seeds within the solution. So the stock of crystal seeds was diluted tenfold with more of the stock solution. Crystals were found and frozen from the JSCG-plus screen in: 0.1 M HEPES pH 7.5, 10 % w/v PEG 8000 and 8 % v/v ethylene glycol; and 0.2 M lithium sulphate, 0.1 M Tris-HCl pH 8.5 and 40 % v/v PEG 400. None of these crystals provided any X-ray diffraction data and no further usable crystals were obtained from this method despite testing the other screens and reducing the protein concentration to 10 mg ml^{-1} in an attempt to create conditions with the protein concentration in the metastable phase for the seeds to propagate crystals. When the first three crystals were analysed for diffraction at the Diamond Synchrotron the data obtained was inconclusive since no spots were seen in the 3-5 Å range so it wasn't salt but none were seen in the 15-20 Å range either so it may not have been disordered protein either. It was not possible to tell if the crystals were either protein or salt from the data.

A further strategy adopted was to try to crystallise the enzyme in the presence of molecules that would increase stability and improve chances of obtaining an ordered crystal. The buffer used for gel filtration and crystallisation was modified by addition of the cofactor, thiamine pyrophosphate, to 50 µM. As well as this, 2 mM calcium chloride was added to the buffer solution, the rationale for this is

that in the structure of the transketolase from *E. coli* there is a calcium ion stabilising a loop close to the cofactor thiamine pyrophosphate, in the structure and there is a high probability based on the fact both enzymes share related structures that the same could be seen here (Littlechild *et al.*, 1995). Microbatch screening yielded crystals at protein concentrations of 12 mg ml^{-1} and 16 mg ml^{-1} in the same conditions as mentioned above with the first screening trials and seeding. In addition, crystals were also seen in the conditions listed in Table 10. Like the other crystals tested, data collection at the Diamond Synchrotron was inconclusive. No further attempts to crystallise were conducted due to time constraints.

Table 10. Crystallisation conditions for TaDXP.

Screen	Well	Ligand	Buffer	pH	Precipitant
MultiXtal	C3	0.5 M sodium chloride	0.1 M sodium phosphate	7	30 % v/v PEG 300
JCSG-Plus	F4	0.2 M calcium chloride dihydrate	0.1 M MES	6.5	30 % v/v PEG 400
	B4	None	0.1 M HEPES	7.5	10 % w/v PEG 8000, 8% v/v ethylene glycol
	D7	0.2 M lithium sulphate	0.1 M Tris-HCl	8.5	40 % v/v PEG 400
	E11	0.16 M calcium acetate	0.08 M sodium cacodylate	6.5	14.4 % w/v PEG 8000, v/v 20 % glycerol
SG1	G10	0.2 M sodium fluoride	None	-	20 % w/v PEG 3350

Since no X-ray structure could be obtained for the TaDXP homology modelling was carried out. The best model produced for the TaDXP protein, based on the Ramachandran plot made by Rampage, came from Phyre2 and was calculated to have a good QMEAN of −2.44. The Ramachandran plot (Fig. 6.5 A) showed 92.6 % of all amino acids in a favoured conformation, 5.4 % in an allowed conformation, and the remaining 2.0 % as outliers. These outliers were identified as: Ser 7, which is the first amino acid of a β-strand; Thr 37, which is on a short loop connecting two anti-parallel α-helices; Gly 38,Gly 112 and Gly 294, considering the flexibility of glycine it is not uncommon to find them in unusual conformations; Gln 305, in the middle of a loop with the side chain facing the solvent; Pro 307, the last amino acid of a loop before an α-helix; Met 334, which is located in the binding pocket where unusual conformations are common in enzymes; Arg 457, on a loop with the side chain facing the solvent, alternative conformations cause a steric clash; Glu 475, on a loop facing the solvent; and finally Asp 493, the first amino acid of a β-strand. The TaDXP possessed an α/β

fold with parallel β-sheets located in the centre of each domain surrounded by α-helices (Fig. 6.5 B). Superposition with the closest homologue identified from a BLAST search of the PDB, a DXP synthase from *D. radiodurans* (PDB code 2O1X) with 43 % sequence identity, gave a RMSD of 1.3 Å. The side chains of the divalent coordinating amino acids identified through Clustal Omega, Asp 143 and Asn 172, aligned well with the equivalent side chains in the DXP synthase from *D. radiodurans* to the extent that there was no visible difference (Fig. 6.5 C). The side chains of the amino acids involved in substrate binding, His 40, Tyr 380, Asp 415 and Arg 465, also showed no visible difference between the superposed models (Fig. 6.5 D). Fig. 6.5 E shows the surface charge of the dimer interface of an individual TaDXP subunit. At the interface between the subunits the surface is mostly charged with little hydrophobicity. This suggests the dimer is stabilised primarily by polar interactions.

A

B

C
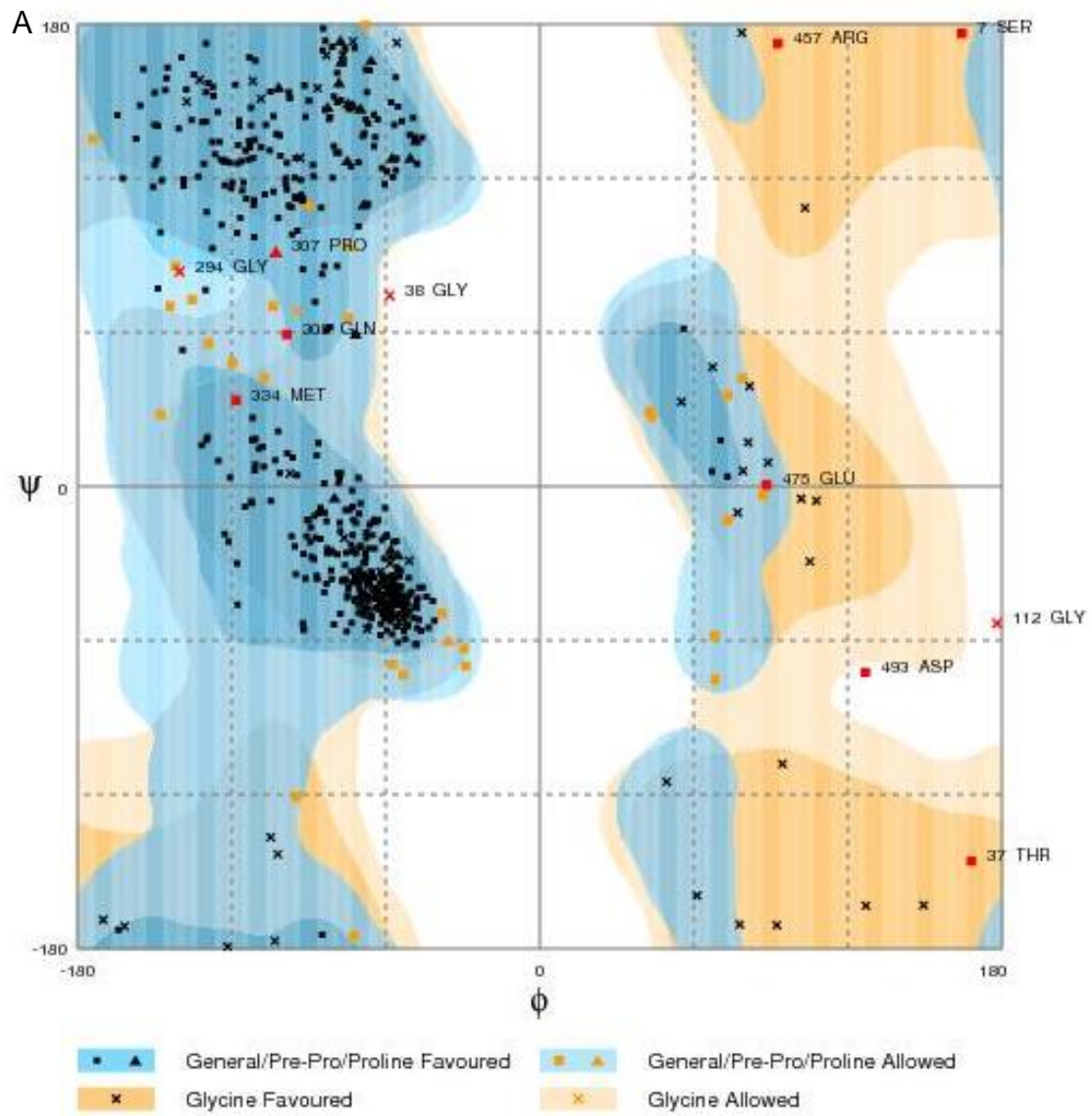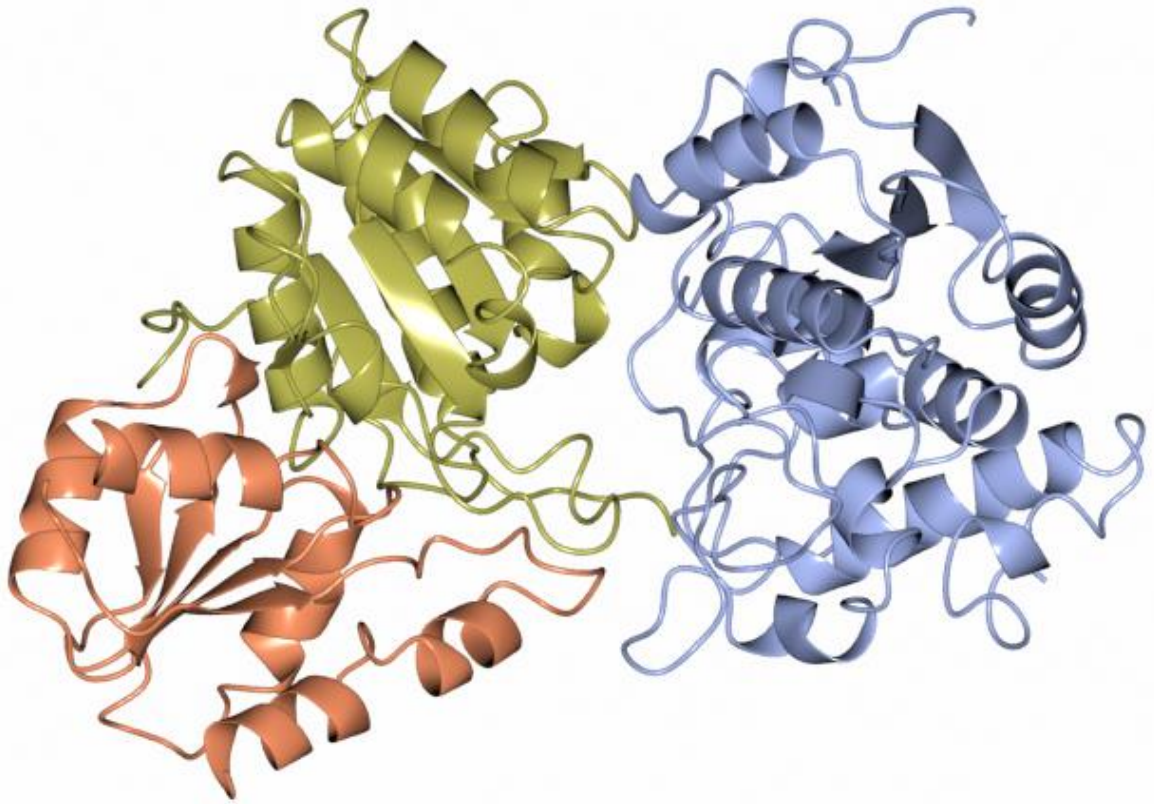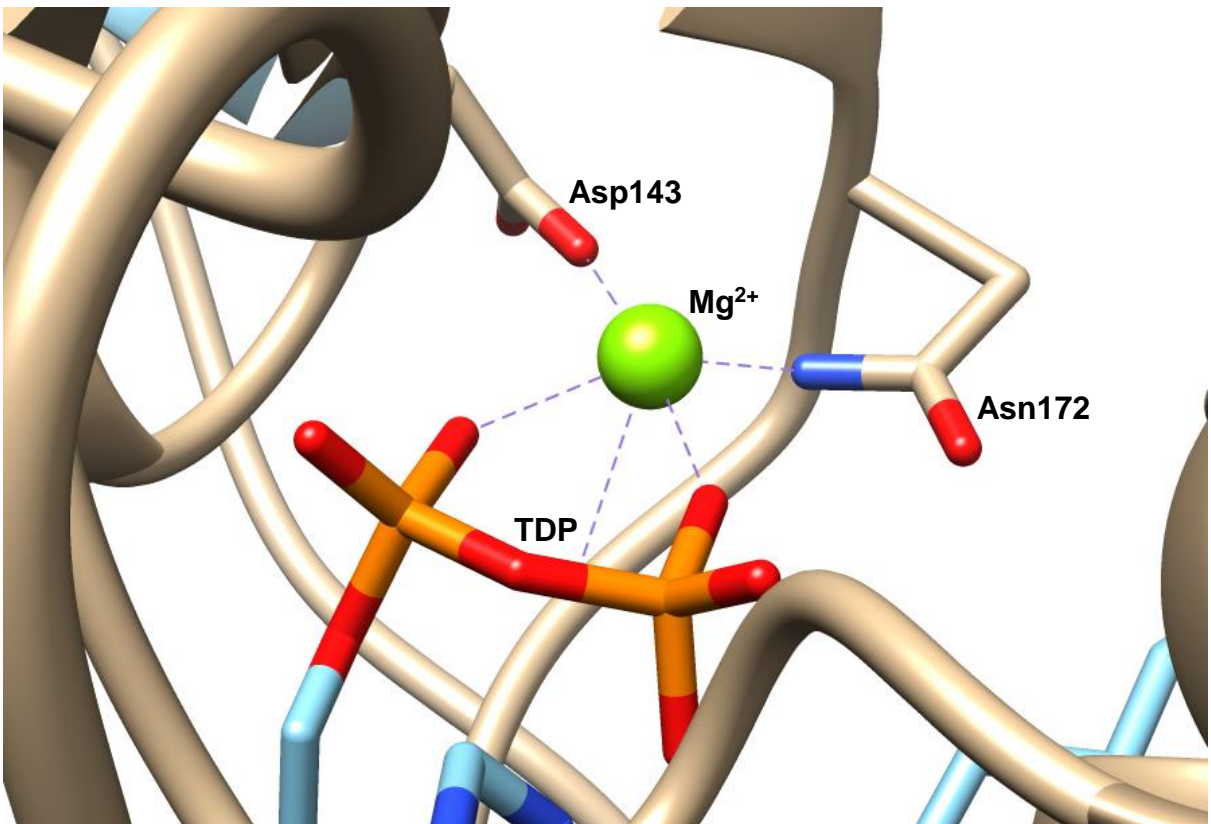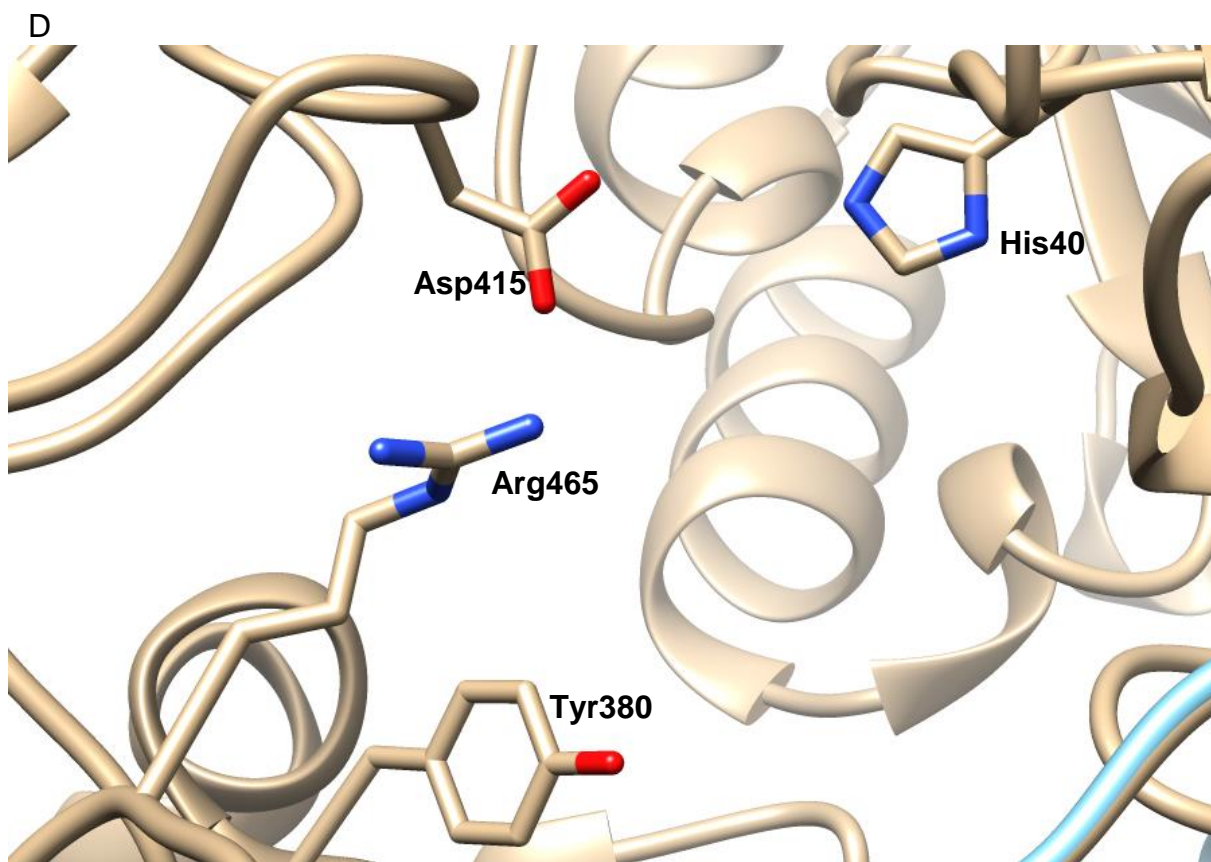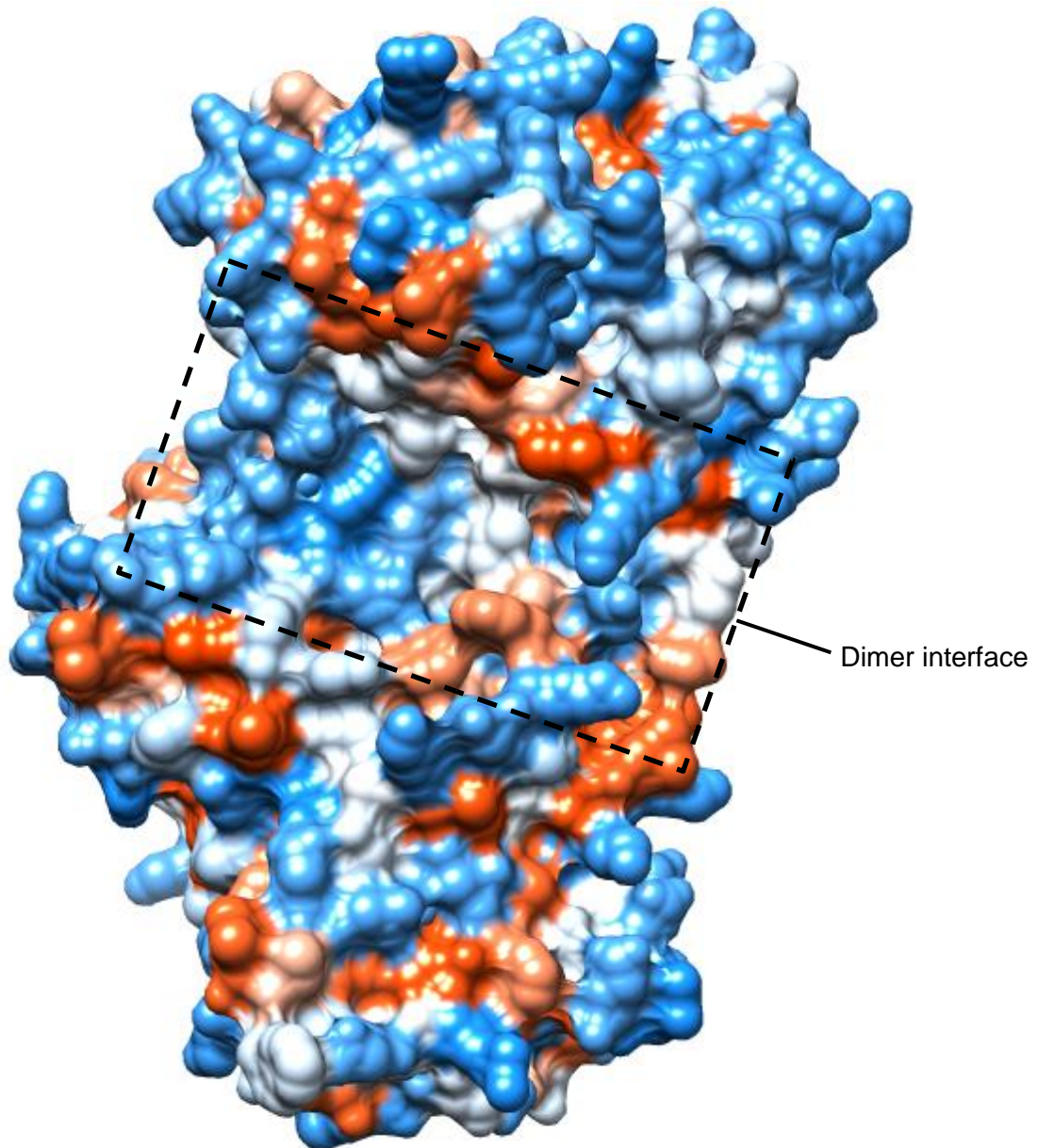
Fig. 6.5. Homology modelling of TaDXP. (A) Ramachandran plot of the model, 92.6 % of residues in a favoured conformation, 5.4 % in an allowed conformation, and 2.0 % as outliers highlighted in red. (B) Structure of the model generated by Phyre2 with each different colour corresponding to domains, made using CCP4mg. It possesses an α/β fold with parallel β-sheets located in the centre of each domain. (C) Divalent coordinating residues Asp 143 and Asn 172, superposed with DXP synthase from *D. radiodurans* (PDB code 2O1X, model in brown, 2O1X in blue) residue side chains were aligned identically. (D) Superposed substrate binding site residues of TaDXP and 2O1X, showing that the position of residues involved in substrate binding is once again conserved in comparison to the known structure. Overall RMSD between the two structures was 1.3 Å. (E) Surface charge of a subunit in the dimer. Blue is positively charged, red is negatively charged, white is hydrophobic, highlighted box shows location of dimer interface

.

Homology modelling was also carried out for ChDXP since no soluble protein could be produced for crystallisation studies. The best model produced for the ChDXP protein, determined by the Ramachandran plot generated by Rampage, also came from Phyre2. It had a good QMEAN of −1.98. The Ramachandran plot (Fig. 6.6 A) showed 92.9 % of all amino acids in a favoured conformation, 5.0 % in an allowed conformation, and the remaining 2.1 % as outliers. These outliers were identified as: Leu 8, which is near the active site where unfavourable bond angles are commonly present; Asn 38, also near the active site; Gly 38,Gly 114 and Gly 296, as glycine is the most flexible amino acid it is common to find them in unusual conformations; Pro 307, which is near the end of β-strand acting as structural disruptor to allow it to progress to a loop; Met 340, located in the active site; Thr 480, located on an external flexible loop region; Asp 498, which is the last amino acid on a loop before a β-strand; Pro 550 and Pro 590, again these are the last amino acids on loops before another β-strand; and Leu 617, the first amino acid of a loop after an α-helix.

The ChDXP possessed an α/β fold with parallel β-sheets located in the centre of each domain surrounded by α-helices (Fig. 6.6 B). Superposition with the closest homologue identified from a BLAST search, found to be the same DXP synthase as before from *D. radiodurans* (PDB code 2O1X), with 46 % sequence homology, gave a mean RMSD of 0.5 Å over all residues (Fig. 6.6 B). The side chains of the divalent coordinating amino acids identified through Clustal Omega, Asp 144 and Asn 173, aligned well with the equivalent side chains in the DXP synthase from *D. radiodurans* and, like the TaDXP model, there was no visible difference observed between the side chains (Fig. 6.5 C). The side chains of the amino acids involved in substrate binding – His 41, Tyr 386, Asp 421 and Arg 471 – also showed no visible difference between the superposed models (Fig. 6.5 D). A total of 10 salt bridges within each subunit were identified within the structure of the ChDXP model and 12 in the TaDXP model. These were actually fewer than in the mesophilic DXP synthase which possessed 18.

The similarities between the two homology models made and the DXP synthase from *D. radiodurans* in terms of the active site residues, where there is no visible difference between any of the superposed structures, suggests that the residues themselves and their positions within the active site are highly conserved

between DXP synthases. To assess this hypothesis, the only other DXP synthase in the PDB at this time was superposed with the homology model generated for TaDXP. This DXP synthase was isolated from *E. coli* (PDB code 2O1S) and when superposed with the TaDXP model had a RMSD of 2.1 Å showing that there was little difference between the overall structures of the two enzymes. Analysis of the active site showed that the divalent coordinating residues were conserved: for the *E. coli* DXP synthase these were Asp 152 and Asn 181, which did not vary much in position from the TaDXP model (Fig. 6.7 A). As for the substrate binding site (Fig. 6.7 B), the amino acids involved were again conserved and the equivalent amino acids in *E. coli* DXP synthase were His 49, Tyr 392, Asp 427 and Arg 478. None of the amino acid side chain positions varied significantly in the two superposed models, apart from the arginine residues, which differed by 3.2 Å. However this difference could be attributed to the high flexibility of the arginine side chains and cannot be explored without high resolution X-ray structures. In general, the homology models of the *T. ammonificans* and *C. hydroformans* DXP synthases show strong similarity to the only two DXP synthases currently available, both in terms of overall structure, with the superposed structures showing reasonable RMSDs, as well as the positions of the active site residues. Fig. 6.6 E shows the surface of the enzyme around the active site. This region is mainly surrounded by loops and is accessible to the solvent.

A

B

C

D

E
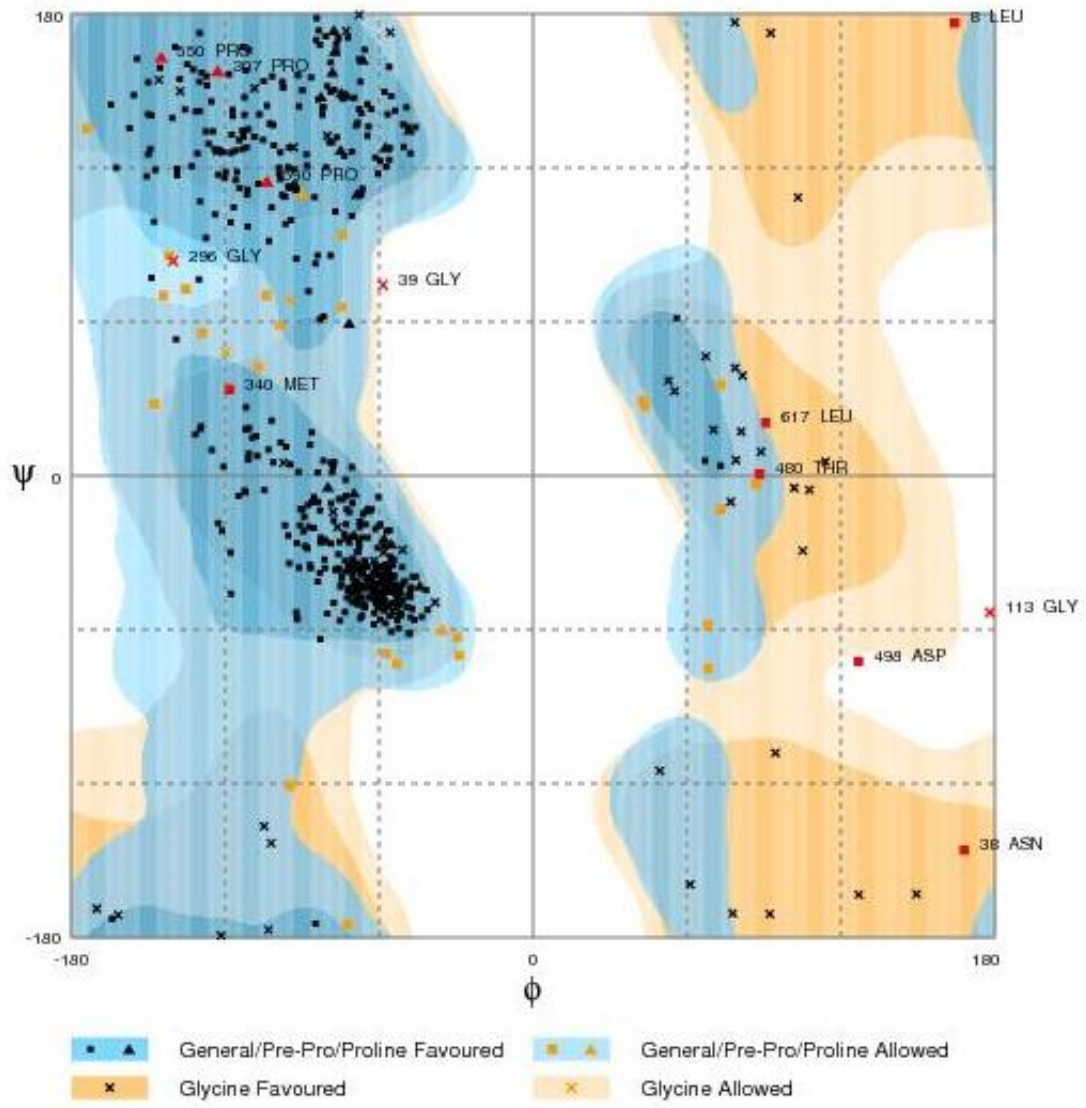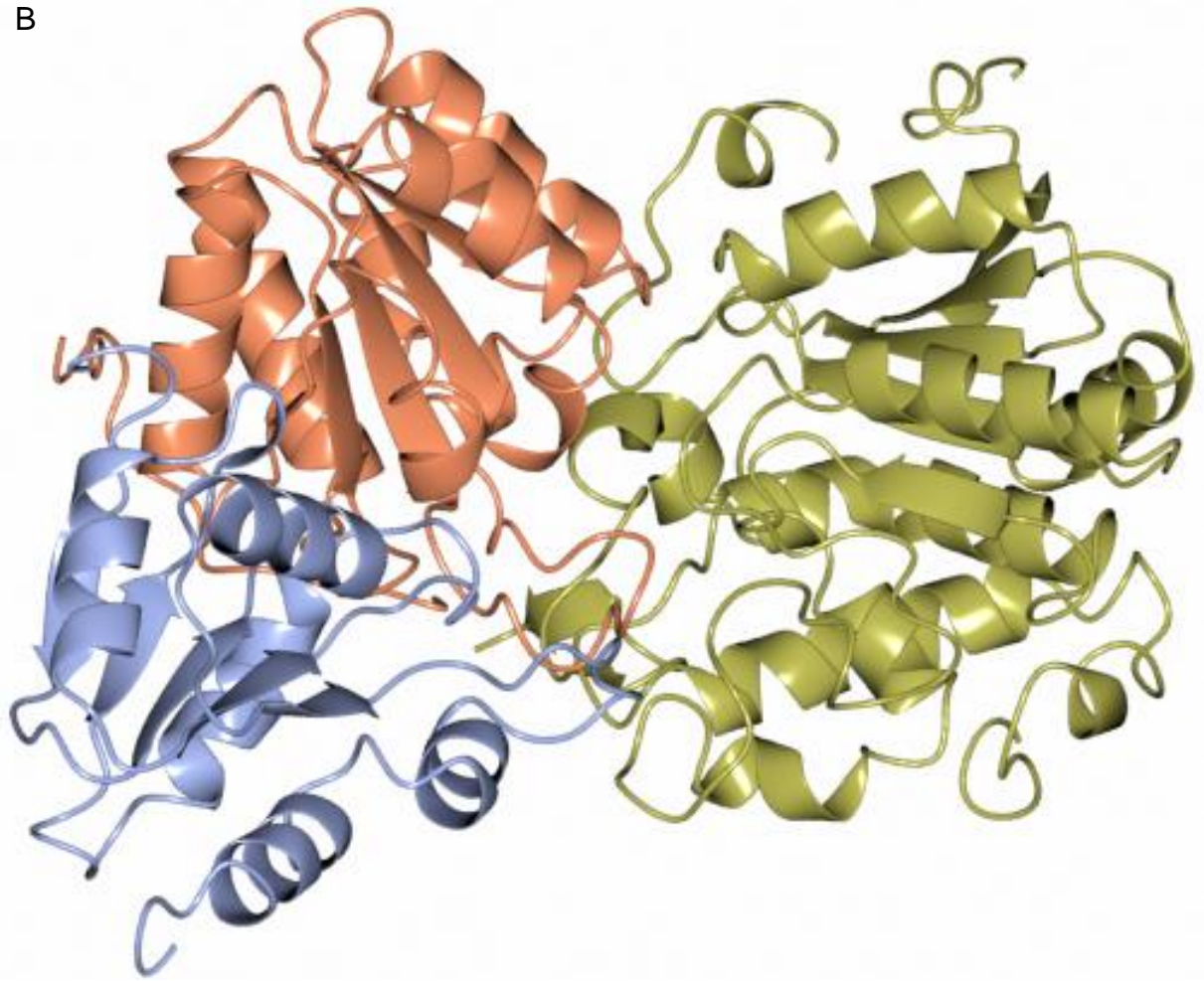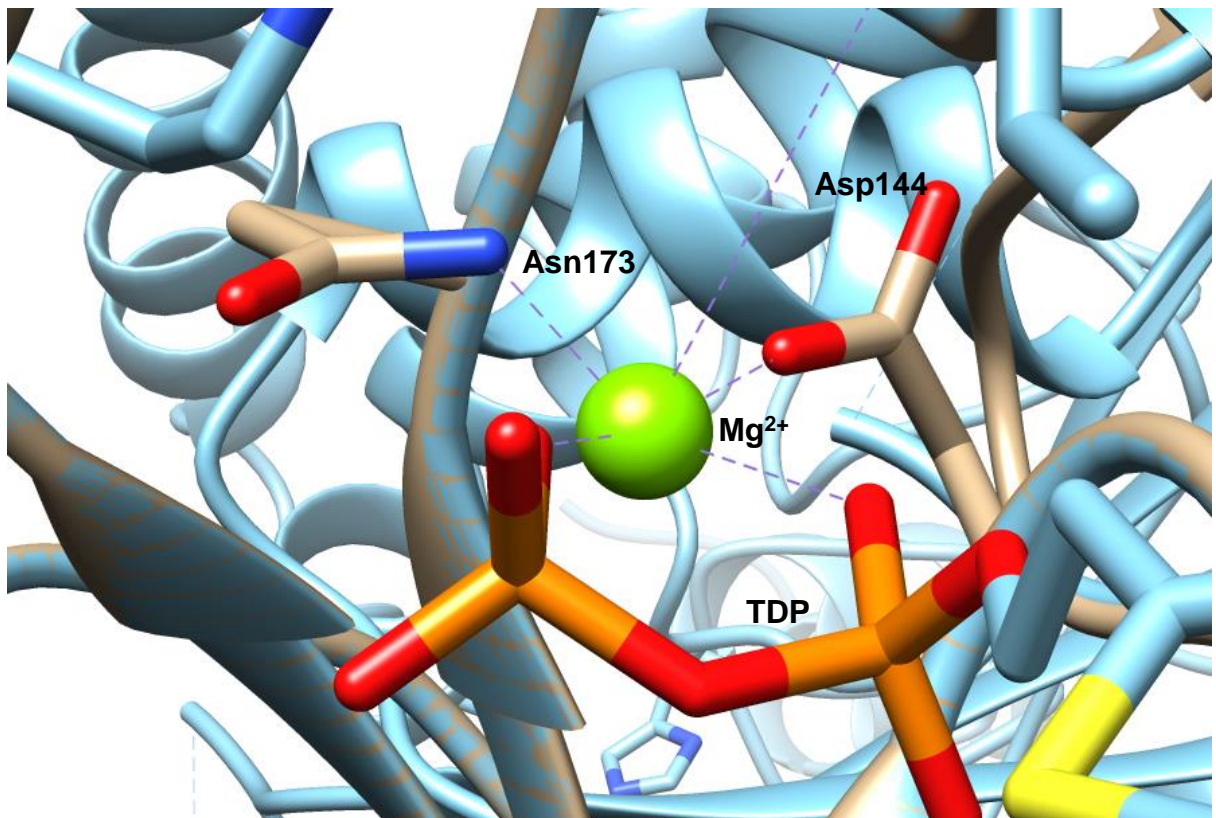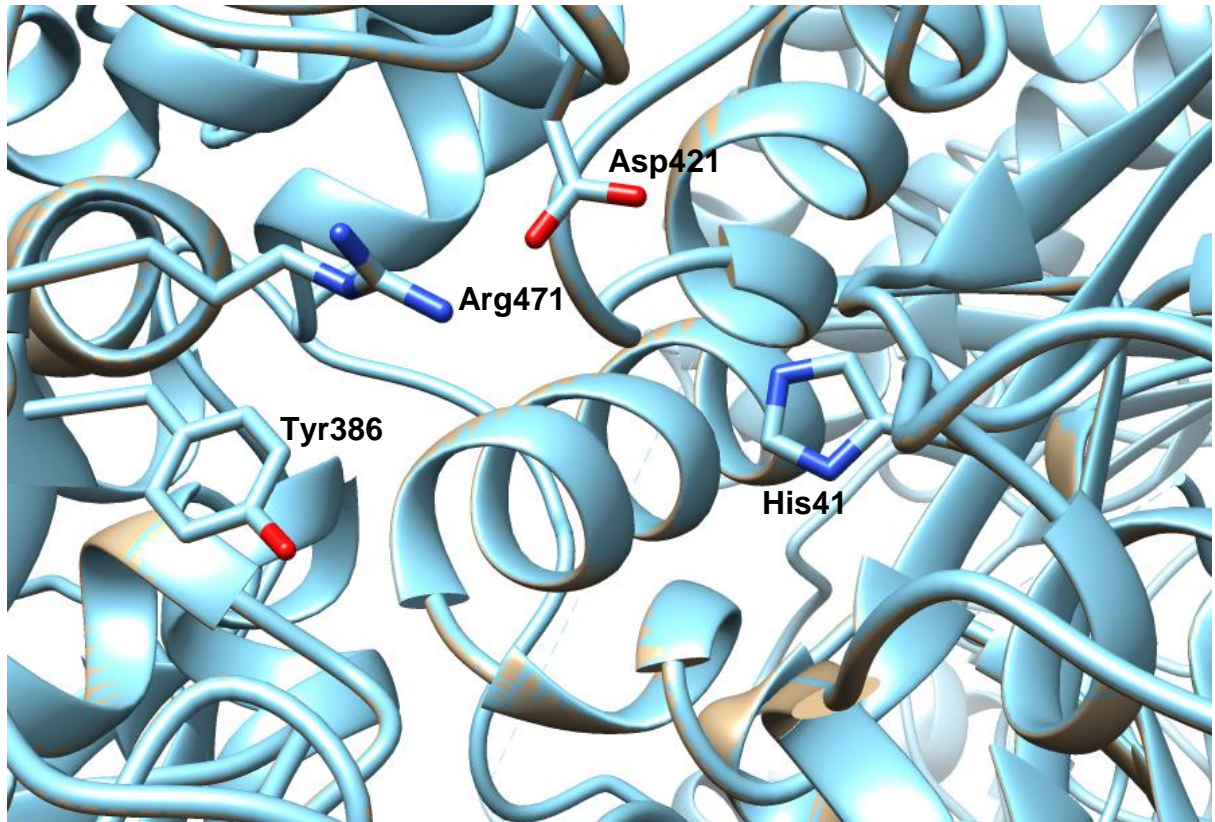
Active site
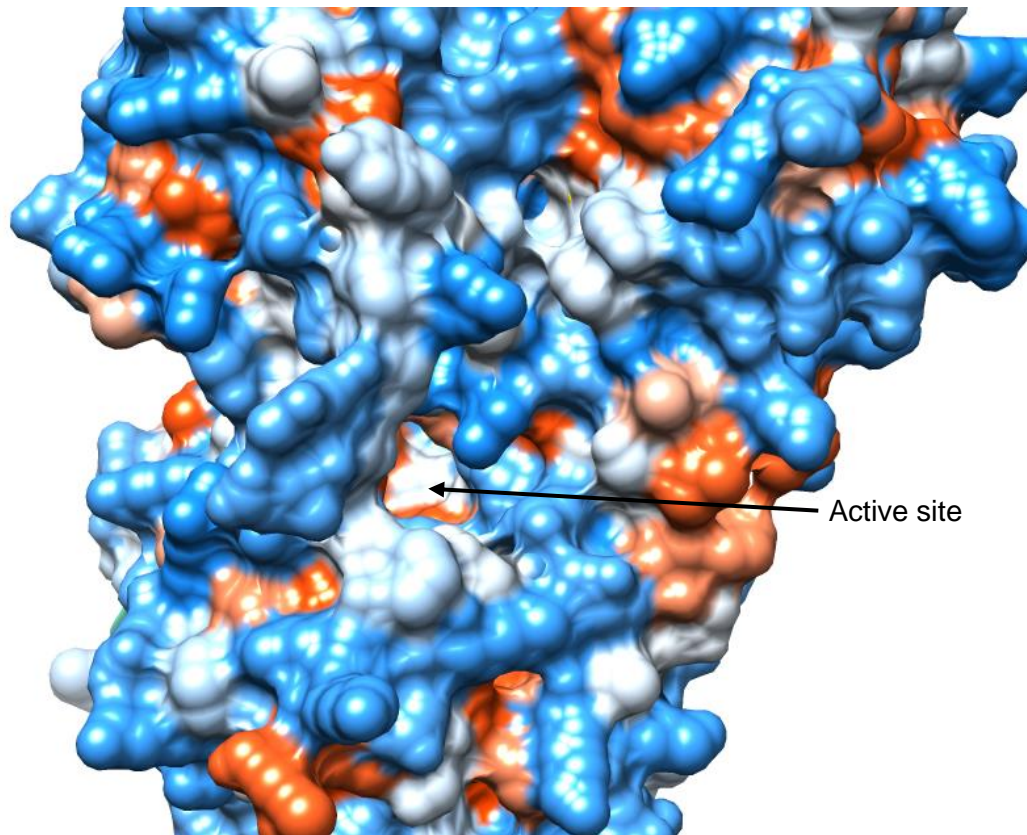
Fig. 6.6. Homology modelling of ChDXP. (A) Ramachandran plot of the model, 92.9 % of residues in a in a favoured conformation, 5.0 % in an allowed conformation, and 2.1 % as outliers shown in red. (B) Model of the monomer generated by Phyre2 with each different colours corresponding to domains, made using CCP4mg. It possess an α/β fold with parallel β-sheets located in the centre of each domain. (C) Divalent coordinating residues D143 and N172, superposed with DXP synthase from *D. radiodurans* (PDB code 2O1X, model in brown, 2O1X in blue) residue side chains were aligned identically. (D) Substrate binding site, position of residues involved in binding is once again conserved in comparison to the known structure. Overall RMSD between the two structures was 0.5 Å. (E) Surface representation of the ChDXP synthase, white is uncharged, red is negatively charged and blue is positively charged. Arrow indicated location of active site.
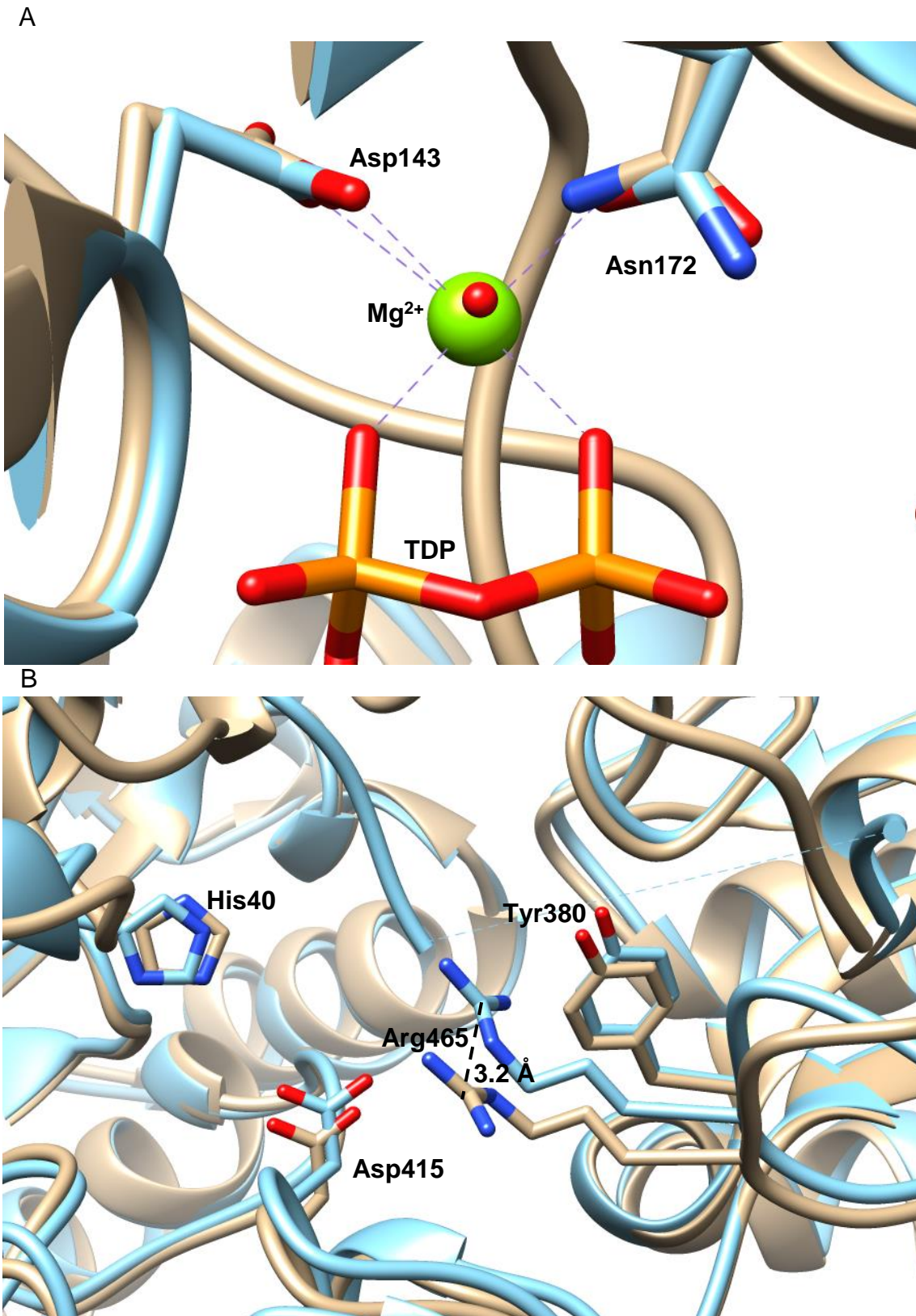
Fig. 6.7. Homology model of TaDXP, coloured in brown, superposed with a DXP synthase from *E. coli* (PDB code 2O1S) coloured in blue. The RMSD between the two structures was calculated to be 2.1 Å. (A) Divalent cation coordinating residues Asp 143 and Asn 172. The residue side chains align reasonably well with only small positional differences. (B) Substrate binding site showing positions of residues involved in substrate binding. These residues have relatively similar positions apart from the arginine.

[157]

The residues responsible for binding TDP are conserved between both of the DXP synthases under investigation and their homologues, this would suggest that the binding affinity for the co-factor is unchanged. Also the residues implicated in substrate binding also remain unchanged with only slight variances in their positions seen. This along with the size, shape and accessibility of the binding pockets to the solvent would suggest very similar substrate specificity and activity. There is currently very little research into the activity and substrate specificity of this class of enzyme. However it can be deduced that the strong similarities would indicate activity towards D-glyceraldehyde 3-phosphate and pyruvate which are shown to be substrates of the *M. tuberculosis* DXP synthase (2O1X) (Basta *et al.,* 2014).

Looking at the primary sequences of the two thermophilic DXP synthases and the percentage composition of amino acids present in each one in comparison to the mesophilic DXP synthase 2O1X from *D. radiodurans*, a trend was seen (Table 11). One of the things usually seen is more proline residues due to their stable structure; interestingly very little difference is seen with the ChDXP and 4.6 % of the TaDXP consists of prolines compared to the 5.6 % seen in 2O1X. In general, there are slightly more hydrophobic residues: for valine there is 1.2 % more for TaDXP and 1.9 % for ChDXP; for isoleucine there is no change seen for TaDXP but 0.7% more for ChDXP; for leucine there was little difference, 0.8 % more for TaDXP and hardly any change for ChDXP; for phenylalanine there is not much change seen in TaDXP and 1.0 % more in ChDXP. These substitutions occur in all secondary structure types; for tyrosine an increase of 1.2 % is seen for TaDXP although there is no difference for ChDXP; for tryptophan very little difference was seen for either enzyme and both showed fewer alanines residues overall. The remaining alanine residues in the TaDXP and ChDXP enzymes were mostly located in α-helices where their presence would help stabilise them due to their small hydrophobic side chains, rather than in loop regions that was the case 2O1X (Fig. 6.8). In their equivalent place in the thermophilic enzymes are more charged and polar residues. Also observed in TaDXP and ChDXP were more valines and fewer threonines in α-helices and β-strands, suggesting that hydrophobic interactions play more of a role than hydrogen bonding in these structures. This information suggests higher levels of hydrophobic forces within

the enzymes and that these are potentially responsible for the increase in thermostability required by thermophilic enzymes. Cysteines in the protein were found to be nearly absent in the ChDXP (0.3 %), and none at all are present in the TaDXP, in comparison to 1.0 % in the 2O1X. The elimination of cysteine residues is commonly seen in thermophilic species as they are highly susceptible to oxidation at high temperatures.

Visual inspection of the TaDXP structure with COOT revealed that there was a slight increase in the number of hydrophobic interactions seen in the structure in comparison to the 2O1X structure. Several instances were seen where hydrophobic interactions replaced hydrogen bonding showing the TaDXP structure relies on a different mechanism from the mesophilic 2O1X structure for stability. There were approximately 20 % more hydrophobic interactions seen in the TaDXP. On the dimer interface there was little hydrophobicity, instead most of the surface was charged, so it is assumed that this is the main element that stabilises the dimeric complex. ChDXP was not examined as it is highly similar to the TaDXP enzyme and would share similar adaptations.

Table 11. Composition of amino acids in DXP synthases.

Amino acids	2O1X	TaDXP	ChDXP
Ala	8.9 %	7.9 %	8.4 %
Arg	4.7 %	5.1 %	4.0 %
Asn	3.1 %	2.3 %	3.1 %
Asp	5.2 %	4.9 %	4.0 %
Cys	1.0 %	0.0 %	0.3 %
Gln	2.4 %	2.0 %	2.4 %
Glu	6.6 %	8.0 %	7.7 %
Gly	9.7 %	9.2 %	9.7 %
His	2.9 %	2.5 %	3.1 %
Ile	6.1 %	6.1 %	6.8 %
Leu	10.5 %	11.3 %	10.2 %
Lys	5.8 %	8.0 %	6.5 %
Met	3.4 %	0.8 %	2.3 %
Phe	3.7 %	4.1 %	4.7 %
Pro	5.6 %	4.6 %	5.3 %
Ser	5.6 %	5.6 %	4.5 %
Thr	4.5 %	5.7 %	4.5 %
Trp	0.6 %	0.7 %	0.3 %
Tyr	3.1 %	3.1 %	3.5 %
Val	6.8 %	8.0 %	8.7 %

A

2O1X
```
        1         10        20        30        40        50        60        70        80
2O1X    MNELPGTSDTPLLDQIHGPKDLKRLSREQLPALTEELRGEIVRVCSRGGLHLASSLGAVDIITALHYVLDSPRDRILFDV
ChDXP   ...........PILERISLPEDIKKLKPSELMALAQELREYIITVASQNGGHLAPSLGVVELTIALHFVFEAPKDKIIWDV
ChDXP
```

2O1X
```
                 90        100       110       120       130       140       150       160
2O1X    GHQAYAHKILTGRRDQMADIKKEGGISGFTKVSESEHDAITVGHASTSLTNALGMALARDAQGKDFHVAAVIGDGSLTGG
ChDXP   GHQAYAHKILTGRKKQFKTLRTFGGLSGFPKRDESPYDAFGVGHSSTSISAALGMALARDLKGEQYEVVAVIGDGALTGG
ChDXP
```

2O1X
```
                 170       180       190       200       210       220       230
2O1X    MALAALNTIGDMGRKMLIVLNDNEMSISENVGAMNKFMRGLQVQKWFQEGEGAGKKAVEAV...SKPLADFMSRAKNSTR
ChDXP   MAFEALNHAGHLQKKLIVVVNDNEMSIAQNVGALSAYLSRIRTDPKYSRGKDELEALIKKIPHIGPTMVKIGERLKDSFK
ChDXP
```

2O1X
```
        240       250       260       270       280       290       300       310
2O1X    HFFDPASVNPFAAMGVRYVGPVDGHNVQELVWLLERLVDLDGPTILHIVTTKGKGLSYAEADPIYWHGPAKFDPATGEYV
ChDXP   YLLVP..GMLFEELGFTYLGPIDGHNIKEMIEVFSRAKTFAGPVVVHVITKKGKGYHWAEENPDGFHGVGKFYISTGEPV
ChDXP
```

2O1X
```
        320       330       340       350       360       370       380       390
2O1X    PSSAYSWSAAFGEAVTEWAKTDPRTFVVTPAMREGSGLVEFSRVHPHRYLDVGIAEEVAVTTAAGMALQGMRPVVAIYST
ChDXP   EAPRVSFTEVFGKALVELAQDRPEVVAITAAMPTGTGLNYFAQNYPERFYDVGIAEQHAVTMAAGMACEGLKPVVAIYST
ChDXP
```

2O1X

```
        400         410         420         430         440         450         460         470
2O1X    FLQRAYDQVLHDVAIEHLNVTFCIDRAGIVGADGATHNGVFDLSFLRSIPGVRIGLPKDAAELRGMLKYAQTHDGPFAIR
ChDXP   FLQRSFDQIIHDVCLQNLPVVFAVDRAGIVGEDGPTHHGIFDLSYLRMIPNLTIMVPRNEDMLRKMLFTALNHSGPVALR
```

ChDXP

2O1X

```
        480         490         500         510         520         530
2O1X    YPRGNTAQVPAGTWPDLKWGEWERLKGGDDVVILAGGKALDYALKAAEDLPGV....GVV
ChDXP   YPRGAAVGVELTPYEQLPIGTAEILKEGSDGVVIGVGRPLNYALKAAQKLENEGISLTVI
```

ChDXP

[162]

B
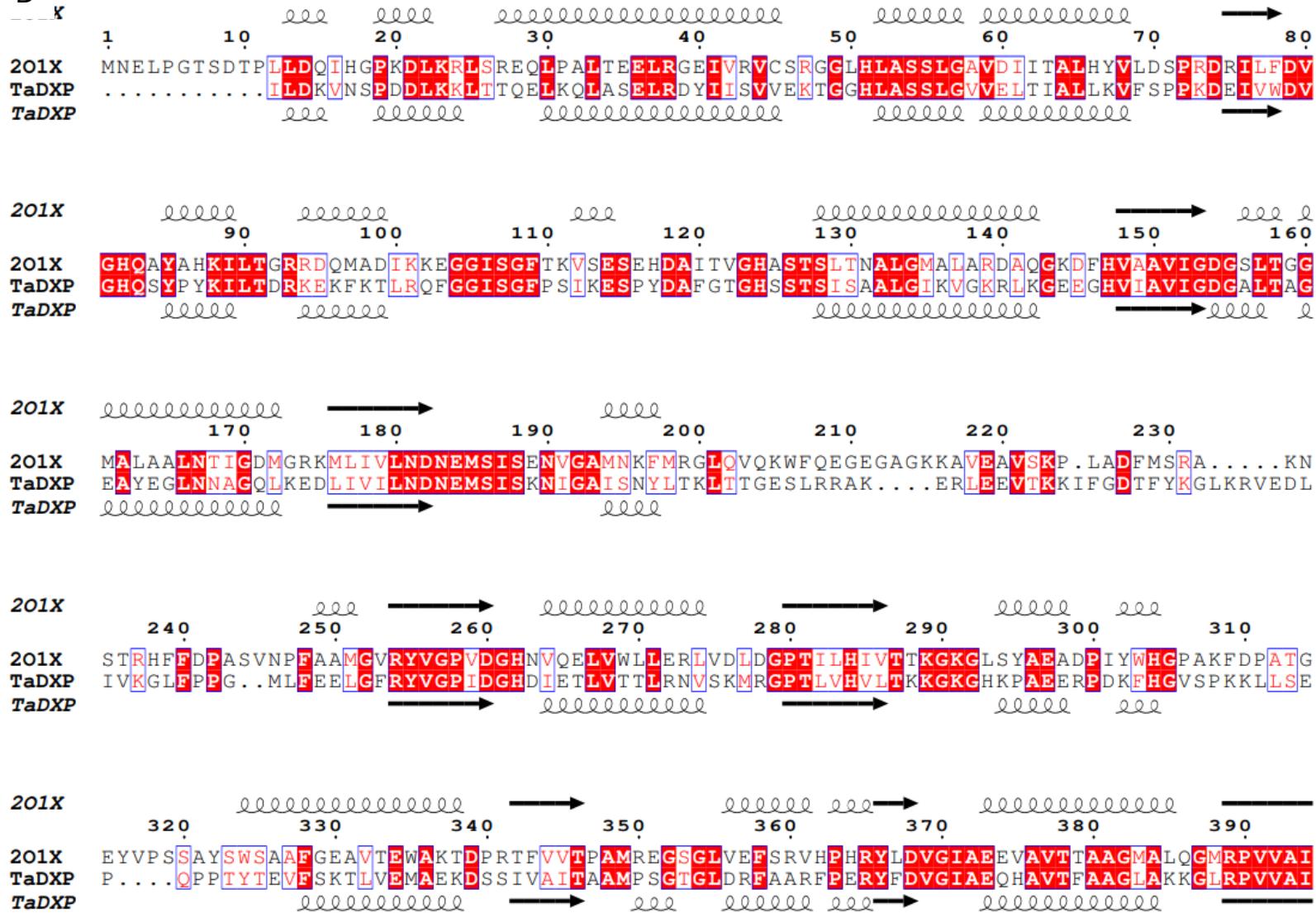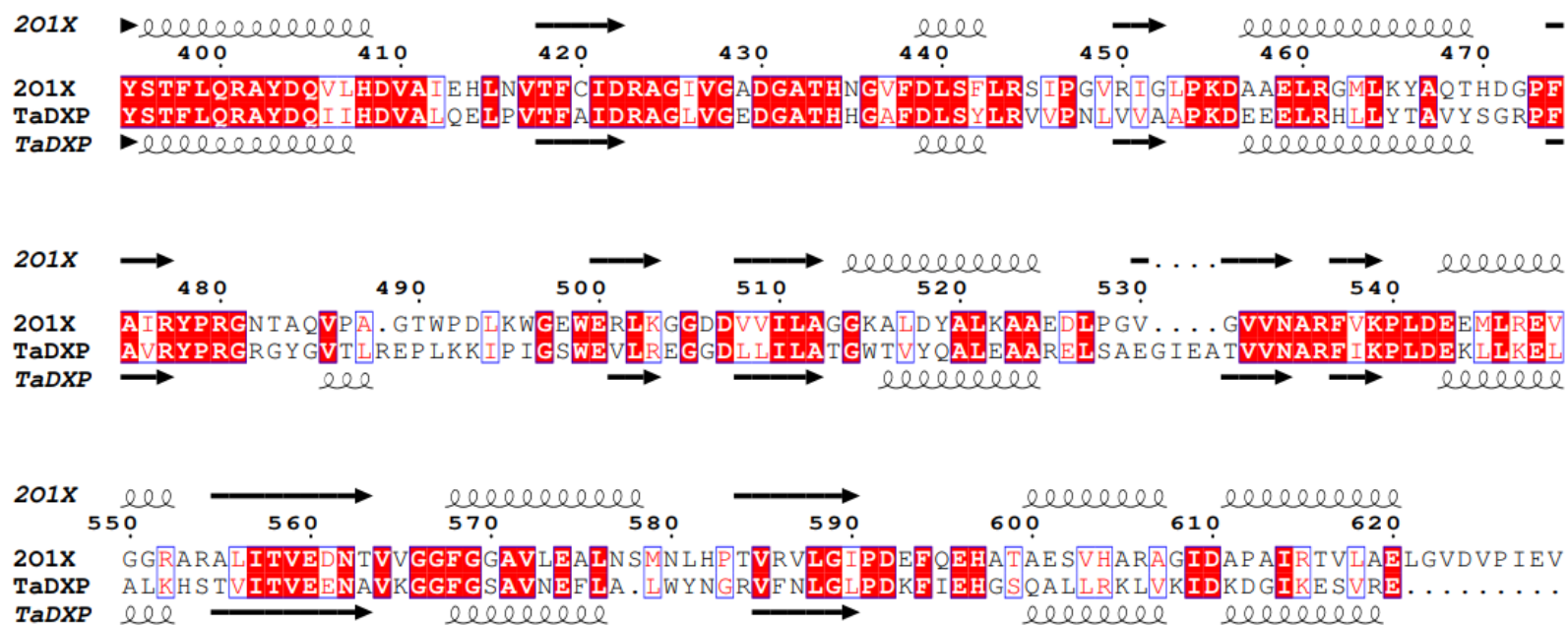
Fig. 6.8. Secondary structure and amino acid sequence alignment between the homology model of ChDXP and the 2O1X DXP synthase (A) and the homology model of TaDXP and the 2O1X DXP synthase (B) made using the ESPript 3 sequence and structure alignment server (Robert and Gouet, 2014). Residues highlighted in red are conserved, residues in blue boxes are conserved substitutions. Black arrows represent β-strands, loops represent α-helices.

DXP synthases and transketolases are evolutionarily related and share related structures and both use the cofactor thiamine pyrophosphate (Lois *et al*., 1998). Where DXP synthases use pyruvate as a substrate, transketolases use hydroxypyruvate in the industrial process. The reason for this difference is not yet understood and the homology models do not provide enough data to draw any conclusions as even good crystal structures would be hard pressed to find a difference since the difference between the two substrates is only one oxygen atom. The requirement for understanding the basics of this mechanism would be three high resolution structures of the enzyme in its native state, bound to a substrate/inhibitor complex.

Most transketolase enzymes and both of the known DXP synthases are known to be dimers as this is usually critical for activity in both of these enzymes (Xiang et al., 2013). With the model generated as a monomer it was necessary for proper comparison to predict its dimeric form. The model for TaDXP was duplicated in COOT and superposed on top of the dimer of the 2O1X structure. No overlaps in the loops were seen (Fig. 6.9 A). To confirm the accuracy of this the monomeric structure was submitted to the FRODOCK 2.0 protein-protein docking server (Ramírez-Aportela et al., 2016). This generated a dimer very similar to the one created in manually in COOT (Fig. 6.9 B). When the overall structures of the dimeric TaDXP synthase homology model and a transketolase from *E. coli* (PDB code 1QGD) (Fig. 6.9 C) are compared it is clear that both enzymes share very similar structures. The RMSD between the monomeric subunits two when superposed was 2.1 Å. The biggest difference comes from the positions of the large and small domains. Both enzymes share the same domains but in the DXP synthase they have been swapped. The transketolase possesses a long loop region that connects the large and small domains causing the C-terminal domain to be positioned directly below the N-terminal domain of the neighbouring subunit (Fig. 6.9 A). The DXP synthase on the other hand has a much shorter loop connecting the domains causing the C-terminal domain to be positioned below the N-terminal domain of the same subunit (Fig. 6.9 B). The active sites of the two DXP synthases, the transketolase from *E. coli* and a transketolase from *Burkholderia pseudomallei* (PDB code 3UTP) were very similar with all active site

residues conserved. The main difference was a conserved substitution of the catalytic tyrosine in the DXP synthases for a phenylalanine in the transketolases (Fig 6.9 D).

While only homology models were obtained for these DXP synthase enzymes, it was shown that the TaDXP protein did crystallise, though poorly, and that given more time it may be possible to eventually find conditions that would produce an ordered crystal. Alternative methods to be tested include lysine methylation, and co-crystallisation trials with the substrates pyruvate and D-glyceraldehyde 3-phosphate. There has also been identified an inhibitor of DXP synthases, butylacetylphosphonate, although this is not currently commercially available and the authors of the paper synthesised it themselves through a series of reactions (Smith *et al.*, 2014). Another potential inhibitor, β-fluoropyruvate, that is commercially available was identified through single-molecule force spectroscopy (Sisquella *et al.*, 2010). The homology models obtained in this project are important in the ongoing research into DXP synthases especially since to date there are only two structures in the Protein Data Bank for this class of enzymes. While it would be better to obtain crystal structures, these models nevertheless provide another step in our understanding of these enzymes.

A

B

C
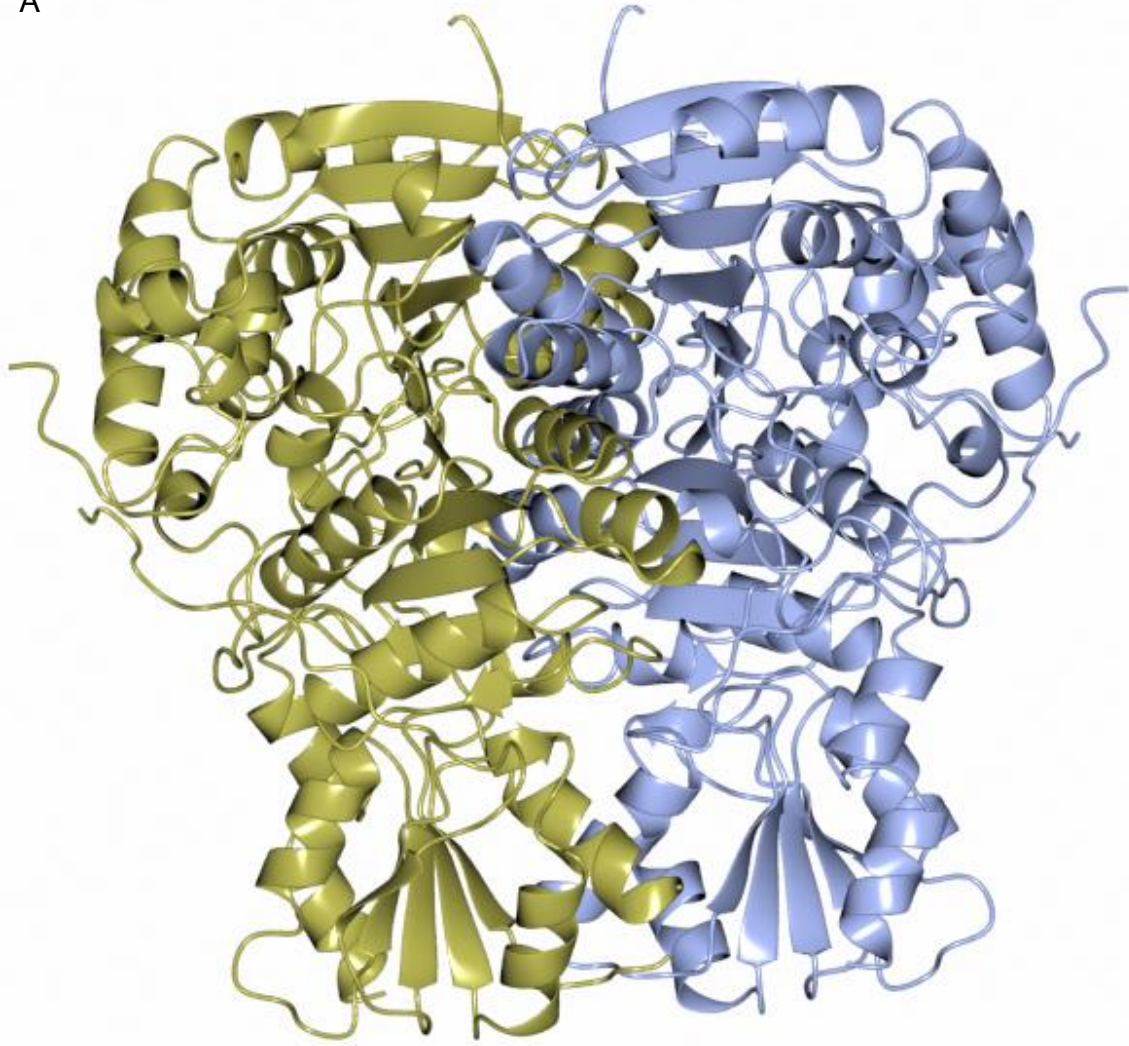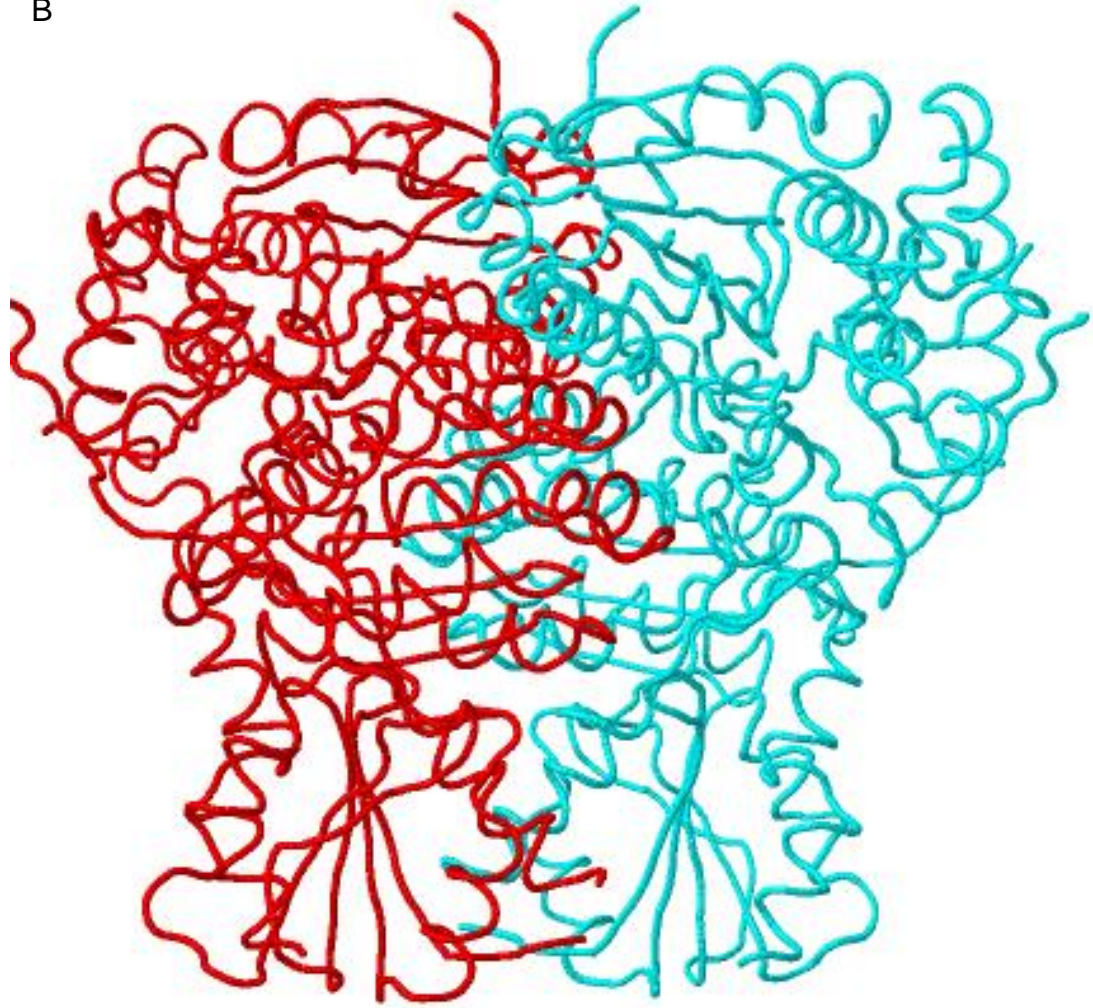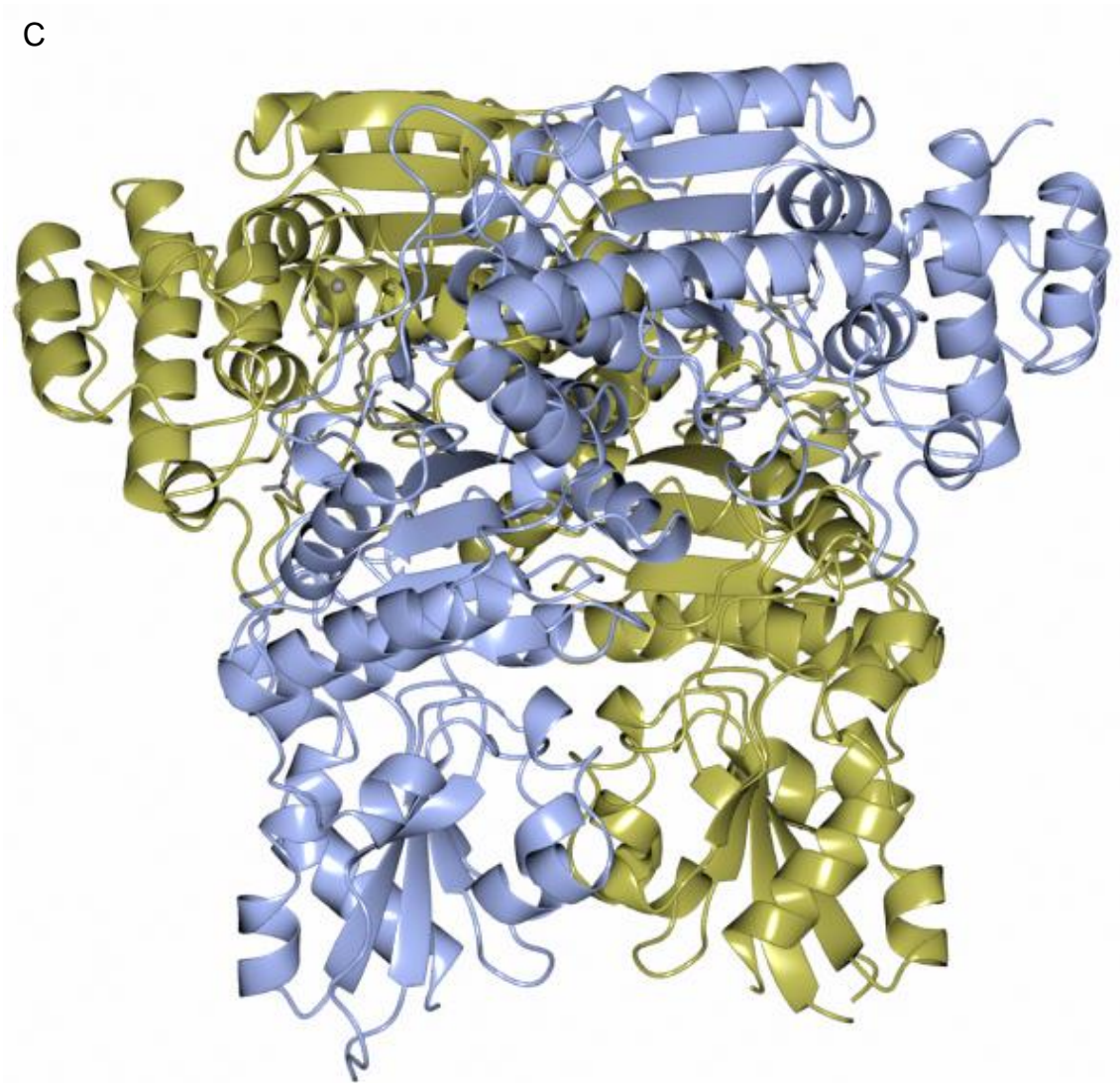
D

```
1QGD    GNYIHYGVREFGMTAIANGISLHGGFLPYTSTFLMFVEYARN    442
3UPT    GNHINYGVREFGMSAAINGLVLHGGYKPFGGTFLTFSDYSRN    489
TaDXP   -RYFDVGIAEQHAVTFAAGLA-KKGLRPVVAIYSTFLQRAYD    389
ChDXP   -RFYDVGIAEQHAVTMAAGMA-CEGLKPVVAIYSTFLQRSFD    395
```

Fig. 6.9. (A) Dimeric structure of the homology model for TaDXP with different subunits in blue and yellow. Images made using COOT and CCP4mg. (B) Dimer formed by the protein docking server FRODOCK (Ramírez-Aportela *et al.*, 2016) (C) Dimeric structure of transketolase from *E. coli* (PDB code 1QGD) with different monomers in blue and yellow. (D) Multiple sequence alignment made with Clustal Omega between the two DXP synthases, the transketolase from *E. coli* and a transketolase from *B. pseudomallei* (PDB code 3UTP). The residues highlighted in red show the conserved substitution of a tyrosine in the DXP synthases for a phenylalanine in the transketolases.

[169]

6.4 Summary

Two DXP synthases from different thermophiles were successfully cloned but only one was expressed as a soluble protein. This TaDXP was then purified and underwent crystallisation trials where several crystals were found growing. Unfortunately, none of these crystals produced X-ray diffraction data that could be analysed and so two homology models for both proteins were made. Using these it was shown that there were more hydrophobic interactions within both thermophilic DXP synthases in comparison to mesophilic equivalents. However without high resolution X-ray diffraction data it is not possible to deduce how the mechanism of this enzyme allows it to use pyruvate as a substrate over hydroxypyruvate that transketolase enzymes use in the industrial process.

7. Conclusion

This project has allowed the successful biochemical characterisation of a quorum sensing lactonase from the thermophilic archaeon *Vulcanisaeta moutnovskia*. This enzyme has been shown to be active towards a broad range of substrates: γ-butyrolactone, γ-valerolactone, γ-caprolactone, whiskey lactone and δ-dodecalactone. This shows differences from the substrates used by a related lactonase from *Sulfolobus.* This difference in substrate specificity is proposed to be attributed to a α-helix near the active site of the enzyme as revealed in the crystal structure, which is observed to be a flexible loop region in other quorum sensing lactonases.

The gene encoding a gluconolactonase called lac11 from a thermophilic *Planctomyces* species was successfully cloned and the enzyme was over-expressed, purified and crystallised. These crystals were of sufficient quality to provide X-ray diffraction data to a resolution of 2.4 Å at the Diamond Synchrotron. This allowed a structure to be determined to an R factor of 0.28 and a free R factor of 0.32 through structural refinement and was observed to be a monomeric six bladed β-propeller with the active site present in the centre of all six blades. The activity and substrate specificity of this enzyme was investigated with commercially available substrates: D-glucono-1,5-lactone, D-glucuronic acid-1,4-lactone, L-gulonic acid-1,4-lactone, D-threono-1,4-lactone, D-arabino-1,4-lactone, L-arabino-1,4-lactone, D-xylano-1,4-lactone, L-fucono-1,4-lactone, D-altrono-1,4-lactone, L-rhamnono-1,4-lactone and D-rhamnono-1,5-lactone. Four other enzymes were also investigated, a second quorum sensing lactonase and an enol lactonase derived from metagenomic DNA sequences, and two DXP synthases from *Thermovibrio ammonificans* and *Carboxydothermus hydroformans*. The quorum sensing lactonase and the DXP synthase from *T. ammonificans* were successfully over-expressed in soluble form and the enzymes purified.

Out of five enzymes that were successfully expressed in soluble form and subsequently purified, only one crystallised to give crystals with high enough quality to collect X-ray diffraction data. This is not unusual when attempting to crystallise proteins as the rate of success for creating good quality crystals is low.

For example, the Robotic Crystallization Facility at the University of Auckland set up 10000 crystallisation drops for 140 different proteins, covering 672 different crystallisation conditions. Out of all of these, 300 conditions provided protein crystals, representing a 3 % success rate, where 56 % of the proteins tested failed to crystallise to high enough quality for X-ray diffraction or failed to crystallise at all (Ivanovic *et al.*, 2009).

All of the proteins examined showed strong similarities in terms of secondary and tertiary structure to their mesophilic counterparts with RMSDs ranging between 0.5 Å and 5.2 Å. The differences in thermal stability are therefore determined by different interactions in the tertiary structure provided by changes in the primary structure between the mesophilic and thermophilic homologues. These changes are usually away from the active site and in loops and on the surface of the protein (Littlechild *et al.*, 2007, 2013). All of the proteins examined, apart from the Tomsk lactonase, showed a greater number of hydrophobic interactions within the structures in comparison to mesophilic variants. In the case of the Tomsk lactonase there were roughly the same number of hydrophobic interactions as its mesophilic variant which would be most likely due to its being obtained from a hot spring at 46 °C; this was the lowest temperature out of all the hot springs from which proteins were examined in this project and was not that much higher than mesophilic conditions. The enzymes that showed the greatest frequency of hydrophobic interactions were the DXP synthases, although the dimer interfaces were shown to be primarily stabilised through ionic interactions

All of the thermophilic lactonase enzymes investigated showed more proline residues in loop regions, and shorter loops, in comparison to the mesophilic proteins, both of which would also aid in the thermostability of the enzymes. A high number of proline residues is especially seen in organisms with DNA with high G-C content such as the thermophilic bacterium *Thermus thermophilus* (Kumar and Kaur, 2014). As well as this there are fewer cysteine residues seen because these are vulnerable to oxidation which is more likely to occur at higher temperatures. For the Tomsk lactonase there were also fewer lysine residues and more arginine residues because the side chain of arginine can form more

electrostatic interactions in comparison to lysine (Sokalingam *et al.*, 2012). And while it did not appear that lysine residues were substituted in this way in the B1 lactonase there was a significantly greater number of arginine residues when compared to a mesophilic lactonase.

A possibility for future work in regards to the DXP synthases would be to attempt to crystallise the TaDXP in the presence of a herbicide called ketoclomazone which has been found to inhibit activity (Matsue *et al.*, 2010). It was found to bind to an unidentified inhibitor binding site, different from both of the substrate binding sites and not blocking the pyruvate binding site or the D-glyceraldehyde 3-phosphate binding site, making it an uncompetitive inhibitor. With an inhibitor bound, the enzyme would in theory be more stable than on its own and would hopefully provide higher quality crystals that could diffract X-rays to a high enough resolution to determine the structure. Structures are required of the native DXP synthase enzyme, the enzyme bound to its substrate and an intermediate to provide insight as to what allows this enzyme to utilise pyruvate instead of hydroxypyruvate as a ketol donor.

8. References

Aghajari, N., Petegem, F. Van, Villeret, V., Chessa, J., Gerday, C., Haser, R., Beeumen, J. Van and I, B. L. (2003) Crystal Structures of a Psychrophilic Metalloprotease Reveal New Insights into Catalysis by Cold-Adapted Proteases, *Proteins,* 647,, 636–647. doi: 10.1002/prot.10264

Altincicek, B., Hintz, M., Sanderbrand, S., Wiesner, J., Beck, E. and Jomaa, H. (2000) Tools for discovery of inhibitors of the 1-deoxy-D-xylulose 5-phosphate (DXP) synthase and DXP reductoisomerase: An approach with enzymes from the pathogenic bacterium *Pseudomonas aeruginosa. FEMS Microbiology Letters*, 190(2), 329–333. doi: 10.1016/S0378-1097(00)00357-8.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403–10. doi: 10.1016/S0022-2836(05)80360-2.

Bains, J., Kaufman, L., Farnell, B. and Boulanger, M. J. (2011) A product analog bound form of 3-oxoadipate-enol-lactonase (PcaD) reveals a multifunctional role for the divergent cap domain. *Journal of Molecular Biology*, 406(5), 649–658. doi: 10.1016/j.jmb.2011.01.007.

Basta, L. A. B., Patel, H., Kakalis, L., Jordan, F. and Meyers, C. L. F. (2014) Defining critical residues for substrate binding to 1-deoxy-D-xylulose 5-phosphate synthase: Active site substitutionsstabilize the pre-decarboxylation intermediate C2α-lactylthiamindiphosphate, *FEBS Journal,* 281(12), 2820–2837 doi: 10.1111/febs.12823

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. and Leslie, A. G. W. (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 67(4), 271–281. doi: 10.1107/S0907444910048675.

Becker, P., Abu-Reesh, I., Markossian, S., Antranikian, G. and Märkl, H. (1997) Determination of the kinetic parameters during continuous cultivation of the lipase producing thermophile *Bacillus sp.* IHI-91 on olive oil. *Applied Microbiology and Biotechnology.*, 48(2), 184–190. doi: 10.1007/s002530051036.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T.,

Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L. and Schwede, T. (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, 42(W1), 252–258. doi: 10.1093/nar/gku340.

Blacker, A. and Holt, R., Collins A. N. (Ed), Sheldrake, G. (Edr), Crosby, J. (Ed), (1997) Chirality in Industry II. *Wiley*

Brammer, L. A., Smith, J. M., Wade and H. Meyers, C. F., (2011) 1-Deoxy-D-xylulose 5-phosphate synthase catalyzes a novel random sequential mechanism. *Journal of Biological Chemistry*., 286(42), 36522–36531. doi: 10.1074/jbc.M111.259747

Broadwater, J. A., Bekeny, P. A., Salmon, J. E., Lee, A. S., Eldridge, A. M., Pett, V. B., Salmon, J. E., Lee, A. N. N. S., Eldridge, A. M. and Pett, V. B. (1994) A structural role for arginine in proteins : Multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Science*, 3541-548 doi: 10.1002/pro.5560030402

Bruggink, A., Roos, E. C. and de Vroom, E. (1998) Penicillin Acylase in the Industrial Production of β-Lactam Antibiotics. *Organic Process Research & Development*, 2(2), 128–133. doi: 10.1021/op9700643.

Bunker, R. D., Dickson, J. M. J., Caradoc-Davies, T. T., Loomes, K. M. and Baker, E. N. (2012) Use of a repetitive seeding protocol to obtain diffraction-quality crystals of a putative human D-xylulokinase. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*. International Union of Crystallography, 68(10), 1259–1262. doi: 10.1107/S1744309112031181.

Cary S. C., McDonald I. R., Barrett J. E. and Cowan D. A., (1982) On the rocks: the microbiology of Antarctic Dry Valley soils. *Science*, 215, 4536, 129-138 doi: 10.1038/nrmicro2281

Chanitnun, K. and Pinphanichakarn, P. (2012) Glucose(xylose) isomerase production by *Streptomyces SP*. CH7 grown on agricultural residues. *Brazilian Journal of Microbiology*, 43(3), 1084–1093. doi: 10.1590/S1517-83822012000300035.

Chen, B. H., Sayar, A., Kaulmann, U., Dalby, P. A., Ward, J. M. and Woodley, J. M. (2006) Reaction modelling and simulation to assess the integrated use of transketolase and ω-transaminase for the synthesis of an aminotriol. *Biocatalysis and Biotransformation*, 24(6), 449–457. doi: 10.1080/10242420601068668.

Chen, C.-N., Chin, K.-H., Wang, A. H.-J. and Chou, S.-H. (2008) The First Crystal Structure of Gluconolactonase Important in the Glucose Secondary Metabolic Pathways. *Journal of Molecular Biology*. Elsevier Ltd, 384(3), 604–614. doi: 10.1016/j.jmb.2008.09.055.

Defoirdt, T., Boon, N. and Bossier, P. (2010) Can bacteria evolve resistance to quorum sensing disruption? *PLoS Pathogens*, 6(7), 1–6. doi: 10.1371/journal.ppat.1000989.

Delano, W. L. (2002) PyMOL : An Open-Source Molecular Graphics Tool.CCP4 Newsletter On Protein Crystallography, (40), 82–92.

Demuynck, C., Bolte, J., Hecquet, L. and Dalmas, V. (1991) Enzyme-catalyzed synthesis of carbohydrates: synthetic potential of transketolase. *Tetrahedron Letters*, 32(38), 5085–5088. doi: 10.1016/S0040-4039(00)93434-8.

Draganov, D. I. (2010) Lactonases with oragnophosphatase activity: Structural and evolutionary perspectives. *Chemico-Biological Interactions*. Elsevier Ireland Ltd, 187(1–3), 370–372. doi: 10.1016/j.cbi.2010.01.039.

Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.*, 2126–2132. doi: 10.1107/S0907444904019158.

Feller, G. (2013) Psychrophilic Enzymes : From Folding to Function and Biotechnology, *Scientifica*, vol. 2013, 28 pages, doi.org/10.1155/2013/512840.

Feller, G., Narinx, E., Arpigny, J. L., Aittaleb, M., Baise, E., Genicot, S. and Gerday, C. (1996) Enzymes from psychrophilic organisms, *FEMS Microbiology Reviews,* 32(41). doi.org/10.1016/0168-6445(96)00011-3

Ferrandi, E. E., Sayer, C., Isupov, M. N., Annovazzi, C., Marchesi, C., Iacobone, G., Peng, X., Bonch-Osmolovskaya, E., Wohlgemuth, R., Littlechild, J. A. and Monti, D. (2015) Discovery and characterization of thermophilic

limonene-1,2-epoxide hydrolases from hot spring metagenomic libraries. *FEBS Journal*, 282(15), 2879–2894. doi: 10.1111/febs.13328.

Fuchs, M., Farnberger, J. E. and Kroutil, W. (2015) The Industrial Age of Biocatalytic Transamination. *European Journal of Organic Chemistry*, 2015(32), 6965–6982. doi: 10.1002/ejoc.201500852.

Guendouze, A., Plener, L., Bzdrenga, J., Jacquet, P., Rémy, B., Elias, M., Lavigne, J. P., Daudé, D. and Chabrière, E. (2017) Effect of quorum quenching lactonase in clinical isolates of *pseudomonas aeruginosa* and comparison with quorum sensing inhibitors. *Frontiers in Microbiology*, 8(FEB), 1–10. doi: 10.3389/fmicb.2017.00227.

Gumerov, V. M., Mardanov, A. V, Beletsky, A. V, Prokofeva, M. I., Bonch-Osmolovskaya, E. A., Ravin, N. V and Skryabin, K. G. (2011) Complete genome sequence of '*Vulcanisaeta moutnovskia*' strain 768-28, a novel member of the hyperthermophilic crenarchaeal genus *Vulcanisaeta*. *J Journal of Bacteriology*, 2355–2356. doi: 10.1128/JB.00237-11.

Guo, F. and Berglund, P. (2017) Transaminase biocatalysis: Optimization and application.*Green Chemistry*. Royal Society of Chemistry, 19(2), 333–360. doi: 10.1039/c6gc02328b.

Guy, J. E., Isupov, M. N. and Littlechild, J. A. (2003) The structure of an alcohol dehydrogenase from the hyperthermophilic archaeon *Aeropyrum pernix*. *Journal of Molecular Biology*, 331(5), 1041–1051. doi: 10.1016/S0022-2836(03)00857-X.

Hense, B. A., Kuttler, C., Müller, J., Rothballer, M., Hartmann, A. and Kreft, J. U. (2007) Does efficiency sensing unify diffusion and quorum sensing? *Nature Reviews Microbiology*, 5(3), 230–239. doi: 10.1038/nrmicro1600.

Hiblot, J., Bzdrenga, J., Champion, C., Chabriere, E. and Elias, M. (2015) Crystal structure of VmoLac, a tentative quorum quenching lactonase from the extremophilic crenarchaeon *Vulcanisaeta moutnovskia*. *Scientific Reports*, 5, 8372. doi: 10.1038/srep08372.

Hiblot, J., Gotthard, G., Chabriere, E. and Elias, M. (2012) Structural and Enzymatic characterization of the lactonase Sis Lac from *Sulfolobus islandicus*, *PLOS ONE,* 7(10). doi: 10.1371/journal.pone.0047028

Hickey, A. M., Ngamsom, B., Wiles, C., Greenway, G. M., Watts, P. and Littlechild, J. A. (2009) A microreactor for the study of biotransformations by a cross-linked γ-lactamase enzyme. *Biotechnology Journal*, 4(4), 510–516. doi: 10.1002/biot.200800302.

Hobbs, G. R., Mitra, R. K., Chauhan, R. P., Woodley, J. M. and Lilly, M. D. (1996) Enzyme-catalysed carbon-carbon bond formation: Large-scale production of *Escherichia coli* transketolase. *Journal of Biotechnology*, 45(2), 173–179. doi: 10.1016/0168-1656(95)00165-4.

Ivanovic, I., Baker, H., Lott, S. and Baker, E. N. (2009) Analysis of overall success of robotic crystallization, 42(2003), 2009.

Kara, S., Schrittwieser, J. H., Hollmann, F. and Ansorge-Schumacher, M. B. (2014) Recent trends and novel concepts in cofactor-dependent biotransformations. *Applied Microbiology and Biotechnology*, 98(4), 1517–1529. doi: 10.1007/s00253-013-5441-5.

Kelly, L. A., Mezulis, S., Yates, C., Wass, M. and Sternberg, M. (2015) The Phyre2 web portal for protein modelling, prediction, and analysis. *Nature Protocols*. Nature Publishing Group, 10(6), 845–858. doi: 10.1038/nprot.2015-053.

Knaus, T., Böhmer, W. and Mutti, F. G. (2017) Amine dehydrogenases: Efficient biocatalysts for the reductive amination of carbonyl compounds. *Green Chemistry*. Royal Society of Chemistry, 19(2), 453–463. doi: 10.1039/c6gc01987k.

Krahe, M., Antranikian, G. and Märkl, H. (1996) Fermentation of extremophilic microorganisms. *FEMS Microbiology Reviews*, 18(2–3), 271–285. doi: 10.1016/0168-6445(96)00018-6.

Kremer-Aach, A., Kläui, W., Bell, R., Strerath, A., Wunderlich, H. and Mootz, D. (1997) Cobalt as a Probe for Zinc in Metalloenzyme Model Compounds? A Comparison of Spectroscopic Features and Coordination Geometry of Four- and Five-Coordinate Complexes. Crystal and Molecular Structures of [Co(η^3-Tp^{Ph})(η^2. *Inorganic Chemistry*, 36(8), 1552–1563. doi: 10.1021/ic961434r.

Krieger, E. and Vriend, G. (2014) YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics (Oxford, England)*,

30(20), 2981–2982. doi: 10.1093/bioinformatics/btu426.

Krissinel, E. and Henrick, K. (2007) Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*, 372(3), 774–797. doi: 10.1016/j.jmb.2007.05.022.

Kristiansen, E. and Zachariassen, K. E. (2005) The mechanism by which fish antifreeze proteins cause thermal hysteresis, *Cryobiology,* 51(3), 262-80

Kuang, Z., Hao, Y., Walling, B. E., Jeffries, J. L., Ohman, D. E. and Lau, G. W. (2011) *Pseudomonas aeruginosa* Elastase provides an Escape from phagocytosis by degrading the pulmonary surfactant protein-A. *PLoS ONE*, 6(11). doi: 10.1371/journal.pone.0027091.

Kumar, A. and Kaur, J. (2014) Primer Based Approach for PCR Amplification of High GC Content Gene: *Mycobacterium* Gene as a Model. *Molecular Biology International*, 2014, 1–7. doi: 10.1155/2014/937308.

Ladoukakis, E., Pilalis, E., Chatziioannou, A. and Kolisis, F. (2014) ANASTASIA a versatile web platform for metagenomic analysis. *New Biotechnology*. doi: 10.1016/j.nbt.2014.05.2042.

Leadbetter, J. R. and Greenberg, E. P. (2000) Metabolism of acyl-homoserine lactone quorum-sensing signals by *Variovorax paradoxus. Journal of Bacteriology*, 182(24), 6921–6926. doi: 10.1128/JB.182.24.6921-6926.2000.

Lee, J., Jeong, K., Jin, B., Ryu, K., Kim, E., Ahn, J. and Kim, Y. (2013) Structural and Dynamic Features of Cold-Shock Proteins of Listeria monocytogenes , a Psychrophilic Bacterium. *Biochemistry,* 52(14), 2492-504 doi: 10.1021/bi301641b.

Littlechild, J., Turner, N., Hobbs, G., Lilly, M., Rawas, A. and Watson, H. (1995) Crystallization and preliminary X-ray crystallographic data with *Escherichia coli* transketolase. *Acta Crystallographica Section D: Biological Crystallography*, 51(6), 1074–1076. doi: 10.1107/S0907444995005415.

Littlechild, J., Novak, H., James, P. and Sayer, C. (2013) Mechanisms of thermal stability adopted by thermophilic proteins and their use in white biotechnology. *Thermophilic Microbes in Environmental and Industrial Biotechnology: Biotechnology of Thermophiles*, 481–507. doi: 10.1007/978-94-

007-5899-5_19.

Littlechild, J. A. (1991) Protein crystallization: Magical or logical: Can we establish some general rules? *Journal of Physics D: Applied Physics*, 24(2), 111–118. doi: 10.1088/0022-3727/24/2/004.

Littlechild, J. A., Guy, J., Connelly, S., Mallett, L., Waddell, S., Rye, C. A., Line, K. and Isupov, M. (2007) Natural methods of protein stabilization: thermostable biocatalysts. *Biochemical Society Transactions*, 35(6), 1558–1563. doi: 10.1042/BST0351558.

Liu, X., Cao, L. C., Fan, X. J., Liu, Y. H. and Xie, W. (2016) Engineering of a thermostable esterase Est816 to improve its quorum-quenching activity and the underlying structural basis. *Scientific Reports*. Nature Publishing Group, 6(July), 1–11. doi: 10.1038/srep38137.

Lois, L. M., Campos, N., Putra, S. R., Danielsen, K., Rohmer, M. and Boronat, a (1998) Cloning and characterization of a gene from *Escherichia coli* encoding a transketolase-like enzyme that catalyzes the synthesis of D-1-deoxyxylulose 5-phosphate, a common precursor for isoprenoid, thiamin, and pyridoxol biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 95(5), 2105–2110. doi: 10.1073/pnas.95.5.2105.

Long, F., Vagin, A. A., Young, P. and Murshudov, G. N. (2007) BALBES: A molecular-replacement pipeline. *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 64(1), 125–132. doi: 10.1107/S0907444907050172.

Lovell, S. C., Davis, I. W., Adrendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. and Richardson, D. C. (2003) Structure validation by C alpha geometry: phi,psi and C beta deviation. *Proteins-Structure Function and Genetics*, 437–450. doi: 10.1002/prot.10286.

Lynn, D. J. (2002) Synonymous codon usage is subject to selection in thermophilic bacteria. *Nucleic Acids Research*, 30(19), 4272–4277. doi: 10.1093/nar/gkf546.

Lytle, F. W. (1966) Determination of interatomic distances from X-ray absorption

fine structure, (2), 398–399. *Advances in X-Ray Analysis*, 398-409, doi: 10.1007/978-1-4684-7633-0_36

Macedo-Ribeiro, S., Darimont, B., Sterner, R. and Huber, R. (1996) Small structural changes account for the high thermostability of 1[4Fe- 4S] ferredoxin from the hyperthermophilic bacterium *Thermotoga maritima*. *Structure*, 4(11), 1291–1301. doi: 10.1016/S0969-2126(96)00137-2.

Maduro, M. (2014) *E. coli* Codon Usage Analyzer 2.1. *Programa computacional*.

Mallick, P., Boutz, D. R., Eisenberg, D. and Yeates, T. O. (2002) Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds. *Proceedings of the National Academy of Sciences*, 99(15), 9679–9684. doi: 10.1073/pnas.142310499.

Mao, Y. J., Sheng, X. R. and Pan, X. M. (2007) The effects of NaCl concentration and pH on the stability of hyperthermophilic protein Ssh10b. *BMC Biochemistry*, 8, 1–8. doi: 10.1186/1471-2091-8-28.

Matsue, Y., Mizuno, H., Tomita, T., Asami, T., Nishiyama, M. and Kuzuyama, T. (2010) The herbicide ketoclomazone inhibits 1-deoxy-D-xylulose 5-phosphate synthase in the 2-C-methyl-D-erythritol 4-phosphate pathway and shows antibacterial activity against *Haemophilus influenzae*. *Journal of Antibiotics*. Nature Publishing Group, 63(10), 583–588. doi: 10.1038/ja.2010.100.

McNicholas, S., Potterton, E., Wilson, K. S. and Noble, M. E. M. (2011) Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 67(4), 386–394. doi: 10.1107/S0907444911007281.

Menzel, P., Gudbergsdóttir, S. R., Rike, A. G., Lin, L., Zhang, Q., Contursi, P., Moracci, M., Kristjansson, J. K., Bolduc, B., Gavrilov, S., Ravin, N., Mardanov, A., Bonch-Osmolovskaya, E., Young, M., Krogh, A. and Peng, X. (2015) Comparative Metagenomics of Eight Geographically Remote Terrestrial Hot Springs. *Microbial Ecology*, 70(2), 411–424. doi: 10.1007/s00248-015-0576-9.

Micaêlo, N. M. and Soares, C. M. (2007) Modeling hydration mechanisms of enzymes in nonpolar and polar organic solvents. *FEBS Journal*, 274(9), 2424–2436. doi: 10.1111/j.1742-4658.2007.05781.x.

Minagawa, H., Yoshida, Y., Kenmochi, N., Furuichi, M., Shimada, J. and Kaneko, H. (2007) Improving the thermal stability of lactate oxidase by directed evolution. *Cellular and Molecular Life Sciences*, 64(1), 77–81. doi: 10.1007/s00018-006-6409-8.

Momb, J., Wang, C., Liu, D., Thomas, P. W., Petsko, G. A., Guo, H., Ringe, D. and Fast, W. (2008) Mechanism of the quorum-quenching lactonase (AiiA) from *Bacillus thuringiensis*. Substrate modeling and active site mutations. *Biochemistry*, 47(29), 7715–7725. doi: 10.1021/bi8003704.

Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica D Biological Crystallography*. Chemistry Department, University of York, Heslington, England., 53(Pt 3), 240–255. doi: 10.1107/S0907444996012255.

Nemoviov, I. and Kut, I. (2012) Alternative Protein Crystallization Technique: Cross-Influence Procedure (CIP). *Crystallization and Materials Science of Modern Artificial and Natural Crystals*. doi: 10.5772/30461.

Ng, F. S. W., Wright, D. M. and Seah, S. Y. K. (2011) Characterization of a Phosphotriesterase-Like Lactonase from *Sulfolobus solfataricus* and Its Immobilization for Disruption of Quorum Sensing. *Applied and Environmental Microbiology*, 77(4), 1181–1186. doi: 10.1128/AEM.01642-10

Nikkola, M., Lindqvist, Y. and Schneider, G. (1994) Refined Structure of Transketolase from *Saccharomyces cerevisiae* at 2.0 Å Resolution. *Journal of Molecular Biology*. Academic Press, 238(3), 387–404. doi: 10.1006/JMBI.1994.1299.

Paoli, M. (2001) Protein folds propelled by diversity. *Progress in Biophysics and Molecular Biology*, 76(1–2), 103–130. doi: 10.1016/S0079-6107(01)00007-4.

Paredes, D. I., Watters, K., Pitman, D. J., Bystroff, C., & Dordick, J. S. (2011). Comparative void-volume analysis of psychrophilic and mesophilic enzymes: Structural bioinformatics of psychrophilic enzymes reveals sources of core flexibility. *BMC structural biology*, 11, 42. doi:10.1186/1472-6807-11-42

Petersen, T. N., Brunak, S., Von Heijne, G. and Nielsen, H. (2011) SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*. Nature Publishing Group, 8(10), 785–786. doi: 10.1038/nmeth.1701.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. doi: 10.1002/jcc.20084.

Popiel, D., Koczyk, G., Dawidziuk, A., Gromadzka, K., Blaszczyk, L. and Chelkowski, J. (2014) Zearalenone lactonohydrolase activity in *Hypocreales* and its evolutionary relationships within the epoxide hydrolase subset of a/b-hydrolases. *BMC Microbiology*, 14(1), 1–12. doi: 10.1186/1471-2180-14-82.

Quellère, P. (2007) Why is alanine such a good α-helix builder. *Acta Crystallographica Section A Foundations of Crystallography*, 63(a1), s113–s113. doi: 10.1107/S0108767307097590.

Ramírez-Aportela, E., López-Blanco, J. R. and Chacón, P. (2016) FRODOCK 2.0: fast protein–protein docking server. *Oxford Academic*, 32(15), 2386–2388. doi: 10.1093/bioinformatics/btw141

Rasko, D. A. and Sperandio, V. (2010) Anti-virulence strategies to combat bacteria-mediated disease. *Nature Reviews Drug Discovery*. Nature Publishing Group, 9, 117. doi: 10.1038/nrd3013

Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1), 320–324. doi: 10.1093/nar/gku316.

Romero, D. L., Morge, R. A., Genin, M. J., Biles, C., Busso, M., Resnick, L., Althaus, I. W., Reusser, F., Thomas, R. C. and Tarpley, W. G. (1993) Bis(heteroaryl)piperazine (BHAP) Reverse Transcriptase Inhibitors: Structure-Activity Relationships of Novel Substituted Indole Analogues and the Identification of l-[(5-Methanesulfonamido-liT-indol-2-yl)- carbonyl]-4-[3-[(l-methylethyl)amino]- pyridinylj, J Med Chem. 1993 May 14-36(10), 1505–1508.

Rupp, B. (2009) *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. 1st editio. Garland Science.

[185]

Rye, C. A., Isupov, M. N., Lebedev, A. A. and Littlechild, J. A. (2009) Biochemical and structural studies of a l-haloacid dehalogenase from the thermophilic archaeon *Sulfolobus tokodaii. Extremophiles*, 13(1), 179–190. doi: 10.1007/s00792-008-0208-0.

Ryu, Y. W. and Ryu, D. D. Y. (1986) Semisynthetic p-lactam antibiotics synthesizing enzyme from *Acetobacter turbidans*: purification and properties. *Enzyme and Microbial Technology*, 9, 339–344. doi: 10.1016/0141-0229(87)90056-1

Samson, M., Porter, N., Orekoya, O., Hebert, J. R., Adams, S. A., Bennett, C. L. and Steck, S. E. (2017) Targeting DXP synthase in human pathogens: enzyme inhibition and antimicrobial activity of butylacetylphosphonate. Progestin, *The Journal of Antibiotics,* 155(1), 3–12. doi: 10.1007/s10549-015-3663-1..

Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W. and Hughes, G. J. (2010) Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science*, 329(5989), 305 LP-309. doi: 10.1126/science.1188934

Sayer, C., Bommer, M., Isupov, M., Ward, J. and Littlechild, J. (2012) Crystal structure and substrate specificity of the thermophilic serine:pyruvate aminotransferase from *Sulfolobus solfataricus. Acta Crystallographica Section D*, 68(7), 763–772. doi: 10.1107/S0907444912011274.

Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M. and Witholt, B. (2001) Industrial biocatalysis today and tomorrow. *Nature*. Institute of Biotechnology, ETH-Honggerberg, Zurich, Switzerland., 409(6817), 258–268. doi: 10.1038/35051736.

Schulz, G. E. (2007) Protein Crystallization. *Comprehensive Medicinal Chemistry II*. Elsevier, 433–446. doi: 10.1016/B0-08-045044-X/00093-6.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D. and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 539. doi: 10.1038/msb.2011.75.

Sisquella, X., de Pourcq, K., Alguacil, J., Robles, J., Sanz, F., Anselmetti, D., Imperial, S. and Fernandez-Busquets, X. (2010) A single-molecule force spectroscopy nanosensor for the identification of new antibiotics and antimalarials. *The FASEB Journal*, 24(11), 4203–4217. doi: 10.1096/fj.10-155507.

Smith, J. M., Warrington, N. V, Vierling, R. J., Kuhn, M. L., Anderson, W. F., Koppisch, A. T. and Freel Meyers, C. L. (2014) Targeting DXP synthase in human pathogens: enzyme inhibition and antimicrobial activity of butylacetylphosphonate. *The Journal of Antibiotics (Tokyo).* 67(1), 77–83. doi: 10.1038/ja.2013.105.

Sokalingam, S., Raghunathan, G., Soundrarajan, N. and Lee, S.-G. (2012) A Study on the Effect of Surface Lysine to Arginine Mutagenesis on Protein Stability and Structure Using Green Fluorescent Protein. *PLOS ONE.* Public Library of Science, 7(7), e40410.

Song, J. A. E. K. and Rhee, J. S. (2000) Simultaneous Enhancement of Thermostability and Catalytic Activity of Phospholipase A 1 by Evolutionary Molecular Engineering, *Applied and Environmental Microbiology.*, 66(3), 890–894. doi: 10.1128/AEM.66.3.890-894.2000

Srinivas, N. R., Barbhaiya, R. H. and Midha, K. K. (2001) Enantiomeric drug development: Issues, considerations, and regulatory requirements. *Journal of Pharmaceutical Sciences*, 90(9), 1205–1215. doi: 10.1002/jps.1074.

Struvay, C. and Feller, G. (2012) Optimization to Low Temperature Activity in Psychrophilic Enzymes, *International Journal of Molecular Sciences* 13(9), 11643–11665. doi: 10.3390/ijms130911643.

Taylor, S. (1998) Chem. Br. *Chemical Education Trust Fund for the Chemical Society*

Taylor, S. J., McCague, R., Wisdom, R., Lee, C., Dickson, K., Ruecroft, G., O'Brien, F., Littlechild, J., Bevan, J., Roberts, S. M. and Evans, C. T. (1993) Development of the biocatalytic resolution of 2-azabicyclo[2.2.1]hept-5-en-3-one as an entry to single-enantiomer carbocyclic nucleosides. *Tetrahedron: Asymmetry*, 4(6), 1117–1128. doi: 10.1016/S0957-4166(00)80218-9.

[187]

Thoma, R., Hennig, M., Sterner, R. and Kirschner, K. (2000) Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from T*hermotoga maritima*. *Structure*, 8(3), 265–276. doi: 10.1016/S0969-2126(00)00106-4.

Valadez-Blanco, R., Ferreira, F. C., Jorge, R. F. and Livingston, A. G. (2008) A membrane bioreactor for biotransformations of hydrophobic molecules using organic solvent nanofiltration (OSN) membranes. *Journal of Membrane Science*, 317(1–2), 50–64. doi: 10.1016/j.memsci.2007.04.032.

Vieille, C. and Zeikus, G. J. (2001) Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiology and Molecular Biology Reviews*, 65(1), 1–43. doi: 10.1128/MMBR.65.1.1-43.2001.

Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., Nettleship, J. E. E., Owens, R. J., Stuart, D. I. I. and Grimes, J. M. (2006) Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. *Structure*, 14(11), 1617–1622. doi: 10.1016/j.str.2006.09.005.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 67(4), 235–242. doi: 10.1107/S0907444910045749.

Winter, G. (2010) xia2: an expert system for macromolecular crystallography data reduction.*Journal of Applied Crystallography*. 43, 186–190. doi: Doi 10.1107/S0021889809045701.

Xiang, S., Usunow, G., Lange, G., Busch, M. and Tong, L. (2013) 1-deoxy-D-xylulose 5-phosphate synthase (DXS), a crucial enzyme for isoprenoids biosynthesis. *Isoprenoid Synthesis in Plants and Microorganisms: New Concepts and Experimental Approaches*, 282(4), 17–28. doi: 10.1007/978-1-4614-4063-5_2.

Yang, Z., Zhang, L., Zhang, Y., Zhang, T., Feng, Y., Lu, X., Lan, W., Wang, J., Wu, H., Cao, C. and Wang, X. (2011) Highly efficient production of soluble proteins from insoluble inclusion bodies by a Two-Step-Denaturing and refolding method. *PLoS ONE*, 6(7), 1–8. doi: 10.1371/journal.pone.0022981.

Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9, 1–8. doi: 10.1186/1471-2105-9-40.

Zhou, X. X., Wang, Y. B., Pan, Y. J. and Li, W. F. (2008) Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. *Amino Acids*, 34(1), 25–33. doi: 10.1007/s00726-007-0589.

9. Appendix
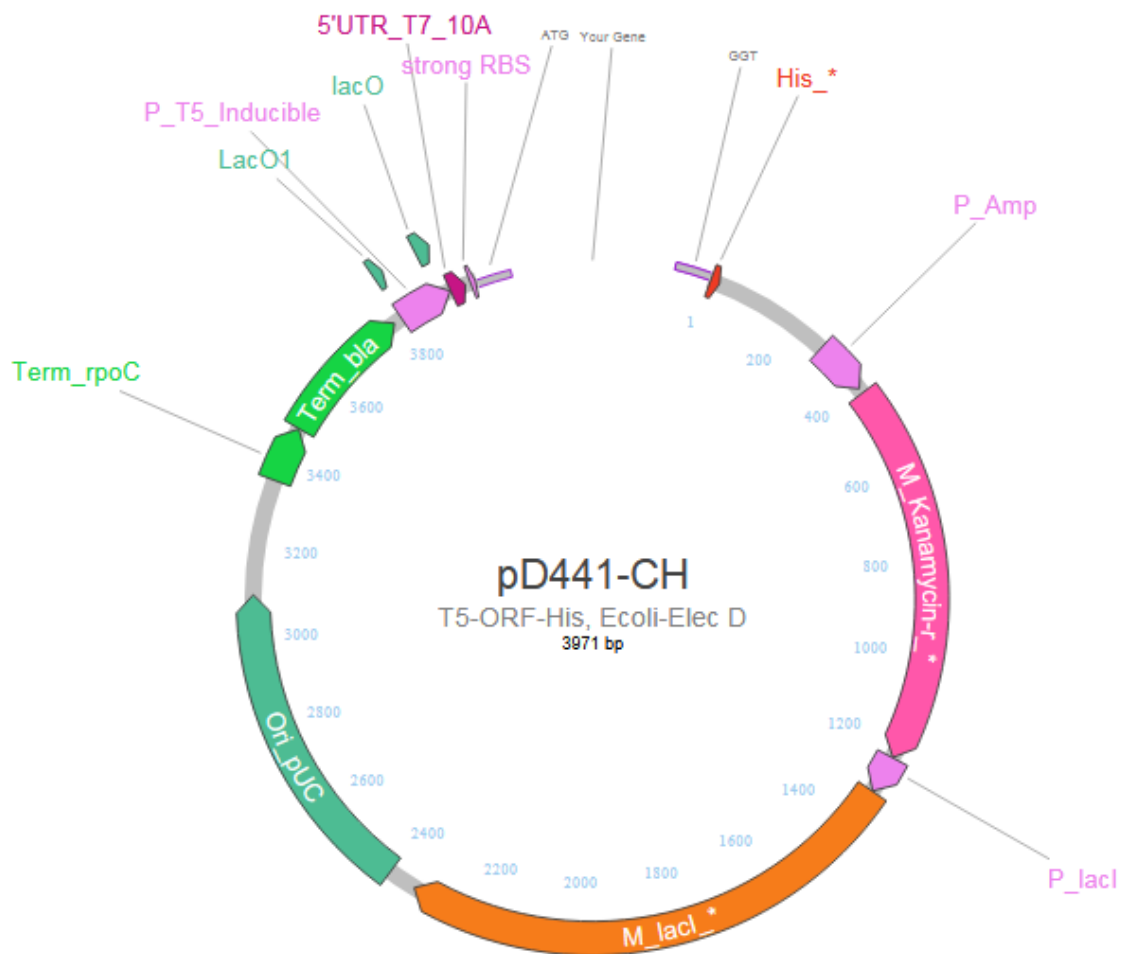
The pD441-CH expression plasmid provided by ATUM (California, USA) and is part of their Electra cloning system. Genes that were synthesised were inserted into this gene containing a T5 promoter with a strong ribosome binding site, a C-terminal histidine tag, and a kanamycin resistance gene to selection purposes. Expression of the gene was controlled by the lacI operon and IPTG inducible.
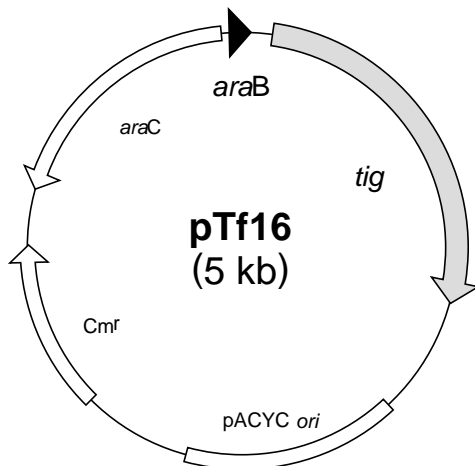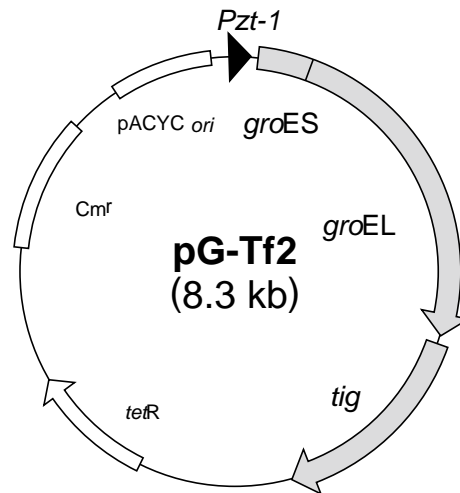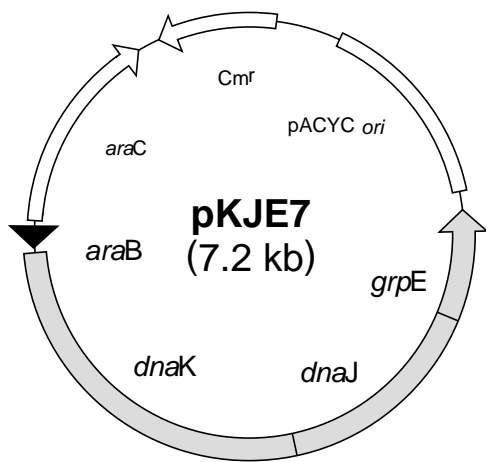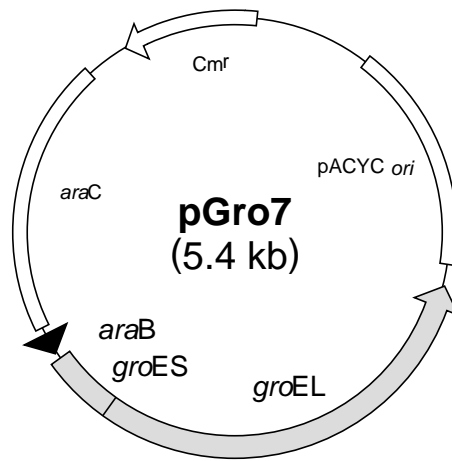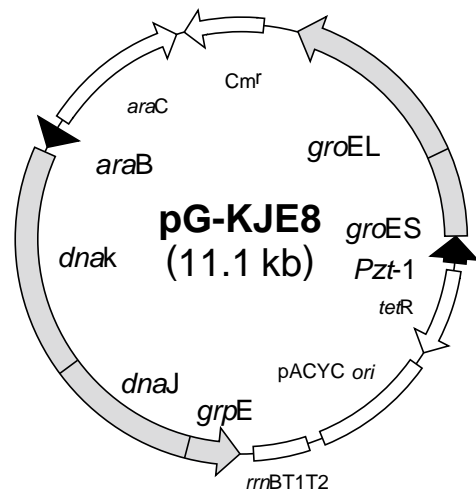
Takara chaperone plasmids. Plasmids contain a chloramphenicol resistance gene and one or more of the following chaperones: GroEL with a size of 60 kDa; GroES at 10 kDa; DnaK at 70 kDa; DnaJ at 40 kDa; Tf at 56 kDa; GrpE at

22 kDa. Gene expression is controlled by arabinose through the araB promoter, or by tetracycline via the Pzt-1 promoter.

Enzyme amino acid sequences

Vulcanisaeta moutnovskia lactonase –

GMVRISIAGGNEIDPGSMGLTLFHEHLRLITEVVRWNWPHLYNEDEELKRAIDA
VNAAKKYGVKTIIDLTVAGIGCDVRFNEKVAKATGVNIIMGTGFYTYTEIPFYFK
NRGIDSLVDAFVHDITIGIQGTNTRAAFVKAVIDSSGLTKDVEMAIRAAAKAHIKT
DVPIITHSFVGNKSSLDLIRIFKEEGVDLARTVIGHVGDTDDISFIEQILREGAFIG
LDRFGLDIYLPLDKRVKTAIELIKRGWIDQLLLSHDYCPTIDWYPPEVVRSTVPD
WTMTLIFEKVIPRMRSEGITEEQINRVLIDNPRRLFTGR

Tomsk lactonase –

MLAARPSAPDGVWHGEVSVAGGKVSFAMAGDGLPLILLHGWTLDHRMWRP
QIGQLSRDFLLVMPDRRGCGASTAPPDLSREAEDVIAIADFLGFERFGLLGLS
QGAVVALDVARKFSSRLTGLVVSGAPLPCLVERDEAIPLDRYRAMVAAGDVA
GMRRDWARHPLMRTHDPDARTLAAAMLADYDGRDLAAVSEPPGLPREVLSH
LAVPVLALAGEHDTPWRRACAAALADCAPRGRHALIGRAGHLANCDNPQDFN
ALAGFLRTCADPRTRANR

B1 lactonase –

MTPKFTVPDHLKPYQRRISANGVGLHLYDSGPAQAHDPTFLLIHGLGDEADS
WRKVFPLLTGRGRVIAPDLPGFGRSEHPRRAYTLNFFADTMAALLESLKVPQA
VLVGSSMGAAVALRVALRRADLVARLVLVDGPPVRGRLNRVQLMFLIPGQGE
KLYNSFRSSQEAAFESLRPYYASLDALPPEDRQFLRERVWDRVWSDDQRRA
FFSTFRWMALESLLGRARLGQVKTPTLLVWGEQDAVIPLEAAKTLQSWIPGSQ
LQVIPSCGHLPQQEKPLELIRLILQ

Lac11 –

MRKLLGSSAIHVLSRRTGYGMYFLILALGVATQVVFGAEPSQNPSEKVAWIER
ADPGLDDLIAPDAQVEVLAEGFEWSEGPVWIPEGGYLLFSDVPKNTIYRWKE
GQGIDIFLKPSGYTGFRERGGESGSNGLALDRQGRLLLCQHGDRRVARWEK

[193]

GCFITLADQYEGKPLNSPNDLVVKSNGDIYFTDPPYGMTPEAQRDPNALGFC
GVYRISADGKLTLLVRDMTRPNGIAFSPDEKTLYVAQSDPQRPLWMAFPVRE
DGTLGEGKVFFDAKPWQQSGLPGLPDGMKVDQKGNLFATGPGGVNIFRPD
GTFLGRIRVKVPTANCAFGDDGSTLYITADMYLLRVKTKTKGLGF

TaDXP –

ILDKVNSPDDLKKLTTQELKQLASELRDYIISVVEKTGGHLASSLGVVELTIALLK
VFSPPKDEIVWDVGHQSYPYKILTDRKEKFKTLRQFGGISGFPSIKESPYDAFG
TGHSSTSISAALGIKVGKRLKGEEGHVIAVIGDGALTAGEAYEGLNNAGQLKED
LIVILNDNEMSISKNIGAISNYLTKLTTGESLRRAKERLEEVTKKIFGDTFYKGLK
RVEDLIVKGLFPPGMLFEELGFRYVGPIDGHDIETLVTTLRNVSKMRGPTLVHV
LTKKGKGHKPAEERPDKFHGVSPKKLLSEPQPPTYTEVFSKTLVEMAEKDSSI
VAITAAMPSGTGLDRFAARFPERYFDVGIAEQHAVTFAAGLAKKGLRPVVAIYS
TFLQRAYDQIIHDVALQELPVTFAIDRAGLVGEDGATHHGAFDLSYLRVVPNLV
VAAPKDEEELRHLLYTAVYSGRPFAVRYPRGRGYGVTLREPLKKIPIGSWEVL
REGGDLLILATGWTVYQALEAARELSAEGIEATVVNARFIKPLDEKLLKELALKH
STVITVEENAVKGGFGSAVNEFLALWYNGRVFNLGLPDKFIEHGSQALLRKLV
KIDKDGIKESVRE

ChDXP –

PILERISLPEDIKKLKPSELMALAQELREYIITVASQNGGHLAPSLGVVELTIALH
FVFEAPKDKIIWDVGHQAYAHKILTGRKKQFKTLRTFGGLSGFPKRDESPYDA
FGVGHSSTSISAALGMALARDLKGEQYEVVAVIGDGALTGGMAFEALNHAGH
LQKKLIVVVNDNEMSIAQNVGALSAYLSRIRTDPKYSRGKDELEALIKKIPHIGP
TMVKIGERLKDSFKYLLVPGMLFEELGFTYLGPIDGHNIKEMIEVFSRAKTFAG
PVVVHVITKKGKGYHWAEENPDGFHGVGKFYISTGEPVEAPRVSFTEVFGKA
LVELAQDRPEVVAITAAMPTGTGLNYFAQNYPERFYDVGIAEQHAVTMAAGM
ACEGLKPVVAIYSTFLQRSFDQIIHDVCLQNLPVVFAVDRAGIVGEDGPTHHGI
FDLSYLRMIPNLTIMVPRNEDMLRKMLFTALNHSGPVALRYPRGAAVGVELTP
YEQLPIGTAEILKEGSDGVVIGVGRPLNYALKAAQKLENEGISLTVIDARFVKPL
DYKLLEEVGSLHKPVITVEENVVAGGFGSAVNEYFSFRGIGTKVVNLGIADEFP
PHGKVEEILNLYGLTEEKLYLKFREILSKL

[194]