BIOCHEMICAL AND STRUCTURAL CHARACTERISATION OF DEHALOGENASES FROM MARINE BACTERIA

Submitted by

Halina Novak

to the University of Exeter as a thesis for the Doctor of Philosophy in Biological Sciences

(September 2011)

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.

Halina Novak

Abstract

An L-haloacid dehalohenase from the psychrophilic marine bacteria *Psychromonas ingrahamii* has been cloned, over-expressed in a bacterial expression system and biochemically characterised. The enzyme is stable at temperatures of up to 60ºC for 90 min and shows highest activity towards substrates with short carbon chains (≤C3). The enzyme is stable in up to 30% ethanol, methanol and DMSO when incubated for 1 h. The K_m for the enzyme is 1.36 mM.

The genome of AQP5750 from the Aquapharm Biodiscovery Ltd Microbial library was sequenced. An L-haloacid dehalohenase and a haloalkane dehalogenase gene were identified within the genome.

The AQP5750 L-haloacid dehalohenase has been cloned, over-expressed in a bacterial expression system, biochemically characterised, crystallized and the native and crystal complex structure with chloropropionic acid (MCP) determined. The L-haloacid dehalohenase from AQP5750 shows highest activity at 55ºC towards brominated substrates with short carbon chains (≤C3). The enzyme shows increased activity of 150% in 40% DMSO and 123% in 30% methanol. The L-haloacid dehalohenase crystal complex structure with covalently bound MCP confirmed Asp 18 as the main catalytic residue. Residues His 183, Asp 186 and Glu 21 in the active site are proposed to be involved in activation of the catalytic water which attacks the ester intermediate in the second part of the S_N2 dehalogenase mechanism.

The AQP5750 haloalkane dehalogenase has been cloned, over-expressed in a bacterial expression system, crystallized and the native and complex structure with 1-hexanol has been determined. Substrate specificity experiments showed that the haloalkane dehalogenase from AQP5750 does not show high activity towards substrates used by other haloalkane dehalogenases with high amino acid sequence identity. The large active site cavity and the presence of Ser 176 and Arg 136 in the hydrophobic binding pocket may alter the binding of substrates tested which could account for the low activity observed.

Acknowledgements

I would like to thank Richard Tennant for his help with sample collection, Dr Hannah Florance for preparing, running and analysing the protein samples on the LC Q-TOF mass spectrometer and for useful discussions, Alex Moorhouse for the preparation of genomic DNA for genome sequencing, Dr Konrad Paszkiewicz for annotating the AQ5750 genome and for useful discussions, Jana Panning for help with the cloning experiments and Dr Michail Isupov for his help with the X-ray crystallography and for useful discussions.

I would like to say a big thank you to Dr Christopher Sayer for teaching me X-ray crystallography and for his help with biochemical characterisation, for useful discussions and continuous support.

I would also like to thank Dr Carrie Rye, Dr Kirsty Line, Dr Anne Marie Hickey and Dr Rob Gibson for their help throughout the project. I would also like to thank everyone in the Biocatalysis centre.

Thank you to the BBSRC and Aquapharm Biodiscovery Ltd for funding and Professor Jenny Littlechild and Dr Dorothee Gotz for supervision.

Finally, I would like to thank my parents for always supporting me and funding my education and to Nick who has been my biggest support over the last four years.

Table of Contents

List of Figures

List of Tables

List of Equations

Abbreviations

APS	Ammonium persulfate
BAM	Benzamidine
BHX	Bromohexane
BLAST	Basic Local Alignment Search Tool
Bp	base pairs
1BO	1-bromohexane
CCP4	Collaborative Computational Project, Number 4
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
ddH_2O	double distilled H_2O
dNTPs	Deoxyribose nucleotides
DTT	Dithiothreitol
EC	Enzyme commission
EDTA	Ethylenediaminetetraacetic acid (disodium salt)
FFQ	Fast flow Q
GF	Gel filtration
GFP	Green fluorescent protein
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIC	Hydrophobic interaction chromatography
His-tag	Poly histidine tag
IPTG	Isopropyl β-D-galactopyranoside
kbp	Kilo base pair
KDa	Kilo Dalton
LB	Luria-Bertani
MAD	Multi wavelength anomalous dispersion
Mbp	Mega base pair
MCP	l-chloropropionic acid
MES	2-(N-morpholino) ethanesulfonic acid
MEOR	Microbial enchanced oil recovery
MWCO	molecular weight cut off
NAC	Nickel affinity chromatography

PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Databank
PHAs	Polyhydroxyalkanoates
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
rRNA	Ribosomal ribonucleic acid
rmp	Revolutions per minute
SAP	Shrimp alkaline phosphatase
SAD	Single wavelength anamalous dispersion
SDR	Short chain reductases
SDS	Sodium dodecyl sulphate
TE	Tris-HCl pH 8.0, 1 mM EDTA buffer
TEMED	N,N,N,N-tetramethylethylene diamide

Chapter 1- Introduction

1.1 Biocatalysts

Biocatalysis refers to the use of enzymes in biotransformation reactions. High turnover and stability are key properties of the ideal biocatalyst (Burton *et al.,* 2002). It is often very difficult to manufacture enantiomerically pure compounds using traditional organic synthesis methods. Enzymes are often regioselective, chemoselective and stereospecific. These characteristics make biocatalysts useful for the production of compounds with chiral centres for industrial and pharmaceutical applications. Biocatalysts can be used as a whole cell system or as isolated enzymes.

The use of biocatalysts in industrial applications is increasing (Schmid *et al*., 2001). Two main factors need to be carefully considered to ensure the viability of a biocatalyst. The first factor is the concentration of the product synthesised, which must be between 50-100 g/L. The second factor is the activity unit of the biocatalyst (amount of product produced), which needs to be 15 for whole cell samples (gram product/gram cells) and 1000 for purified enzyme samples (gram product/gram enzyme) (Pollard and Woodley, 2006). The activity unit is much lower for whole cell systems. This is due to protein purification dramatically increasing the cost of using a biocatalyst.

Although biocatalysts are used in numerous industrial applications, they have the biggest role in the pharmaceutical sector (Buckland *et al*., 2000). Pharmaceutical companies work under huge pressure from the outset of the discovery of a novel compound with medicinal applications, to the point of marketing the product due to the short life-span of patents. A number of enzymes along with organic synthesis methods are often combined to produce a pharmaceutical compound (Davis and Boyer, 2001). Xylose isomerase is an example of a biocatalyst which is very stable and can be used over many months (Schoemaker *et al*., 2003). After the initial identification of xylose isomerase in *Pseudomonas hydrophila*, the enzyme was found to be produced by many bacteria (Marshall *et al*., 1957). The xylose isomerase from *Thermus aquaticus* shows high activity and stability at

temperatures up to 70ºC (Lehmacher *et al*., 1990). Xylose isomerase enzymes have been used for the production of D-glucose and D-fructose. High fructose corn syrup can be used as a substitute for sugar.

There are many advantages of using biocatalysts to produce compounds rather than using organic synthesis methods. Biocatalysts show high activity in mild conditions and in aqueous solutions. Of all solvent consumption, over 80% is used by industrial and pharmaceutical companies (Woodley, 2008). These solvents can be toxic and harmful. The use of enzymes not only decreases solvent consumption but also allows disposal and solvent extraction methods to be avoided. Biocatalysts can be reused over long periods of time and the steps in the synthesis of a compound can be reduced. These advantages can dramatically reduce the cost of synthesising a compound. These reasons make the biocatalytic route a greener process in comparison to traditional organic synthesis methods (Ran *et al*., 2007, Tao *et al*., 2009).

The disadvantages of using biocatalysts are often due to the instability of the enzyme. Conditions in biological systems are tightly controlled so that enzymes are able to work efficiently. When enzymes are used under harsh conditions, such as high temperatures, pH and in the presence of solvents they may become unstable (Polizzi and Bommarius, 2007). However, some enzymes naturally show very high stability. For example, the lipase from *Staphylococcus saprophyticus* M36 shows high activity in organic solvents (Fang *et al*., 2006).

Limitations of biocatalysts can be overcome using a number of methods. Advances in recent years in genetic engineering, molecular biology techniques and in recombinant methods have allowed the production of large quantities of biocatalysts at relatively low costs. Numerous techniques have been developed to try and improve the activity or substrate specificity of an enzyme with a particular industrial application in mind. This includes random mutagenesis of a target gene followed by screening and direct protein engineering. The latter technique is dependent on the determination of structural information which is

used to predict stabilising mutations (Cherry and Fidantsef, 2003, Eijsink *et al*., 2005, Morley *et al*., 2005).

Enzyme immobilisation has also been used to overcome instability issues and has allowed biocatalysts to be reused more efficiently (Mateo *et al.*, 2007). Immobilisation techniques were first developed in the 1950's. An example of an efficiently immobilised enzyme that is used in a chemo-enzymatic process is the nitrilase from *Acidovorax facilis* in the production of glycolonitrile. The recombinant enzyme was immobilised in carrageenan (linear sulphated polysaccharides, extracted from red algae) and is able to synthesise 1 kg of product for every 1 g of dry cell weight used (Panova *et al*., 2007, Beloqui *et al*., 2008). Glycolic acid (which can be used as a precursor molecule in pharmaceutical applications) is obtained from glycolonitrile by ion exchange chromatography.

To overcome temperature instability, enzymes from thermophiles have been used as they can show enhanced stability at high temperatures. Enzymes from extremophiles are often more robust and stable at high pressures and in solvents.

With the need to produce a wide range of structurally diverse molecules in pharmaceutical and chemical industries, the discovery of novel enzymes and proteins is becoming increasingly important. High throughput screening methods allow the identification of novel proteins. Automated methods with easily interpreted results are favoured, such as colorimetric assays which can be measured spectroscopically. Increased research into high throughput screening techniques will allow higher numbers of novel enzymes to be identified.

The commercial availability of enzymes with desired activity is one of the biggest limitations when considering using a biocatalyst. The production of recombinant proteins in high quantities, protein engineering, immobilisation and the use of proteins from extremophiles will increase the use of biocatalysts in the future.

1.2 The marine environment

Approximately 71% of the Earth's surface is covered by aquatic environments. This provides a huge habitat for marine life where a diverse range of plants, invertebrates, vertebrates and marine microorganisms have adapted to survive (Kennedy *et al*., 2008). The marine environment is considered an untapped resource, which over the last 20 years has been recognised as a potential source of novel enzymes and metabolites (Trincone, 2011).

Marine biotechnology has been described as "the exploration of the capabilities of marine organisms at the whole, cell, or molecular level, to provide solutions to today's problems, with the use of technology to advance the understanding and accessibility of marine biological material" (Lee and Burrill, 1994). This includes the discovery of pharmaceutically active molecules, the search for enzymes and biopolymers for use in industrial applications, research into biofuels from algae, research into sustainable fisheries and the management of marine ecosystems (Querellou *et al*., 2010).

1.2.1 Biopolymers from marine sources

Biopolymers from marine sources are of great interest in pharmaceutical, agricultural and medical industries. For example, laminarin from red algae is used to enhance crop growth. There is also a considerable interest in bioplastics such as polyhydroxyalkanoates (PHAs) which are synthesised by a range of bacteria, such as *Ralstonia eutropha* (Jacquel *et al*., 2008). PHAs are high in molecular weight, non-toxic and biodegradable.

1.2.2 Energy from marine sources

There is a worldwide demand for the discovery of new energy supplies. Microbial enhanced oil recovery (MEOR) is a biological technique used to improve the efficiency of fossil fuel recovery. The process aims to enhance fossil fuel recovery by using microorganisms to reduce the viscosity of oil in petroleum reservoirs (Al-Sulaimani *et al*., 2011). By reducing the viscosity, the recovery of oil can be greatly increased (Querellou *et al*., 2011). There are two methods used in MEOR. The first method is the injection of nutrients which increases the

growth rate of native microbes. The second method is the injection of specific bacterial strains. MEOR is considered controversial and extensive research into nutrient development and microbial selection is needed (Sen, 2008).

Many worldwide research programmes are looking into biofuels from microalgae (Hossain *et al*., 2008). Unlike the research into the production of biodiesel from terrestrial crops, the production of biofuels from microalgae is still in its infancy.

1.2.3 Pharmaceutically active molecules from the marine environment

Although the majority of marine research is focussed around microorganisms, plants and animals also contain valuable enzymes and secondary metabolites which can show interesting pharmacological activity. These enzymes and novel compounds may be useful for the production of precursors or analogues of drugs. The first pharmaceutically active molecule from the marine environment was identified in the Caribbean sponge *Tethya crypta*. This led to the development of Arabinofuranosyl Cytidine (Ara-C) in 1969, which is a drug used to treat non-Hodgkins lymphoma and leukaemia (Mulder and Harrap, 1965). Since this discovery, many pharmaceutically active compounds from marine species have been identified. To date, approximately 15 natural marine products are currently in clinical trials (Querellou *et al*., 2010). The majority of these compounds are anti-cancer drugs, for example Trabectedin (ecteinascidin 743) which is an anti-tumour drug isolated from a tube worm (Izbicka *et al*., 1998). Acyclovir is another example of a drug isolated from a marine source, this antiviral drug acts against the herpes virus (type 1 and 2) and varicella-zoster virus (shingles and chicken pox). Acyclovir was modelled on a sponge derived spongothymidine that was originally isolated from *Cryptotethya crypta* (it was later found to be a natural product from a gorgonian (a type of soft coral) *Eunibcella cavolini* species which inhabited the *C. crypta*) (Schaeffer and Krylov, 2000). Acyclovir is more commonly known as ZOVIRAX®.

1.2.4 Enzymes and proteins isolated from the marine environment

To date, the majority of enzymes have been isolated from terrestrial environments (Querellou *et al*., 2010). Although only a small number of proteins from marine organisms have been characterised, some of these proteins are extremely important. For example, Shrimp Alkaline Phosphatase (SAP), DNA polymerases from Thermococcales and green fluorescent protein (GFP) from *Aequorea victoria* are widely used and considered essential in molecular and cell biology research (De backe *et al*.,2002, Niehaus *et al*., 1997, Prasher *et al*., 1992, Cubitt *et al*., 1995).

Marine algae have been recognised as a potential source of enzymes and natural products (Faulkner, 2001). Many novel enzymes from marine animals are of great interest but a lot of research is focused on marine microorganisms. This is because marine microorganisms living with plants and animals (intracellularly and extracellularly) must contain complex numbers of enzymes and pathways to be viable candidates for symbiosis. A direct relationship between salt tolerance and enzyme stability in organic solvents has been observed. This makes biocatalysts from halophiles (which are able to grow at concentrations ≥0.2 M NaCl) attractive for use in industrial processes (Marhuenda-Egea *et al*., 2002, Trincone, 2011). Oxidoreductase enzymes isolated from a marine microbe showed interesting stereoselectivity in comparison to similar enzymes isolated from terrestrial environments (Trincone, 2010). These enzymes display higher activity and thermostability in high salt conditions. This example highlights the novel characteristics of biocatalysts sourced from the marine environment.

1.2.4.1 The ‘great plate count anomaly’

Approximately 99% of microorganisms cannot be cultivated *in vitro.* This is one of the oldest microbiological problems and is referred to as the ‘great plate count anomaly’. This means that only a fraction of microbial species present in bacterial communities have been identified. This problem has made culturing and isolating marine microorganisms very challenging. To overcome this problem, a significant amount of microbiological research into the optimisation of culture media (varying the concentrations of nutrients, chemicals, carbon sources and the addition of

hormones) has been carried out (Sharma *et al*., 1995). Other techniques have also been developed which involve mimicking the natural environment that a microorganism is isolated from by constructing diffusion chambers. In these diffusion chambers substances from environmental samples diffuse across a membrane. This technique has allowed the cultivation of previously uncultivable microorganisms (Nichols *et al*., 2008).

Molecular techniques have been designed to identify unculturable microorganisms. This includes the identification of wild type species by amplifying the bacterial 16S rRNA gene by PCR. Metagenomics is also used to analyse genomic information from mixed microbial communities. This technique involves the construction of DNA libraries from mixed microbial environmental samples. The libraries are analysed further to identify the expression of particular phenotypes and are screened for sequences of known interest (Riesenfeld *et al*., 2004).

1.2.5 Extremophiles from the marine environment

Acid lakes, deep sea hydrothermal vents, brine pockets in icebergs and the Dead Sea are examples of extreme marine environments which are inhabited by unique microorganisms. These include bacteria, archaea and fungi which have adapted to survive. Of the marine environments, it is the extreme habitats which have been most extensively studied (Synnes, 2007). Marine microorganisms isolated from these habitats often possess unique properties and are referred to as extremophiles. This includes thermophiles (which are able to grow between 60ºC and 120ºC), psychrophiles (which are able to grow between 10ºC and -15ºC), barophiles (which are able to grow at pressures exceeding 38 MPa), acidophiles (which are able to grow at a pH ≤3), alkaliphiles (which are able to grow at a pH ≥9) and halophiles (which are able to grow at concentrations ≥0.2 M NaCl). Microorganisms that have adapted and evolved over time to live in these extreme conditions often contain novel enzymes which possess desirable characteristics and are highly stable (Champdoré *et* al., 2007).

Sea ice provides one of the coldest habitats on Earth and can vary between -1.8ºC and -30ºC. The sea ice interface has been identified as a key environment to isolate halophilic and psychrophilic microorganisms which may possess salt tolerant and cold adapted enzymes. Marine environments are not always rich in nutrients. Many microbial species have adapted and are able to survive in a state of dormancy or have developed other survival strategies (Hall-Stoodley *et al*., 2004).

1.2.5.1 Psychrophiles

Psychrophiles require adaptations to overcome two main problems associated with survival at low temperatures. These are the high viscosity of aqueous solutions (the viscosity of aqueous solution increases by over two fold between 37ºC and 0ºC) and controlled regulation of cell dynamics (lower membrane fluidity, regulation of nutrients and waste, decreased rates of transcription, translation, cell division and lower enzyme activity) at low temperatures. Research into the structural properties of proteins from psychrophiles show they have enhanced structural flexibility at low temperatures (Feller, 2003). The fine balance between activity, stability and flexibility of proteins which control enzyme kinetics has been reviewed by Georlette *et al.* (2006). The direct comparison of proteins from psychrophiles with equivalent proteins from mesophiles has shown they have decreased ionic interactions, hydrogen bonds and fewer hydrophobic residues. Enzymes from psychrophiles are considered a potential source of novel biocatalysts and have shown high specific activity at low temperatures (Cavicchioli *et al*., 2002, Siddiqui *et al*., 2006).

Research into the genomics of psychrophiles is being carried out with the aim of understanding how these organisms survive at low temperatures (Riley *et al*., 2008). The expression of highly regulated cold shock proteins are known to be involved in transcription, translation and protein folding (Phadtare *et al*., 1999). The production of anti-freeze proteins has also been studied. These proteins bind to ice crystals, allowing psychrophiles to grow at lower temperatures (Jia and Davies, 2002). The synthesis of exopolysaccharides (which have been detected

at high concentrations in Arctic sea ice) are thought to prevent proteins denaturing and aggregating (D'Amico *et al*., 2006).

Biocatalysts with high activity at low temperatures have the potential to be used in detergents (cellulases, lipases, amylases and proteases), bioremediation processes and industrial applications (Gerday *et al*., 2000). The use of psychrophilic enzymes in industrial processes allows instability issues with reactants and products to be avoided and a reduction in cost because of lowered energy consumption. Psychrophiles are also often exposed to other factors such as high salt concentrations, pressure and oxidative stress. These factors can make proteins isolated from psychrophiles very stable (Gomes *et al*., 2004). A cold adapted esterase (which retains 50% activity at 0ºC) from *Pseudoalteromonas arctica* was cloned, over-expressed and purified (Al Khudary *et al*., 2010). The protein has broad substrate specificity towards short chain fatty acids (between C2-C8 in length) and shows highest activity at 25ºC, pH 7.5.

1.3 Naturally occurring halogenated compounds

The word halogen originates from the Greek words halos (meaning salt) and genes (meaning production). Halogens are very reactive and halogenated metabolites produced by plants are known to be involved in host defence mechanisms (Paul *et al*., 2006).

Although fungi are known to produce halometabolites (for example, *Basidiomycetae* produce chlorinated anisyl and chlorinated hydroquinone metabolites) marine algae are the largest source of organohalogens (Field *et al*., 1995, Valverde *et al.*, 2004). Algae are known to produce a diverse range of biologically active compounds including halogenated terpenes (a variety of organic compounds primarily produced by plants, which are often used as biosynthetic building blocks). These compounds are produced by vanadium bromoperoxidase containing species such as *Corallina, Laurencia* and *Plocamium* sp. (Butler, 2005). Halogenated terpenes are of interest in the pharmaceutical industry, as they often show antimicrobial and antifeedant properties (Kurata *et al*., 1998, Cabrita *et al*., 2010). The red alga *Bonnemaisonia*

hamifera harbours more than 20 organobromine compounds. This includes the compound lachrymatory that gives *B. hamifera* its sweet odour (Gribble, 1999). The edible red seaweed *Asparagopsis taxiformis* found in Hawaii also contains almost 40 organochlorines (Gribble, 1996).

Halogenated compounds synthesised by algae are thought to deter ingestion by herbivores and may inhibit the colonisation and growth of microorganisms on the algal surface (Nightingale *et al.*, 1995). The toxic characteristics of halogenated metabolites are thought to be responsible for driving the evolution of algal defence mechanisms. The abundance of organohalogens in the marine environment makes algae and marine microbes a potential source of novel halogenase and dehalogenase enzymes.

1.3.1 Using halogens to improve drug activity

It was calculated that in 2006, 20% of all pharmaceuticals were halogenated (Yarnell, 2006). The ability of drugs to diffuse across cell membranes depends upon two things: the membrane composition and the properties of the drug. The diffusion of drugs across the cell membrane can be enhanced by replacing a hydrogen atom with a chlorine atom or a trifluoromethyl group (Gerebtzoff *et al*., 2004). Metoclopramide is an example of a drug containing a halogen. This drug is a dopamine receptor antagonist which is used to treat vomiting and nausea. It is prescribed to patients with gastric stasis to aid gastric emptying (Harrington *et al*., 1983).

The halogenation of a compound also has the potential to increase, reduce or modify a compound's biological activity. The halogenation of drugs has been used to increase potency. For example, anesthetic drugs halogenated with a bromide have shown over 3-fold the activity in comparison to the chlorine form of the drug (Targ *et al*., 1989). The naturally chlorinated anti-cancer agent salinosporamide A from *Salinispora tropica* also shows amplified activity is comparison to its non-chlorinated form (Feling *et al*., 2003). This makes halogenases an interesting target as biocatalyst in the pharmaceutical industry as reviewed by Eustáquio *et al*. 2008.

Research into the use of natural halogenated marine products as pharmaceutical agents is still in its infancy.

1.4 Bioremediation

Bioremediation is the process used to describe the metabolism of pollutants by microorganisms. Bioremediation can either take place on site (*in situ*) or the contaminated mass can be moved elsewhere to be treated (*ex situ*). Many agricultural products such as herbicides, fungicides and insecticides are halogenated xenobiotic pollutants (Perelo *et al*., 2010). Examples of halogenated pollutants include the pesticide 2,4-dichlorophenoxyacetic acid and the insecticide 2,4,6-trichlorophenol.

Halogenated xenobiotic pollutants are also formed when free halogens are used in organic synthesis reactions in chemical and pharmaceutical industrial processes. These products are toxic priority pollutants which can persist and bioaccumulate in the environment such as, pentachlorophenol and polychlorine biphenol.

These recalcitrant and carcinogenic compounds have driven scientific research to discover microbial dehalogenases which can degrade these xenobiotic pollutants. Some microorganisms that contain dehalogenases can metabolise halogenated carbon compounds as a sole carbon source. For example, *X. autotrophicus* GJ10 contains a haloalkane dehalogenase that degrades 1,2-dichloroethane, using it as a sole carbon source (Keuning *et al*., 1985). A second example is the microbial degradation of hexachlorocyclohexane, which is a xenobiotic pollutant used as a pesticide and in malaria control programmes (Li *et al*., 1999). *S. paucimobilis* contains two proteins (LinA and LinB) which are part of a pathway that degrades hexachlorocyclohexane (Nagata *et al*., 2005). LinA converts hexachlorocyclohexane to 2,3,5,6-tetrachloro-1,2-cyclohexadiene. LinB then converts this compound to 3,6-dichloro-2,5-dihyroxy-1,2-cyclohexadiene (Scott *et al*., 2008). This makes microbial dehalogenases an important research target in the bioremediation industry.

1.5 The synthesis of α-halogenated compounds

Short chain α-halocarboxylic acids can be used as intermediates in the fine chemical and pharmaceutical industry. A number of chemical synthesis methods for the production of α-halocarboxylic acids have been described. Enzymes could also be used for the synthesis of chiral α-halocarboxylic acids. This includes the potential use of halogenases and dehalogenases that can resolve racemic mixtures leaving one enantiomer of the compound unreacted.

1.5.1 The chemical synthesis of α-halogenated compounds

The Hell-Zelinsky halogenation reaction was the first known method used to synthesise α-halogenated carboxylic acids (Hell, 1881, Volhard, 1888). Figure 1.1 shows the reaction for the regiospecific (non-stereospecific) Hell-Zelinsky halogenation. PBr_3 is used as the catalyst to replace the OH group with a bromide to produce an acyl bromide. The acyl bromide is then tautomerized to an enol which can further react (for the second time) with Br_2 at the α-carbon position. In acidic solutions the hydrolysis of the α-bromoacyl halide occurs and α-bromocarboxylic acid is produced.

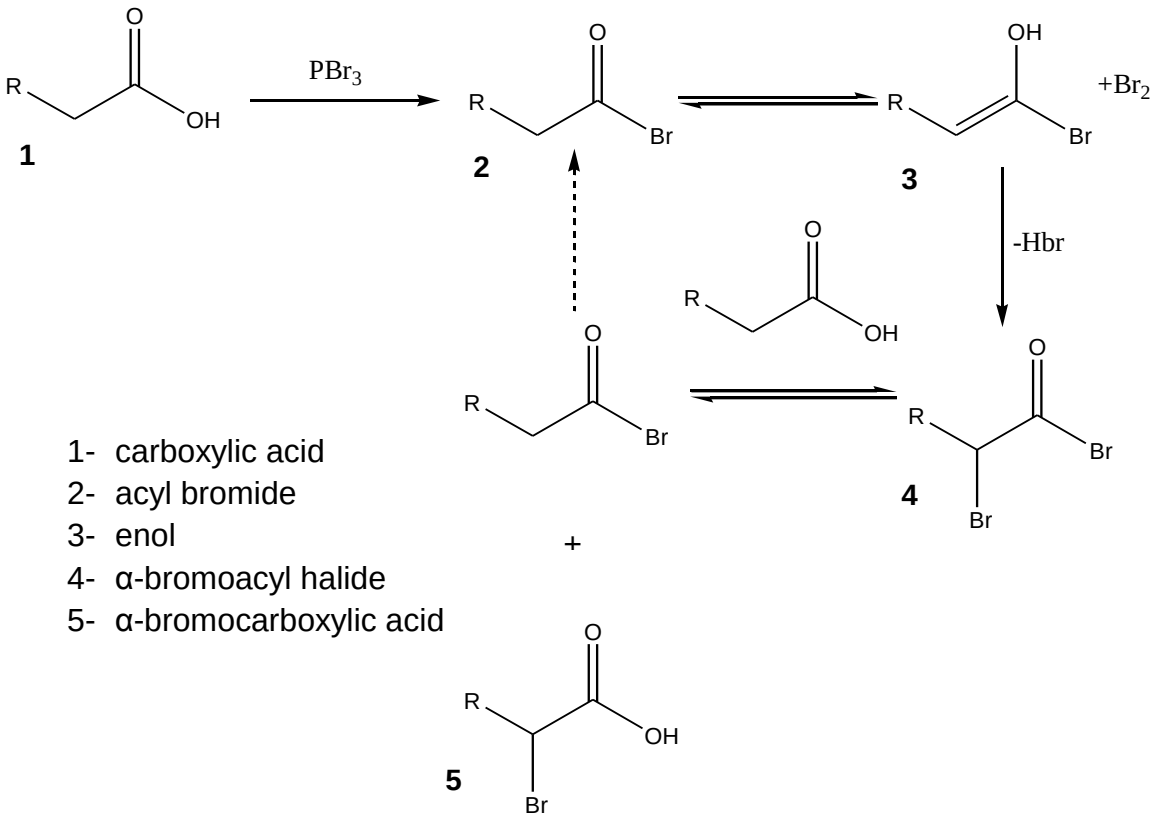


Figure 1.1- The reaction catalysed by the Hell-Zelinsky halogenation.

More recently, a more efficient and faster method of producing highly pure α-halocarboxlyic acid has been reported (Zhang *et al*., 1998). This method involves heating the carboxylic acid to 85ºC with N-bromosuccinimide in trifluoroacetic acid and sulphuric acid. Although this method is more efficient then the Hell-Zelinsky halogenation reaction, it also proceeds in a non-stereospecific manner.

1.6 Enzymes investigated in this project

1.6.1 Haloperoxidases (EC 1.11.1.18)

The enzymatic halogenation of a compound is split into two classes: haloperoxidases and halogenases. Haloperoxidases are further split into two groups: haem-containing haloperoxidases and vanadium containing haloperoxidases. These enzymes are able to catalyse the halogenation of organic molecules in the presence of hydrogen peroxide (H_2O_2) and halide ions (F , Cl , Br). Both types of haloperoxidase catalyse other reactions such as sulfoxidation and aromatic hydroxylation (Burton, 2003). A non-cofactor

dependent esterase that can carry out bromination in low pH environments has also been identified (Pelletier and Altenbucher, 1995). Haloperoxidases (which were first isolated in marine organisms) belong to the oxidoreductase superfamily (Colonna *et al*., 1999). Halogenated secondary metabolites such as indole derivatives and acetogenins are produced by algae which contain haloperoxidase (Butler and Franklin, 2003). In this project, microbial libraries were assayed for bromoperoxidase activity.

1.6.1.1 Haem-containing haloperoxidase

The first haem-containing chloroperoxidase was identified in the marine fungus *Caldariomyces fumago*. The enzyme was biochemically characterised and the crystal structure elucidated at a resolution of 1.9 Å (Sundaramoorthy *et al.*, 1998). The structure showed a novel fold, with eight helical segments. The mechanism for haem-containing haloperoxidases has been established and the residue Glu 183 identified as the acid-base catalyst which cleaves the O-O peroxide bond. This is unusual as other peroxidases use a histidine as the acid-base catalyst. Both hydrogen peroxide and halide ions are needed for the chloroperoxidase reaction to occur (Wagenknect and Woggon, 1997).

Some species of polychaete (tube worms) such as *Capietellidae, Spionidae* and *Cirratulidae* are known to produce a range of structurally diverse halogenated compounds (Fielman *et al*., 2001). Many of these compounds are complex molecules such as 2,4,6-tribromophenol and bromooctylpyrrole. The capitellid polychaete (tube worm) *Notomastus lobatus* contains a haem-containing chloroperoxidase. This chloroperoxidase has two dissociable moieties, a flavoprotein which contains 4 identical subunits and a haem-protein, which contains two copies of each subunit (Roach *et al.*, 1997). This chloroperoxidase is unusual in two senses. It has a smaller subunit size in comparison to other haem-containing haloperoxidases and a flavoprotein which is essential for the peroxidase activity.

1.6.1.2 Vanadium containing haloperoxidases

Vanadium containing haloperoxidases have been isolated from a variety of organisms including algae (red, brown and green), fungi and lichens (Simons *et al*., 1995). The crystal structure of a vanadium chloroperoxidase isolated from the pathogenic fungus *Curvularia inaequalis* was solved at a resolution 2.1 Å (Messerschmidt and Wever, 1996). The main structural motif consists of two four-helix bundles. Vanadium is bound as hydrogen vanadate to three oxygen atoms and a histidine residue in a trigonal bipyrimidal coordination. Halogenated organic compounds produced by the *C. inaequalis* haloperoxidases are strong bactericidal oxidising agents. These organohalogens are proposed to aid plant penetration by oxidising cell wall material (Van Pée, 2001).

The crystal structure of a multi subunit bromoperoxidase from *Corallina officinalis* has also been solved (Brindley *et al.*, 1998). Structural studies with crystals grown in the presence of ammonium phosphate showed that phosphate competes with vanadate in the active site. The crystal structure of the dodecameric form of the protein was determined at a resolution of 2.3 Å (Isupov *et al*., 2000). The haloperoxidases from *C. inaequalis* and *C. officinalis* have a conserved protein scaffold for vanadium, which provides a rigid oxyanion binding site (Butler and Franklin, 2004). The conserved motif consists of P(S/A)YPSGHAT. Figure 1.2 shows the crystal structure of *C. officinalis* bromoperoxidase.

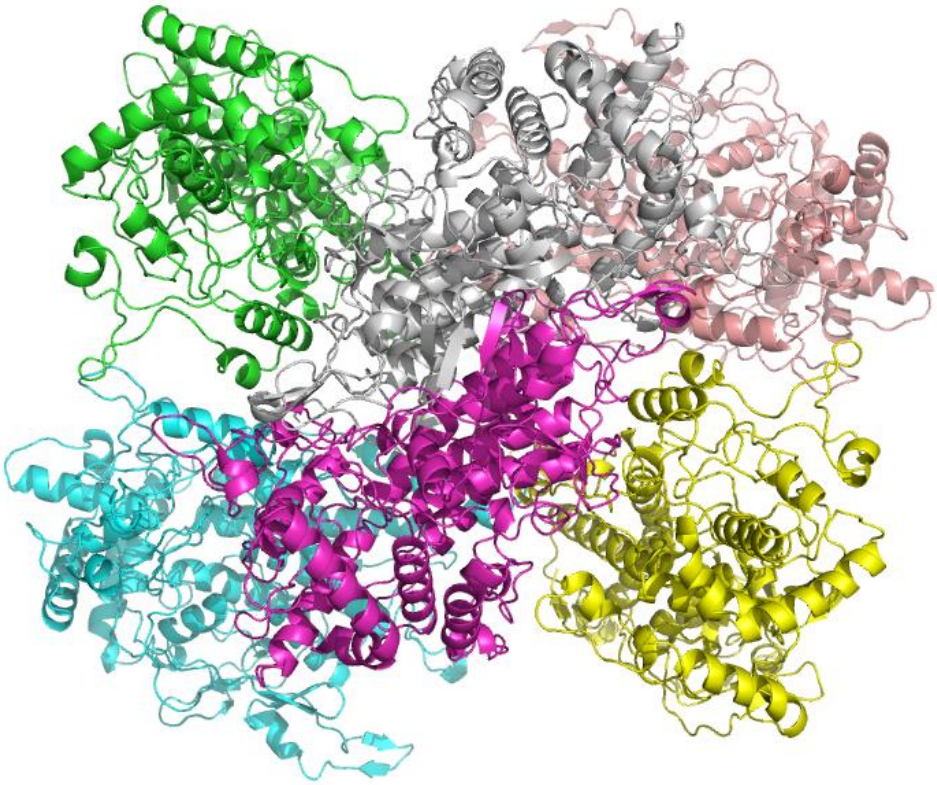


Figure 1.2- The crystal structure of the dodecameric form of the bromoperoxidase from *C. officinalis*, with each subunit coloured separately (PDB: 1QHB). The Figure was produced using the program PyMOL (DeLano Scientific).

1.6.2 Dehalogenases

Dehalogenation is the process where one or more halogen atoms are removed from a halogenated organic compound. There are many naturally occurring organohalogens produced by a variety of organisms (Gribble *et al*., 2003). During the dehalogenation mechanism the most important step is the physical removal of the halide ion.

There are a number of factors which will affect the susceptibility of a halogenated organic compound to undergo dehalogenation. These factors include the structural complexity of the organohalogen, the location, the number and the type (iodine, bromide, chlorine, fluorine) of halogen attached (Allpress *et al.*, 1998). Dehalogenases are present in both anaerobic and aerobic microorganisms (Wohlfarth *et al*., 1997, Nagata *et al*., 2007).

In this project, microbial libraries were assayed for haloalcohol dehalogenase, haloacid dehalogenase and haloalkane dehalogenase activity.

1.6.2.1 Haloalcohol dehalogenases (EC 3.8.1)

Numerous microorganisms have evolved to degrade halogenated aliphatics as a sole carbon source. Haloalcohol dehalogenase (also known as halohydrin dehalogenase) have received considerable attention because of their broad substrate specificity and enantioselectivity towards aromatic haloalcohol aliphatics. These enzymes convert haloalcohols to their corresponding epoxide, halide ion and proton (in a co-factor independent manner). The reaction proceeds via an S_N2 type nucleophilic substitution (De Jong *et al.,* 2003). Haloalcohol dehalogenases are also able to catalyse the reverse reaction (Assis *et al*., 1998).

Based on amino acid sequence identity, there are three subclasses of haloalcohol dehalogenases: A, B and C. Groups A and B show some enantioselectivity whereas group C are strictly enantioselective. Haloalcohol dehalogenases in all three subclasses are able to degrade 1,3-dichloro-2-propanol. This compound is a xenobiotic pollutant used in industrial processes. The haloalcohol dehalogenase (HheC) and epoxide hydrolase (EH) from *Agrobacterium radiobacter* AD1 are able to degrade 1,3-dichloro-2-propanol to non-toxic products which can be used as a sole carbon source (Lutje Spelberg *et al*., 1999). Figure 1.3 shows the mechanism for degradation of 1,3-dichloro-2-propanol in *A. radiobacter* AD1 (De Jong *et al*., 2006).

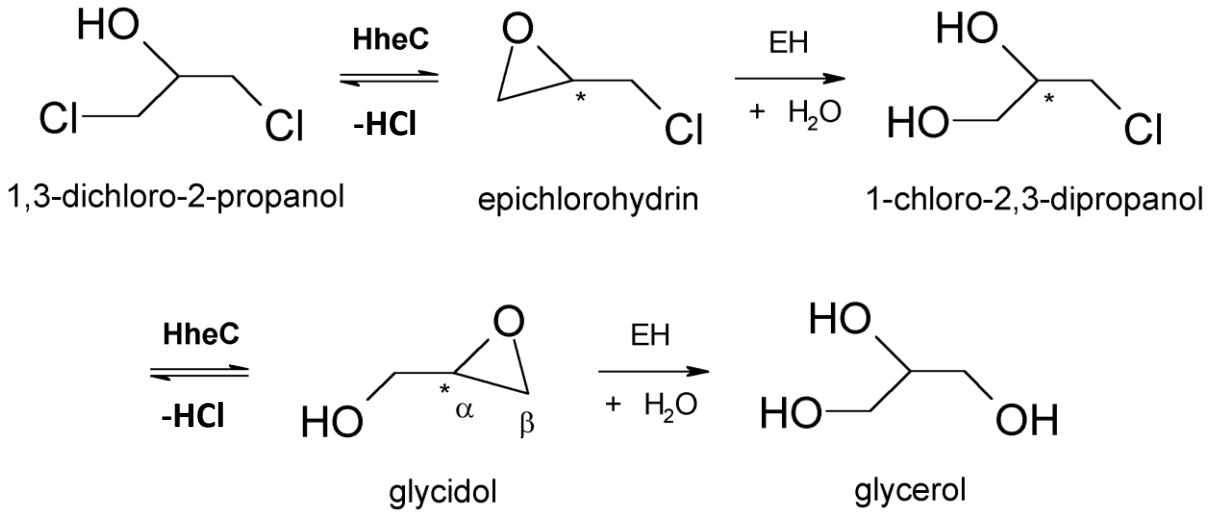


Figure 1.3- The degradation pathway of 1,3-dichloro-2-propanol by the haloalcohol dehalogenase and epoxide hydrolase from *A. radiobacter* AD1. The haloalcohol dehalogenase removes one halogen at a time from the substrate via an S_N2 type nucleophilic substitution. This produces the epoxide product which is subsequently hydrolysed by the expoxide hydrolase. This Figure is modified from De Jong *et al.* (2003).

Haloalcohol dehalogenases share sequence similarity to short-chain reductases/dehydrogenase proteins (SDR) (≤25% amino acid sequence identity). Cofactor dependent (NAD(H)/NADP(H)) SDR family proteins catalyse the oxidation and reduction of alcohols and ketones. SDRs have a highly conserved catalytic triad which consists of Ser, Tyr and Lys residues. Haloalcohol dehalogenases have the highly conserved Ser and Tyr but the Lys is replaced by Arg (De Jong *et al.*, 2003).

To date, two haloalcohol dehalogenase crystal structures have been solved. The crystal structure for the first (group C) haloalcohol dehalogenase (HheC) from *Agrobacterium radiobacter* was solved at a resolution 2.7 Å (De Jong *et al*., 2003). The second (group A) crystal structure of a haloalcohol dehalogenase ($HheA_{AD2}$) from *Arthrobacter* sp. was solved at a resolution 2.0 Å (De Jong *et al*., 2006). $HheA_{AD2}$ has a larger active site cavity in comparison to HheC. Both of

these haloalcohol dehalogenases are tetramers with a subunit molecular weight between 26 and 27 kDa.

1.6.2.2 Haloalkane dehalogenases (EC 3.8.1.5)

In this project, both the native and complex crystal structures of a haloalkane dehalogenase from a marine bacterium (in the Aquapharm Biodiscovery Ltd Microbial library) were determined.

Haloalkane dehalogenases catalyse the conversion of halogenated alkanes to their hydrogen halide and alcohol compounds. Haloalkane dehalogenases act on both chlorinated and brominated substrates and have the potential to be used in the bioremediation of waste streams (Janssen *et al.*, 2004). Several haloalkane dehalogenases have been characterised and together they show activity towards a broad range of substrates including long and short chain molecules (plus halogenated amines, ethers and cyclic structures). Some substrates are halogenated in multiple positions (Koudelakova *et al*., 2011).

Haloalkane dehalogenases do not show activity towards fluorinated substrates but do share structural and mechanistic characteristics with the biochemically characterised fluoroacetate dehalogenase from *Moraxella* species (Liu *et al*., 1998). A haloalkane dehalogenase gene has been found on a catabolic plasmid; this suggests that the gene could be distributed among microbes by horizontal gene transfer (Janssen *et al.*, 1994).

The haloalkane dehalogenase from *X. autotrophicus* was the first to be biochemically and structurally characterised (Van der Ploeg *et al*., 1991, Verschueren *et al*., 1993). These enzymes belong to the α/β hydrolase fold family and have two domains: a core domain and a cap domain. The active site cavity is located between the two domains. Since then, numerous haloalkane dehalogenases from a variety of microorganisms have been identified and structurally characterised. These include the haloalkane dehalogenases from *Rhodococcus rhodochrous, Mycobacterium tuberculosis and S. paucimobilis* (Newman *et al.,* 1999, Marek *et al*., 2000, Mazumdar *et al*., 2008). Figure 1.4

shows a diagram outlining the general topology of haloalkane dehalogenases. Site-directed mutagenesis and crystal soaking experiments (to obtain structural complexes) allowed the confirmation of a catalytic pentad. Three of the residues in the catalytic pentad are directly involved in catalysis (Asp, His, Asp/Glu). The two other residues are involved in stabilisation of the halide leaving group (Trp, Trp/Asn) (Janssen *et al.*, 2004).

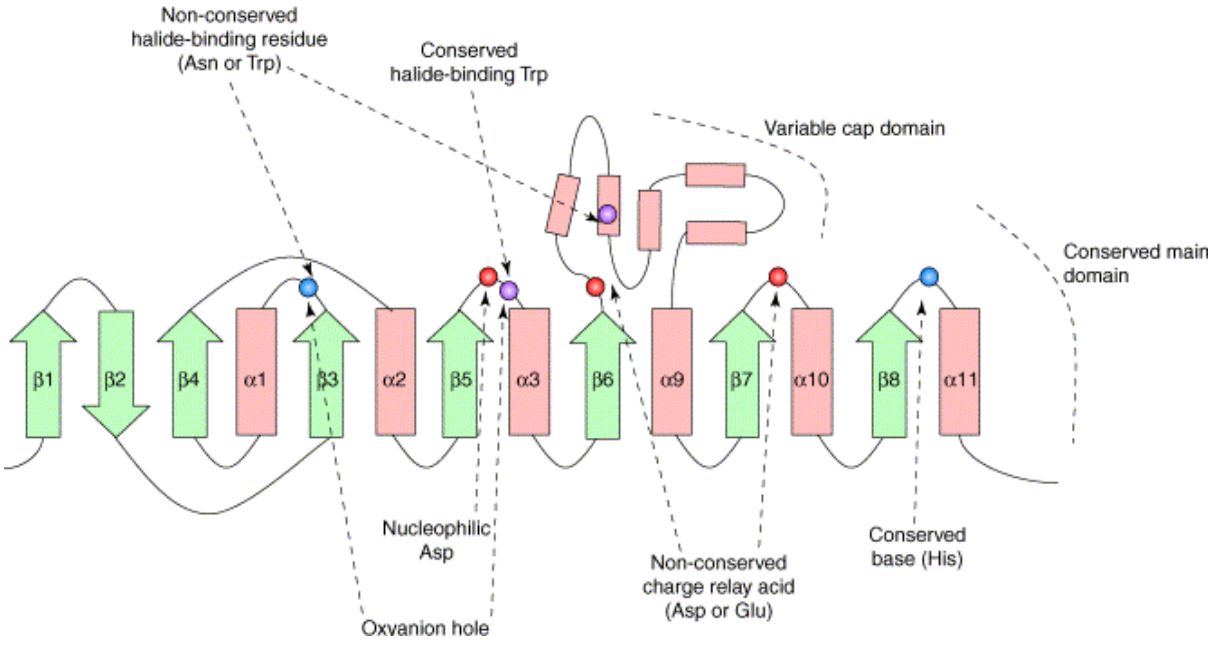


Figure 1.4- Diagram outlining the general topology of haloalkane dehalogenases. This Figure was taken from Janssen *et al.* (2004).

Figure 1.5 shows the haloalkane dehalogenase S_N2 type nucleophilic substitution reaction mechanism. The first step in the catalytic mechanism is the nucleophilic attack on the carbon attached to the halogen by the carboxyl group of the conserved Asp residue. This produces an ester intermediate complex, where the oxygen in the ester bond is incorporated from the enzyme. The second step is the nucleophilic attack of the carboxyl group of the ester intermediate by a water molecule (which is activated by a histidine). The last step in the mechanism is the release of the halide and product. Stop flow fluorescent spectroscopy experiments with the haloalkane dehalogenase from *X. autothophicus* confirmed that the last step in the reaction is the slowest. This is most likely due to slow

conformational changes in the cap domain needed to allow the halide molecule to exit the active site and a water molecule to enter (Jenssen *et al*., 1996, Schanstra *et al*., 1996,).

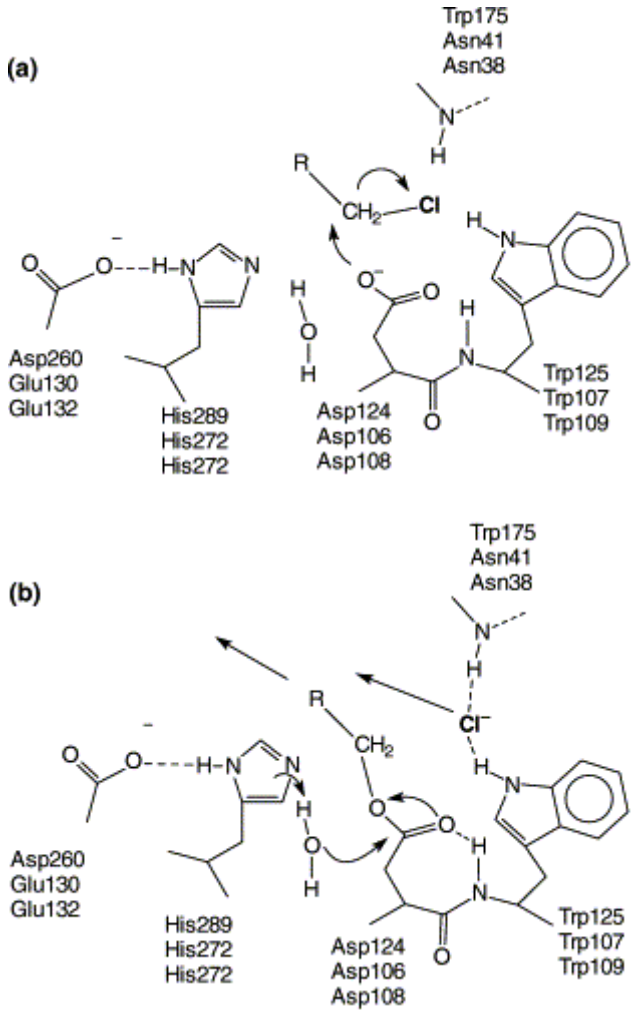


Figure 1.5- Haloalkane dehalogenase catalytic mechanism, (a) the formation of the ester intermediate and (b) the hydrolysis of the ester intermediate. Residue numbers refer to the dehalogenases from *X. autothophicus, Rhodococcus erythropolis* and *S. paucimobilis*. This Figure was taken from Janssen *et al.* (2004).

Although haloalkane dehalogenases have the potential to be used in bioremediation processes, they do not show high activity towards pollutants such as 1,2-dichloropropane and trichloroethane. Two processes were employed to try to increase the activity of haloalkane dehalogenases towards recalcitrant

compounds. The first method was the cultivation of wild type microbes in the presence of pollutants, with the aim that the haloalkane dehalogenase would degrade the compound as a sole carbon source. The second method was the site-directed mutagenesis of amino acids to improve substrate specificity and activity. Mutagenesis experiments of Leu 177 (which partially blocks the active site tunnel) in the *S. paucimobilis* haloalkane dehalogenase were carried out (Chaloupkova *et al*., 2003). Mutations to smaller, non-polar amino acids (Ala and Gly) generally increased the activity of the enzyme, while mutagenesis to charged or polar residues (Asp, Arg, Ser and Thr) reduced the activity.

The size and shape of the active site cavity in haloalkane dehalogenases can significantly affect the substrate specificity of the enzyme. The haloalkane dehalogenase from *X. autotrophicus* has a smaller active site pocket in comparison to the *S. paucimobilis* haloalkane dehalogenase. The larger activite site pocket in the *S. paucimobilis* haloalkane dehalogenase lowers the enzymes activity towards smaller substrates such as 1,2-dichloroethane (Marek *et al*., 2000). Figure 1.6 shows the crystal structure of the haloalkane dehalogenase from *S. paucimobilis*.

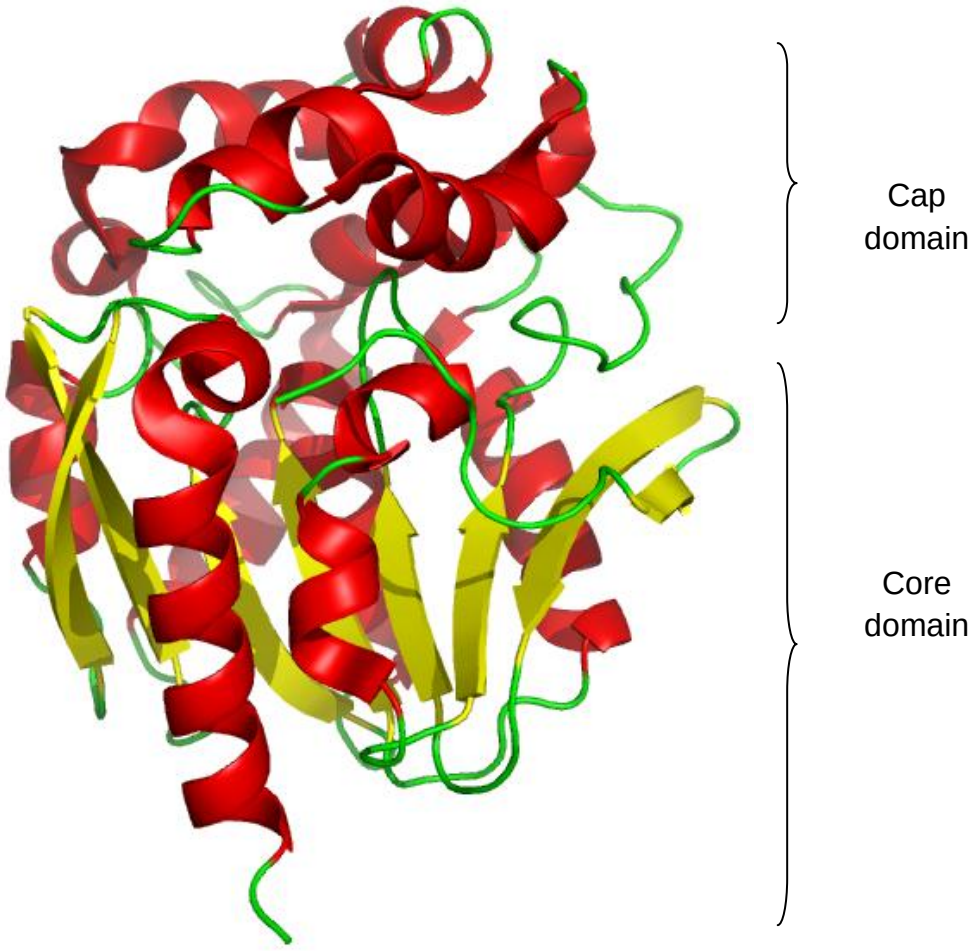


Figure 1.6- The crystal structure of the haloalkane dehalogenase from *S. paucimobilis* (PDB: 2bfn). The Figure was produced using the program PyMOL (DeLano Scientific).

1.6.2.3 Fluoroacetate dehalogenases (EC 3.8.1.3)

Fluoroacetate is a highly toxic compound (Gribble, 1973). The synthesis of flouroacetate by indigenous plants in Australia is thought to deter ingestion by herbivores (Twigg and King, 1991). Fluoroacetate dehalogenases have been used in the recovery of animals which have ingested these poisonous plants. For example, the fluoroacetate dehalogenase from *Moraxella* species was cloned and transformed into *Butyrivibrio fibrisolvens* (Gregg *et al*., 1998). Sheep inoculated with the recombinant *B. fibrisolvens* showed reduced toxicological symptoms after ingestion of the plants which synthesise fluoroactete.

Many fluorinated compounds used in industrial, pharmaceutical and agricultural applications are xenobiotic pollutants (Parsons *et al*., 2009). This is because fluorinated compounds are highly stable and are resistant to metabolic breakdown. The carbon-fluorine bond has one of the highest dissociation bond energies in all naturally occurring compounds. Fluoroacetate dehalogenases have been identified in numerous bacterial strains and are one of the few enzymes which are able to cleave the carbon fluorine bond (Natarajan *et al*., 2005).

The first fluoroacetate dehalogenase was identified in a *Moraxella* species. This enzyme shows highest activity towards fluoroacetate followed by chloroacetate, bromoacetate and iodoacetate (Liu *et al*., 1998). Site-directed mutagenesis and O^{18} incorporation experiments revealed Asp 105 as the main catalytic residue (Ichiyama *et al*., 2004). The first step in the mechanism is the nucleophilic attack of the α-carbon and the formation of an ester intermediate. The second step is the hydrolysis of the ester intermediate by a water molecule which is activated by His 272. The fluoroacetate dehalogenase mechanism is shown in Figure 1.7. Fluoroacetate dehalogenases and haloalkane dehalogenases have similar catalytic amino acids. The catalytic Asp and His residues in both fluoroacetate dehalogenases and haloalkane dehalogenases perform the same role.

More recently, the fluoroacetate dehalogenase from *Burkholderia* sp. strain FA1 has been identified and the crystal structure solved to 1.5 Å (Jitsumori *et al.,* 2009). This enzyme displayed 100% activity towards fluoroacetate in comparison to 2.6% activity towards chloroacetate. This enzyme has an α/β hydrolase fold with two domains: the core and cap domain. The active site is located between the two domains. Site-directed mutagenesis confirmed that the completely conserved residues Asp 104, His 271 and Asp 128 form the catalytic triad and Trp 150 was proposed to be the halide ion acceptor.

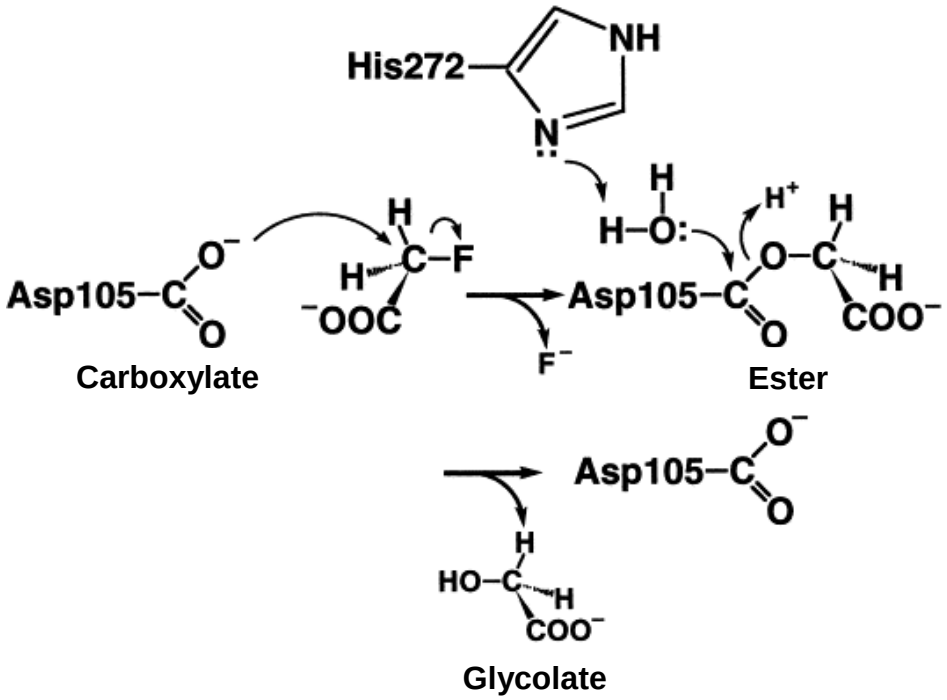


Figure 1.7- The mechanism for fluoroactete dehalogenases from the *Moraxella* species. This Figure has been adapted from Ichiyama *et al.* (2004).

1.6.2.4 2-haloacid dehalogenases

2-haloacid dehalogenases are hydrolytic enzymes that are split into three sub classes: D-haloacid dehalogenases, L-haloacid dehalohenases and D/L-haloacid dehalohenases. D-haloacid dehalogenases and L-haloacid dehalohenases are specific to one enantiomer and always yield products that have inverted chirality at the α-carbon position. D/L-haloacid dehalohenases use both enantiomers of the substrate and proceed with inversion of configuration at the α-carbon position. In some organisms more than one haloacid dehalogenase has been detected. Haloacid dehalogenases shows activity towards short chain substrates with the halogen atom attached at the α-carbon.

2-haloacid dehalogenases are sub divided into two groups: I and II. Group I contains D/L-haloacid dehalohenases and D-haloacid dehalogenases. Group II contain L-haloacid dehalohenases. The two groups do not share any amino acid sequence similarity, suggesting they are evolutionarily unrelated (Nardi-Dei *et al*.,

1997). The crystal structures of numerous group I haloacid dehalogenases have been characterised.

In this project, two L-haloacid dehalohenases from marine bacteria were characterised.

1.6.2.4.1 D-haloacid dehalogenases (EC 3.8.1.9)

D-haloacid dehalogenases in *Rhizobium* sp. and *Pseudomonas putida* AJ1/23 have been identified (Leigh *et al*., 1988, Smith *et al*., 1990). The 31.8 kDa D-haloacid dehalogenase from *P. putida* strain AJ1/23 is a tetramer and shows optimal enzyme activity at 30ºC. This enzyme shows highest activity towards bromoacetate followed by chloroacetate, D-bromopropionic acid and D-chloropropionic acid. The enzyme shows low activity (≤13%) towards multiple halogenated and longer chain substrates. To date, no crystal structures of D-haloacid dehalogenases have been elucidated.

1.6.2.4.2 D/L-haloacid dehalohenases (EC 3.8.1.11)

The D/L-haloacid dehalohenase from *Pseudomonas* sp. strain 113 has been biochemically and structurally characterised. D/L-haloacid dehalohenases and D-haloacid dehalogenases (group I) have relatively high amino acid sequence identities (≥23%). O^{18} incorporation experiments and the crystal structure allowed a potential mechanism to be proposed (Nardi *et al*., 1997, Schmidberger *et al*., 2008). Unlike D-haloacid dehalogenases, L-haloacid dehalohenases and haloalkane dehalogenases, the mechanism proceeds without the formation of an ester intermediate. Residues Asp 189 and Asn 114 activate a water molecule which attacks the chiral carbon of the substrate and an S_N2 transition-state intermediate is formed. The halide leaving group is cleaved, followed by the formation of the hydroxylated product (Schmidberger *et al*., 2008). The reaction proceeds with inversion of configuration at the α-carbon. The D/L-haloacid dehalohenase reaction mechanism is shown in Figure 1.8.

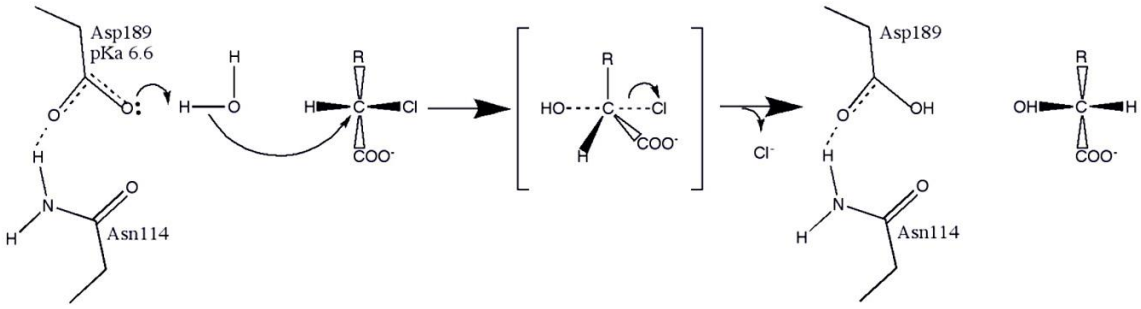


Figure 1.8- The reaction mechanism for the *Pseudomonas* sp. strain 113 D/L-haloacid dehalohenases. This Figure was taken from Schmidberger *et al.* (2008).

1.6.2.4.3 L-haloacid dehalohenases (EC 3.8.1.2)

L-haloacid dehalohenases catalyse the conversion of short chain l-2-haloalkanoic acids to D-2-hydroxyalkanoic acids. L-haloacid dehalohenases from numerous microorganisms have been biochemically and structurally characterised. These include the L-haloacid dehalohenases from *Pseudomonas* sp. strain YL, *X. autotrophicus, Burkholderia cepacia* and *Sulfolobus tokodaii* (Hisano *et al*., 1996, Ridder *et al*., 1999, Schmidberger *et al*., 2007, Rye *et al*., 2007). These L-haloacid dehalohenases have ≥20% amino acid sequence identity. The reaction mechanism was determined by obtaining native and crystal complex structures and by performing site-directed mutagenesis and O^{18} incorporation experiments.

The first characterised L-haloacid dehalohenase was from *Pseudomonas* sp. YL (Hisano *et al*., 1996). This enzyme shows activity towards chloroacetic acid and iodoacetic acid in aqueous solution and longer chain substrates such as 2-bromohexadecane in n-heptane. The *Pseudomonas* sp. YL L-haloacid dehalohenase is solvent stable (retaining 100% of activity in 30% DMSO) and shows optimal activity at 65ºC, pH 9.5 (Hasan *et al*., 1991). O^{18} incorporation experiments identified Asp 10 as the main catalytic residue (Liu *et al*., 1994). The first step in the reaction is the nucleophilic attack on the α-C by the carboxyl group of Asp 10 and the formation of ester intermediate. The second step is the nucleophilic attack of the carboxyl group of the ester intermediate by an activated water molecule. The crystal complex structure with chloroacetate bound to the

catalytic Asp 10 was elucidated at a resolution of 1.8 Å (Li *et al*., 1998). The L-haloacid dehalohenase reaction mechanism is shown in Figure 1.9.

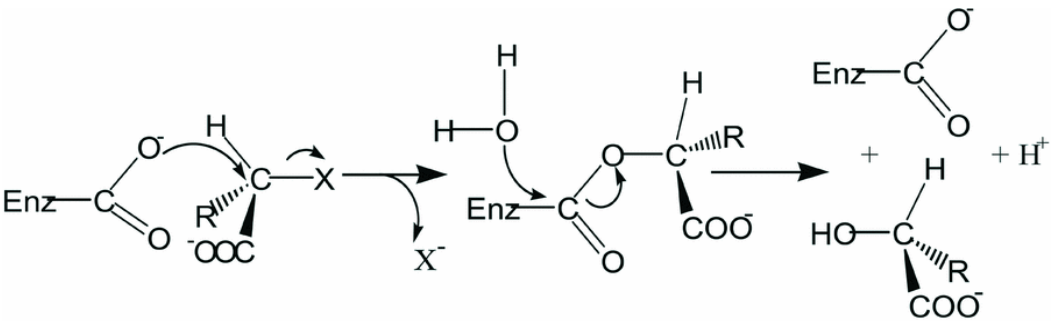


Figure 1.9- The L-haloacid dehalohenases reaction mechanism. This Figure was taken from Liu *et al.* (1994).

Haloacid dehalogenases belong to the haloacid dehalogenase (HAD) superfamily. HAD superfamily enzymes are dimers with two domains in each subunit. The core domain is a Rossmann-fold-like six stranded parallel β-sheet, flanked by five α-helices. The cap domain consists of a four helix bundle. The active site is located between the two domains. Proteins in the HAD superfamily share the Rossmann fold core domain but variations in the cap domain give rise to the enzymatic diversity within the superfamily (Burroughs *et al.,* 2006). Figure 1.10 shows the dimeric *B. cepacia* L-haloacid dehalohenase. HAD superfamily proteins of known function fall into one of five subfamilies: L-haloacid dehalohenases (hydrolysis of the C-halogen bond), phosphate monoesterases (hydrolysis of the P-OC bond), ATPases (hydrolysis of the P-OP bond), phosphonoacetaldehyde hydrolases (hydrolysis of the P-C bond) and phosphomutases (hydrolysis of the P-OC bond via phosphoryl group transfer) (Lahiri *et al*., 2004). Although L-haloacid dehalohenase, haloalkane dehalogenase and fluoroacetate dehalogenase enzymes proceed via a nucleophilic substitution reaction (via an ester intermediate), they show no amino acid sequence identity.

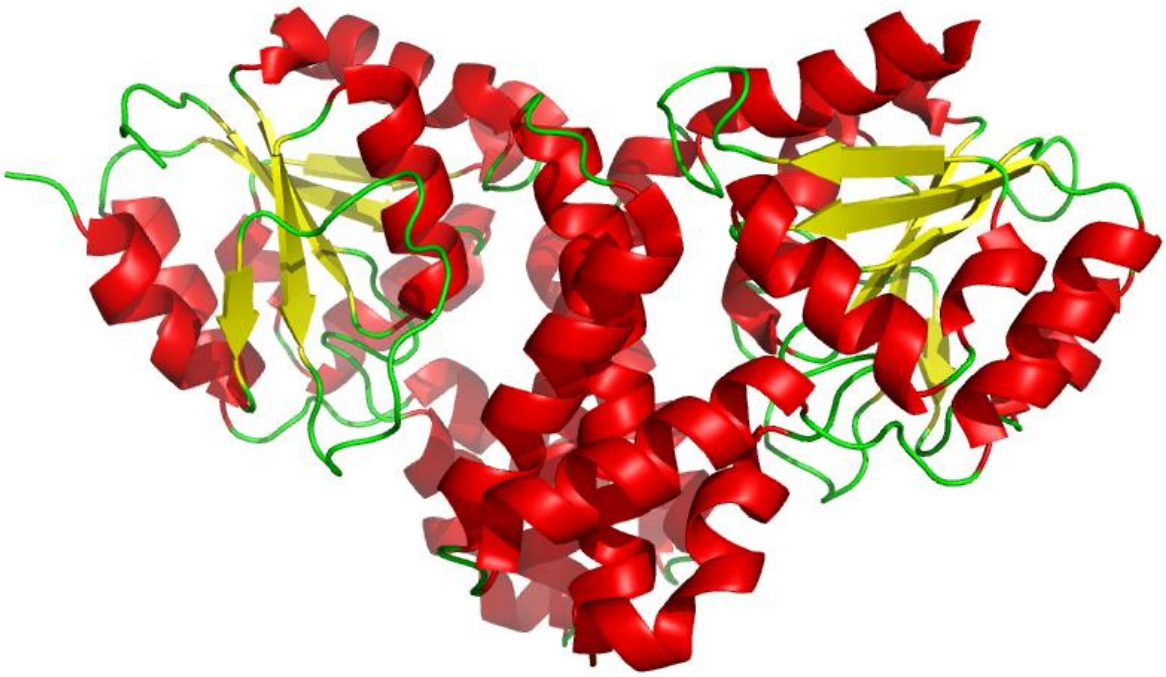


Figure 1.10- L-haloacid dehalohenase from *B. cepacia* shown as a dimer (PDB: 1NO5). The Figure was produced using the program PyMOL (DeLano Scientific).

Figure 1.11 shows the amino acid sequence alignments of the L-haloacid dehalohenases from *B. cepacia* and *Pseudomonas* sp. strain YL using ClustalW, (European Bioinformatics Institute- http://www.ebi.ac.uk/Tools/msa/clustalw2). The two sequences are 38% identical and the completely conserved amino acid residues that are essential for catalysis are highlighted. Point mutation experiments of these residues in the L-haloacid dehalohenase from *Pseudomonas* sp. YL resulted in a loss of activity between 70-100% (Kurihara *et al*., 1995).

```
sequence1       MVDSLRACVFDAYGTLLDVHSAVMRNADEVGASAEALSMLWRQRQLEYSWTRTLMHQYAD 60
sequence2       -MDYIKGIAFDLYGTLFDVHSVVGRCDEAFPGRGREISALWRQKQLEYTWLRSLMNRYVN 59
                 :* ::. .** ****:****.* *  : . . .. :* ****:****:* *:**::*.:

sequence1       FWQLTDEALTFALRTYHLEDRKGLKDRLMSAYKELSAYPDAAETLEKLKSAGYIVAILSN 120
sequence2       FQQATEDALRFTCRHLGLDLDARTRSTLCDAYLRLAPFSEVPDSLRELKRRGLKLAILSN 119
                * * *::** *: *   *:     :. * .** .*:.:.:..::*.:**  *  :*****

sequence1       GNDEMLQAALKASKLDRVLDSCLSADDLKIYKPDPRIYQFACDRLGVNPNEVCFVSSNAW 180
sequence2       GSPQSIDAVVSHAGLRDGFDHLLSVDPVQVYKPDNRVYELAEQALGLDRSAILFVSSNAW 179
                *. : ::*.:. : *   :*  **.* :::**** *:*::* : **:: . : *******

sequence1       DLGGAGKFGFNTVRINRQGNPPEYEFAPLKHQVNSLSELWPLLAKNVTKAA-- 231
sequence2       DATGARYFGFPTCWINRTGNVFEEMGQTPDWEVTSLRAVVELFETAAGKAEKG 232
                *  **  *** *  *** **  *    . . :*.**  :  *: . . **
```

Figure 1.11- ClustalW amino acid sequence alignment of the L-haloacid dehalohenase from *B. cepacia* (sequence 1) and *Pseudomonas* sp. strain YL (sequence 2). The residues essential for catalysis are highlighted in yellow.

Comparisons of native and complex L-haloacid dehalohenase crystal structures have given an insight into the amino acid residues which are essential for catalysis. The highly conserved residues Arg 42 (Arg 41 and Arg 39 in *Pseudomonas* sp. YL and *X. autotrophicus*) and Asn 178 (Asn 177 and Asn 173) in the *B. cepacia* L-haloacid dehalohenase are proposed to be involved in substrate binding (Schmidberger *et al*., 2007). This was concluded when Arg 42 was observed adopting an open position when the active site is empty in comparison to a more closed position upon binding of a substrate. In the closed position Arg 42 is hydrogen bonded to Asn 178 and to the main chain carboxyl oxygen of Tyr 13 (Tyr 12 and Tyr 10). This “lock down” mechanism would be opened after the ester intermediate was hydrolysed causing the bond between Asn 178 and Arg 42 to be weakened. The completely conserved Asp 180 in the L-haloacid dehalohenase from *Pseudomonas* sp. strain YL is thought to stabilise the rotation of the catalytic Asp 10 residue and to act as the base that activates the catalytic water (Nakamura *et al*., 2009).

The L-haloacid dehalohenase from thermophilic archaeon *S. tokodaii* has been structurally and biochemically characterised (Rye *et al*., 2007, Rye *et al*., 2009). The enzyme displays high thermal and solvent stability with maximum activity at 60ºC.

2-Haloacid dehalogenases can produce optically pure 2-hydroxyalkonoic acids, which can be used in industrial applications. Thermostable and solvent stable enzymes are attractive for use in biotransformation reactions because of their robust nature. Chiral hydroalkanoic acids can be used as substrates for the production of pharmaceuticals. This makes l-2-haloacid dehalogenases an attractive class of biocatalysts (Schoemaker *et al*., 2003, Haki and Rakshit, 2004).

1.7 Marine bacteria containing dehalogenases

Dehalogenase enzymes from the following marine bacteria were intestigated in this project.

1.7.1 AQP4626

AQP4626 from the Aquapharm Biodiscovery Ltd library was isolated from a sponge of unknown species collected from Linnhe Marine, Argyll, UK. The microorganism was identified (using 16S rRNA PCR) as having 99.7% sequence identity to *Psychromonas arctica*.

P. arctica was isolated using enrichment cultures containing starch from the sea-ice water interface in Spitzberg, Arctic (Groudieve *et al*., 2003). The strain grows between 0º and 25ºC in aerobic conditions and has a pH optimum of 8.7. Figure 1.12 shows the electron micrograph of negatively stained *P. arctica* cells. In a recent study, a cold shock protein (CspA) from *P. arctica* has been over-expressed in *Eschericha coli (*Jung *et al.,* 2010*).* Upon induction of the cold shock protein, cell growth was stunted and morphological elongation was observed. This increased the bacterial cold resistance by more than 10 fold. The introduction of CspA into microorganisms used in industrial processes has the potential to increase their capabilities and aid their survival for a longer period of time at lower temperatures.

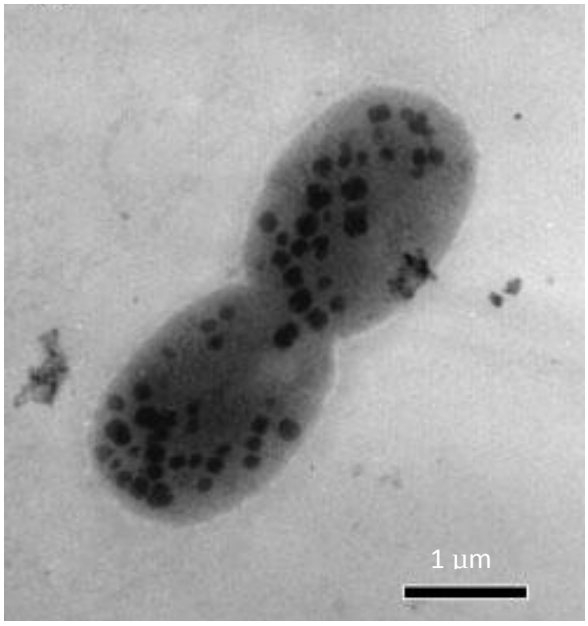


Figure 1.12- The electron micrograph of negatively stained *P. arctica* cells. This Figure was taken from Groudieva *et al*., 2003.

1.7.2 AQP5750

AQP5750 from the Aquapharm Biodiscovery Ltd library was isolated from a Polychaeta (tube worm) collected from Tralee beach, Argyll, UK. The microorganism was identified (using 16S rRNA PCR) as having 99.0% sequence identity to *Rhodobacteraceae* family bacteria. *Rhodobacteraceae* are classified in the alpha subclass of Gram-negative proteobacteria and are predominantly present in marine environments. Figure 1.13 shows a photograph of Tralee beach, Argyll, UK.

Figure 1.13- A photograph of Tralee beach, Argyll, UK. This image was taken from Google maps.

1.7.3 *Psychromonas ingrahamii*

P. ingrahamii was isolated in 1991 from Elson Lagoon (Figure 1.14), Point Barrow Alaska (Gosink *et al*., 1993). Original samples were collected from the sea ice interface, where temperatures can reach -10ºC. *P. ingrahamii* strain 37 has the lowest recorded growth temperature of an organism at -12ºC with a generation time of 240 h (Breezee *et al*., 2004). Figure 1.15 shows a phase contrast micrograph of *P. ingrahamii* strain 37 cells. *P. ingrahamii* has an optimal growth temperature of 5ºC, with a generation time of 12 h. The non-motile, large, rod shaped bacterium can utilise glycerol as a sole carbon source, which was added to liquid growth media to prevent freezing at temperatures below 0ºC (Breezee *et al*., 2004). The genomic analysis of *P. ingrahamii* determined that there are numerous genes encoding enzymes which synthesise polysaccharides and betaine choline. These compounds are thought to aid survival at low temperatures (Riley *et al*., 2008). The production of polysaccharides may help lower the freezing point of liquid surrounding the cell and could aid water

retention. At temperatures lower then 0ºC betaine choline is thought to act as an osmolyte to help balance the osmotic pressure in the cells.

Figure 1.14- A photograph of Elson Lagoon in Point Barrow Alaska, this image was taken from Google maps.

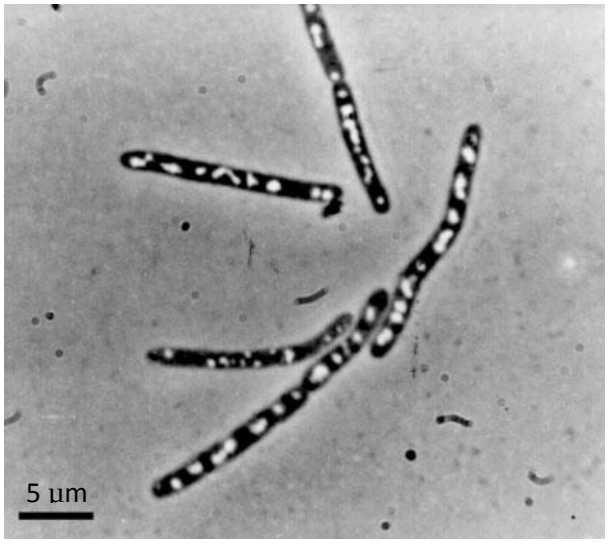


Figure 1.15- Phase contract micrograph of *P. ingrahamii* strain 37 cells. This Figure was taken from Auman *et al*. 2006.

1.8 Aims

This project aims to identify marine microorganisms which express bromoperoxidases, haloacid dehalogenases, haloalcohol dehalogenases or haloalkane dehalogenases. These enzymes have the potential to be used in industrial and pharmaceutical applications and in bioremediation processes. Using biochemical, molecular and genomic techniques, the DNA sequences for the enzymes will be determined. The enzymes will then be cloned, over-expressed in *E. coli* and purified. Biochemical characterisation will be carried out with the aim of determining the substrate specificity, thermostability, temperature optimum and solvent stability of the enzyme. Michaelis-Menten kinetics will also be determined. Protein crystallization will be attempted, with the aim of elucidating the three-dimensional structure of the enzymes using X-ray crystallography. The biochemical and structural characterisation of these proteins will allow assessment of their potential role in industrial, pharmaceutical and bioremediation processes.

Chapter 2- General Materials and Methods

All chemicals, unless otherwise stated, were obtained from Sigma-Aldrich (Poole, UK) and were of the highest analytical grade. Reagents for molecular biology were obtained from Fermentas Ltd (York, UK). All solutions were made using Purite double distilled purified water (ddH_2O), unless otherwise stated. All media and buffers used in microbiology were sterilised by autoclaving at 121ºC for 20 min at 15 psi, or by filter sterilisation using a Sartorius Stedim, minisar 0.2 µm single use filter unit.

2.1 Microbiology

2.1.1 Growth media

Media	**Components (L)**
Sea salt buffer	-33 g Crystal sea marinemix (marine enterprises international)
Marine broth	-5 g Peptone extracted from meat -1 g Yeast extract (Melford Ltd) -33 g Crystal sea marinemix
Marine agar	-1 L Marine broth - 8 g Agar
Luria-Bertani (LB) broth	-25 g premixed LB powder (Melford Ltd)
Luria-Bertani (LB) agar	-1 L LB broth -8 g Agar (Melford Ltd)
Super optimal broth (SOC)	-20 g Tryptone (Melford Ltd) -5 g Yeast extract -0.5 g NaCl -0.186 g KCl -3.6 g Glucose -0.95 g $MgCl_2$

Table 2.1- A list of components in buffers, media and agar used for microbial growth.

2.1.2 Growth measurement

Cell growth was measured by aliquoting a 1 ml sample of culture into a disposable plastic cuvette and recording the $OD_{595\ nm}$ using a WPA *CO 8000* Biowave Cell Density Meter.

2.1.3 Marine bacterial growth

An overnight starter culture was prepared by adding 20 µl of bacterial glycerol stock to 10 ml of Marine broth (Table 2.1). The culture was incubated at an appropriate temperature of 4^{o}C, 25^{o}C or 37^{o}C at 220 rpm for 16 h. The 10 ml overnight culture was added to 1 L of Marine broth (Table 2.1) followed by incubation at the appropriate temperature at 220 rpm.

2.1.4 *E. coli* growth

An overnight starter culture was prepared by adding 20 µl of bacterial glycerol stock to 10 ml of LB broth (Table 2.1). The culture was incubated at 37ºC at 220 rpm for 16 h. The 10 ml culture was added to 1 L of LB broth (Table 2.1) followed by incubation at 37ºC, at 200 rpm.

2.1.5 Bacterial preservation

Equal volumes of cell culture and 50% glycerol stock solution were mixed, flash frozen in liquid nitrogen and stored at -80ºC.

2.1.6 Bacterial harvesting

All bacterial cultures were harvested by centrifugation at 12,000 x *g* for 20 min.

2.2 Molecular biology

2.2.1 LB agar/antibiotics plates

1 µl/ml of antibiotic stock solutions (Ampicillin 100 mg/ml or Kanamycin 50 mg/ml) were added to the LB agar once the media had cooled below 50ºC.

2.2.2 DNA gel electrophoresis

Agarose gel electrophoresis was used to analyse DNA fragments based on their size.

2.2.2.1 TAE running buffer (50X) stock solution

- 18.6 g Ethylenediaminetetraacetic acid (disodium salt) (EDTA)
- 57.1 ml of glacial acetic acid
- 242 g Tris-base
- 900 ml of ddH_2O
- pH 8.0

2.2.2.2 Preparation of agarose gels

All agarose gels were made to a final concentration of 1% with 1X TAE buffer (section 2.2.2.1). The gel was placed into the electrophoresis tank and run using Bio-rad mini-sub cell GT horizontal gel apparatus with a Shandon Vokam 500-500 power supply.

2.2.2.3 Running of agarose gels

For analysis DNA samples were mixed with 5X DNA loading buffer (Qiagen Ltd, West Sussex) and loaded onto the gel. 5 µl of DNA hyperladder I (Bioline, see Appendix, 11.1) was also loaded onto the gel, allowing the size of DNA between 0.2-10 kilobase (kbp) to be estimated. A 100 V electric current was passed through the gel for 50 min causing the negatively charged DNA to migrate towards the positive electrode. DNA was visualised using ethidium bromide staining at a concentration of 0.04 µl/ml of gel.

2.2.3 DNA extraction

From overnight cultures genomic DNA was extracted using a GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich). The manufacturer's instructions were followed and all processes carried out at room temperature. Bacterial cells were first lysed with a solution containing chaotrophic salts. With the addition of ethanol, the DNA was bound to a silica membrane by centrifugation. After washing the membrane bound DNA to remove contaminants, the DNA was eluted in 200 µl of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and stored at -20ºC.

2.2.4 PCR reactions

2.2.4.1 Primer design and incorporation of restriction sites

The online program, www.restrictionmapper.org was used to determine appropriate restriction sites which could be incorporated into primer sequences.

All primers were ordered from Eurofins MWG Operon as a 10 nM salt-free pellet. Primers were resuspended in ddH_2O to make a 100 µM solution and stored at -20ºC.

2.2.4.2 PCR cycle 1

Step	Temperature (ºC)	Time (min)
Initial denaturation	95	5
15 cycles of -	-	-
Denaturation	95	0.5
Annealing	50-60	0.5
Extension	72	1
Final Extension	72	5
Stop	4	-

Table 2.2- The standard PCR cycle used.

2.2.4.3 PCR composition 1

- 1.0 unit of Deep Vent polymerase (NEB Ltd, Hertfordshire)
- 5 µl Thermopol Buffer (200 mM Tris-HCl, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 20 mM $MgSO_4$, 1% Triton X-100, pH 8.8)
- 200 µM dNTPs
- 4 mM $MgSO_4$
- 200 ng DNA template
- 0.2 µM of each primer
- 2% DMSO

The PCR reaction was made up to a final volume of 50 µl with ddH_2O.

2.2.4.4 PCR composition 2

- 0.5 units of Deep Vent polymerase (NEB Ltd, Hertfordshire)
- 5 µl Thermopol Buffer (200 mM Tris-HCl, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 20 mM $MgSO_4$, 1% Triton X-100, pH 8.8)
- 200 µM dNTPs
- 8 mM $MgSO_4$
- 400 ng DNA template
- 0.8 µM of each primer
- 2% DMSO

The PCR reaction was made up to a final volume of 50 µl with ddH_2O.

2.2.5 DNA purification and concentration

SureClean Plus (Bioline Ltd, London) was used to purify and concentrate DNA. The manufacturer's instructions were followed and all procedures carried out at room temperature. SureClean Plus removes primer, primer dimers, dNTPs and restriction enzymes.

2.2.6 Gel extraction

DNA gel extraction was performed using a QIAquick Gel Extraction Kit (QiagenLtd, West Sussex). The manufacturer's instructions were followed and all procedures carried out at room temperature. The DNA bands of interest were excised from the gel using a sterile scalpel. The DNA was dissolved in a high salt buffer then bound to a silica membrane by centrifugation. The bound DNA was washed with ethanol to remove impurities, eluted in a low salt buffer and stored at -20ºC.

2.2.7 Competent *E. coli* strains

Table 2.3 summarises the chemically competent cell lines (Stratagene, Cheshire) used and the heat shock time needed at 42ºC for transformation reactions.

***E. coli* strain**	**Type of *E. coli* strain**	**Heat shock Time (sec) for transformation**
XL 10-Gold®	Cloning	30
ArcticExpress™	Expression	20
BL21-CodonPlus (DE3)- RIPL	Expression	30
BL21-CodonPlus (DE3)- *Rosetta*2	Expression	30

Table 2.3- A summary of the chemically competent *E. coli* cell lines used and the heat shock time needed for transformation reactions.

2.2.7.1 Producing chemically competent *E. coli* cell lines

Chemically competent cells were manufactured using the method described in Hanahan *et al.* (1983). An overnight starter culture was prepared by adding 20 µl of bacterial glycerol stock to 10 ml of LB broth (Table 2.1). The culture was incubated at 37ºC at 220 rpm for 16 h. 1 ml of the overnight culture was added to 100 ml of LB broth (Table 2.1), which was incubated at 22ºC at 150 rpm, until the mid log phase (OD_{595} of 0.4-0.6) was reached. Cells were then harvested by centrifugation at 6,000 x *g* for 10 min at 4ºC. The supernatant was discarded and the cell pellet resuspended in 15 ml of ice cold 1XRbCa TXN salts (Anachem Ltd) and stored on ice for 15 min. The resuspended cells were then centrifuged at 4,000 x *g* for 10 min at 4ºC. The supernatant was discarded and the cell pellet gently resuspended in 4 ml of 1XRbCa TXN salts. The resuspended cells were incubated at 4ºC overnight to produce ultra competent cells and preserved as glycerol stocks (section 2.1.5).

2.2.8 Cloning

2.2.8.1 Ligation into pJET1.2/blunt

For ligation into the cloning vector pJET1.2/blunt (see Appendix, 11.2), the ligation reactions in Table 2.4 were used. Reaction mixtures were incubated for 1 h at 22ºC and stored at 4ºC before transformation. Negative controls with no insert were used.

	Reaction 1 (µl)	Reaction 2 (µl)	Reaction 3 (µl)
pJET1.2/blunt	1	1	1
Insert	2	4	8
Reaction Buffer (x2)	10	10	10
T4 DNA Ligase	1	1	1
ddH_2O	6	4	0

Table 2.4- Ligation reactions used with pJET1.2/blunt.

2.2.8.2 Ligation into pET-28a (Novagen Ltd, Nottingham)

For ligation into the expression vector pET-28a (see Appendix, 11.3) the following ligation reactions in Table 2.5 were used. Reaction mixtures were incubated for 1 h at 22ºC and stored at 4ºC before transformation.

	Reaction 1 (µl)	Reaction 2 (µl)	Reaction 3 (µl)	Reaction 4 (µl)	Reaction 5 (µl)	Reaction 6 (µl)
pET-28a	1	1	1	1	2	3
Insert	1	3	5	8	1	1
T4 DNA Ligase	1	1	1	1	1	1
Reaction Buffer (x10)	2	2	2	2	2	2
ddH_2O	15	13	11	8	14	13

Table 2.5- Ligation reactions used with pET-28a.

2.2.8.3 Transformation protocol

A heat shock protocol was used for the transformation of the plasmids pJET1.2/blunt and pET-28a into chemically competent *E. coli* cloning strains (Table 2.3). 2.5 µl of the ligation reaction was added to 50 µl of the competent *E. coli* cell line. The sample was incubated on ice for 20 min, followed by heat shock exposure for the appropriate amount of time (Table 2.3) at 42ºC. The cells were immediately placed on ice for 5 min. The sample was mixed with 200 µl of pre-

warmed SOC media (Table 2.1) and incubated at 37ºC, 200 rpm for 1 h. The culture was plated out onto an LB/antibiotic agar plate (section 2.2.1) and incubated at 37ºC overnight.

As a control 1 µl of the pUC18 control plasmid was transformed into 50 µl of the competent *E. coli* cells and the transformation protocol followed to test the transformation efficiency.

2.2.8.4 Plasmid extraction

Overnight cultures were incubated at 37ºC at 220 rpm. 4.5 ml samples of these overnight cultures were used for plasmid extractions using a GeneJet^{TM} Plasmid Miniprep kit. The manufacturer's instructions were followed and the procedure carried out at room temperature. The kit is based on the alkaline lysis method (Birnboim & Doly, 1979). Cells were ruptured on mixing with a lysis buffer. The solution was neutralized and centrifuged to remove proteins and cell debris. The plasmid DNA was bound to a silica membrane and washed several times to remove impurities. Plasmid DNA was then eluted in 10 mM Tris-HCl, pH 8.5 and stored at -20ºC.

2.2.8.5 pJET1.2/blunt plasmid digestion

The following protocol was used to liberate the DNA insert from pJET1.2/blunt.

- 5 µl MiniPrep plasmid
- 1 µl *Not*I restriction enzyme (Fast Digest)
- 1 µl *Nde*I restriction enzyme (Fast Digest)
- 2 µl 10x Fast Digest Green Buffer
- 11 µl ddH_2O

Digestion samples were mixed gently and incubated at 37ºC for 15 min.

2.2.8.6 pET-28a plasmid digestion

- 20 µl MiniPrep plasmid
- 4 µl *Not*I restriction enzyme (Fast Digest)
- 4 µl *Nde*I restriction enzyme (Fast Digest)
- 4 µl 10x Fast Digest Green Buffer
- 7.6 µl ddH_2O
- 0.4 µl BSA

Digestion samples were mixed gently and incubated at 37ºC for 4 h.

2.2.8.7 SAP treatment of linearised pET-28a

To prevent self-annealing of pET-28a, the plasmid was treated with SAP (Roche Ltd), which catalyses the 5' dephosphorylation of DNA. 0.5 µl of SAP and 2.5 µl of SAP buffer were added to the plasmid digest mixture (section 2.2.8.7) followed by an incubation at 37ºC for 15 min and heat inactivation at 65ºC for 20 min.

2.2.8.8 Sanger sequencing

To ensure that the gene of interest was inserted into the vectors, plasmid miniprep samples were sent to Geneservice, London for Sanger sequencing. T7 Forward and T7 Reverse (Table 2.6) primers were used for pET-28a sequencing and the primers supplied with the CloneJET™ PCR Cloning Kit (Table 2.6) were used for pJET1.2/blunt sequencing.

Primer	**Sequence**
T7 Forward	5'-TAATACGACTCACTATAGGG-3'
T7 Reverse	5'-GCTAGTTATTGCTCAGCGG-3'
pJET1.2/blunt Forward	5'-CGACTCACTATAGGGAGAGCGGC-3'
pJET1.2/blunt Reverse	5'-AAGAACATCGATTTTCCATGGCAG-3'

Table 2.6- Forward and reverse PCR primers used for Sanger sequencing of pJET1.2 and pET-28a.

2.3 Biochemistry

2.3.1 Induction of protein expression

An overnight starter culture was prepared by adding 20 µl of *E. coli* glycerol stock to 10 ml of LB broth (Table 2.1). The cultures were incubated at 37ºC at 220 rpm for 16 h. 1 ml of the overnight cultures were added to 100 ml of LB broth (Table 2.1).

Cultures were incubated under various conditions to determine the best over-expression system. Conditions that were varied included induction at $OD_{595\ nm}$ at readings between 0.6-1.2; incubation temperature at 37ºC, 28ºC and 12ºC; final IPTG concentrations ranging between 0.5-2 mM and time after induction ranging between 2-20 h. For analysis of the over-expressed protein, 1.5 ml samples of the culture were taken and the bacterial paste harvested at 12,000 x *g*.

2.3.2 Cell lysis

2.3.2.1 Sonication

Bacterial paste was resuspended in an appropriate homogenisation buffer to a final concentration of 10% (w/v). The solution was sonicated using a Soniprep 150 Sonicator (Sanyo) at 10 amplitude microns on ice for a total of 4 x 1 min, with 4 x 1 min breaks in between to prevent overheating. Cell lysis samples were centrifuged at 20,000 x *g* for 20 min to pellet the insoluble fraction.

2.3.2.2 Cell lysis using BugBuster

BugBuster (Novagen, UK) was used to lyse bacterial cells when cell pellets were collected from 1.5 ml of culture or less. BugBuster contains non-ionic detergents which rupture bacterial cells walls without denaturing proteins. The cell pellet was resuspended in 300 µl of a 1/10 dilution of BugBuster in appropriate buffer. The sample was incubated on a shaking platform for 20 min at room temperature. Cell lysis samples were centrifuged at 13,000 x *g* for 20 min to pellet the insoluble fraction.

2.3.3 Protein purification

All protein samples were purified using an ÄKTA purifier (GE Healthcare). Prior to protein purification, samples were filtered with a Sartorius Stedim, minisar 0.2 µm single use filter unit and buffers were vacuum filtered with a 0.2 µm nylon membrane filter (Whatman). Throughout protein purification on chromatography columns, the protein concentration was determined by measuring the absorbance at 280 nm.

2.3.3.1 Purification Buffers

Buffer	Component
FFQ buffer A	0.1 M Tris-H_2SO_4 0.5 mM EDTA 1 mM benzamidine (BAM) 1 mM Phenylmethylsulfonyl fluoride (PMSF) pH 8.2
FFQ buffer B	0.1 M Tris-H_2SO_4 1 M $NaCl_2$ 5 mM EDTA 1 mM BAM 1 mM PMSF pH 8.2
Hydrophobic interaction chromatography (HIC) buffer A	0.1 M Tris-H_2SO_4 1 M Ammonium sulfate 5 mM EDTA 1 mM BAM 1 mM PMSF pH 8.2
Hydrophobic intereaction chromatography (HIC) buffer B	0.1 M Tris-H_2SO_4 5 mM EDTA 1 mM BAM 1 mM PMSF pH 8.2
Nickel affinity chromatography (NAC) buffer A	0.1 M Tris-H_2SO_4 0.1 M NaCl 0.01 Imidazole 5 mM EDTA 1 mM BAM 1 mM PMSF pH 8.2
Nickel affinity chromatography (NAC) buffer B	0.1 M Tris-H_2SO_4 0.2 M NaCl 0.5 M Imidazole 5 mM EDTA 1 mM BAM 1 mM PMSF pH 8.2
Gel filtration chromatography (GF) buffer	0.1 Tris-H_2SO_4 0.1 M NaCl 5 mM EDTA 1 mM BAM 1 mM PMSF pH 8.2

Table 2.7- A list of purification buffers and their components.

2.3.3.2 Dialysis

Protein samples were dialysed using a cellulose dialysis tubing membrane (Sigma and Aldrich, Poole, UK) with a molecular weight cut off (MWCO) of 8,000 kDa. The dry cellulose membrane was boiled in ddH_2O for 10 min and stored in 0.05% sodium azide for long term storage. Dialysis membranes were rinsed with ddH_2O prior to use and the protein dialysed in 5 L of GF buffer (Table 2.7) for 4 h at 4ºC with gentle stirring; this was repeated twice. For volumes lower than 500 µl a Slide-A-Lyzer Mini Dialysis Unit (Thermo Scientific) with a MWCO of 10,000 kDa was used and the protein dialysed against 1 L of GF buffer (Table 2.7) for 4 h at 4ºC with gentle stirring; this was repeated twice.

2.3.3.3 Ammonium sulfate precipitation trials

After sonication soluble proteins in the supernatant were treated with ammonium sulfate in 10% increments from 20-80% saturation. With each increment, ammonium sulfate was added over a period of 1 h at 4ºC, followed by centrifugation at 12,000 x *g* for 20 min at 4ºC to remove precipitated proteins.

2.3.3.4 Ammonium sulfate precipitation

After sonication the soluble proteins in the supernatant were treated to a 50% ammonium sulfate salt cut over 1 h at 4ºC, followed by centrifugation at 12,000 x *g* for 20 min at 4ºC to remove precipitated proteins. The supernatant was then treated to a 80% ammonium sulfate salt cut over 1 h at 4ºC, followed by centrifugation at 12,000 x *g* for 20 min at 4ºC to precipitate the remainder of the proteins in solution. The precipitated proteins were then resuspended in the appropriate homogenisation buffer to a final concentration of 10% (w/v).

2.3.3.5 Hydrophobic interaction chromatography (HIC)

A 70 ml Phenyl Sepharose column (GE Healthcare) was equilibrated with 3 column volumes of HIC buffer A (Table 2.7). The ammonium sulfate salt pellet (section 2.3.3.4) was resuspended in HIC buffer A and the protein loaded onto the column using a Superloop (GE Healthcare). Unbound protein was washed off with 2 column volumes of HIC buffer A. To elute bound proteins a 4 column volume gradient to 100% HIC buffer B (Table 2.7) was applied followed by a

further 2 column volumes of HIC buffer B. 5 ml fractions were collected throughout purification and the protein concentration determined by measuring the absorbance at 280 nm.

2.3.3.6 Anion exchange chromatography

A 70 ml FFQ column (GE Healthcare) was equilibrated with 3 column volumes of FFQ buffer A (Table 2.7). The protein was loaded onto the column using a Superloop (GE Healthcare) and the unbound protein was washed off with 2 column volumes of FFQ buffer A. To elute bound proteins a 4 column volume gradient to 100% FFQ buffer B (Table 2.7) was applied followed by a further 2 column volumes of FFQ buffer B. 5 ml fractions were collected throughout purification. The fractions corresponding to the active protein were pooled together and concentrated using a 10 kDa membrane (Vivaspin 20; Viva science) at 3,000 x *g*, at 4ºC until the final volume reached 1 ml.

2.3.3.7 Gel filtration (GF) chromatography

A 120 ml Superdex 200 or 75 GF column (GE Healthcare) was equilibrated with 1 column volume of GF buffer (Table 2.7). The 1 ml protein sample was applied to the column, using a 1 ml sample loop (GE Healthcare) and the protein eluted over 1 column volume of GF buffer. 1 ml fractions were collected throughout purification.

2.3.3.8 Nickel affinity chromatography

A 70 ml nickel affinity chromatography column (GE Healthcare) was equilibrated with 3 column volumes of NAC buffer A (Table 2.7). The protein was loaded using a Superloop (GE Healthcare) and the unbound protein was washed off with 2 column volumes of NAC buffer A. To elute the bound protein a 4 column volume gradient to 100% NAC buffer B (Table 2.7) was applied followed by a further 2 column volumes of NAC buffer B. 5 ml fractions were collected throughout purification. The fractions corresponding to the active protein were pooled together and concentrated using a 10 kDa membrane (Vivaspin 20; Viva science) at 3,000 x *g*, at 4ºC until the final volume reached 1 ml.

2.3.4 Protein parameter calculations

Protein parameters were calculated using the online program ProtParam (http://expasy.org/tools/protparam.html). This online program makes a number of calculations about proteins based on their amino acid sequence. The programme calculates the amino acid composition, theoretical pI, extinction coefficient and the theoretical stability of the protein.

2.3.5 Protein concentration determination

Protein concentration was calculated by measuring the absorbance at 280 nm using a Shimadzu UV-VIS 2100 spectrophotometer. A 200 µl aliquot of protein solution was placed in a quartz cuvette with a path length of 1 cm and a blank reading also measured with the buffer. The protein concentration was then determined using Beer-Lambert law.

$$A = \varepsilon l c$$

Equation 2.1- Beer-Lambert law: A is the absorbance at 280 nm, ε is the extinction coefficient, l is the path length and c is the protein concentration.

2.3.6 SDS-PAGE

SDS-PAGE is based on the method of Laemmli (1970) and is a common method of separating proteins based on their molecular weight. Dependent on the size of the protein of interest, the concentrations of the acrylamide and the crosslinker (acrylamide/bisacrylamide) can be varied to obtain the best separation of protein in the sample. All SDS-PAGE gels consisted of a 6.0% stacking gel and a 12.5% separating gel. Gels were run using a Bio-Rad mini-protean-Tetra cell (Bio-Rad Laboratories Ltd).

2.3.6.1 SDS-PAGE buffers

Buffer	Composition
SDS-PAGE 10X running buffer	18 g/ L Tris-HCl 6 g/L SDS 86.4 g/L glycine
SDS buffer A	30% acrylamide/bisacrylamide (29:1)
SDS buffer B	1.5 M Tris-HCl, pH 8.6
SDS buffer C	0.5 M Tris-HCl, pH 6.8
SDS buffer D	10% SDS w/v
SDS-PAGE loading buffer	100 mM Tris-HCl, pH 8.0 2% β-mercaptoethanol 20% glycerol v/v 4% SDS w/v 0.2% bromophenol blue w/v
SDS-PAGE microwave stain	- 1 g Brillant Blue Coomassie (G250) - 800 ml of ddH_2O - 200 ml glacial acetic acid - 1 L methanol

Table 2.8- A list of SDS-PAGE buffers, stains and their components.

2.3.6.2 SDS-PAGE preparation

Separating gel (12.5%)

- 4.2 ml SDS buffer A
- 2.5 ml SDS buffer B
- 1 ml SDS buffer D
- 2.3 ml ddH_2O

Stacking gel (6.0%)

- 2 ml SDS buffer A
- 2.5 ml SDS buffer C
- 1 ml SDS buffer D
- 4.5 ml ddH_2O

To initiate polymerisation 100 µl of ammonium persulfate (10% w/v) and 10 µl of tetramethylethylenediamine were added to the mixtures. SDS-PAGE apparatus were set up according to manufacturer's instructions. The gel was cast in-between 2 glass plates with a 1 mm gap. The separating layer was filled up to two thirds of the way to the top and covered with water saturated 1-butanol until the separating layer was set to prevent the gel from drying out. The 1-butanol saturated water was removed and the stacking gel poured to the top of the cast. The wells were created by inserting a comb and the stacking gel left to set.

2.3.6.3 Sample preparation

10 µl of protein was mixed with 10 µl of SDS-PAGE loading buffer (Table 2.8) and the sample incubated at 100ºC for 10 min to ensure protein denaturation.

2.3.6.4 SDS-PAGE running

Once the gel was set the comb was removed and the gel placed in the electrophoresis tank. The tank was filled with 1X SDS-PAGE running buffer (Table 2.8) and 10 µl of the protein ladder (Bio-Rad precision plus, see Appendix, 11.4) was loaded onto the gel to allow accurate measurement of protein sizes in the sample. The protein samples were loaded onto the gel and the gel run at 200 V for 50 min.

2.3.6.5 SDS-PAGE staining

The gel was removed from the cast, immersed in SDS-PAGE gel staining solution (Table 2.8) and microwaved for 3 min. The gel was rinsed, placed in 500 ml of ddH_2O and microwaved for 20 min to destain.

2.3.6.6 Pre-Cast SDS-PAGE for protein mass spectrometry.

To reduce contamination of protein samples, pre-cast NuPAGE gels at a concentration of 10% Bis-Tris (Invitrogen) were used for protein mass spectrometry. All protein samples run on NuPAGE gels for mass spectrometry were treated aseptically after purification to avoid contamination.

2.3.6.7 NuPAGE running

The comb was removed and the NuPAGE apparatus were set up according to manufacturer's instructions.

The tank was filled with 1X 2-(N-morpholino) ethanesulfonic acid (MES) running buffer (Invitrogen) and 10 µl of the protein ladder (Spectra Multi Colour Broad Range, see Appendix, 11.5) was loaded onto the gel to allow accurate measurement of protein size in the sample. The protein samples were loaded onto the gel and the gel run at 200 V for 35 min.

2.3.6.8 NuPAGE staining

The NuPAGE gel was removed from the cast and rinsed with 20 ml of ddH_2O with gentle shaking; this was repeated three times. The gel was then covered with 20 ml of Simple Blue Safe Stain (Invitrogen) and incubated at room temperature for 1 h with gentle shaking. The NuPAGE gel was incubated with 100 ml of ddH_2O for 1 h with gentle shaking to destain; this was repeated twice.

2.3.7 Liquid chromatography mass spectrometry sample (LC-MS) preparation

All samples were prepared aseptically in a laminar flow hood to prevent contamination and all necessary solutions and equipment sterilised by autoclaving at 121ºC for 20 min at 15 psi.

2.3.7.1 Reagents for protein destaining and reduction

- LC-MS grade acetonitrile
- 50 mM ammonium bicarbonate
- 50 mM ammonium bicarbonate in 50% acetonitrile
- 10 mM dithiothreitol (DTT) in 50 mM ammonium bicarbonate
- 50 mM iodoacetamide in 50 mM ammonium bicarbonate
- porcine trypsin (Promega, Southampton)

2.3.7.2 Reagents for protein peptide extraction

- LS-MS 10% acetonitrile/5% formic acid
- 50 mM ammonium bicarbonate/5% acetonitrile/0.1% formic acid

2.3.7.3 In gel tryptic digest protocol

Protein bands were excised from the gel using a sterile scalpel and transferred to an Eppendorf tube.

2.3.7.4 Protein destaining

The protein bands were mixed with 300 µl of 50 mM ammonium bicarbonate in 50% acetonitrile and incubated at 30ºC for 30 min with mixing every 5 min. This was repeated three times.

2.3.7.5 Protein reduction

The gel was removed from the destaining solution and rehydrated with 200 µl of 50 mM ammonium bicarbonate for 10 min. The 50 mM ammonium bicarbonate was removed and the gel slice dehydrated with 200 µl of acetonitrile for 10 min. The acetonitrile was removed and the gel slice rehydrated and reduced by incubating with 200 µl of 10 mM DTT in 50 mM ammonium bicarbonate for 45 min at 50ºC. The DTT was removed and the sample alkylated with 200 µl of 50 mM iodoacetamide in 50 mM ammonium bicarbonate for 45 min at room temperature in the dark. The gel slice was removed and incubated with 200 µl 50 mM ammonium bicarbonate for 3 min followed by incubation with 200 µl of acetonitrile, this was repeated twice. The gel slice was removed and placed in an Eppendorf tube and left to dry in a laminar flow for 30 min.

2.3.7.6 Tryptic digest

The gel slice was rehydrated with 30 µl of 50 mM porcine trypsin in 50 mM ammonium bicarbonate for 20 min on ice. 40 µl of 50 mM ammonium bicarbonate was added and the sample incubated overnight at 37ºC. Protein samples were then placed in a sonication bath (Branson 2510) for 15 min followed by centrifugation for 15 min at 12,000 x *g*. The supernatant was decanted into a clean Eppendorf tube and centrifuged for a further 15 min at 12,000 x *g*. The top 10 µl of the supernatant was decanted and used as the peptide sample for peptide mass spectrometry.

2.3.8 Activity assays

2.3.8.1 Bromoperoxidase assay

The bromoperoxidase assay is a colorimetric assay based on the conversion of phenol red to bromophenol blue, in the presence of hydrogen peroxide and bromide (De Boer *et al*,. 1987). The bromoperoxidase assay solution contained 100 ml of 100 mM K_2HPO_4/KH_2PO_4 pH 6.5, 3 ml of phenol red solution (0.002%) and 100 ml of 100 mM KBr. To initiate the assay, 150 µl of the assay solution was mixed with 25 µl of 10 mM H_2O_2 and 25 µl of cell extract or purfied protein. The assay was performed in a 96 well ELISA Microplate (Greiner Bio-One) and the progress of the reaction was visualised as the substrate (phenol red) was turned over and the buffer changed from a yellow to purple. Positive controls were performed using a purified sample of the vanadium dependant bromoperoxidase from *C. officinalis* (stock prepared 'in house') and negative controls performed using buffer without protein.

2.3.8.2 Dehalogenase assay

The dehalogenase assay is a quantitative and colorimetric assay based on the method of Holloway *et al.* (1998). The dehalogenase assay solution had a final concentration of 1 mM HEPES, 1 mM EDTA, 20 mM sodium sulfate, 10 mM substrate, 20 µg/ml phenol red, pH 8.2. To initiate the assay, 180 µl of dehalogenase assay solution was mixed with 20 µl of cell extract. There is a reduction in the pH of the weakly buffered solution when the dehalogenase

enzymes turn over the substrate to produce protons and halide ions. This can be visualised from a colour change from red to yellow. The assay was either performed in a 96 well ELISA Microplate (Greiner bio-one) and measured at 540 nm using a FLUO star OPTIMA spectrophotometer from BMG LABTECH or in a plastic cuvette and measured on a Shimadzu UV-VIS 2100 Spectrophotometer.

2.3.9 Biochemical characterisation

All standard curves and activity assays were performed in triplicate with negative controls using a plastic cuvette and measured with a Shimadzu UV-VIS 2100 Spectrophotometer. All reactions unless otherwise stated were carried out at 25ºC. The final protein concentration used in all biochemical characterisation experiments was 0.2 mg/ml in GF buffer (Table 2.7).

2.3.9.1 Standard curves

To produce a standard curve the dehalogenase assay solution (section 2.3.8.2) was mixed with 1 M HCl to final concentrations between 0-2 mM in a total volume of 200 µl. Temperature had a slight effect on the activity assay, so standard curves were performed at 10ºC intervals between 25-65ºC.

2.3.9.2 Thermostability

Enzyme thermostability was determined by incubating samples at temperatures between 30-70ºC for 30, 60 and 90 min followed by incubation on ice for 10 min. The standard dehalogenase assay (section 2.3.8.2) was then carried out.

2.3.9.3 Solvent stability

Solvent stability of the enzyme was determined by incubation of the samples with ethanol, methanol, acetonitrile and DMSO at concentration between 10-70% for 1 h at room temperature. The standard dehalogenase assay (section 2.3.8.2) was then carried out.

2.3.9.4 Substrate specificity

Substrate specificity of the enzyme was determined using the standard dehalogenase assay (section 2.3.8.2) was carried out with a range of substrates. Substrates had varying chain lengths, halides and halide positions (see structures in Appendix, 11.6).

2.3.9.5 Temperature optimum

Optimum temperature of the enzyme was determined using the standard dehalogenase assay (section 2.3.8.2) which was carried out at temperatures ranging between 25-65ºC. The assay solution was pre-incubated at the set temperatures for 5 min before the addition of the enzyme.

2.3.9.6 Michaelis-Menten kinetics

Enzyme kinetics were determined by carrying out the standard dehalogenase assay (section 2.3.8.2) with the best substrate at the optimum temperature for the enzyme reaction. Substrate concentration varied from 0.1-10 mM and the final protein concentration used was at 0.2 mg/ml in GF buffer (Table 2.7). GraFit version 7 was used to calculate the K_m and V_{max}.

Chapter 3- Microbial Isolation from the marine environment

3.1 Introduction

Sea water has high bromide and chloride concentrations, therefore microbes able to survive in sea water have resistance to these halogens. Depending on the geographical location, the approximate concentrations of iodide, bromide, chloride are 1 µM, 1 mM and 0.5 M respectively (Butler *et al*., 1993). Marine organisms are known to utilise chloride and bromide by incorporating them into primary and secondary metabolites. Halogenated secondary metabolites produced by algae and sponges are involved in host defence mechanisms. For example, *Asparagopsis armata* produces dibromoacetic acid and bromoform which are secreted by gland cells onto the extracellular tissue. These halogenated metabolites inhibit the colonisation and formation of biofilms on the surface of sponges and algae (Paul *et al*., 2006). Microbes which are able to colonise algal surfaces may express dehalogenases which can detoxify the antimicrobial agents by directly removing the halogen atom. This would allow microbial surface colonisation and biofilm formation on the algal surface and potential utilisation of the dehalogenated compound as a carbon source (Torz *et al*., 2007). Marine environments are enormously diverse and can vary in salinity, pH, temperature and pressure. Marine microorganisms have adapted to survive in these extreme conditions and often possess novel enzymes which are highly stable under these conditions (Champdoré *et al*., 2007).

This chapter describes the collection of a library of microorganisms from marine environments. This microbial collection is referred to as the Exeter microbial library. The Exeter microbial library was assayed for bromoperoxidase, haloacid dehalogenase, haloalcohol dehalogenase and haloalkane dehalogenase activity. High throughput screening of the wild type microorganisms allows specific enzymatic activities to be identified for a range of conditions (e.g. temperature and substrates).

3.2 Materials and Methods

3.2.1 Collection of marine microbes

Numerous marine samples of algae, sediment and crustacean were collected from Devon (Landram bay) and Lochaber (Loch Leven). From the site of collection, samples were stored and transported in sea water. 2.5 g of each sample were gently rinsed with 40 ml of Sea salt buffer (Table 2.1) to collect loosely attached microbes. The same 2.5 g samples were washed and vortexed for 1 min with 40 ml of fresh Sea salt buffer to collect tightly attached microbes present in biofilms. 100 µl of the marine microbial solutions were spread onto Marine agar plates (Table 2.1) and incubated at 4^{o}C, 25^{o}C and 37^{o}C for 72 h. Some samples were cultured with supplemented media, with the addition of either 5 mM 1,2-dichloroacetic acid, 1,2-dichloropropane or 1,2-dichloropropanol (in the agar or liquid media).

3.2.2 Isolation of marine microbial monocultures

Single colonies from the mixed culture agar plates were picked and the streak plate technique was used to isolate single microbial species. The Marine agar streak plates (Table 2.1) were incubated for 48 h at the temperature the microbes were originally isolated. 10 ml vials of Marine broth were inoculated with single colonies and incubated at correct growth temperatures for 48 h at 220 rpm. Bacteria were preserved as glycerol stocks (section 2.1.5). All microbes isolated on supplemented media were always cultured with the addition of 5 mM of the substrate (1,2-dichloroacetic acid, 1,2-dichloropropane or 1,2-dichloropropanol).

3.2.3 Marine microbial growth and enzymes assays

All microbial strains in the Exeter microbial library were cultured as described in section 2.1.3 with incubation at appropriate growth temperatures for 48 h. The microbial paste was harvested (section 2.1.6) and the cells lysed with BugBuster (section 2.3.2).

The cell lysate samples were assayed for bromoperoxidase, haloacid dehalogenase, haloalcohol dehalogenase and haloalkane dehalogenase activity (sections 2.3.8.2 and 2.3.8.3). Substrates used in the dehalogenase activity assay were chloroacetic acid, 2-chloropropionic acid, 2-bromopropionic acid, 3-bromopropionic acid, 1,2-dichloroethane, 1-bromohexane, 1,2-dichloropropane, 1,6-dichlorohexane, 1,3-dichloro-2-propanol and 2-bromoethanol (structures shown in Appendix, 11.6). The positive controls were performed using the the L-haloacid dehalohenase from *S. tokodaii* and the bromoperoxidase from *C. officinalis*.

3.2.4 Identification of the microorganisms in the Exeter microbial library

All microbial strains from the Exeter microbial library were incorporated into the Aquapharm Biodiscovery Ltd microbial collection. Strains were sent to Geneservice, Source Bioscience and the microbial species identified using 16S rRNA PCR.

3.3 Results and Discussion

A total of 14 samples of algae were collected from Landram Bay, Devon, England. A total of 9 samples of algae, 1 crustacean and 1 sediment sample were collected from the shore of Lock Leven, Lochaber, Scotland. Algae samples included *Laminaria saccharina* (Figure 3.1), *Ulva lactuca* (Figure 3.2), *C. officinalis* (Figure 3.3), *Laminaria hyperborea*, *Laminaria digitata*, *Fucus serratus*, *Halidrys siliquosa* and *Pelvetia canaliculata*.

3.3.1 Identification of microorganism in the Exeter microbial library

In the Exeter microbial library a total of 56 different bacterial species were identified using 16S rRNA PCR. The FASTA results were analysed using nucleotide Basic Local Alignment Search Tool (BLAST) (Altschul *et al*., 1990). Table 3.1 displays the Aquapharm Ltd code numbers, strain identifications, location and type of the marine samples collected, growth temperatures and the presence of supplements in the growth media. A total of 14 *Pseudoalteromonas* spp., 6 uncultured/marine bacteria, 4 *Arthrobacter* spp., 2 *Bacillus* spp., 2 *Halomonas* spp., 2 *Zobellia* spp., 2 *Psychrobacter* spp., 2 *Photobacterium* spp., 2

Pseudorhodobacter spp., 1 *Cobetia* sp., 1 Alpha *proteobacterium* sp., 1 Gamma *proteobacterium* sp., 1 *Vibrio* sp., 1 *Streptomyces* and 1 *Shewanella* sp. were identified.

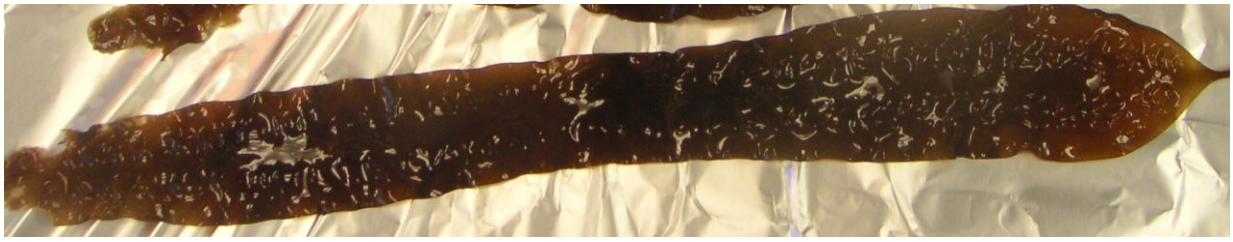


Figure 3.1- Photograph of the algal sample *L. saccharina*, collected from Landram bay, Devon, England. Pseudoalteromonas spp. AQP8276, AQP8280 and AQP8291 were isolated from *L. saccharina*.

Figure 3.2- Photograph of the algal sample *U. lactuca*, collected from Landram bay, Devon, England. *Halomonas halodurans*, AQP8273 was isolated from *U. lactuca.*

Figure 3.3- Photograph of the algal sample *C. officinalis*, collected from Landram bay, Devon, England. Bacterium QM35, AQP8286 was isolated from *C. officinalis*.

AQP number	Microbial strain identification	Location of marine sample	Marine sample	Supplemented media (5 mM)	Growth temperature (ºC)
AQP8273	*Halomonas halodurans* (99%)	Devon	Algal- *U. lactuca*	N/A	37
AQP8276	*Pseudoalteromonas* sp. S3178 (99%)	Devon	Algal- *L. saccharina*	N/A	25
AQP8277	*Cobetia* sp. QDHT-10 (99%)	Devon	Algal- *Corallina* sp.	N/A	25
AQP8280	*Pseudoalteromonas* sp. AMGP1 (99%)	Devon	Algal- *L. saccharina*	N/A	25
AQP8281	North Sea bacterium H7 (99%)	Devon	Algal- *L. hyperborea*	N/A	25
AQP8286	Bacterium QM35 (99%)	Devon	Algal- *C. officinalis*	N/A	25
AQP8291	*Pseudoalteromonas* sp. SM9913 (98%)	Devon	Algal- *L. saccharina*	N/A	25
AQP8292	*Zobellia galactanovora* (98%)	Devon	Algal- *Corallina* sp.	N/A	25
AQP8297	Marine proteobacterium MS-85 (99%)	Devon	Algal- *F. serratus*	N/A	25
AQP8298	*Halomonas marina* (99%)	Devon	Algal- *Corallina* sp.	N/A	25
AQP8299	*Zobellia amurskyensis* (95%)	Devon	Algal- *H. siliquosa*	N/A	25
AQP8300	Marine bacterium B27 (99%)	Devon	Algal- *Corallina* sp.	N/A	25

AQP8301	*Pseudoalteromonas citrea* CIP (99%)	Devon	Algal- *Corallina* sp.	N/A	25
AQP8308	*Arthrobacter* sp. TM4_3 (99%)	Devon	Algal- *P. canaliculata*	N/A	25
AQP8312	*Pseudoalteromonas* sp. S1692 (98%)	Devon	Algal- *L. hyperborea*	N/A	25
AQP8313	*Alpha proteobacterium* KY8-1 (99%)	Devon	Algal- *L. digitata*	N/A	25
AQP8315	*Bacillus* sp. MACL11B (99%)	Devon	Algal- *Corallina* sp.	N/A	25
AQP8316	*Pseudoalteromonas tetraodonis* S1002 (94%)	Devon	Algal- *H. siliquosa*	N/A	25
AQP8319	*Bacillus* sp. YTM5 (99%)	Lochaber	Crustacean	1, 2-dichloroacetic acid	25
AQP8320	Uncultured bacterium SBS-FW-030 (99%)	Lochaber	Algal	1, 2-dichloroacetic acid	25
AQP8321	*Pseudoalteromonas* sp. BSw20058 (98%)	Lochaber	Crustacean	N/A	25
AQP8322	*Pseudoalteromonas marina* S411 (99%)	Lochaber	Sediment	N/A	25
AQP8324	*Vibrio casei* WS 4540 (99%)	Lochaber	Algal	1, 2-dichloropropanol	25
AQP8325	*Psychrobacter* sp. Z1 (99%)	Lochaber	Algal	1, 2-dichloropropanol	25
AQP8326	*Photobacterium* sp. COX (98%)	Lochaber	Algal	1, 2-dichloropropanol	25

AQP8327	*Psychrobacter* sp. 4Dc (98%)	Lochaber	Algal	1, 2-dichloropropanol	25
AQP8337	*Photobacterium* sp. P16011A (95%)	Lochaber	Algal	1, 2-dichloropropanol	25
AQP8339	*Pseudorhodobacter incheonensis* KOPRI 13537 (99%)	Lochaber	Algal	1, 2-di chloropropane	25
AQP8340	*Pseudorhodobacter incheonensis* (99%)	Lochaber	Algal	1, 2-dichloropropane	25
AQP8341	*Gamma proteobacterium* SOMBO36 (99%)	Lochaber	Algal	1, 2 dichloropropane	25
AQP8346	*Pseudoalteromonas* sp. P108 (99%)	Lochaber	Algal	1, 2-dichloropropane	25
AQP8347	Uncultured *Pseudoalteromonas* sp. H02C48-3 (98%)	Lochaber	Sediment	1, 2-dichloropropanol	25
AQP8349	*Pseudoalteromonas* sp. 114Z-11 16S (99%)	Lochaber	Sediment	1, 2-dichloropropanol	25
AQP8352	Pseudoalteromonas sp. H02C48-52 (99%)	Lochaber	Crustacean	1, 2-dichloropropanol	25
AQP8353	*Pseudoalteromonas tetraodonis* JAM-K142 (99%)	Lochaber	Sediment	1, 2-dichloropropanol	25
AQP8354	*Arthrobacter* sp. I_GA_A_1_10 (99%)	Lochaber	Sediment	1, 2-dichloropropanol	25

AQP8355	*Pseudoalteromonas* sp. SM9913 (98%)	Lochaber	Algal	1, 2-dichloropropanol	25
AQP8358	*Microbacterium* sp. OS-6 (100%)	Lochaber	Sediment	N/A	25
AQP8359	*Streptomyces* sp. strain RHH34 (99%)	Lochaber	Algal	1, 2-dichloroacetic acid	25
AQP8360	*Shewanella baltica* OS678 (99%)	Lochaber	Algal	1, 2-dichloroacetic acid	25
AQP8361	*Pseudomonas koreensis* SSG6	Lochaber	Sediment	N/A	25
AQP8362	*Arthrobacter* sp. (99%)	Lochaber	Sediment	N/A	25
AQP8363	*Arthrobacter* sp. 18/4 (99%)	Lochaber	Algal	N/A	25

Table 3.1- A table displaying the Aquapharm Biodiscovery Ltd code numbers, strain identifications, location and type of the marine samples collected, growth temperatures and the presence of supplements in the growth media. N/A stands for not applicable (indicating that the media is not supplemented).

3.3.2 Activity assays

A range of microbes from the surface of algae, sediments and crustaceans were isolated and cultured as described in section 3.2.3. Bacterial cell pellets were lysed using bugbuster and the soluble fractions were assayed for bromoperoxidase, haloacid dehalogenase, haloalcohol dehalogenase and haloalkane dehalogenase activity. The dehalogenase assays were performed with a range of substrates (structures shown in Appendix, 11.6) utilised by previously characterised dehalogenase proteins, as reviewed by Kurihara *et al*., 2008 (listed in section 3.2.3). Figure 3.4 shows an activity assay plate used to test the strains in the Exeter microbial libary for haloacid dehalogenase activity. No dehalogenase or bromoperoxidase activity was detected.

There is a large change in environmental conditions when culturing wild type microorganisms *in vitro*. This can cause the microorganisms to go into 'shock' which can dramatically alter the expression of proteins, causing genes to be turned off. Silent genes are present in all organisms and often specific compounds (sometimes environmental) are needed to induce the expression of these proteins. This could account for not detecting any dehalogenase or bromoperoxidase activity in the microorganisms in the Exeter microbial library.

In an attempt to upregulate dehalogenase expression in wild type microorganisms, growth trials in the presence of dehalogenase substrates were attempted. AQP5750 and AQP4626 growth trials in Marine broth at concentrations between 10-200 mM chloroacetic acid and chloropropionic acid were conducted (with and without adjustment of media to pH 7.5 with NaOH). AQP5750 and AQP4626 showed significantly stunted or no growth in the presence of substrates.

All microorganisms in the Exeter microbial library have been incorporated into the proprietary marine microorganism collection in Aquapharm Biodiscovery Ltd. This collection exceeds over 7,000 microbes and is screened for anti-cancer, anti-oxidant, anti-inflammatory and anti-microbial compounds.

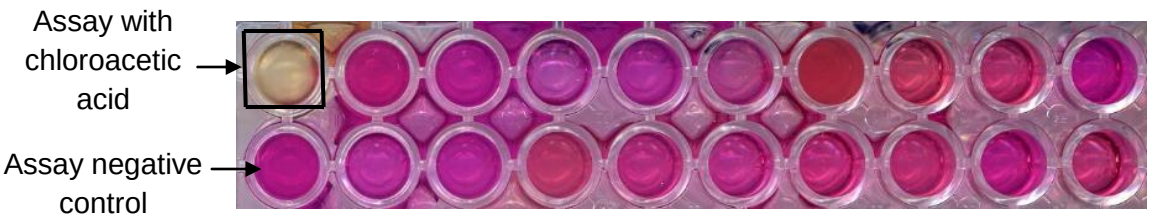


Figure 3.4- Activity assay plate used to test the strains in the Exeter microbial libary for haloacid dehalogenase activity. Activity in the soluble fraction is detected by a colour change from red to yellow. The positive control (L-haloacid dehalohenase from *S. tokodaii*) is highlighted by the black box.

Chapter 4- Purification of L-haloacid dehalohenases from wild type bacteria

4.1 Introduction

The company Aquapharm Biodiscovery Ltd aim to discover novel pharmaceutical compounds including anti-microbials, antioxidants, anti-inflammatories and novel biocatalysts by exploiting the chemical diversity of marine microorganisms. Aquapharm Biodiscovery Ltd has a microbial library of over 7,000 marine bacteria and fungi which are referred to as the Aquapharm Biodiscovery Ltd Library.

4.1.1 Background

The microbial strains from the Aquapharm Biodiscovery Ltd Library were previously assayed for dehalogenase and bromoperoxidase activity. Microbial strains which tested positive for dehalogenase and bromoperoxidase activity were identified using 16S rRNA PCR by Aquapharm Biodiscovery Ltd. The closest species to the two strains which expressed most dehalogenase activity were identified as having 99.7% and 99.0% sequence identity to *Rhodobacteraceae* family bacteria (AQP5750) and *P. arctica* (AQP4626) respectively. AQP5750 was isolated from a Polychaeta (tube worm) from Tralee beach, Argyll, UK and AQP4626 was isolated from a sponge of unknown species from Linnhe Marine, Argyll, UK. AQP5750 and AQP4626 tested positive for L-haloacid dehalohenase activity only, with activity towards chloroacetic acid and l-chloropropionic acid. The genome sequences for AQP5750 and AQP4626 have not yet been determined.

4.1.2 Determination of dehalogenase protein sequences

Milligram quantities of protein are needed for crystallization and biochemical characterisation of the respective proteins. In order to obtain sufficient quantities of the haloacid dehalogenase, the gene needs to be cloned and over-expressed in a bacterial expression system. To clone the genes of interest ideally the DNA sequences need to be determined. However, the semi-purified proteins from the wild type bacteria can be analysed by mass spectrometry, allowing identification of the protein and some protein sequence information to be determined. This

information can be used to design degenerate PCR primers so the gene of interest can be amplified.

4.1.3 Mass spectrometry

Mass spectrometry is a quantitative and qualitative method used to determine the mass to charge ratio of ionized molecules. Mass spectrometry applications include accurate determination of the mass and the elucidation of the composition of molecules. Shotgun proteomics refers to the method of combining liquid chromatography with mass spectrometry to analyse complex protein samples. This method has been successfully used to determine unknown protein sequences (Marcotte, 2007).

'Bottom up' proteomics is one method of elucidating the protein sequences (Loo *et al.,* 2005). This approach involves purifying the protein of interest from the wild type organism. The semi-purified protein samples are then analysed on an SDS-PAGE gel. The protein bands of interest are extracted from the gel and enzymatically digested at sequence specific locations into smaller peptides. The peptide mixtures are analysed on a LC quantitative time of flight (Q-TOF) mass spectrometer (Henzel *et al*., 1993). The mass to charge ratios for peptides detected are then compared to a database of theoretically digested proteins from homologous species.

Tandem mass spectrometry (MS/MS) involves fragmenting the peptide ions to help with primary sequence elucidation. Peptides generated *in silico* with proteins from homologous species are also fragmented, so the experimental MS/MS product ion scans are matched against these. If the proteins have high sequence similarity, the molecular mass of the resulting peptides can be determined by comparing the results to known databases. Once some primary protein sequence information has been elucidated, degenerate oligonucleotide primers can be designed to amplify the gene of interest.

In LC-MS the protein needs to be vaporised and ionized. Electrospray ionisation is used to disperse the protein sample as an aerosol consisting of very small highly charged droplets. The ionized molecules are then separated by an electromagnetic field and a quadruple mass analyzer is used to select ions with a specific mass to charge ratio. TOF mass spectrometry operates by applying an electromagnetic field of a set strength causing the velocity of ions of the same charge to vary according to their mass. Accurate determination of the peptide sizes can then be calculated by measuring the time they take to reach the detector.

This chapter describes the purification of the L-haloacid dehalohenases from wild type bacteria AQP4626 and AQP5750, with the aim to elucidate amino acid sequence information for the proteins using peptide mass spectrometry.

4.2 Materials and Methods

4.2.1 Growth and purification of the L-haloacid dehalohenase from AQP4626 and AQP5750

Wild type AQP4626 and AQP5750 were cultured (section 2.1.3) at 22ºC and the bacterial paste harvested (section 2.1.6). Cells were resuspended in HIC buffer A (Table 2.7) and bacterial cells lysed (section 2.3.2). The soluble protein fractions were subjected to ammonium sulfate precipitation trials (section 2.3.3.3) and the fractionation of the L-haloacid dehalohenase at set ammonium sulfate concentrations was determined. The protein was then purified by ammonium sulfate precipitation (section 2.3.3.4), HIC (section 2.3.3.5), anion exchange chromatography (section 2.3.3.6) and GF chromatography (section 2.3.3.7) with a Superdex 200 chromatography column. Protein elution peaks from each chromatography column were dialysed (section 2.4.3) into GF buffer (Table 2.7) and assayed for dehalogenase activity (section 2.3.8.2) with chloroacetic acid as the substrate. Throughout the purification process, fractions that tested positive for L-haloacid dehalohenase activity were analysed by SDS-PAGE (section 2.3.6). The fractions corresponding to the active protein from the GF chromatography column were concentrated using a 10 kDa membrane filter

(Vivaspin 20; Viva science), at 3,000 x *g*, at 4ºC until the final volume reached 20 µl.

4.2.2 Protein mass spectrometry

Partially purified and concentrated haloacid dehalogenase from wild type bacteria AQP4626 and AQP5750 were analysed on a NuPAGE gel (section 2.3.6). Protein bands between 20-40 kDa were excised from the NuPAGE gel with a sterile scalpel. Protein bands were prepared for mass spectrometry analysis and digested with trypsin (section 2.3.7). Peptide samples were analysed using the 'in house' LC Q-TOF mass spectrometer (Agilent).

4.3 Results

4.3.1 Purification of the L-haloacid dehalohenase from AQP4626

Soluble protein from AQP4626 was initially subjected to ammonium sulfate fractionation trials. The L-haloacid dehalohenase precipitated at 60% saturation (Figure 4.1). To provide sufficient amounts of L-haloacid dehalohenase for peptide mass spectrometry, it was necessary to purify the enzyme from 12 L of wild type AQP4626. The crude protein extract was initially purified by removing all proteins that precipitated below 50% ammonium sulfate. The remaining protein was purified using HIC with a Phenyl Sepharose chromatography column (Figure 4.2). Fractions 64-77 (320-385 ml) from the HIC column assayed positive for dehalogenase activity. These fractions were analysed using SDS-PAGE (Figure 4.3). All active fractions from the HIC column were pooled together and purified by anion exchange chromatography with a FFQ chromatography column (Figure 4.4). Fractions 54-58 (270-290 ml) from the FFQ chromatography column assayed positive for activity and were analysed using SDS-PAGE (Figure 4.5). All active fractions from the FFQ chromatography were pooled together, concentrated to 1 ml and purified by GF chromatography using a Superdex 200 chromatography column (Figure 4.6). Fractions 74-82 from the GF chromatography column assayed positive for activity (Figure 4.7) and were analysed on a NuPAGE gel (Figure 4.8).

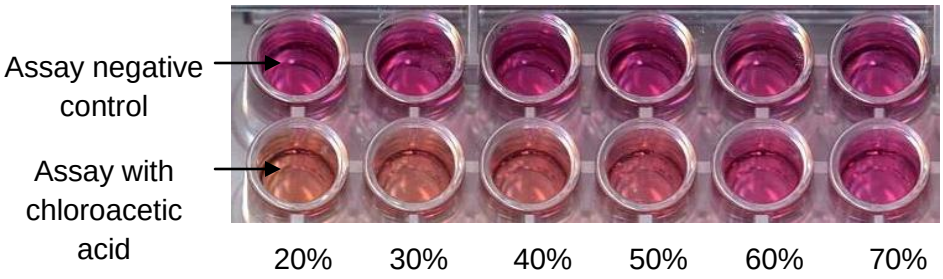


Figure 4.1- Activity assay plate from the ammonium sulfate fractionation trials showing the AQP4626 L-haloacid dehalohenase precipitating at 60% saturation. Activity in the soluble fraction is detected by a colour change from red to yellow.

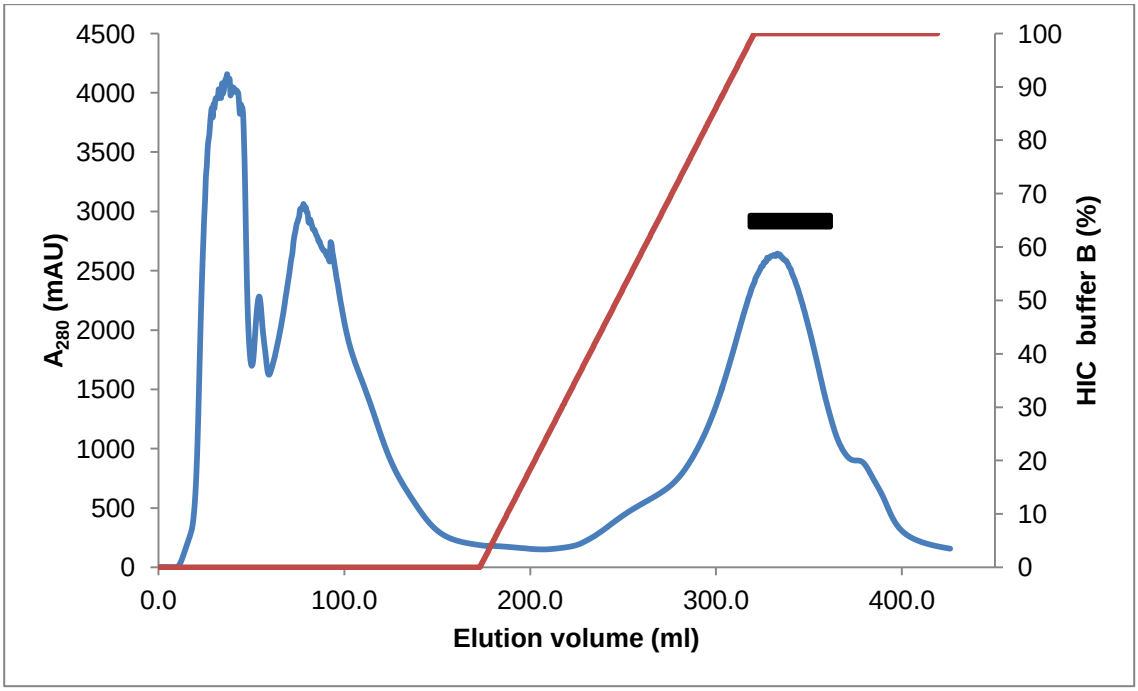


Figure 4.2- Elution profile from the HIC column. The AQP4626 L-haloacid dehalohenase eluted between 320-385 ml (fractions 64-77) which is highlighted by the black bar.

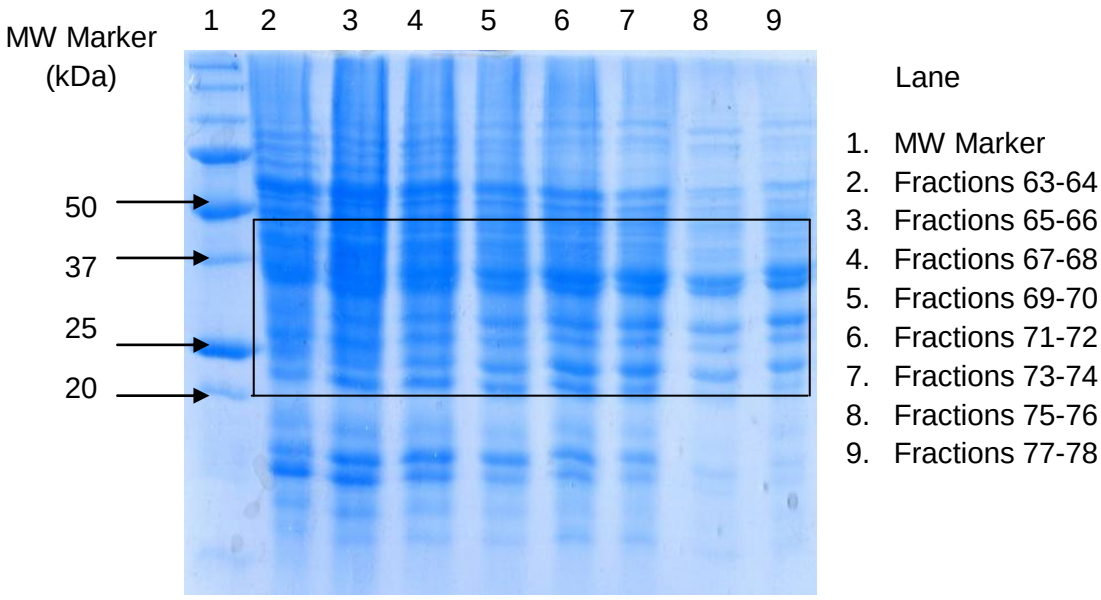


Figure 4.3- SDS-PAGE gel of fractions 66-77 that tested positive for AQP4626 L-haloacid dehalohenase activity, purified by HIC. Protein bands of interest between 20-45 kDa are highlighted by the black box.

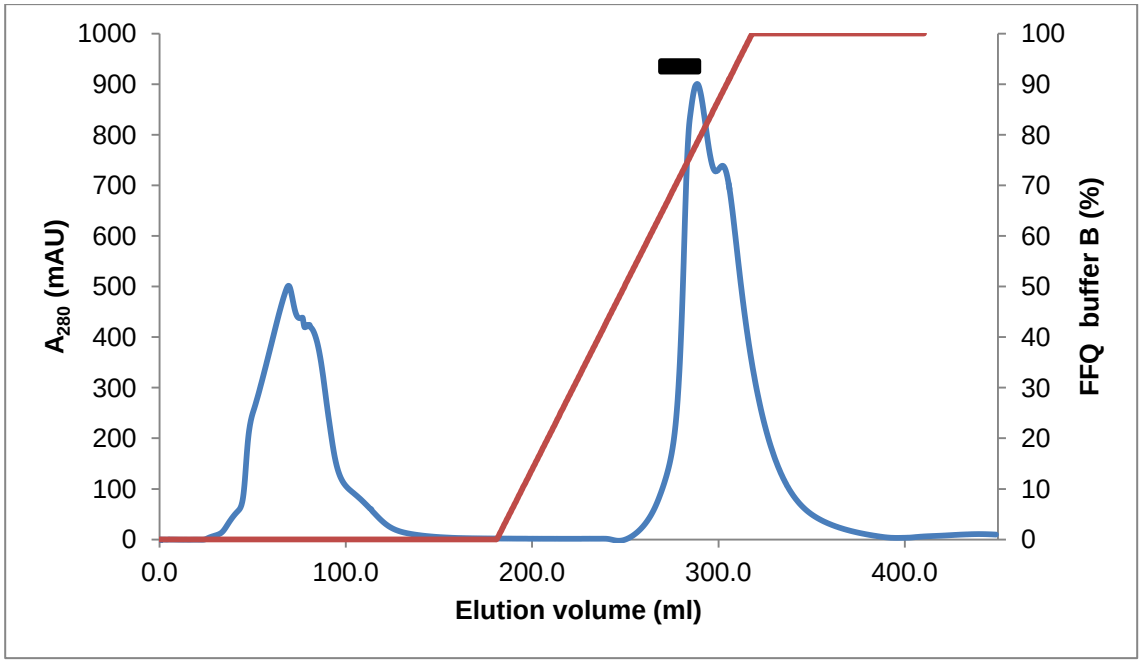


Figure 4.4- Elution profile from the FFQ chromatography column. The AQP4626 L-haloacid dehalohenase eluted between 270-290 ml (fractions 54-58) which is highlighted by the black bar.

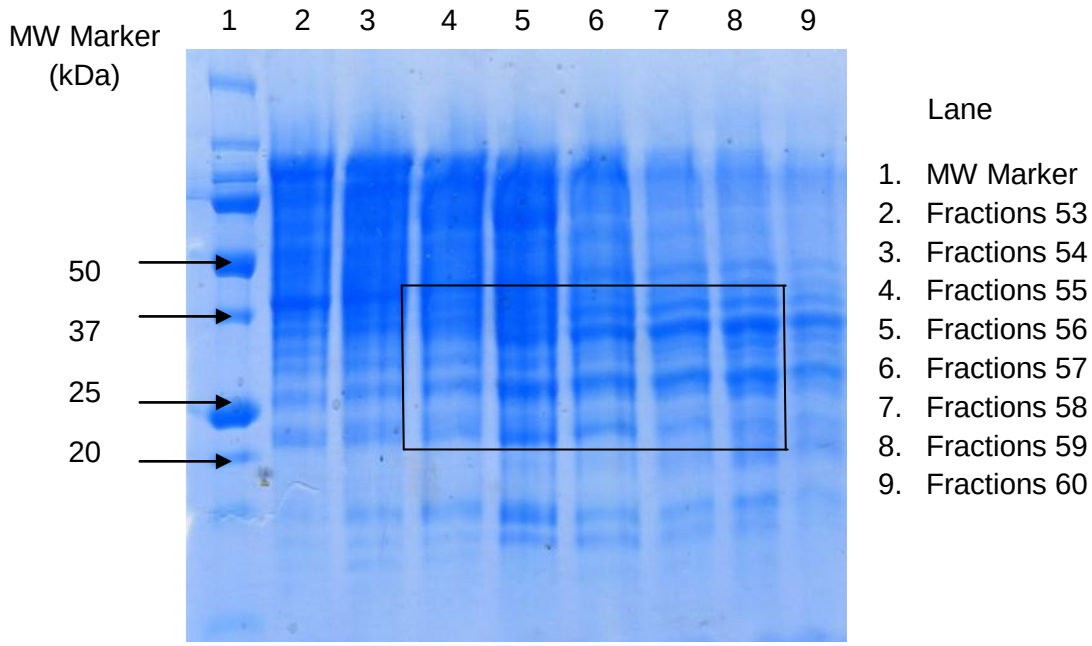


Figure 4.5- SDS-PAGE gel of fractions 54-58 that tested positive for AQP4626 L-haloacid dehalohenase activity, purified by FFQ chromatography. Protein bands of interest between 20-45 kDa are highlighted by the black box.

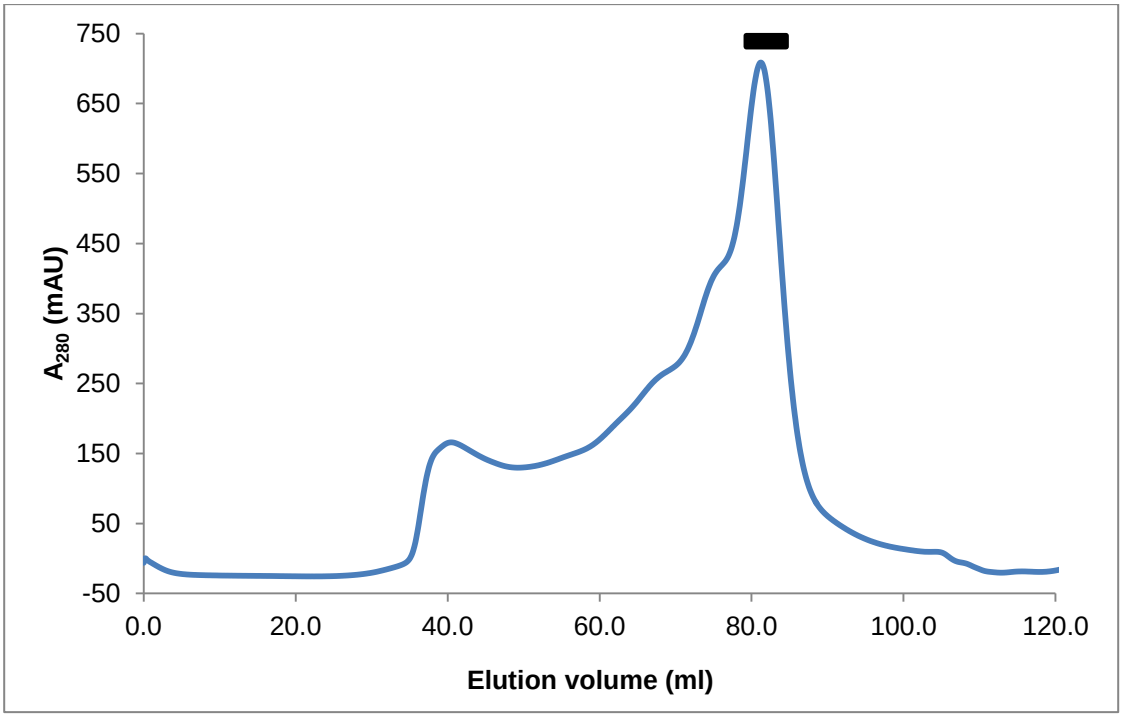
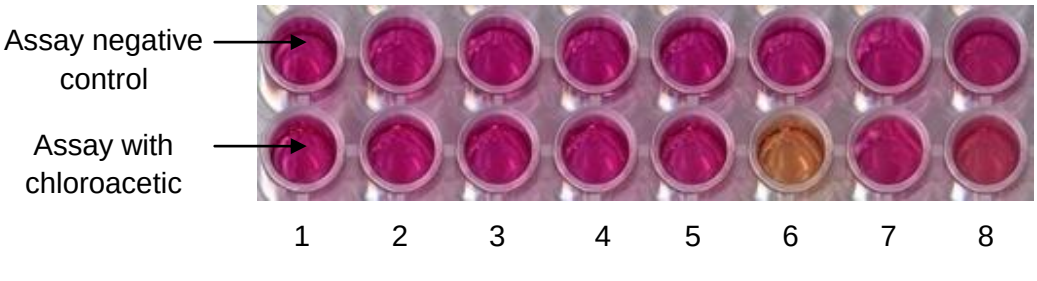


Figure 4.6- Elution profile from the Superdex 200 GF chromatography column. The AQP4626 L-haloacid dehalohenase eluted in fractions 74-82 which are highlighted by the black bar. 1 ml fractions were collected throughout the GF chromatography.



Lane

1. Fractions 30-38
2. Fractions 39-47
3. Fractions 48-56
4. Fractions 57-64
5. Fractions 65-73
6. Fractions 74-82
7. Fractions 83-90
8. Fractions 91-98

Figure 4.7- Activity assay plate showing the purified and concentrated fractions 30-98 from the Superdex 200 GF chromatography column with the AQP4626 L-haloacid dehalohenase. L-haloacid dehalohenase activity was detected in fractions 74-82 which can be visualised by a colour change from red to yellow.

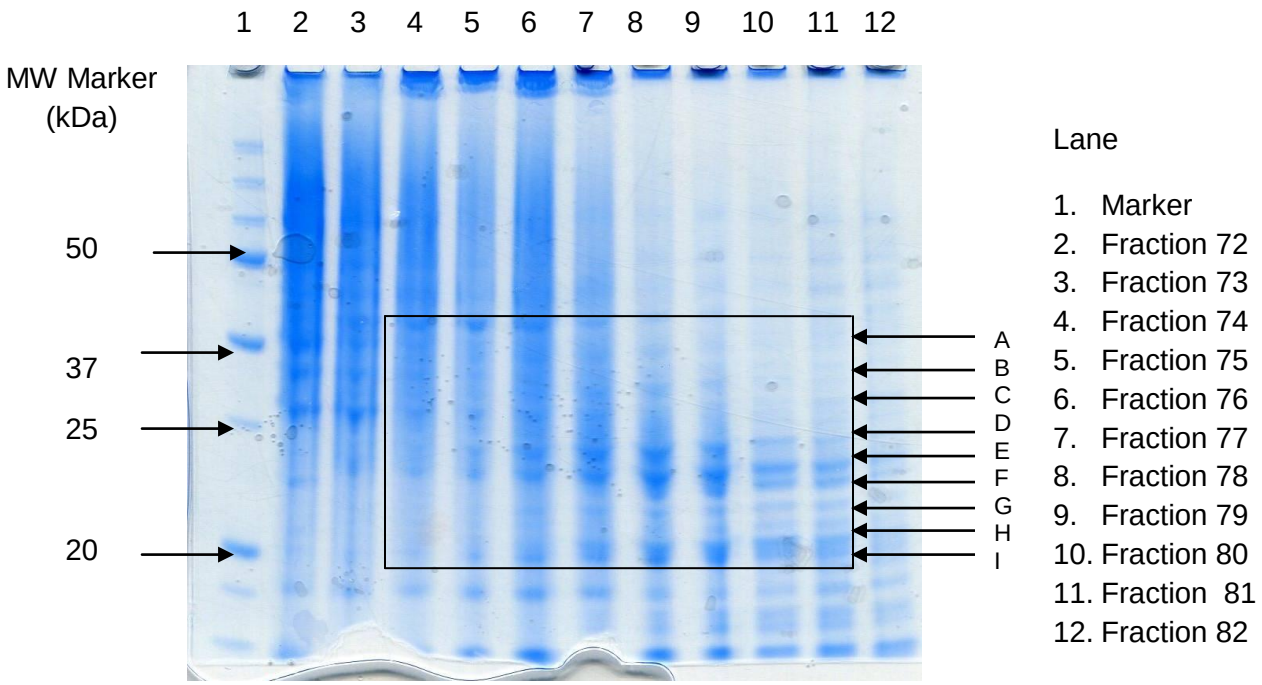


Figure 4.8- NuPAGE gel of concentrated fractions 74-82 containing the partially purified AQP4626 L-haloacid dehalohenase from the Superdex 200 GF chromatography column. All 9 protein bands labelled A-I, between 20-45 kDa from the NuPAGE were sent for peptide mass spectrometry. Protein bands were cut out from the gel across several lanes which are highlight by the black box.

4.3.2 Purification of the L-haloacid dehalohenase from AQP5750

Soluble protein from AQP5750 was initially subjected to ammonium sulfate fractionation trials. The L-haloacid dehalohenase precipitated at 60% saturation (Figure 4.9). To provide sufficient amounts of L-haloacid dehalohenase for peptide mass spectrometry, it was necessary to purify the enzyme from 12 L of wild type AQP5750. The crude protein extract was initially purified by removing all proteins that precipitated below 50% ammonium sulfate. The remaining protein was purified using HIC with Phenyl Sepharose chromatography column (Figure 4.10). Fractions 68-77 (340-385 ml) from the HIC column assayed positive for activity. These fractions were analysed using SDS-PAGE (Figure 4.11). All active fractions from the HIC column were purified by anion exchange chromatography using a FFQ chromatography column (Figure 4.12). Fractions 10-14 (50-70 ml) from the FFQ chromatography column assayed positive for

activity and were analysed using SDS-PAGE (Figure 4.13). All active fractions from the FFQ chromatography column were pooled together, concentrated to 1 ml and purified by GF chromatography using a Superdex 200 chromatography column (Figure 4.14). Fractions 78-84 off the GF chromatography column tested positive for L-haloacid dehalohenase activity (Figure 4.15) and were analysed on a NuPAGE gel (Figure 4.16).

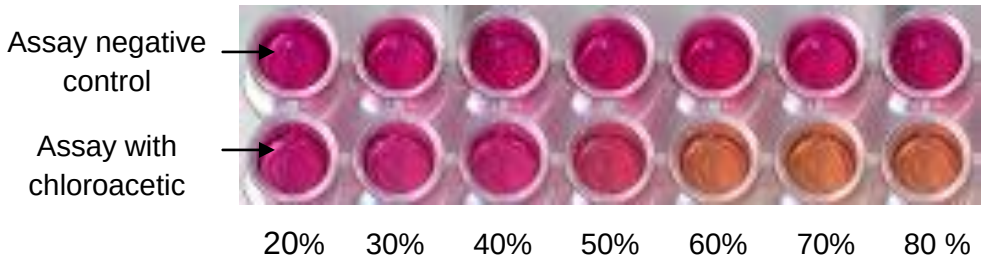


Figure 4.9- Activity assay plate from the ammonium sulfate fractionation trials showing the AQP5750 L-haloacid dehalohenase precipitating at 60% saturation. Activity in the insoluble fraction is detected by a colour change from red to yellow.

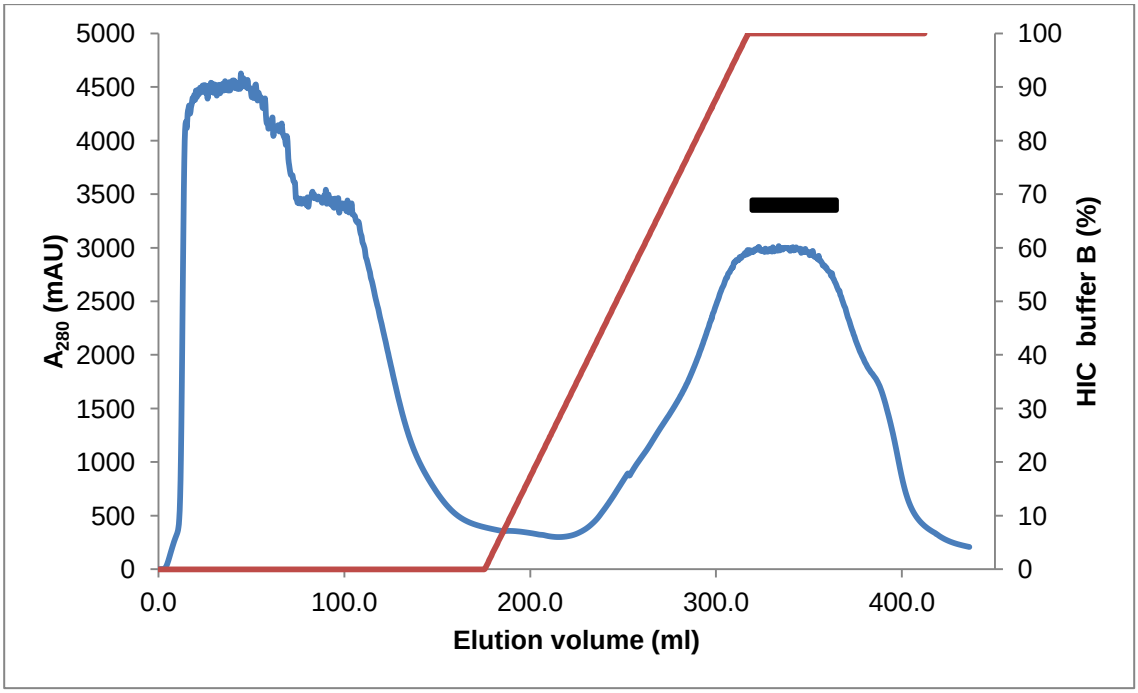


Figure 4.10- Elution profile from the HIC column. The AQP5750 L-haloacid dehalohenase eluted between 340-385 ml (fractions 68-77) which are highlighted by the black bar.

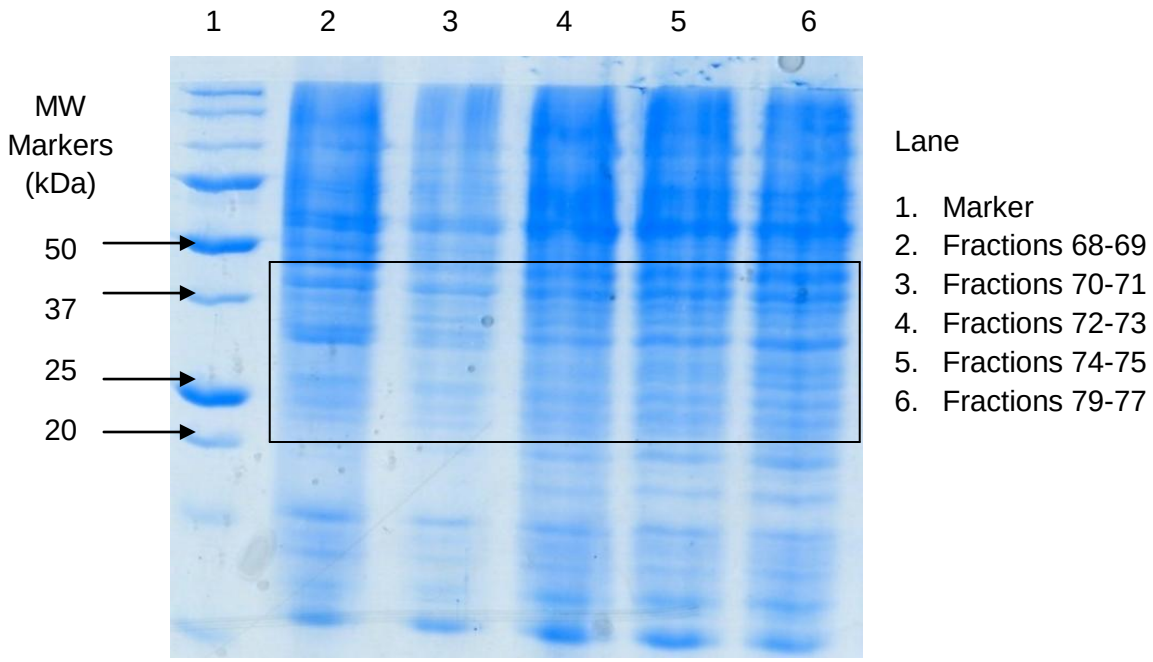


Figure 4.11- SDS-PAGE gel of fractions 68-77 that tested positive for AQP5750 L-haloacid dehalohenase activity, purified by HIC. Protein bands of interest between 20-45 kDa are highlighted by the black box.

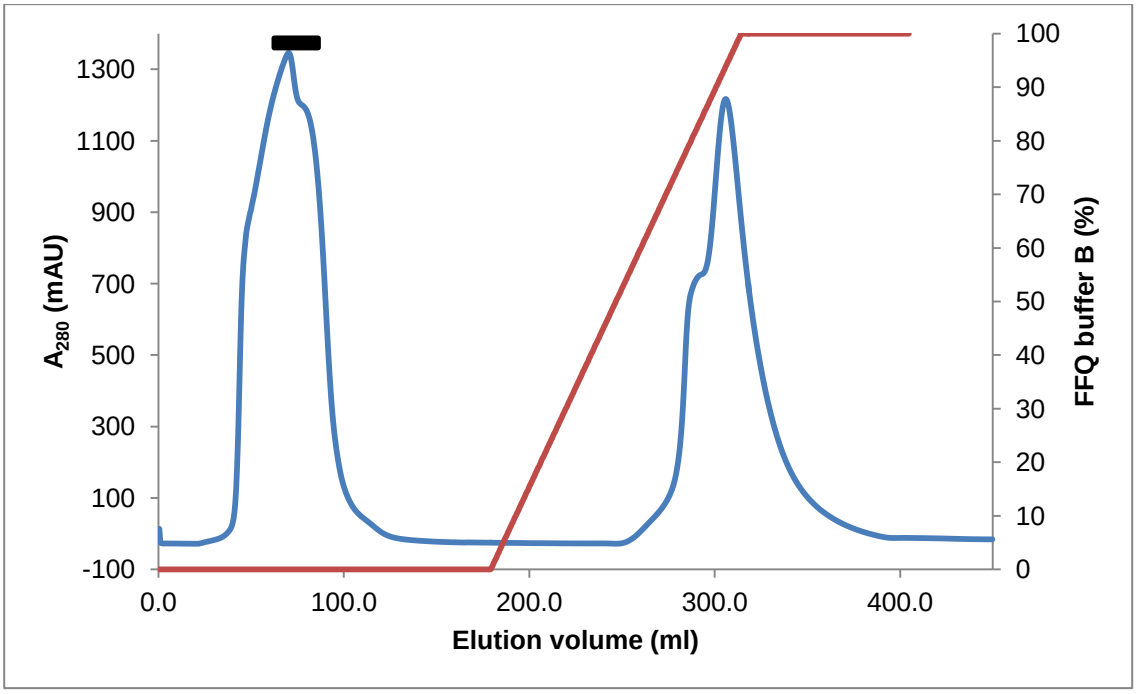


Figure 4.12- Elution profile from the FFQ chromatography column. The AQP5750 L-haloacid dehalohenase eluted between 50-70 ml (fractions 10-14) which are highlighted by the black bar.

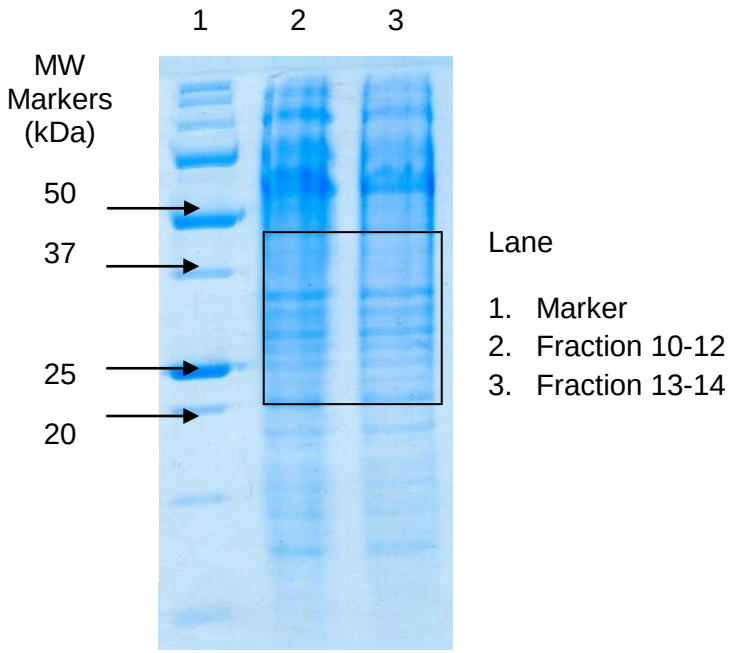


Figure 4.13- SDS-PAGE gel of fractions 10-14 that tested positive for AQP5750 L-haloacid dehalohenase activity, purified by FFQ chromatography. Protein bands of interest between 20-45 kDa are highlighted by the black box.

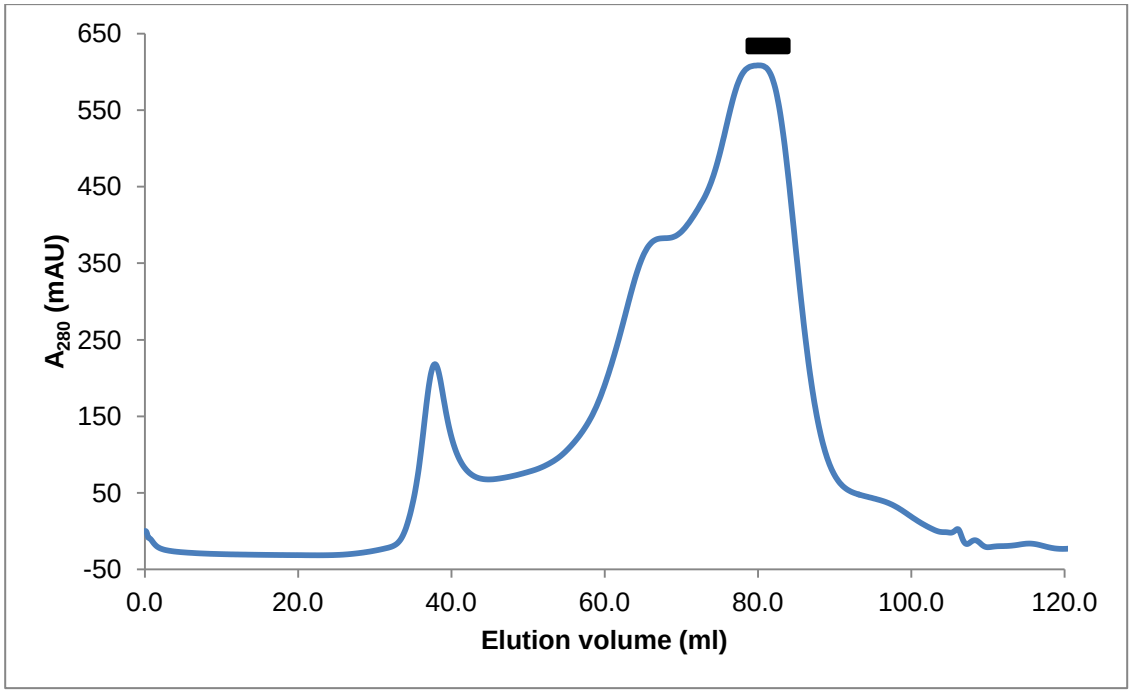


Figure 4.14- Elution profile from the Superdex 200 GF chromatography column. The AQP5750 L-haloacid dehalohenase eluted in fractions 78-84 which are highlighted by the black bar. 1 ml fractions were collected throughout GF chromatography.

Figure 4.15- Activity assay plate showing the partially purified and concentrated fractions 57-105 from the Superdex 200 GF chromatography column with the AQP5750 L-haloacid dehalohenase. L-haloacid dehalohenase activity was detected in fractions 78-84 which can be visualised by a colour change from red to yellow.

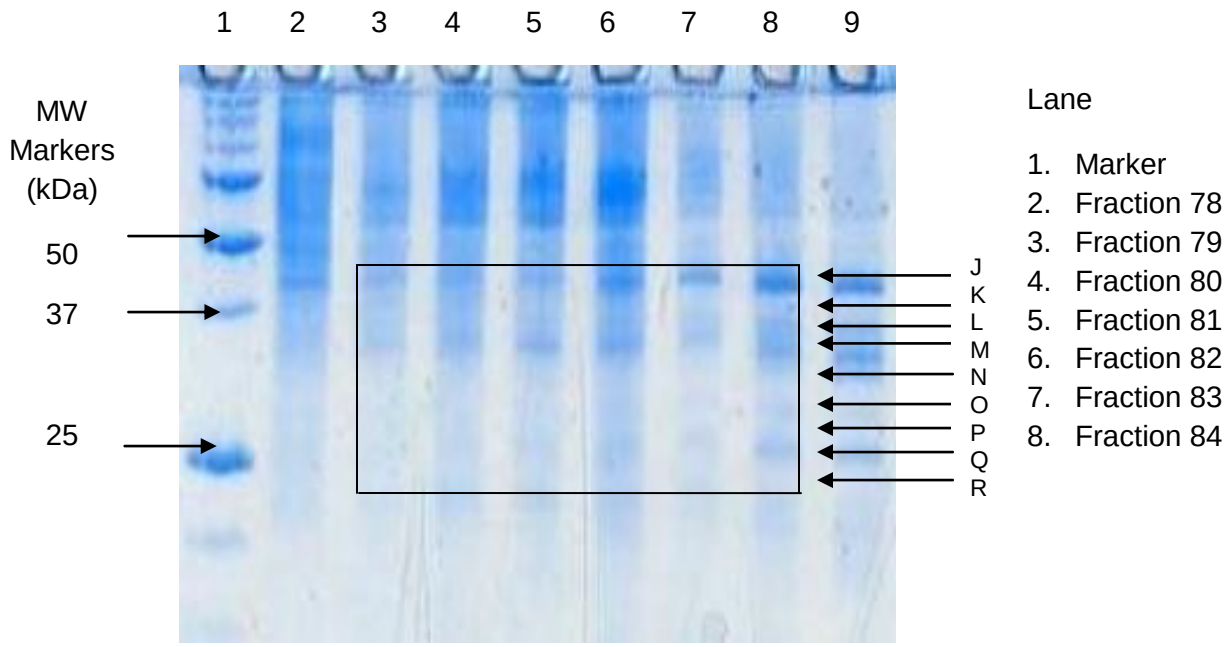


Figure 4.16- A NuPAGE gel of concentrated fractions 78-84 containing partially purified AQP5750 L-haloacid dehalohenase from the Superdex 200 GF chromatography column. All 9 protein bands labelled J-R between 20-45 kDa from the NuPAGE were sent for peptide mass spectrometry. Protein bands were cut out from the gel across several lanes which are highlighted by the black box.

4.3.3 Peptide mass spectrometry

All characterised L-haloacid dehalohenases to date are between 23-30 kDa. To allow for variations in protein sizes, all bands from the partially purified AQP4626 and AQP5750 samples between 20-45 kDa were excised from a NuPAGE gel. The protein samples were digested into peptides with trypsin (which cleaves at the C-terminal end of lysine and arginine, unless a proline follows either of the residues). The peptide samples were analysed on a LC Q-TOF mass spectrometer. Mass spectrometry results were processed using the University of California protein prospector (Baker, P.R. and Clauser, K.R. http://prospector.ucsf.edu).

4.3.3.1 Mass spectrometry with the partially purified proteins from AQP4626

The final NuPAGE gel, after the last stage of purification was taken for further analysis using peptide mass spectrometry. A total of 9 bands (A, B, C, D, E, F, G H and I) were visualised on a NuPAGE gel between 20-45 kDa (Figure 4.8). These bands are highlighted by the arrows and the black box indicated the several lanes which the protein bands were taken from. The mass spectrometry results showed bands A, B, C, D, E, F, G, H had between 5-42% sequence identity to adenylate kinase proteins from *Psychromonas, Vibro* and *Burkholderia* species (Table 4.1). Band I showed 23% identity to a DNA-binding response regulator ChvI from *Roseobacter denitrificans* (Table 4.1). The presence of several bands in the SDS-PAGE gel all showing identity towards adenylate kinases is thought to be due to possible degredation of the protein. This is despite the presence of a serine protease inhibitor (PMSF), a metal chelating inhibitor (EDTA) and a trypsin-like enzyme inhibitor (BAM) in the purification buffers. Some proteins may still have been susceptible to degradation by cysteine and aspartic acid proteases. The mass spectrometry results with the partially purified proteins from AQP4626 did not indicate the presence of an L-haloacid dehalohenase.

Sample	Protein	Species	Sequence identity (%)
A	adenylate kinase	*P. ingrahamii* 37	28
B	adenylate kinase	*Vibrio* sp.	42
C	adenylate kinase	*Vibrio* sp.	28
D	adenylate kinase	*Psychromonas* sp. CNPT3	22
E	adenylate kinase	*P. ingrahamii* 37	19
F	adenylate kinase	*P. ingrahamii* 37	18
G	adenylate kinase	*Psychromonas sp.*	5
H	adenylate kinase	*Burkholderia cenocepacia* HI2424	5
I	DNA-binding response regulator ChvI, putative	*R. denitrificans*	23

Table 4.1- Proteins identified from the LC Q-TOF mass spectrometry results with the partially purified protein samples from AQP4626. Information includes sample label, protein identified, species that the protein was identified from and sequence identity.

4.3.3.2 Mass spectrometry with the partially purified proteins from AQP5750

In the partially purified protein sample from AQP5750 a total of 9 bands (J, K, L, M, N, O, P, Q and R) were visualised on a NuPAGE gel between 20-45 kDa (Figure 4.16). These bands are highlighted by the arrows and the black box indicated the several lanes which the protein bands were taken from. Identity to 5 different proteins from Proteobacteria were observed (Table 4.2). Protein band J at 40 kDa had 4% amino acid sequence identity to a putative protein from Alteromonadales sp. Protein bands K, L and M at 35 kDa, 32 kDa and 27 kDa had 23%, 8% and 15% amino acid sequence identity to a DNA-binding response regulator ChvI protein in *Rhodobacteraceae* family bacteria. Protein bands N and O at 24 kDa and 23 kDa had 3% and 4% amino acid sequence identity to a HAD family hydrolase from a *Rhodobacteraceae* and *Hyphomonadaceae* species. This hydrolase is in the same HAD superfamily of proteins as L-haloacid dehalohenases. ATPases, epoxide hydrolases and a number of phosphatases are also in the HAD superfamily. All these enzymes contain a conserved aspartic acid residue in the catalytic triad. Figure 4.17 shows the amino acid sequence alignment of the HAD-like hydrolase and an L-haloacid dehalohenase from *P. ingrahamii*, with the conserved aspartic acid residue highlight in red. The hydrolases from *Rhodobacteraceae* and *Hyphomonadaceae* were aligned to L-haloacid dehalohenases from homologous species. The results showed low homology with a sequence alignment of 13% or less. Protein band P at 22 kDa had a 10% amino acid sequence identity to a transporter solute receptor protein from *Pseudoalteromonas atlantica*. Protein bands Q and R at 21 kDa and 20 kDa had 13% and 10% amino acid sequence identity to a C4-dicarboxylate-binding periplasmic protein in *Vibro* sp. As previously discussed above, multiple protein bands visualised at different sizes that are identified as the same protein are likely to be a result of protein degradation. The mass spectrometry results with the partially purified proteins from AQP5750 did not reveal the identity of the L-haloacid dehalohenase.

Sample	Protein	Species	Sequence identity (%)
J	putative enzyme	*Alteromonadales* bacterium-7	4
K	DNA-binding response regulator ChvI, putative	*Roseobacter denitrificans* OCH-114	23
L	DNA-binding response regulator ChvI	*Ruegeria pomeroyi*	8
M	DNA-binding response regulator ChvI	Rhodobacterales bacterium HTCC2150	15
N	HAD-superfamily hydrolase	*Ruegeria pomeroyi*	3
O	HAD-superfamily hydrolase,	*Pelagibaca bermudensis*	4
P	TRAP transporter solute receptor	*Pseudoalteromonas atlantica*	10
Q	C4-dicarboxylate-binding periplasmic protein	*Vibrio parahaemolyticus*	13
R	C4-dicarboxylate-binding periplasmic protein	*Vibrio parahaemolyticus*	10

Table 4.2 - Proteins identified from the LC Q-TOF mass spectrometry results with the partially purified protein samples from AQP5750. Information includes sample label, protein identified, species that the protein was identified from and sequence identity.

```
sequence1       MSVILAFDVYGTLINTHGIVTILEKWLGKNAQAFSQTWRDKQLEYSFRRGLMQNYQSFAV 60
sequence2       MLKALFIDLSGVLYEGHNVIPGAVAAIKKARASQLQLRFVTNTSRRTRTQLLTDLQNLGF 60
                *   * :*: *.* : *.::.     : *   :  *    .: .   *  *: : *.:..

sequence1       CTRHALDYTCNFYRVSLSKEQ------------KQALMDGYKTLP------------AFD 96
sequence2       DLQKKELFTAPVAVHAWLQEKKLRPYCLIHHNIKSEFADLLQAMPNAVVIGDAEQNFCYD 120
                  ::   :*. .   :  :*:            *. : *  :::*            .:*

sequence1       DAANALAQLKGAGFKLYAFSNGTKEAVEELLTGAG-IRAFFDGIVSCDDLKSFKPNPAVY 155
sequence2       KLNRAFQLCQQGAVLVGIGYNRYFKLEGQLLLDAGPFIKAIEFAALTQAIIIGKPSKDFF 180
                .  .*:   : ... :    *   :   :** .** :   ::  .  : :   **.  .:

sequence1       CHFLRESGAKHNQAWLISSNPF-DVIGAISTGMLSAWVQRSQDAIFDPWDVDPTAIITS- 213
sequence2       LQALASTGLSADQVLMIGDDIYGDIEGAINAGLLAGLVRTGKYQTGDEHKISAAHLTFNS 240
                 : * .:* . :*. :*..: : *: ***.:*:*:. *: .:    *  .:..: :  .

sequence1       -LKELKGVLPKPD-- 225
sequence2       IVDAVDYALSDQSNN 255
                 :. :. .*.. .
```

Figure 4.17- ClustalW amino acid sequence alignment between the L-haloacid dehalohenase (sequence 1) and the HAD-like hydrolase (sequence 2) from *P. ingrahamii*.

lar amino acid.

4.4 Discussion

The mass spectrometry results indicate that there was no amino acid sequence identity towards L-haloacid dehalohenases. MS/MS was also performed using target MS/MS to help identify peptides with a given mass. Targeted MS/MS did not improve results and peptides detected showed no homology to L-haloacid dehalohenases.

The 'Bottom up' proteomics approach relies on high sequence homology between the unknown protein samples and proteins in the database used for analysis. Biochemically and structurally characterised L-haloacid dehalohenase from *Xanthobacter autotrophicus, S. tokodaii, B. cepacia* and *Pseudomonas* sp. YL have conserved amino acid residues required for catalysis and motifs which are highly similar. Despite taking this into consideration, it is possible that the L-haloacid dehalohenases in AQP5750 and AQP4626 are significantly different and may not have high enough sequence identity to L-haloacid dehalohenases in the homologous species.

Although the L-haloacid dehalohenases were partially purified using 4 different techniques, many protein bands were visualised in the final samples. The L-haloacid dehalohenase activity was detectable but protein concentrations may have been too low to be visualised on an SDS-PAGE gel. Both of these reasons could account for the negative results observed.

Chapter 5- Determination of L-haloacid dehalohenase gene sequences- the molecular approach

5.1 Introduction

5.1.1 Determination of L-haloacid dehalohenase gene sequences by degenerate PCR and genome sequencing.

Milligram quantities of pure protein are needed for crystallization and biochemical characterisation. In order to obtain sufficient quantities of protein, the L-haloacid dehalohenase gene would potentially need to be cloned and over-expressed in a bacterial expression system. Proteins from related species often have high sequence identity. This information can be used to design degenerate oligonucleotide primers in order to amplify the gene using degenerate PCR. Successful amplification of the gene of interest relies on high sequence identity between the DNA sequences used to design primers and the target gene sequence. Degenerate PCR has previously been used for the identification and analysis of a reductive dehalogenase gene (Holscher *et al*., 2004).

In the last 5 years genome sequencing has become widely accessible due to the reduction in cost and the development of sequencing techniques and instruments. The next generation sequencing platform on the Illumina GA2 allows high throughput 72 base pair (bp) reads. This gives high sequence coverage allowing genomes to be fully annotated. The Illumina GA2 uses sequencing by synthesis technology which has been employed by all next-generation sequencing platforms. This technology follows the activity of a polymerase as it synthesises DNA (which is attached to a solid surface) by detecting nucleotides with fluorescent labels. For *de novo* sequence assembly the genomic DNA is extracted and cut into fragments. The DNA fragments are then ligated to short nucleotide adapter sequences and amplified by PCR. The genomic DNA fragments with adapter sequences on both ends are then amplified by bridge PCR to create clonal clusters of each single DNA strand. This amplifies the fluorescent signal in the base by base sequencing system.

AQP5750 and AQP4626 strains from the Aquapharm Biodiscovery Ltd Library tested positive for L-haloacid dehalohenase activity as decribed in chapter 4. This chapter describes the molecular approach to determining the L-haloacid dehalohenase gene sequences.

5.2 Materials and methods

5.2.1 Bioinformatics

The National Center for Biotechnology Information (NCBI) database was used to identify proteins present in bacterial species and the BLAST used to identify and analyse homologous DNA and protein sequences (Altschul *et al*., 1990). ClustalW, European Bioinformatics Institute was used for multiple sequence alignment (http://www.ebi.ac.uk/Tools/msa/clustalw2). Velvet (Zerbino *et al*., 2008) was used for genomic *de novo* sequence assembly (designed to remove errors when possible and assemble short DNA sequencing reads into contigs).

5.2.2 Degenerate oligonucleotide primer design

Where possible, degenerate primers were designed on the beginning, end and conserved regions of L-haloacid dehalohenase protein sequences.

5.2.3 Degenerate oligonucleotide primer design to amplify the L-haloacid dehalohenase in AQP5750

Two methods were employed when designing degenerate primers. The first method used the L-haloacid dehalohenase gene sequences from the 5 closest related bacteria to AQP5750 (*Jannaschia* sp. CCS1, *Dinoroseobacter shibae* DFL 12, *Roseobacter denitrificans* OCh 114, *Ruegeria* sp. TM1040 and *Paracoccus denitrificans* PD1222). The second method employing degenerate primers used the 5 closest homologues to the L-haloacid dehalohenases from the closest species to AQP5750 (*Jannaschia* sp. CCS1, *Roseovarius nubinhibens* ISM, Rhodobacterales bacterium HTCC2150, *Roseobacter* sp. CCS2 and *Roseobacter* sp. AzwK-3b). All primers designed were at least 14 bp long and were no more than 256 fold degenerate (because of these restrictions

only certain primer combinations could be used). The different combinations of wobble bases used in degenerate PCR primers are shown in Appexdix, 11.7. Forward and reverse degenerate primers for the amplification of the L-haloacid dehalohenase gene are shown in Table 5.1.

Primer	**Sequence**
Forward 1	5’- A T G M C C A W M A Y S A C C W S M -3’
Forward 2	5’- A T G R C C A W W A W G A C M W S M -3’
Forward 3	5’- A T G A C S A W M A W R A C C W S -3’
Forward 4	5’- A T G M C M A W M A Y R A C C W S -3’
Forward 5	5’- A T G W C C A W W A W R R Y -3’
Forward 6	5’- A T G A Y S A W R M Y G A Y C A C-3’
Forward 7	5’- A T G R C Y A W M A Y R A C C W -3’
Forward 8	5’- G C Y T A T GG Y T B T T Y G A Y G T B -3’
Reverse 1	5’- T Y A Y R Y Y C C Y R C A A K -3’
Reverse 2	5’- T C A C R S M C C S A M M A S S -3’
Reverse 3	5’- T C A S S S C S C C R C M A G -3’
Reverse 4	5’- T C A S G C M M S C W C M R G Y -3’
Reverse 5	5’- T C A S S C C M C C R Y M A G C K -3’
Reverse 6	5’- A A Y G G Y G G G A Y G C S G S G S V -3’

Table 5.1- Degenerate PCR primers designed to amplify the L-haloacid dehalohenase gene from AQP5750.

5.2.4 Degenerate oligonucleotide primer design to amplify the L-haloacid dehalohenase in AQP4626

Two methods were employed when designing degenerate primers. The first method used L-haloacid dehalohenases gene sequences from the 5 closest related bacteria to AQP4626 (*P. ingrahamii* 37, *Vibrio vulnificus*, *Vibrio* sp. Ex25, *Colwellia psychrerythraea* 34H and *Vibrio vulnificus* MO6-24/O). The second method employed to design degenerate primers used the 5 closest homologues to the L-haloacid dehalohenase from the closest species to AQP4626 (L-haloacid dehalohenase sequences from *P. ingrahamii, Geobacter lovleyi* SZ, *Sideroxydans lithotrophicus* ES-1, *Chlorobium phaeobacteroides* BS1 and *Prosthecochloris aestuarii* DSM 271). Forward and reverse degenerate primers for the amplification of the L-haloacid dehalohenase gene are shown in Table 5.2. Primers based on the L-haloacid dehalohenase from *P. ingrahamii* 37 only were also designed (Table 5.3) and tested.

Primer	**Sequence**
Forward 1	5’- A T G W M A K T K W W W C Y -3’
Forward 2	5’- A T G W M A S K R A Y W M T -3’
Forward 3	5’- A T G R M A W T Y W M K C K C -3’
Forward 4	5’- A T G A M R W T K W M Y C K -3’
Reverse 1	5’- T T A R W M M R G Y K T G R -3’
Reverse 2	5’- T T A R T C M Y S Y T T K K S -3’
Reverse 3	5’- T Y A S T C Y G R Y T T G S S T -3’

Table 5.2- Degenerate PCR primers designed to amplify the L-haloacid dehalohenase gene from AQP4626.

Forward primer	5'- A T G T C A G T G A T A C T T G C T T T T G A T G -3'
Reverse primer	5'- T T A G T C C G G CT T G G G T A G C -3'

Table 5.3- PCR primers designed to amplify the L-haloacid dehalohenase gene from AQP4626 based on the L-haloacid dehalohenase gene sequence from *P. ingrahamii 37*.

5.2.5 Degenerate PCR compositions and programs

All degenerate PCR reactions with AQP4626 and AQP5750 were initially trialled using PCR composition 1 (section 2.2.4.3) with PCR program 1 (Table 2.2). All combinations of forward and reverse primers were tested. To try and improve the results the annealing temperature range of unsuccessful PCR reactions were varied (45-55ºC, 55-65ºC and 65-75ºC). The PCR reaction composition was also varied by altering the amounts of Deep Vent polymerase (0.2-1 units), $MgSO_4$ concentration (5-40 mM), dNTPs concentration (100-300 µM), DNA template concentration (50-300 ng), primer concentration (0.1-0.3 µM) and DMSO concentration (0-4%). PCR reactions were run on a 1% agarose gel (section 2.2.2) to visualise amplified DNA. DNA bands were extracted (section 2.2.3) and sent for Sanger sequencing with the primers used in the PCR.

5.2.6 Genome sequencing of AQP5750

The manufacturer's instructions for the preparation of genomic DNA for sequencing using the Illumina GA2 were followed. Genomic DNA from AQP5750 was extracted (section 2.2.3) and the DNA sheared into fragments of less than 800 bp using the Bioruptor sonication method (7 x 1 min, with 7 x 1 min breaks in between). A T4 DNA polymerase was used to blunt the DNA fragments by removing the 3' overhangs and filling in the 5' overhangs. The DNA fragments were then prepared for ligation by adding an adenine (A) base to the 3' end of the DNA fragment. DNA fragments were ligated to the short adapter sequences which have a thymine (T) base on the 3' overhang. The ligated fragments were purified by gel extraction (section 2.2.6), size selected to obtain fragments of 500 bp and the DNA fragments attached to a flowcell. To increase DNA concentrations, fragments were amplified by PCR (using primers complementary

to the adapter sequences) with fluorescently labelled nucleotides and small clusters of thousands of identical DNA strands were formed. Complementary oligonucleotide probes were used to obtain 72 bp paired-end reads of the genomic DNA. *De novo* sequence assembly was then carried out and the data assembled into contigs using Velvet, version 0.7.63 (Zerbino *et al*., 2008). A parameter sweep of k-mer lengths and expected coverage values were used to optimise the N50 value. A k-mer length of 39, expected coverage of 43x and coverage cutoff parameter of 7.95 were used. An N50 value of 12929 bp was obtained with a total assembly size of 3.77 Mbp across 795 contigs.

5.3 Results and Discussion

5.3.1 Amplification of the L-haloacid dehalohenases from AQP4626 and AQP5750 using degenerate PCR

Amplification of the L-haloacid dehalohenase genes was attempted using degenerate PCR as previously described in sections 5.2.2, 5.2.3 and 5.2.4. Where possible, degenerate primers were designed on the beginning, end and conserved regions of L-haloacid dehalohenase protein sequences. These regions are highlighted in Figure 5.1 using the L-haloacid dehalohenase from *P. ingrahamii* as an example. Amplification of the L-haloacid dehalohenase genes was unsuccessful. Successful degenerate PCR relies on high sequence homology between the genes used for primer design and the gene to be amplified. Degenerate primers designed were between 128-256 fold degenerate resulting in a large mixture of oligonucleotide primer sequences. Low sequence identity and high primer degeneracy could account for the unsuccessful degenerate PCR reactions.

5.3.2 Genome sequencing

AQP5750 genomic DNA was extracted and the genome sequenced using the ‘in house’ Illumina GA2. *De novo* assembly was used to arrange the 72 bp paired-end reads into 1082 contigs. The online Galaxy/Exeter Bioscience bioinformatics server was used to search for dehalogenases within the genome using the NCBI BLAST tool (Altschul *et al*., 1990). An L-haloacid dehalohenase was identified

within the 4.5 Mbp genome. The L-haloacid dehalohenase gene sequence had a number of misreads in the middle of the gene, but the coverage was high enough to confirm that the protein had the main catalytic residue. An haloalkane dehalogenase was also identified within the 4.5 Mbp genome, This gene has had 100% coverage. The AQP5750 organism had tested positive for L-haloacid dehalohenase activity but haloalkane dehalogenase activity was not detected. This could indicate the haloalkane dehalogenase protein is only expressed under specific conditions. It is possible that certain haloalkanes present in the environment may 'switch on' the gene and induce expression of the protein. It is also possible that the haloalkane dehalogenase is expressed but the enzyme shows low activity towards the substrates tested.

The L-haloacid dehalohenase had 62% amino acid sequence identity to the L-haloacid dehalohenase from *Sulfitobacter* sp. EE-36 (Figure 5.1) and 22% sequence identity to the L-haloacid dehalohenase from *B. cepacia* (Figure 5.2). The *B. cepacia* L-haloacid dehalohenase has been biochemically characterised and the crystal structure solved at a resolution of 1.93 Å (Schmidberger *et al*., 2007). The L-haloacid dehalohenase gene sequences are considerably different to the L-haloacid dehalohenase gene sequences used to design degenerate PCR primers. This indicates that degenerate PCR is most likely to have failed because of low sequence homology between the genes used for primer design and the target gene to be amplified.

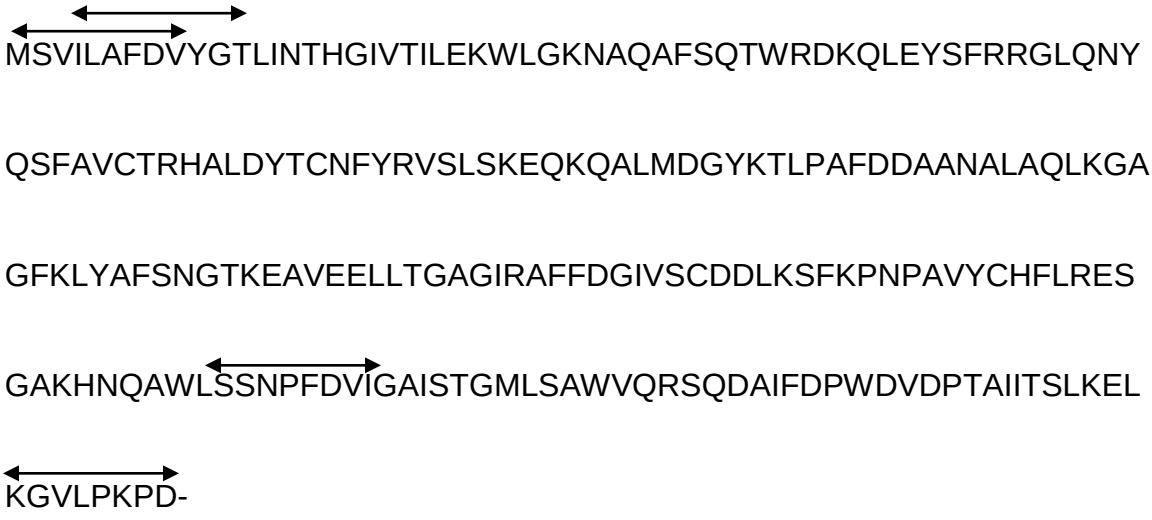


Figure 5.1- A diagram highlighting the beginning, end and some conserved regions of the *P. ingrahamii* L-haloacid dehalohenase. When possible, these regions were used to design degenerate primers.

```
sequence1       MTPSHPARPSRSGILVFDVNETLLDLTSLSPLFERVFGDAKVLREWFPELILYSQTLTLT 60
sequence2       ----MPDRTDFS-TIVFDVNETLLDIATLEPFFDRVFGNAAVLREWFPELILYSQTVTLS 55
                     * *.. *  :**********:::*.*:*:****:* ***************:**:

sequence1       GLYRPFGEIAAAVFEMVAANHQAKVTPDDIAELKTRLTSMPAYPDVAPALTRLQDAGFRL 120
sequence2       GLYAPFGDLAGGVLRMVGDNRNIAIQDSDVSELKSLIGSMPAHADVAPALSKLRDAGFTL 115
                *** ***::*..*:.**. *::  :  .*::***: : ****:.******::*:**** *

sequence1       VTLTNSAPSPAPSPLEKAGIASFFEAHLTVHSSQRFKPHPSVYDSTAETLGAKPEELCMI 180
sequence2       VTLTNSAPSPAPTPLEKAGIAALFDHHFSVAEVGHFKPHPSTYQMVADRLGVDVSSLCMV 175
                ************:********::*: *::* .  :******.*: .*: **.. ..***:

sequence1       ACHIWDTIGAQARGWRGGFVARPHNTPLTLAEVPQPDFIGRDMGELADQLIASLTA-- 236
sequence2       ACHIWDTIGAQSVGCRGAFIARPHNSILHVPGVPEPDFTSHDLAKLADQIIQAQKLAG 233
                ***********: * **.*:*****: * :. **:*** .:*:.:****:* : .
```

Figure 5.2- ClustalW amino acid sequence alignment between the L-haloacid dehalohenase from AQP5750 (sequence 1) and *Sulfitobacter* sp. EE-36 (sequence 2). The conserved aspartic acid in all characterised L-haloacid dehalohenase is highlighted in red. An '*' indicates 100% sequence identity, ':' indicates a highly similar amino acid and a '.' indicates a slightly similar amino acid.

```
sequence1       --MTPSHPARPSRSGILVFDVNETLLDLTSLSPLFERVFGDAKVLR--EWFPELILYSQT 56
sequence2       MDYKDDDDKLVDSLRACVFDAYGTLLDVHSAVMRNADEVGASAEALSMLWRQRQLEYSWT 60
                   . ..    .     ***.  ****: *        .* :       *  . : ** *

sequence1       LTLTGLYRPFGEIAAAVFEMVAANHQAKVTPDDIAELKTRLTSMPAYPDVAPALTRLQDA 116
sequence2       RTLMHQYADFWQLTDEALTFALRTYHLEDRKGLKDRLMSAYKELSAYPDAAETLEKLKSA 120
                 **   *  * :::  .: :.  .:: :   .   .* :  ..:.****.* :* :*:.*

sequence1       GFRLVTLTNSAPSPAPSPLEKAGIASFFEAHLTVHSSQRFKPHPSVYDSTAETLGAKPEE 176
sequence2       GYIVAILSNGNDEMLQAALKASKLDRVLDSCLSADDLKIYKPDPRIYQFACDRLGVNPNE 180
                *: :. *:*.  .   :.*: : :  .::: *:... : :**.* :*: :.: **.:*:*

sequence1       LCMIACHIWDTIGAQARGWRGGFVARPHNTPLTLAEVPQPDFIGRDMGELADQLIASLTA 236
sequence2       VCFVSSNAWDLGGAGKFGFNTVRINRQGNPPE--YEFAPLKHQVNSLSELWPLLAKNVTK 238
                :*::::.: **  **   *:.   : *  *.*    *..  ..  ..:.**   *  .:*

sequence1       --
sequence2       AA 240
```

Figure 5.3- ClustalW amino acid sequence alignment between the L-haloacid dehalohenase from AQP5750 (sequence 1) and *B. cepacia* (sequence 2). The conserved aspartic acid in all characterised L-haloacid dehalohenase is highlighted in red. A '*' indicates 100% sequence identity, ':' indicates a highly similar amino acid and a '.' indicates a slightly similar amino acid.

The haloalkane dehalogenase had 64% and 59% amino acid sequence identity to haloalkane dehalogenase from *S. paucimobilis* (Figure 5.3) and *M. tuberculosis* respectively (Figure 5.4). Both of these haloalkane dehalogenases have been biochemically characterised and the three-dimensional structures solved using X-ray crystallography. The haloalkane dehalogenase from *S. paucimobilis* was solved at a resolution of 1.58 Å (Marek *et al.,* 2000) and haloalkane dehalogenase from *M. tuberculosis* solved at a resolution of 1.2 Å (Mazumdar *et al.,* 2008).

```
sequence1       -MTTSFRDKKKFATVHGKQMAYIEEGTGDPIVFLHGNPMSSYLWRNIMPHLAGKGRLIAP 59
sequence2       SLGAKPFGEKKFIEIKGRRMAYIDEGTGDPILFQHGNPTSSYLWRNIMPHCAGLGRLIAC 60
                 : :.  .:***  ::*::****:*******:* **** ********** ** *****

sequence1       DLIGMGDSDKLDNSGPDSYTFAEHCTYLFALLEQLGVTENVTLVIHDWGSGLGFHWAHTH 119
sequence2       DLIGMGDSDKLDPSGPERYAYAEHRDYLDALWEALDLGDRVVLVVHDWGSALGFDWARRH 120
                ************ ***: *::***  ** ** * *.: :.*.**:*****.***.**: *

sequence1       SDAVKGIAFMEAIVETRESWDAFPERAREMFQALRSPAGEEMVLEKNLFVEALVPGSILR 179
sequence2       RERVQGIAYMEAIAMPIE-WADFPEQDRDLFQAFRSQAGEELVLQDNVFVEQVLPGLILR 179
                 : *:***:****. . * *  ***: *::***:** ****:**:.*:*** ::** ***

sequence1       DLTEEEMNEYRRPFANAGEDRRPTLTFPRQVPIEGQPKDVTELVDAYVDWLGQTSIPKLF 239
sequence2       PLSEAEMAAYREPFLAAGEARRPTLSWPRQIPIAGTPADVVAIARDYAGWLSESPIPKLF 239
                 *:* **  **.**  *** *****::***:** * * **. :.  *..**.::.*****

sequence1       INADPGVLITGEVRDRVRSWPNLTEVTVAGLHFIQEDSPDEIGAAVRDWHAS1--- 292
sequence2       INAEPGALTTGRMRDFCRTWPNQTEITVAGAHFIQEDSPDEIGAAIAAFVRRLRPA 295
                ***:**.* **.:**  *:*** **:**** **************:  :   *
```

Figure 5.4- ClustalW amino acid sequence alignment between the haloalkane dehalogenases from AQP5750 (sequence 1) and *S. paucimobilis* (sequence 2). Amino acid residues essential for catalysis are highlighted in red. A '*' indicates 100% sequence identity, ':' indicates a highly similar amino acid and a '.' indicates a slightly similar amino acid.

```
sequence1       MTTSFRD---KKKFATVHGKQMAYIEEGTGDPIVFLHGNPMSSYLWRNIMPHLAGKGRLI 57
sequence2       MTAFGVEPYGQPKYLEIAGKRMAYIDEGKGDAIVFQHGNPTSSYLWRNIMPHLEGLGRLV 60
                **:   :   : *:  : **:****:**.**.*** **** ************ * ***:

sequence1       APDLIGMGDSDKLDNSGPDSYTFAEHCTYLFALLEQLGVTENVTLVIHDWGSGLGFHWAH 117
sequence2       ACDLIGMGASDKLSPSGPDRYSYGEQRDFLFALWDALDLGDHVVLVLHDWGSALGFDWAN 120
                * ****** ****. **** *::.*:  :**** : *.: ::*.**:*****.***.**:

sequence1       THSDAVKGIAFMEAIVETRESWDAFPERAREMFQALRSPAGEEMVLEKNLFVEALVPGSI 177
sequence2       QHRDRVQGIAFMEAIVTP-MTWADWPPAVRGVFQGFRSPQGEPMALEHNIFVERVLPGAI 179
                 * * *:********* .  :*  :*  .* :**.:*** ** *.**:*:*** ::**:*

sequence1       LRDLTEEEMNEYRRPFANAGEDRRPTLTFPRQVPIEGQPKDVTELVDAYVDWLGQTSIPK 237
sequence2       LRQLSDEEMNHYRRPFVNGGEDRRPTLSWPRNLPIDGEPAEVVALVNEYRSWLEETDMPK 239
                **:*::****.*****.*.********::**::**:*:* :*. **: * .** :*.:**

sequence1       LFINADPGVLITGEVRDRVRSWPNLTEVTVAGLHFIQEDSPDEIGAAVRDWHASl----- 292
sequence2       LFINAEPGAIITGRIRDYVRSWPNQTEITVPGVHFVQEDSPEEIGAAIAQFVRRLRSAAG 299
                *****:**.:***.:** ****** **:**.*:**:*****:*****: ::   *

sequence1       -
sequence2       V 300
```

Figure 5.5- ClustalW amino acid sequence alignment between the haloalkane dehalogenase from AQP5750 (sequence 1) *M. tuberculosis* (sequence 2). Amino acid residues essential for catalysis are highlighted in red. A '*' indicates 100% sequence identity, ':' indicates highly similar amino acid and '.' indicates a slightly similar amino acid.

5.4 Summary

The degenerate PCR method to determine the L-haloacid dehalohenases gene sequences in AQP4626 and AQP5750 was unsuccessful. Genome sequencing of AQP5750 allowed the identification an L-haloacid dehalohenase and a haloalkane dehalogenase gene. Both dehalogenase genes were cloned, over-expressed in a bacterial expression system, purified, biochemically characterised and the three-dimensional structure solved using X-ray crystallography (chapters 7 and 8).

Chapter 6- L-haloacid dehalohenase from *P. ingrahamii*

6.1 Introduction

At its maximum, more than 7% of the Earth's surface is covered with sea ice that contains unique microbial communities (Staley *et al*., 1999). Psychrophilic microbes which live in sea ice are collectively referred to as sea ice microbial communities (SIMCO) (Staley *et al*., 1999). Unlike thermophilic marine environments which have been extensively studied, research into the microbial diversity of microbes which live in sea ice is still in its infancy. Sea ice develops when sea water reaches about -1.8ºC in the winter months. As the ice blocks form, a salting out process occurs, causing the formation of cavities filled with brine. These brine pockets provide a habitat for psychrophlic microbes to colonise, between temperatures ranging from -1.8ºC to -30ºC (Breezee *et al.,* 2004). Most microbial communities are found at the sea-ice interface of the underlining ocean water. Microbes that are found at this interface include bacteria, fungi and protozoa. Psychrophiles isolated from sea ice will often be tolerant to high salt concentrations and may also be classified as halophiles.

6.1.1 Background

As discussed in chapter 4, AQP4626 from the Aquapharm Biodiscovery Ltd Microbial library tested positive for L-haloacid dehalohenase activity. Aquapharm Biodiscovery Ltd determined that AQP4626 has 99.7% sequence identity to *P. arctica*. Using Protein BLAST, an L-haloacid dehalohenase was identified in *P. ingrahamii* (Altschul *et al*., 1990). To date, no L-haloacid dehalohenases from psychrophilic microorganisms have been characterised.

This chapter describes the cloning, over-expression, purification and biochemical characterisation of the *P. ingrahamii* L-haloacid dehalohenase.

6.2 Materials and Methods

6.2.1 Bioinformatics

To confirm the presence of catalytic amino acids, protein sequence alignments of the L-haloacid dehalohenase from *P. ingrahamii* against the biochemically and structurally characterised L-haloacid dehalohenase from *Pseudomonas* sp. YL were performed using ClustalW, European Bioinformatics Insitute (http://www.ebi.ac.uk/Tools/msa/clustalw2).

6.2.2 *P. ingrahamii* growth

P. ingrahamii was purchased from DSMZ, Germany. 200 ml of Marine broth (Table 2.1) was inoculated with a single colony from a slope plate. The culture was incubated at 4ºC at 200 rpm for 7 days. The bacterial paste was harvested by centrifugation (section 2.1.6) and the genomic DNA extracted (section 2.2.3).

6.2.3 Cloning of the L-haloacid dehalohenase from *P. ingrahamii*

PCR primers were designed (section 2.2.4.1) to amplify the L-haloacid dehalohenase gene (see Appendix, 11.8), with the incorporation of the restriction sites *Not*I and *Nde*I for cloning into the pET-28a vector (Table 6.1). PCR reactions were performed using PCR composition 2 (section 2.2.4.4) with PCR program 1 (Table 2.2). The PCR product was run on a 1% agarose gel to visualise the amplified DNA (section 2.2.2). The PCR band was gel extracted (section 2.2.6) and ligated into the cloning vector pJET1.2/blunt (section 2.2.7.1). The DNA from the ligation reactions were transformed into *E. coli* XL 10-Gold® chemically competent cells (sections 2.2.8.3 and 2.2.7). The pJET1.2/blunt construct and the pET-28a vector were double digested with *Not*I and *Nde*I restriction enzymes (section 2.2.8.5 and 2.2.8.6) and the digested pET-28a was SAP treated (section 2.2.8.7). The ligation reaction into the expression vector pET-28a was performed (section 2.2.8.2) and the reaction transformed (section 2.2.8.3) into *E. coli* XL 10-Gold® competent cells (section 2.2.7). Gel extracted PCR bands (section 2.2.6) and plasmid constructs (section 2.2.8.4) were sent for Sanger sequencing (section 2.2.8.8).

Forward primer	*Nde*I 5'- **CATATG**ATGTCAGTGATACTTGCTTTTGATG -3'
Reverse primer	*Not*I 5'- **GCGGCCGC**TTAGTCCGGCTTGGGTAGC -3'

Table 6.1- Forward and reverse primer sequences with incorporated restriction sites *Nde*I and *Not*I for the amplification of the L-haloacid dehalohenase in *P. ingrahamii*. The restriction sites are highlighted in bold.

6.2.4 Over-expression of the L-haloacid dehalohenase from *P. ingrahamii*

The extracted pET-28a construct (section 2.2.8.4) from *E. coli* XL 10-Gold® was transformed into the following competent *E. coli* cell lines: ArcticExpress^{TM}, BL21-CodonPlus (DE3)- RIPL**,** BL21-CodonPlus (DE3)- *Rosetta*2 and BL21-CodonPlus (DE3)-pLysS (section 2.2.7). The plasmids were extracted (section 2.2.8.4) and sent for Sanger sequencing (section 2.2.8.8). Positively transformed strains of *E. coli* were preserved as glycerol stocks (section 2.1.5). Protein over-expression conditions were optimised (section 2.3.1) and the resultant bacterial cell pellets were lysed with BugBuster (1/10 dilution with Tris-H_2SO_4 pH 8.2) as described in section 2.3.2. The soluble and insoluble protein fractions were analysed using SDS-PAGE (section 2.3.6).

6.2.5 Purification of the L-haloacid dehalohenase from *P. ingrahamii*

A 3 L culture of transformed ArcticExpress^{TM} *E. coli* was grown (section 2.1.4). The protein expression was induced (section 2.3.1) and the bacterial paste harvested (section 2.1.6). The bacterial cell pellet was resuspended in NAC buffer A (Table 2.7) to a final concentration of 10% (w/v) and the cells lysed by sonication (section 2.3.3.1). The soluble protein was purified by nickel affinity chromatography (section 2.3.3.8). Fractions corresponding to the recombinant protein were concentrated by an 80% ammonium sulfate precipitation at 4ºC, with gentle stirring over 6 h. The precipitated protein was harvested by centrifugation at 20,000 x *g* and the pellet resuspended in 1 ml of GF buffer (Table 2.7). The concentrated protein was further purified on a Superdex 75 GF chromatography column (section 2.3.3.7).

6.2.6 Biochemical characterisation of the L-haloacid dehalohenase from *P. ingrahamii*

Protein parameters including amino acid composition, theoretical pI and extinction coefficient were calculated (section 2.3.4) based on the L-haloacid dehalohenase protein sequence (see Appendix, 11.9). The purified protein concentration was determined (section 2.3.5) and the protein diluted to 0.2 mg/ml. This stock solution was stored at 4ºC for no longer than 72 h and was used in all biochemical characterisation reactions.

Section 2.3.9 describes the methods used for the biochemical characterisation of the *P. ingrahamii* L-haloacid dehalohenase. The thermostability (section 2.3.9.2), solvent stability (section 2.3.9.3), substrate specificity (section 2.3.9.4) and the temperature optimum (section 2.3.9.5) were determined. Michaelis-Menten kinetics were also calculated (section 2.3.9.6) using bromoacetic acid as the substrate and the dehalogenase reaction carried out at 45ºC.

6.3 Results and Discussion

6.3.1 Bioinformatics

The L-haloacid dehalohenase from *P. ingrahamii* has 53% amino acid sequence identity to the L-haloacid dehalohenase from *Pseudomonas* sp. YL. Protein sequence alignments between the two L-haloacid dehalohenases were performed using ClustalW, European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/clustalw2) (Figure 6.1). The conserved amino acid residues are highlight in red.

```
sequence1       --MSVILAFDVYGTLINTHGIVTILEKWLGKNAQAFSQTWRDKQLEYSFRRGLMQNYQSF
sequence2       MDYIKGIAFDLYGTLFDVHSVVGRCDEAFPGRGREISALWRQKQLEYTWLRSLMNRYVNF
                      :***:****::.*.:*   :: :  ..: :*  **:*****:: *.**:.* .*

sequence1       AVCTRHALDYTCNFYRVSLSKEQKQALMDGYKTLPAFDDAANALAQLKGAGFKLYAFSNG
sequence2       QQATEDALRFTCRHLGLDLDARTRSTLCDAYLRLAPFSEVPDSLRELKRRGLKLAILSNG
                  .*..** :**..  :.*. . :.:* *.*  *..*.:..::* :**  *:**  :***

sequence1       TKEAVEELLTGAGIRAFFDGIVSCDDLKSFKPNPAVYCHFLRESGAKHNQAWLISSNPFD
sequence2       SPQSIDAVVSHAGLRDGFDHLLSVDPVQVYKPDNRVYELAEQALGLDRSAILFVSSNAWD
                : :::: ::: **:*  ** ::* * :: :**:  **    :  * .:.   ::***.:*

sequence1       VIGAISTGMLSAWVQRSQDAIFDPWDVDPTAIITSLKELKGVLPKPD------
sequence2       ATGARYFGFPTCWINRTG-NVFEEMGQTPDWEVTSLRAVVELFETAAGKAEKG
                . **   *: :.*::*:   :*:  .  *   :***: :  :: ..
```

Figure 6.1- Amino acid sequence alignment between the L-haloacid dehalohenase from *P. ingrahamii* (sequence 1) and *Pseudomonas* sp. YL (sequence 2) using ClustalW.

6.3.2 Cloning and over-expression of the L-haloacid dehalohenase from *P. ingrahamii*

PCR was used to amplify the L-haloacid dehalohenase gene and a band of the correct size was observed on a 1% agarose gel (Figure 6.2). The amplified gene was then cloned into the cloning vector pJET1.2/blunt and the expression vector pET-28a with the incorporation of an N-terminal His-tag. Over-expression trials of the L-haloacid dehalohenase were conducted by varying the *E. coli* expression strain and the induction conditions. All expression strains contain the T7 promoter and are induced by IPTG.

Proteins from organisms other than *E. coli* may contain rare codons. The presence of these rare codons can limit the expression of the proteins, especially if they occur next to each other in the amino acid sequence. The *P. ingrahamii* L-haloacid dehalohenase has 14 rare codons coding for amino acids (3 Arg, 1 Leu, 5 Ile and 3 Pro). No rare codons for the same amino acid were repeated next to each other. Expression cell lines BL21-CodonPlus (DE3)-RIPL and BL21-CodonPlus (DE3)- *Rosetta*2 contain extra copies of the genes that code for the tRNAs that are complementary to rare codons. This allows higher levels of recombinant protein expression. BL21-CodonPlus (DE3)-pLysS suppresses basal expression of proteins by producing a lysozyme which inhibits T7 RNA polymerase and can be used to improve over-expression of toxic proteins. Over-

expression of the L-haloacid dehalohenase from *P. ingrahamii* was low in all *E. coli* cell lines. The best over-expression was observed in ArcticExpress^{TM} induced with 1 mM IPTG, at an $OD_{595\ nm}$ 0.8 for 24 h, at 12ºC with agitation at 220 rpm (Figure 6.3, lane 5). Small amounts of over-expressed protein were observed in the insoluble fraction (Figure 6.3, lane 4). ArcticExpress^{TM} cells are designed to help overcome the expression of misfolded and insoluble proteins in inclusion bodies. ArcticExpress^{TM} cell lines contain Cpn10 and Cpn60 chaperonins from the psychrophilic bacterium *Oleispira antarctica*. These chaperonins have higher protein refolding activities (compared to the naturally existing chaperonins in *E. coli*) at temperatures between 4-12ºC. It is an advantage that protein over-expression in ArcticExpress^{TM} mimics the closest conditions to the host organisms (*P. ingrahamii*) natural habitat.

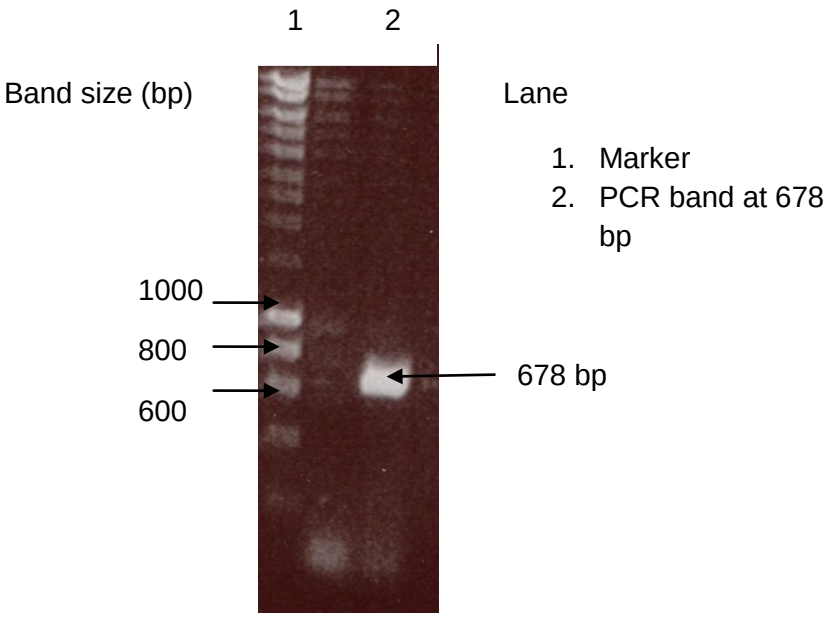


Figure 6.2- 1% agarose gel showing the amplified L-haloacid dehalohenase gene from *P. ingrahamii.*

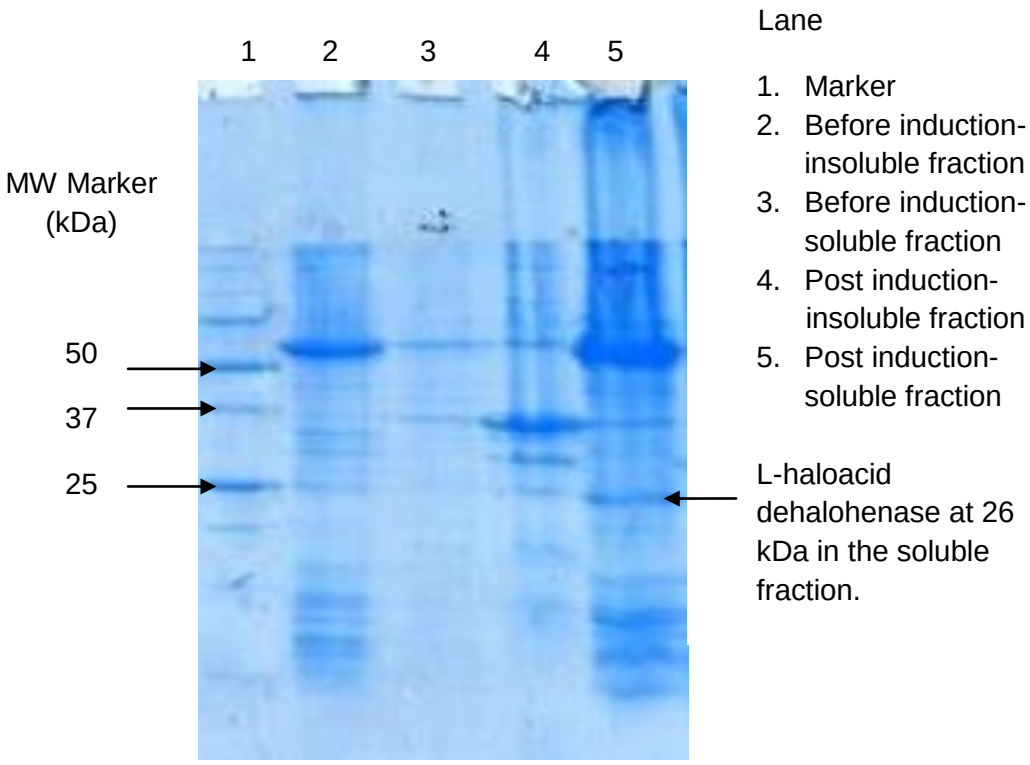


Figure 6.3- SDS-PAGE gel showing the best over-expression of the *P. ingrahamii* L-haloacid dehalohenase in ArcticExpress™ induced with 1 mM IPTG, at an $OD_{595\ nm}$ 0.8 for 25 h, at 12ºC with agitation at 220 rpm.

6.3.3 Purification of the L-haloacid dehalohenase from *P. ingrahamii*

5 g of cell paste was harvested from 3 L of recombinant *E. coli* by centrifugation (section 2.1.6). Bacterial cells were lysed by sonication and the soluble protein purified by nickel affinity chromatography (Figure 6.4). Fractions 50-65 assayed positive for L-haloacid dehalohenase activity and 10 µl samples of these fraction were analysed using SDS-PAGE (Figure 6.5). The fractions that tested positive for activity from the nickel affinity chromatography were pooled together and concentrated by 80% ammonium sulfate precipitation. The concentrated protein sample was purified on a Superdex 75 GF chromatography column (Figure 6.6). Fractions eluting between 39-49 ml assayed positive for dehalogenase activity and 10 µl samples of these fraction were analysed using SDS-PAGE (Figure 6.7). The peak between fractions 39-49 suggests the eluting L-haloacid

dehalohenase is likely to be in the form of a dimer. 8 mg of purified *P. ingrahamii* L-haloacid dehalohenase was obtained from a 3 L culture of recombinant ArcticExpress™.

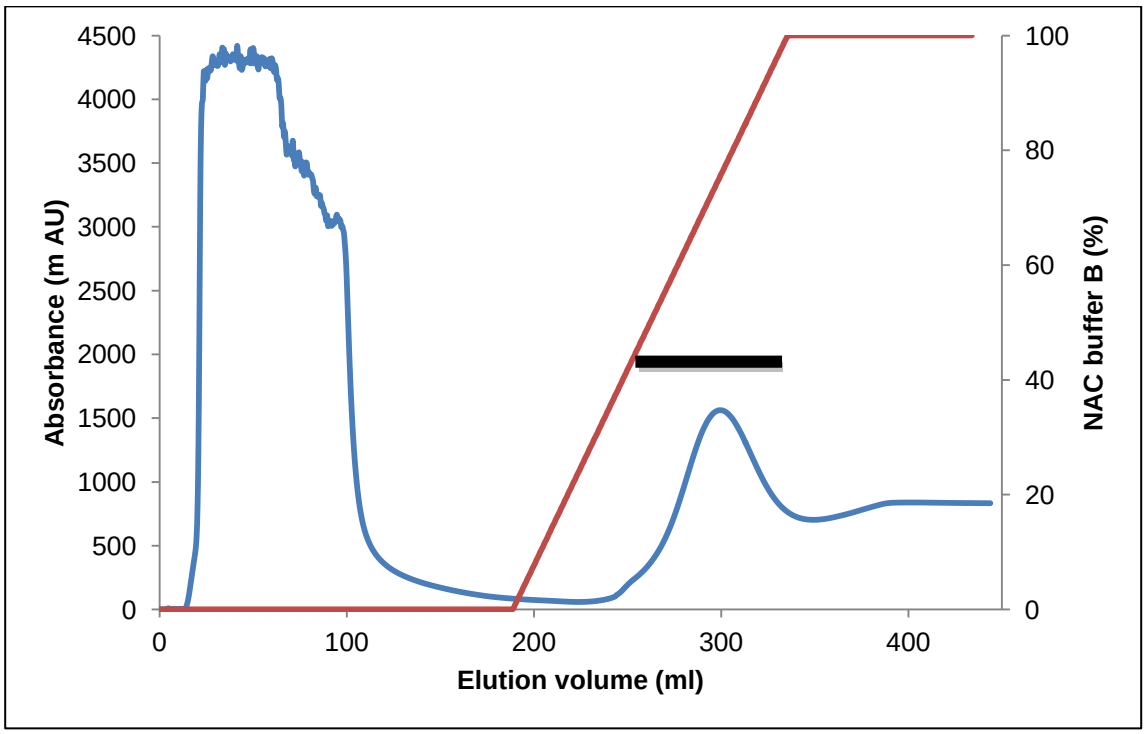


Figure 6.4- Elution profile from the purification of the *P. ingrahamii* L-haloacid dehalohenase from the nickel affinity chromatography column. The red line shows the concentration of NAC buffer B. Fractions 55-65 (275-325 ml) containing the L-haloacid dehalohenase are highlighted by the black bar.

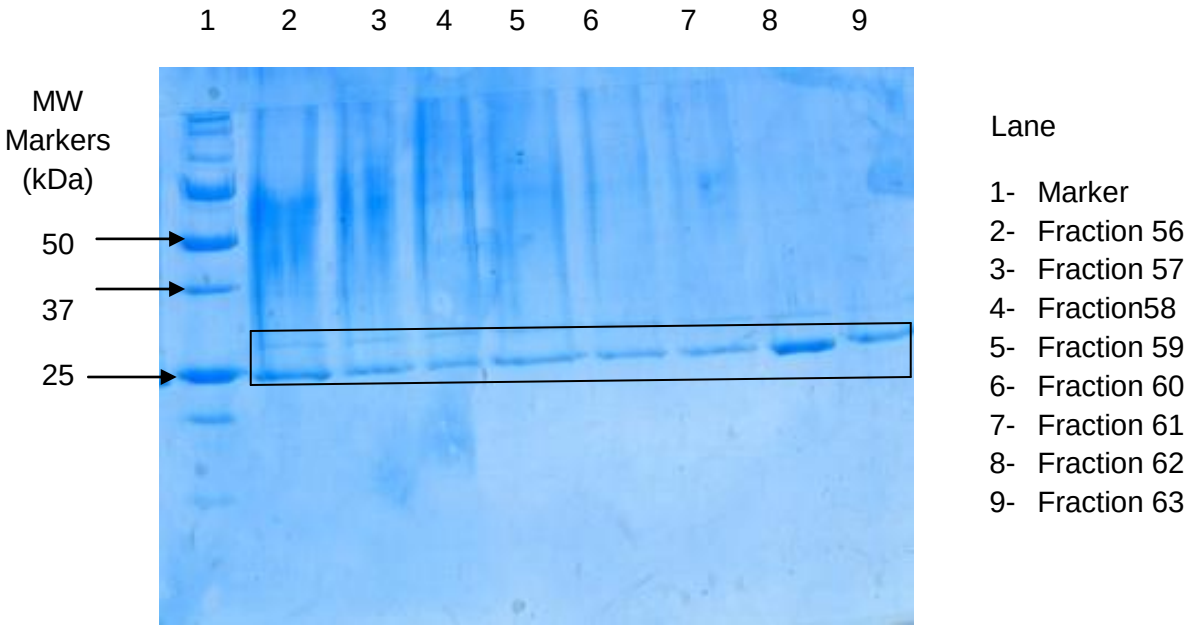


Figure 6.5- SDS-PAGE gel of the fractions containing the *P. ingrahamii* L-haloacid dehalohenase (highlighted by the black box) from the nickel affinity chromatography column.

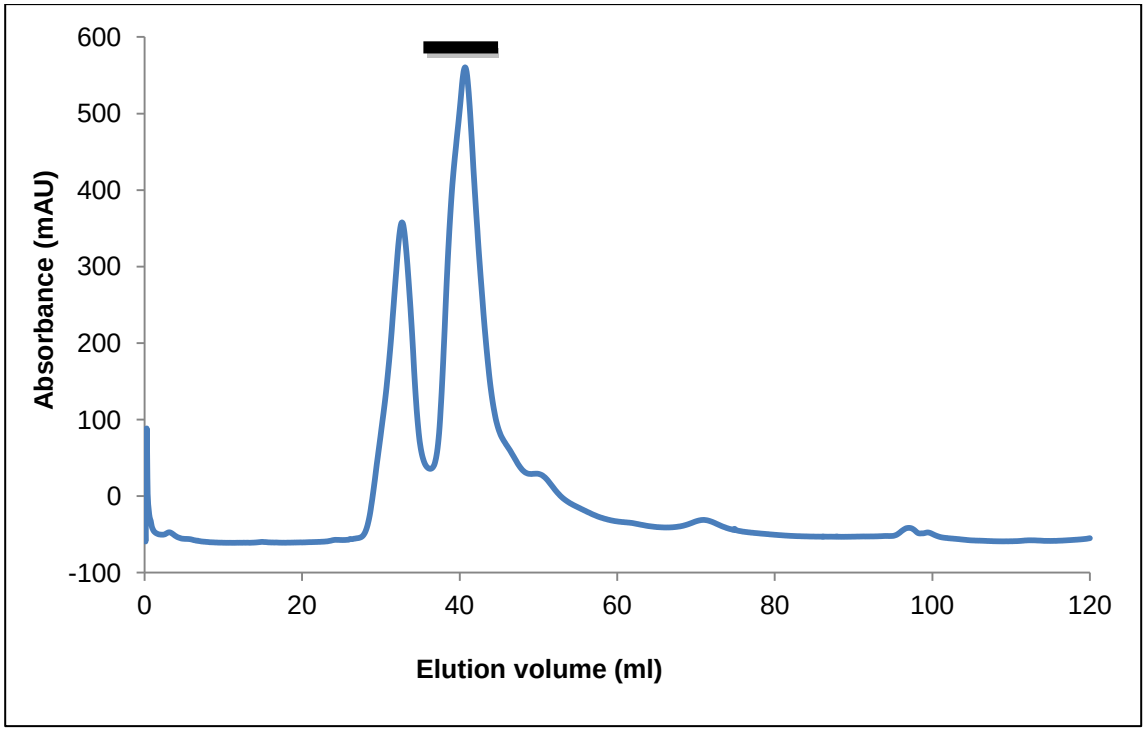


Figure 6.6- Elution profile from the purification of the *P. ingrahamii* L-haloacid dehalohenase from the GF chromatography column (Superdex 75). Fractions 39-49 containing the L-haloacid dehalohenase are highlighted by the black bar. 1 ml fractions were collected throughout GF chromatography.

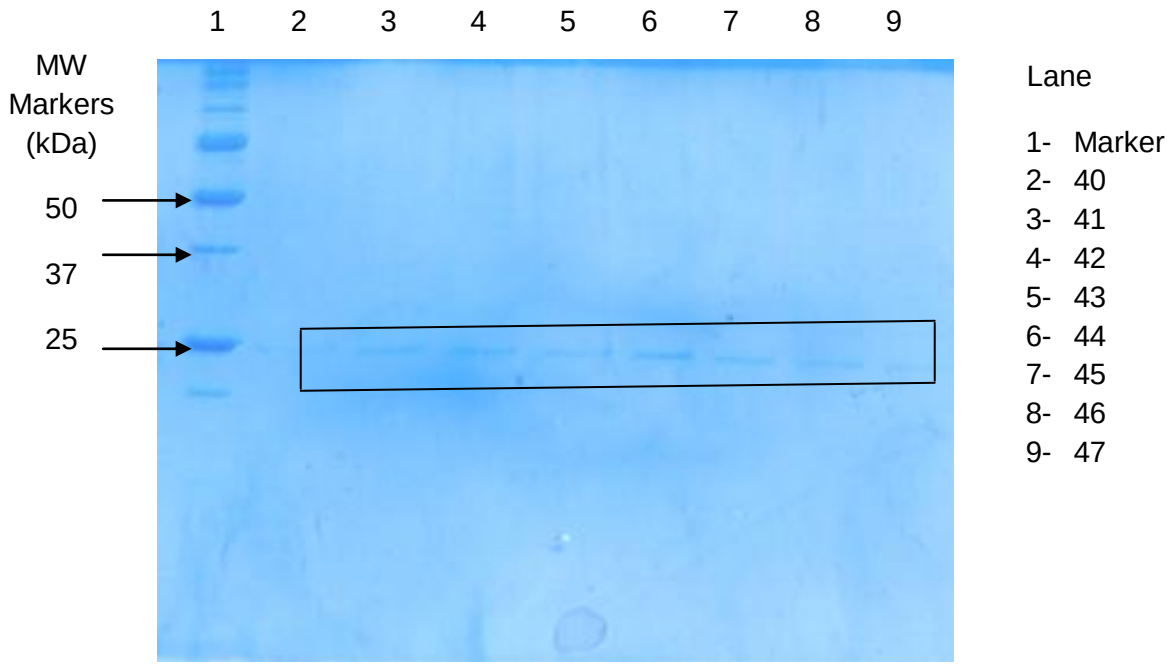


Figure 6.7- SDS-PAGE gel of the fractions containing the *P. ingrahamii* L-haloacid dehalohenase (highlighted by the black box) from the GF chromatography column (Superdex 75).

6.3.4 Biochemical Characterisation of the L-haloacid dehalohenase from *P. ingrahamii*

The colorimetric dehalogenase assay is based on the method described by Holloway *et al.* (1998). It measures the decrease in pH as the dehalogenase turns over the substrate and a halide/proton product is formed. This can be visualised by eye as the assay solution changes from red at pH ≥8 to yellow at pH ≤4. The assay can also be followed spectrophotometrically at 540 nm. Enzyme activity is either shown as a percentage compared to a control reaction or in Enzyme Units (EU). 1 EU is defined as the production of 1 mM product/min/mg of enzyme.

6.3.4.1 Substrate specificity of the L-haloacid dehalohenase from *P. ingrahamii*

The activity of the L-haloacid dehalohenase from *P. ingrahamii* towards substrates (at a concentration of 10 mM) with varying halogens (chlorine, bromine and iodine), carbon chain lengths and both enantiomers were tested (Figure 6.8) (structures shown in Appendix, 11.6). Like all other characterised L-haloacid dehalohenases, the recombinant enzyme from *P. ingrahamii* shows activity towards only the l-enantiomer of the substrate, with the halogen atom attached at the α-carbon only. The *P. ingrahamii* L-haloacid dehalohenase shows highest activity towards monobromoacetic acid (this reading was taken as 100% in comparison to all other substrates). Following monobromoacetic acid, the L-haloacid dehalohenase showed decreasing activity towards monochloroacetic acid (62%), S-chloropropionic acid (42%), S-bromopropionic acid (31%), dichloroacetic acid (28%) and 2-chlorobutyric acid (10%). This concluded that the *P. ingrahamii* L-haloacid dehalohenase has highest activity towards substrates with a shorter carbon chain length (≤C3), without preference towards a chlorine or bromine in the α-carbon position. The *P. ingrahamii* L-haloacid dehalohenase showed less than 10% activity towards 2-chlorobutyric acid and does not show any activity towards trichloroacetic acid, sodium fluoroacetate, 3-bromopropionic acid, R-chloropropionic acid, R-bromopropionic acid.

In comparison, the L-haloacid dehalohenases from *Pseudomonas* sp. strain YL (53% amino acid sequence identity) and *S. tokodaii* (23%) showed highest activity towards l-chloropropionic acid (Liu *et al*., 1994, Rye *et al*., 2009). These enzymes also showed activity towards longer chain substrates such as 2-chlorobutryic acid, whereas the L-haloacid dehalohenase from *X. autotrophicus* (26%) prefers substrates with a smaller carbon chain length and shows highest activity towards dibromoacetic acid and dichloroacetic acid (Van der Ploeg *et al*., 1991). The L-haloacid dehalohenase from *B. cepacia* (29%) shows highest activity towards brominated substrates. This L-haloacid dehalohenase has more than double the activity towards monobromoacetic acid and 2-bromopropionic

acid in comparison to monochloroacetic acid and 2-chloropropionic acid (Tsang *et al*., 1999).

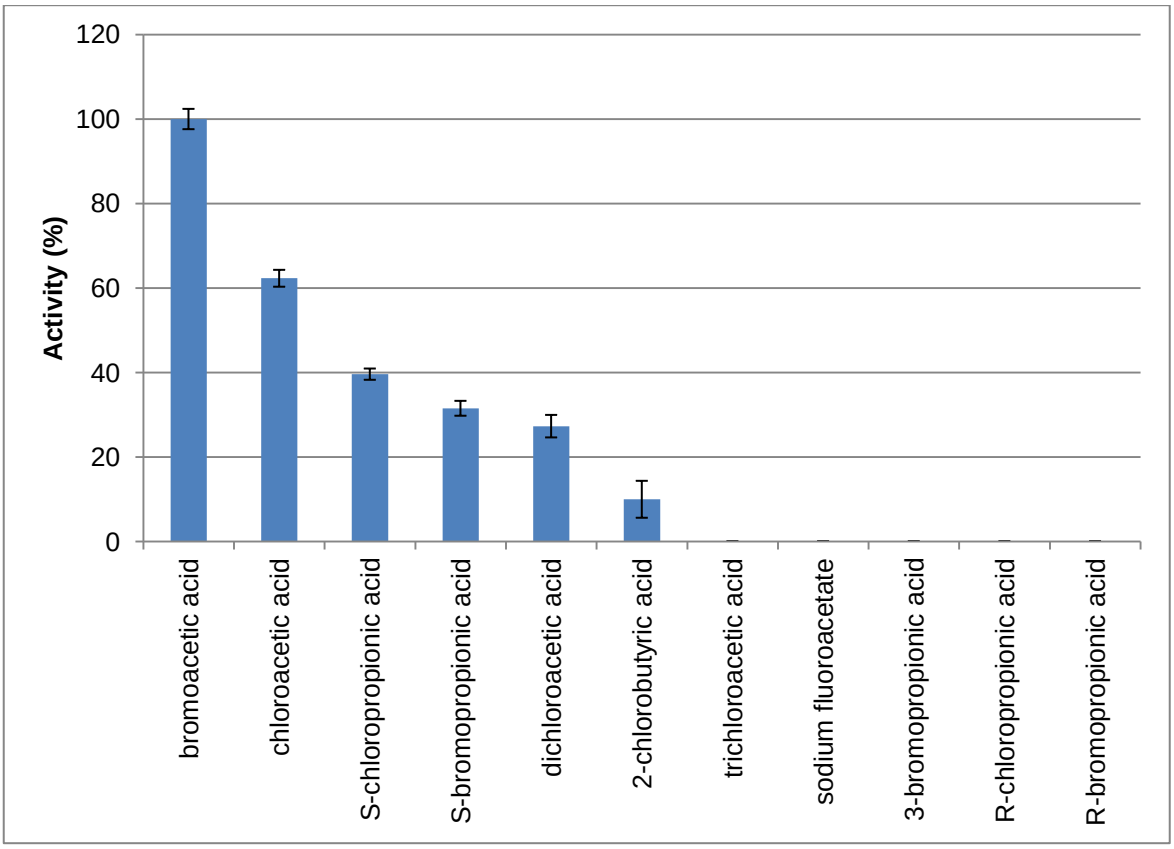


Figure 6.8- Graph showing the substrate specificity of the L-haloacid dehalohenase from *P. ingrahamii*.

6.3.4.2 The thermostability of the L-haloacid dehalohenase from *P. ingrahamii*

Thermostability is an important factor when deciding whether an enzyme is suitable for use in industrial or pharmaceutical applications. For this reason, many enzymes from thermophilic microorganisms have been exploited. Thermal stability assays were carried out using pre-incubated assay solutions, with 10 mM of chloroacetic acid as the substrate. The thermostability of the L-haloacid dehalohenase from *P. ingrahamii* was determined by incubating the enzyme between 30-70ºC for 30, 60 and 90 min (Figure 6.9). Enzyme activity is displayed as a percentage in comparison to untreated protein. The L-haloacid dehalohenase from *P. ingrahamii* displays high thermal stability retaining 87% of

activity when incubated at 50ºC for 90 min. The enzyme started to lose activity when incubated at higher temperatures and was almost completely denatured when incubated at 70ºC for 90 min with a loss of 92% of activity.

In comparison, the L-haloacid dehalohenase from *S. tokodaii* also shows equally high thermal stability, retaining 92% of activity after incubation at 50ºC for 90 min (Rye *et al*., 2009). This is to be expected as *S. tokodaii* is a thermophilic archaeon isolated from the volcanic hot springs of Oita Prefecture in Kyushu, Japan (Inatomi *et al*., 1983). On the other hand the L-haloacid dehalohenase from mesophilic *Pseudomonas* sp. YL is also relatively thermostable, retaining 100% of activity when incubated at 60ºC for 30 min (Liu *et al*., 1994).

The temperature at which the *P. ingrahamii* L-haloacid dehalohenase starts to denature is 60ºC higher than the optimal growth temperature (5ºC) of *P. ingrahamii* (Auman *et al*., 2006). The thermostability of *P. ingrahamii* L-haloacid dehalohenase provides evidence that enzymes from extreme psychrophiles may be thermostable.

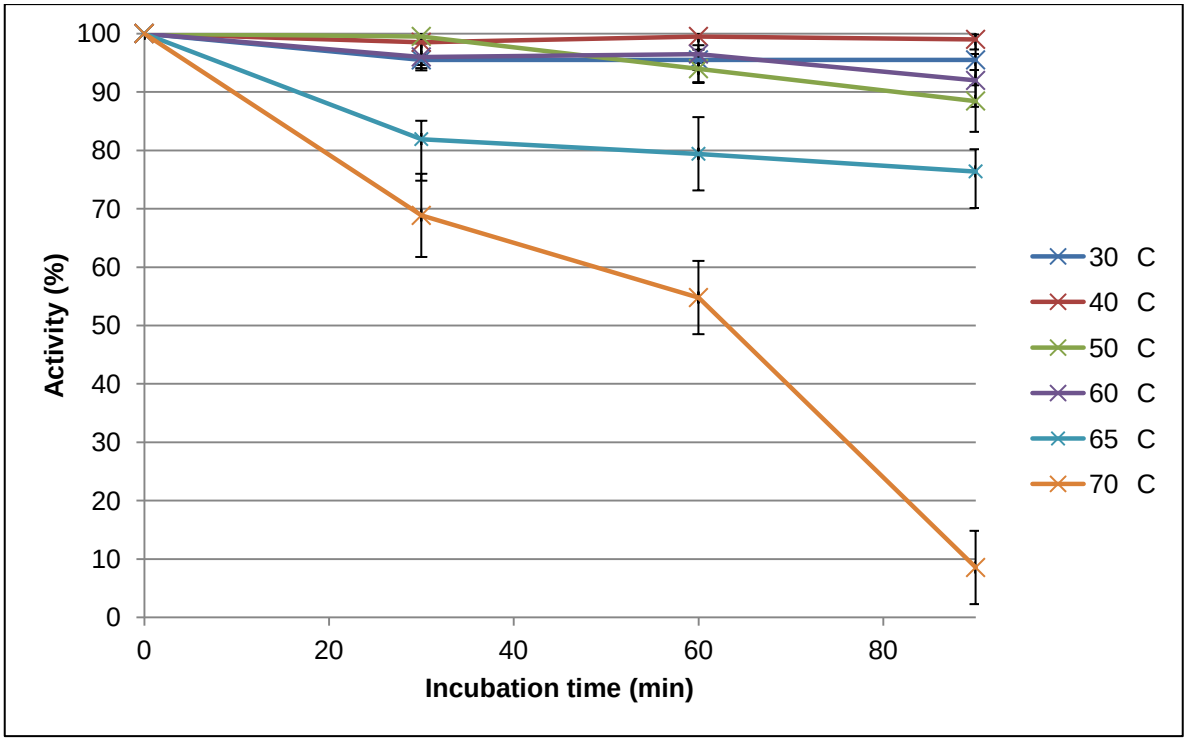


Figure 6.9- Graph showing the thermostability of the L-haloacid dehalohenase from *P. ingrahamii*.

6.3.4.3 Solvent stability of the L-haloacid dehalohenase from *P. ingrahamii*

Chemical processes in industrial and pharmaceutical applications are often carried out in the presence of solvents. Although not all enzymes are stable in organic solvents, there are many advantages to using solvent stable enzymes. This includes improved substrate specificity and solubility (Schmid *et al*., 2001). Therefore, it is an advantage and often necessary to determine if biocatalysts are solvent stable. To characterise the solvent stability of the *P. ingrahamii* L-haloacid dehalohenase, the enzyme was incubated in ethanol, methanol, DMSO and acetonitrile at varying concentrations for 1 h (Figure 6.10). Following incubation, the enzyme was assayed for L-haloacid dehalohenase activity with 10 mM chloroacetic acid and the results displayed as a percentage compared to untreated enzyme.

The L-haloacid dehalohenase from *P. ingrahamii* retained 95%, 75% and 55% of activity in up to 40% methanol, DMSO and ethanol respectively. The enzyme was more unstable in acetonitrile in comparison to the other solvents tested, with a

loss of 69% of activity in 40% acetonitrile. The protein was almost completely denatured when incubated with 50% ethanol, methanol, DMSO and acetonitrile with a loss of activity ranging between 80-99%. Overall, this protein is relatively stable in organic solvents.

In comparison, the L-haloacid dehalohenase from *S. tokodaii* is less stable in organic solvents, with a loss of 40% and 60% of activity towards l-chloropropionic acid in 30% DMSO and 10% ethanol, methanol and acetonitrile (Rye *et al*., 2009). The L-haloacid dehalohenase from *P. putida* has higher activity towards substrates with longer carbon chain lengths in the presence of anhydrous DMSO, with high activity towards 2-bromohexadecanoic acid. Although the activity towards substrates with longer carbon chain lengths was increased, the activity towards l-chloroproponic acid dropped by 94%, in comparison to untreated protein (Hasan *et al*., 1991).

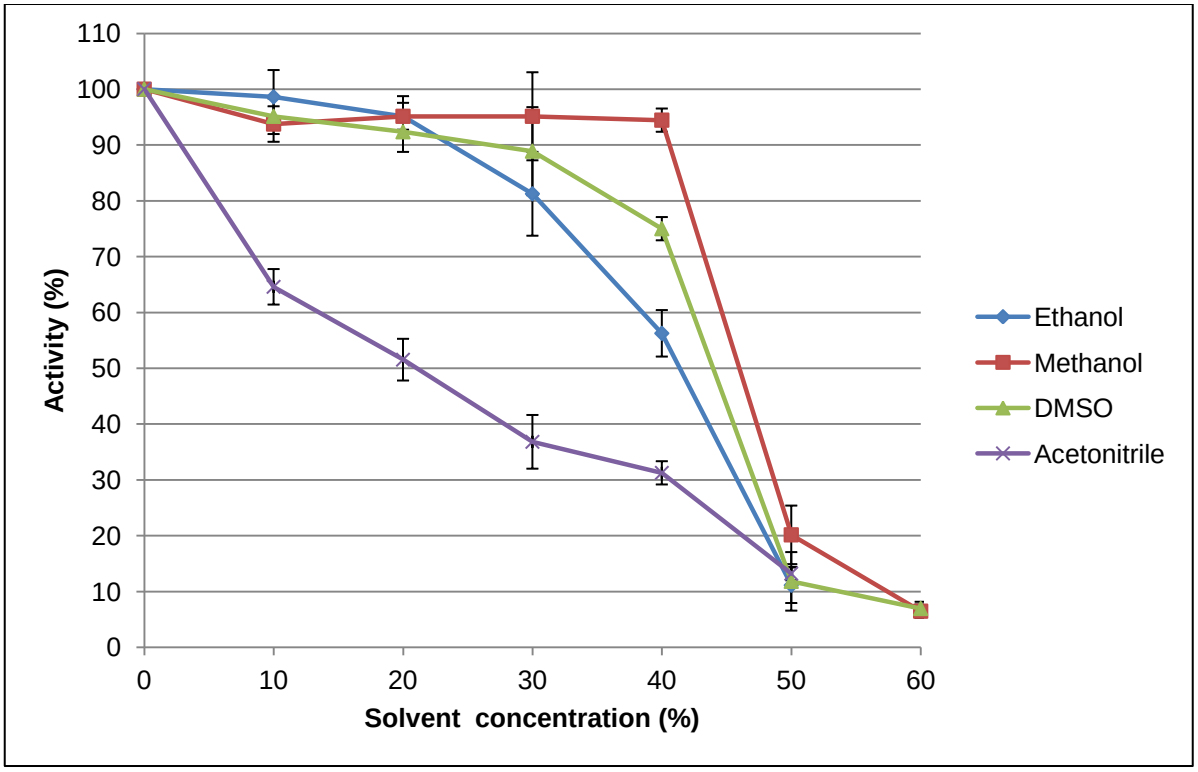


Figure 6.10- Graph showing the solvent stability of the L-haloacid dehalohenase from *P. ingrahamii*.

6.3.4.4 Temperature optimum of the L-haloacid dehalohenase from *P. ingrahamii*

It is important to determine the optimal rate of reaction in organic synthesis and biotransformation reactions. When considering the optimal rate of reaction, industrial and pharmaceutical companies need to take into account the necessary compounds for the reaction and the optimal temperature for the rate of reaction. The optimal compounds and temperature chosen will often be a compromise between the best observed rates and the most cost effective route. It is therefore important to determine the optimal rate of reaction at given temperatures so that the industrial application of the enzyme can be assessed. As the temperature rises, the kinetic energy of the molecules increases. This rise in kinetic energy of the molecules increases the chance of collisions, which in turn will speed up the rate of reaction. In living organisms this process is limited, as proteins will undergo significant structural changes at high temperatures which alter their tertiary structure. Denatured proteins which have lost their structural integrity may be inactive and can precipitate out of solution. As a result of this, the optimal activity of an enzyme is dependent on the reaction conditions and time (as the time increases, so will the denaturation of proteins at high temperatures).

The temperature optimum of the *P. ingrahamii* L-haloacid dehalohenase was determined by assaying the enzyme between 25-65ºC (Figure 6.11). Results are displayed in EU. The optimal temperature of the L-haloacid dehalohenase from *P. ingrahamii* is 45ºC. The optimal enzyme temperature of the L-haloacid dehalohenase is 40ºC higher than the optimal growth temperature of *P. ingrahamii* at 5ºC (Breezee *et al*., 2004). The optimal temperature of the L-haloacid dehalohenase is 25ºC lower than the temperature at which the protein starts to lose activity when incubated at 70ºC for 90 min (Figure 6.9). This confirms that there is not always a direct correlation between thermostability and the optimal temperature of an enzyme. It has previously been observed that a decrease in the activity of an enzyme above the optimum temperature can be reversible and enzyme activity can be recovered as the temperature decreases

back to the optimal temperature range (Eisenthal *et al*., 2006). This was observed with citrate synthase from an Antarctic bacterium, Strain DS2-3R (Gerike *et al*., 1997).

In comparison, the optimum temperature for the L-haloacid dehalohenase from the thermophilic archaeon *S. tokodaii* is 60ºC. This is to be expected at the optimum growth temperature of wild type organism is between 75-80ºC (Rye *et al*., 2009). This dehalogenase is thermostable, as discussed in the previous section. The L-haloacid dehalohenase from *Pseudomonas* sp. YL has a temperature optimum of 65ºC despite the host organism optimally growing at 30ºC.

Thermostable enzymes, such as aldehyde dehydrogenase and aspartase from psychrophilic bacterium *Cytophaga* sp. KUC-1 have previously been reported. They show highest activity at 55ºC (Oikawa *et al*., 2003). *Cytophaga* sp. KUC-1 was isolated from Antarctic seawater and has an optimal growth temperature of 15ºC, which is 10ºC higher than that of *P. ingrahamii* at 5ºC.

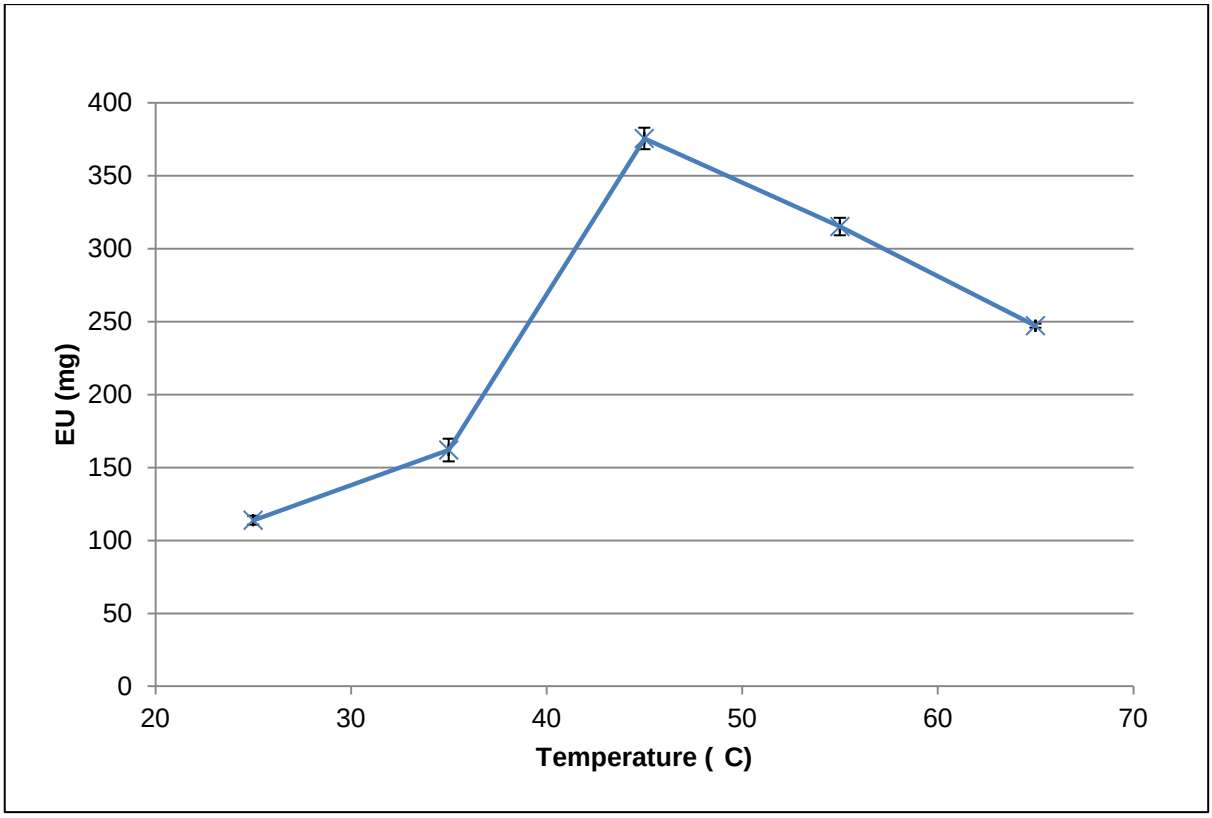


Figure 6.11- Graph showing the temperature optimum of the L-haloacid dehalohenase from *P. ingrahamii*.

6.3.4.5 The Michaelis-Menten kinetics of the L-haloacid dehalohenase from *P. ingrahamii*

The Michaelis-Menten kinetics of the L-haloacid dehalohenase from *P. ingrahamii* was determined by assaying the enzyme at 45ºC, with bromoacetic acid at concentrations between 0.2-5 mM. The results from this are shown in Figure 6.12. The V_{max} and K_m were calculated to be 0.6 µM min^{-1} mg^{-1} and 1.36 mM respectively. In comparison, the L-haloacid dehalohenases from *Pseudomonas* sp. YL and *B. cepacia* have a slightly lower K_m value of 1.1 mM and 1.13 mM (Liu *et al*., 1997, Tsang *et al*., 1999). The L-haloacid dehalohenase from *S. tokodaii* has a higher K_m value of 6.23 mM (Rye *et al*., 2009).

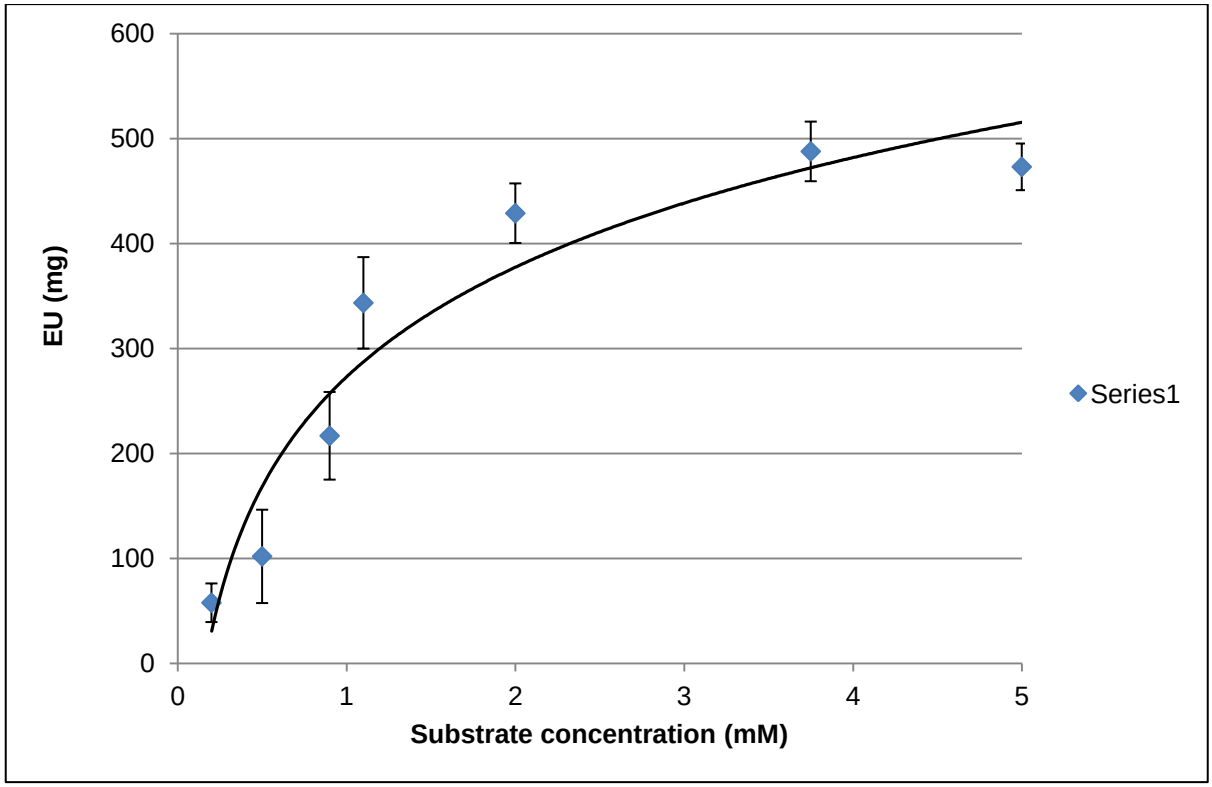


Figure 6.12- Graph showing the activity of the L-haloacid dehalohenase from *P. ingrahamii* with increasing bromoacetic acid concentrations (0.2-5 mM).

6.4 Summary

Thermostable L-haloacid dehalohenase from *P. ingrahamii* shows highest activity towards substrates with short carbon chains (≤C3) and is stable in up to 30% ethanol, methanol and DMSO when incubated for 1 h. The enzyme is stable at temperatures up to 60ºC and has a temperature optimum of 45ºC. The K_m of the enzyme is 1.36 mM.

Chapter 7- L-haloacid dehalohenase from AQP5750

7.1 Introduction

Aquapharm Biodiscovery Ltd determined that AQP5750 has 99.0% sequence identity to *Rhodobacteraceae* family bacteria. Chapter 5 describes the genome sequencing and the identification of an L-haloacid dehalohenase gene in AQP5750. To date, no dehalogenases from bacteria in the *Rhodobacteraceae* family have been structurally characterised. In a recent study, *Rhodobacteraceae* family bacteria (isolated from the marine sponge, *Hymeniacidon perlevis*) tested positive for haloacid dehalogenases (Huang *et al*., 2011).

This chapter describes the cloning, over-expression, purification, biochemical characterisation, crystallization and structural analysis of the L-haloacid dehalohenase from AQP5750.

7.1.1 Introduction to crystallography

7.1.1.1 Protein crystallization

Protein crystallization occurs when the concentration of the protein reaches a point of supersaturation in aqueous solution. This is achieved by varying a number of crystallization conditions including the concentration of the protein, the precipitant, additives/buffers and by varying the pH and temperature. Protein crystallization occurs in two stages. The first stage is the formation of a microscopic bundle of protein molecules called the nucleus. Protein supersaturation is required to overcome the high activation energy needed for nucleation. The second stage is crystal growth from the nucleus. Phase diagrams can be used to describe the state of the protein as a function of the variable crystallization conditions. The two dimensional phase diagram (Figure 7.1, taken from Asherie, 2004) shows how the solubility of a protein can decrease as the prepipitant concentration increases. There are four zones of saturation; the precipitation (supersaturation), labile, metastable and the unstaurated zone. Nucleation occurs in the supersaturation zone. As the protein crystals form, the concentration of the protein in the droplet decreases, pushing the conditions in the droplet into the metastable zone. Post nucleation, the metastable zone

provides the ideal conditions for the formation of large protein crystals. Below this is the undersaturated zone where protein concentrations are too low for the formation of protein crystals (Asherie, 2004).

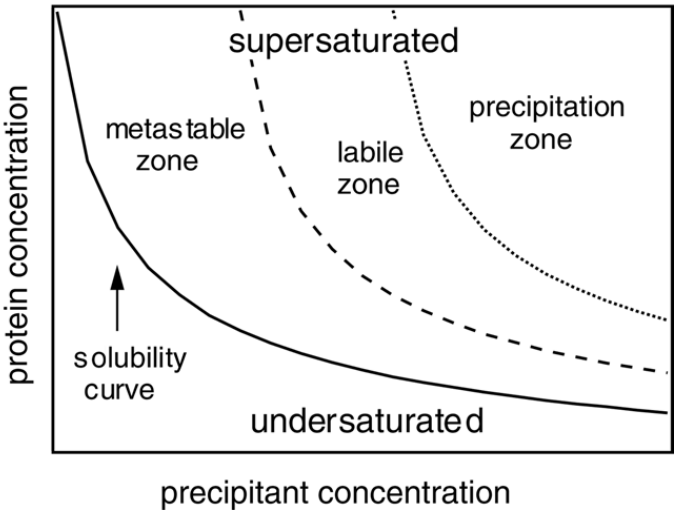


Figure 7.1- A schematic phase diagram showing the solubility of the protein as a function of precipitant concentration. This Figure was taken from Asherie, 2004.

There are a number of methods used to generate the conditions needed for protein crystallization. In microbatch crystallization small amounts of a highly concentrated protein solution (≥10 mg/ml), precipitant and buffer are directly mixed in a well. The crystallization droplets are then covered with silicon and/or paraffin oils which allow varying amounts of solvent/water diffusion from the droplet (Chayen *et al*., 1992). A highly concentrated protein sample is needed so that nucleation can take place. Vapour diffusion (sitting drop or hanging drop) is another method of crystallization where a protein droplet containing a small amount of precipitant is equilibrated with a reservoir of solution containing higher concentrations of the precipitant. As the solutions equilibrate, water diffuses out of the protein droplet causing the precipitant concentration in the protein droplet to increase. The protein will now be supersaturated, allowing nucleation to occur.

7.1.1.2 X-ray Crystallography

X-ray crystallography is a technique which uses the diffraction of X-rays to determine the three-dimensional structures of proteins. X-rays have comparable wavelengths to the interatomic distances in proteins. The X-ray beam is focused on the protein crystal and two dimensional images of the diffracted X-rays are collected. From the series of images, the intensities of the diffraction spots along with phases (determined separately) are used to calculate a model of the electron density. This is because of the direct relationship between the diffracted X-rays and the three-dimensional structure of the crystallized protein using the mathematical Fourier transform method.

Protein crystals are made up of repeating units which are referred to as the unit cell. The protein crystal can be divided into many theoretical planes in different orientations. The planes will differ in both orientation and spacing. The unit cell is described by its three lengths (a, b, c) and its three angles (α, β, γ). The diffraction patterns seen in X-ray crystallography are related to the various lattice planes in the crystal.

Bragg's Law (equation 7.1) describes the conditions that need to be satisfied in order to obtain a diffraction pattern from a crystal, where θ is the angle between the X-ray and the diffraction spot, n is the integer number of wavelengths, d is the distance between the lattice planes and λ is the wavelength (equation 7.1). There is a direct relationship between the distance (d) and the angle at which the incident X-ray is scattered. In order to obtain constructive interference of the diffracted X-rays, the difference in the path length ($2d\sin\theta$) must be equal to an integer (n) multiple of the wavelength (λ) (Figure 7.2). The shorter the distance between the crystal planes (higher resolution) the larger angle of the diffracted X-rays.

$$n\lambda=2d\sin\theta$$

Equation 7.1- Bragg's Law: n is the integer number of wavelengths, d is the distance between the lattice planes, λ is the wavelength and θ is the incident angle.

The reflections are described by a set of Miller indices (h, k, l) which describe the spacing between the lattice planes of the crystal in reciprocal space. Experimental information including the wavelength and the scattering angle (θ) allows the unit cell parameters to be determined. The diffraction spots observed define the reciprocal lattice. The reciprocal lattice has a reciprocal relationship with the unit cell in the crystal lattice. In any one orientation of a crystal only some of the reciprocal lattice points will be satisfied by Bragg's law.

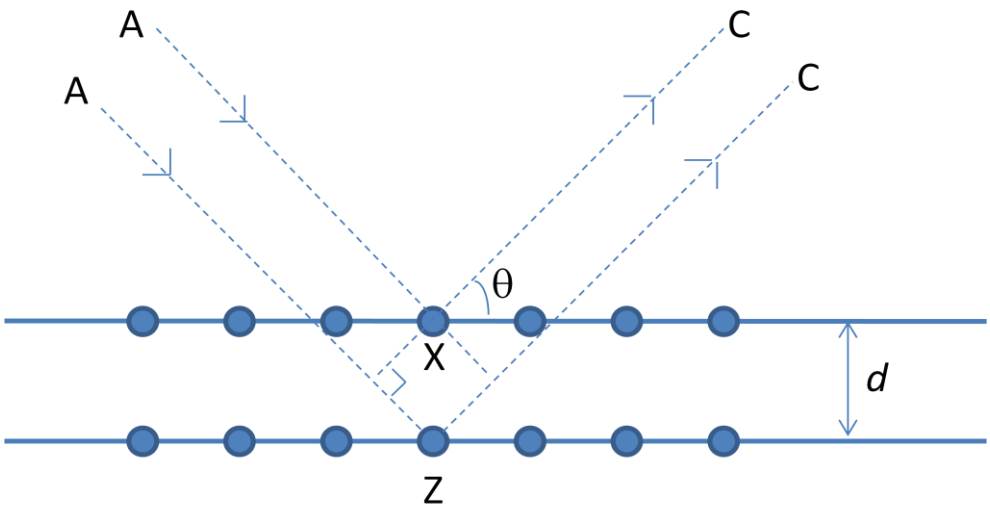


Figure 7.2- A diagram showing the conditions needed to satisfy Bragg's law.

Ewald's sphere is used to identify points in reciprocal space which are satisfied by Bragg's law (Figure 7.3) in any one orientation. The crystal is at the origin of the Ewald's sphere with a radius of $1/\lambda$. Reciprocal lattice points which lie on the spheres surface will satisfy Bragg's law and therefore produce a diffraction spot. Therefore, the crystal needs to be rotated around the origin in order to produce diffraction spots at all points in the reciprocal lattice.

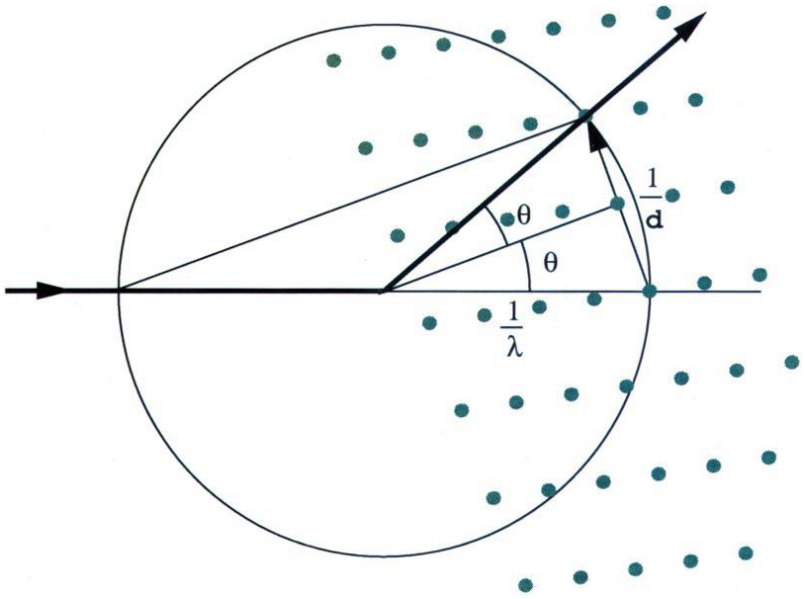


Figure 7.3- The Ewald construction showing that in any one orientation of a crystal only some of the reciprocal lattice points will be satisfied by Bragg's law. The trigonometric condition $1/d = (2/\lambda)\sin\theta$ is satisfied when the reciprocal lattice points are on the surface of the sphere. This Figure was taken from the review by Dauter, 1999.

X-ray crystallography aims to calculate the electron density for all atoms throughout the unit cell. To calculate the electron density, both the amplitudes and phases for the diffracted X-rays need to be determined. The amplitudes can be calculated from experimental information as they are the square root of the intensities. The phases of the diffracted waves cannot be obtained directly during data collection but can be calculated using a number of methods which are discussed later. Once the phases and amplitudes for each of the Miller indices (*h*, *k*, *l*) have been calculated, they are combined to give the structure factor. The structure factors and the electron density are related by a mathematical Fourier transform. The total structure factor equation is used to describe the sum of all atomic structure factors for particular Miller indices (equation 7.2). The inverse Fourier transform of the structure factors produces electron density maps at all points in direct space and therefore for all the electrons within the unit cell (equation 7.3).

Equation 7.2- The structure factor equation.

Equation 7.3- Calculation of the electron density in a crystal.

A number of methods can be used to calculate the phases that cannot be obtained directly during X-ray crystallography. Molecular replacement can be used to deduce the phases, if the protein of unknown structure has approximately 20% or greater sequence identity to a structure of a known protein. The structure of the known protein used in molecular replacement is called the model. The Patterson function of the protein of unknown structure and that of the model are calculated by performing a Fourier transform using only the amplitudes. The Patterson maps formed consist of interatomic self vectors of the atoms within the unit cell. The Patterson map for the model is rotated (rotation function) in relation to the Patterson map of the unknown protein in all possible orientations for the best fit and given a score (Crowther and Blow, 1967). Following rotation, translation (translation function) of the model Patterson map in relation to the protein of unknown structure Patterson map is moved in all possible positions in the unit cell (Rossmann and Blow, 1964). A statistical correlation for each translation is calculated and the translation is repeated for each rotation peak found, as described in the recent reviews by Taylor, 2003 and Evans and McCoy, 2007.

Multiple isomorphous replacement can also be used to deduce the phases. Co-crystallized or crystal soaking experiments can be used to incorporate heavy atoms, such as mercury, lead and uranium. Data sets are collected for the native and the derived heavy atom crystal. The incorporation of heavy atoms alters the scattering of the diffracted X-rays. Patterson maps calculated for both datasets allow the location of the heavy atoms in the unit cell to be determined. This

information is used to calculate the contribution of the heavy atoms to the structure factors so that the phases can be computed.

Anomalous dispersion can also be used to solve the phase problem. This technique also relies on the presence of heavy atoms within the crystal such as zinc, copper or iron. Selenium is often used in anomalous dispersion as methionine residues can be replaced by selenomethionine in cloned proteins during translation. In multi wavelength anomalous dispersion (MAD), multiple data sets from one crystal are collected at different wavelengths (peak, edge and remote). Comparison of the data sets allows phase calculations because the structure factors will be significantly altered at the absorption edge of the data set. This is because Friedel's law (F(h,k,l)=F(-h,-k,-l)) is broken when the wavelength is similar to the transition energy needed to bring an atom to an excited state. In single wavelength anomalous dispersion (SAD) one dataset is collected using a wavelength close to the absorption peak on the higher energy side and the phases are calculated using other techniques.

Once the phases have been computed the F_o-F_c and $2F_o$-F_c electron density maps for the unit cell are calculated. The F_o-F_c subtracts the calculated structure factor amplitudes away from the observed amplitudes. The $2F_o$-F_c electron density map subtracts the calculated structure factor amplitudes away from multiple numbers of the observed amplitudes (equation 7.4). Using both of the electron density maps, atoms are built into the electron density as well as possible. Structural refinement with geometrical restraints on bond lengths, angles and planarity is then performed. Throughout refinement of the structure the positions of atoms, B-factors and occupancy are altered, this should improve the phases.

As the structure is refined, R-factors are calculated. The R-factor is a statistical measurement of how correct the calculated amplitudes in the electron density maps are, in comparison to the observed amplitudes in the diffraction pattern (equation 7.5). The closer the R-factor is to 0 the better the calculated electron density map is to the observed amplitudes. The R-factor is biased because the

intensity measurements used to calculate the R-factor are also used in data processing and refinement. A second statistical R-factor called the R_{free} is calculated to overcome this problem. To calculate the R_{free} the amplitudes of 5% of reflection spots are used which are not included in data processing.

Once the R-factors are no longer improving, refinement is said to have converged. The structure is then validated to ensure that bond lengths, angles and stereochemistry are correct in comparison to calculated parameters.

Equation 7.4- The electron density $2F_o$-F_c map summation.

Equation 7.5- The R-factor equation.

7.2 Materials and Methods

7.2.1 Bioinformatics

To confirm the presence of conserved catalytic residues, amino acid sequence alignments of the L-haloacid dehalohenase from AQP5750 against the biochemically and structurally characterised L-haloacid dehalohenase from *B. cepacia* were performed using ClustalW, European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/clustalw2).

7.2.2 Culturing of AQP5750

Section 2.1.3 describes the bacterial growth of AQP5750. The bacterial paste was harvested (section 2.1.6) and the genomic DNA extracted (section 2.2.3).

7.2.3 Cloning of the L-haloacid dehalohenase from AQP5750

PCR primers were designed (section 2.2.4.1) to amplify the L-haloacid dehalohenase gene (see Appendix, 11.10), with the incorporation of the restriction sites *Not*I and *Nde*I for cloning into the pET-28a vector (Table 7.1). PCR reactions were performed using PCR composition 2 (section 2.2.4.4) with PCR program 1 (Table 2.2.4.2). The PCR product was run on a 1% agarose gel to visualise the amplified DNA (section 2.2.2). The PCR band was gel extracted (section 2.2.6) and ligated into the cloning vector pJET1.2/blunt (section 2.2.8.1). The DNA from the ligation reactions were transformed (section 2.2.8.3) into *E. coli* XL 10-Gold® chemically competent cells (section 2.2.7). The pJET1.2/blunt construct and the pET-28a vector were double digested with *Not*I and *Nde*I restriction enzymes (section 2.2.8.5 and 2.2.8.6) and the digested pET-28a was SAP treated (section 2.2.8.7). The ligation reaction into the expression vector pET-28a was performed (section 2.2.8.2) and the reaction transformed (section 2.2.8.3) into *E. coli* XL 10-Gold® competent cells (section 2.2.7). Gel extracted PCR bands (section 2.2.6) and plasmid constructs (section 2.2.8.4) were sent for Sanger sequencing (section 2.2.8.8).

Forward primer	*Nde*I 5'- **CATATG**ATGACCCCAAGCCACCCTGC -3'
Reverse primer	*Not*I 5'- **GCGGCCGC**TCAGGCGGTGAGGCTCGC -3'

Table 7.1- Forward and reverse primer sequences with incorporated restriction sites *Nde*I and *Not*I for the amplification of the L-haloacid dehalohenase in AQP5750. The restriction sites are highlighted in bold.

7.2.4 Over-expression of the L-haloacid dehalohenase from AQP5750

The extracted pET-28a construct (section 2.2.8.4) from *E. coli* XL 10-Gold® was transformed into BL21-CodonPlus (DE3)-RIPL (section 2.2.7). The plasmid was extracted (section 2.2.8.4) and sent for Sanger sequencing (section 2.2.8.8). The positively transformed strain of *E. coli* was preserved as glycerol stocks (section 2.1.5). Protein over-expression conditions were optimised (section 2.3.1) and the

resultant bacterial cell pellets were lysed with BugBuster (1/10 dilution with Tris-H_2SO_4 pH 8.2) as described in section 2.3.2. The soluble and insoluble protein fractions were analysed using SDS-PAGE (section 2.3.6).

7.2.5 Purification of the L-haloacid dehalohenase from AQP5750

A 2 L culture of transformed *E. coli* strain BL21 (DE3)-RIPL was grown (section 2.1.4). The protein expression was induced (2.3.1) and the bacterial paste harvested (section 2.1.6). The bacterial cell pellet was resuspended in NAC buffer A (table 2.7) to a final concentration of 10% (w/v) and the cells lysed by sonication (section 2.3.2). The soluble protein was purified by nickel affinity chromatography and the fractions corresponding to the L-haloacid dehalohenase were concentrated (section 2.3.3.8). The concentrated protein was further purified using a Superdex 75 GF chromatography column (section 2.3.3.7).

7.2.6 Biochemical characterisation of the L-haloacid dehalohenase from AQP5750

Protein parameters including amino acid composition, theoretical pI and extinction coefficient were calculated (section 2.3.4) based on the L-haloacid dehalohenase protein sequence (see Appendix, 11.11). The purified protein concentration was determined (section 2.3.5) and the protein diluted to 0.2 mg/ml. This stock solution was stored at 4ºC for no longer than 72 h and was used in all biochemical characterisation reactions.

Section 2.3.9 describes the methods used for the biochemical characterisation of the L-haloacid dehalohenase from AQP5750. The thermostability (section 2.3.9.2), solvent stability (section 2.3.9.3), substrate specificity (section 2.3.9.4) and the temperature optimum (section 2.3.9.5) were determined. Michaelis-Menten kinetics were also calculated (section 2.3.9.6) using bromoacetic acid as the substrate and the standard dehalogenase reaction was carried out at 55ºC.

7.2.7 Crystallization

7.2.7.1 Crystal screening

Hampton 96 well plate microbatch crystallization screens were conducted using an Oryx 6 automated crystallization robot (Douglas Research, UK). Commercial screens JCSG, PACT PREMIER, SIGMA, pHClear and pHclear II were used to screen for protein crystallization. The final droplet consisted of 0.5 µl of protein in GF buffer (Table 7.2) and 0.5 µl of crystallization reagent. The droplets were covered in Al's oil, (1:1 ratio of silicon oil and paraffin oil) incubated at 18ºC and checked regularly for crystal growth using a light microscope.

7.2.7.2 Preparing crystals for data collection

Crystals were removed from the droplet and placed in a cryo-protectant before being directly frozen in liquid nitrogen using a nylon loop.

7.2.7.3 Crystal soaks

Crystals were removed from the droplet and placed in an appropriate cryo-protectant (Table 7.2) containing 25 mM of l-chloropropionic acid (MCP) for 1 h before being directly frozen in liquid nitrogen on a nylon loop.

7.2.8 X-ray data collection

Data were collected at beamlines I02, I03 and I04-1 at the Diamond Light Synchrotron Radiation Source (www.diamond.ac.uk). The I02 beamline has a ADSC Q315r detector (charged couple detector), the I03 beamline has a Pilatus 6M-F detector and the I04-1 beamline has a Pilatus dectris P2M detector. Table 7.2 displays the X-ray data collection conditions for the L-haloacid dehalohenase crystals.

Crystal	L-haloacid dehalohenase	L-haloacid dehalohenase complex with sulfate	L-haloacid dehalohenase complex with MCP
Cryo-protectant	0.2 M LiCl 0.1 M Tris-H_2SO_4 18% PEG 6000 25% PEG 400 pH 8	0.2 M LiCl 0.1 M Tris-H_2SO_4 18% PEG 6000 25% PEG 400 25 mM MCP pH 4.5	0.2 M LiCl 0.1 M Tris-HCl 18% PEG 6000 25% PEG 400 25 mM MCP pH 4.5
Soaking time	N/A	1 h	1 h
Beam	Diamond I02	Diamond I04-1	Diamond I03
Wavelength (Å)	0.92	0.97	0.98
No. Frames	360	3600	600
Oscillation in degrees	0.5	0.1	0.5

Table 7.2- X-ray data collection conditions for the AQP5750 L-haloacid dehalohenase crystals.

7.2.9 Structure determination

7.2.9.1 Data processing

Data were processed using the program Xia2 version 0.3.3.0 (Winter *et al*., 2010), which uses the data reduction software programs XDS (Kabsch, 1993), MOSFLM (Leslie, 1992), Labelit (Sauter *et al*., 2004), Pointless (Evans, 2005) and CCP4 (Collaborative Computational Project, Number 4. 1994).

7.2.9.2 Phase determination

The program MOLREP, version 10.2 was used for molecular replacement (Vagin and Teplyakov, 1997). MOLREP carries out rotational and translational searches to map the model structure to the electron density calculated for the experiment.

7.2.9.3 Model building and refinement

The $2F_o$-F_c and F_o-F_c electron density maps were calculated and the protein structure was built to give the best fit. Maximum likelihood refinement was carried out in REFMAC, version 5.6 (Murshudov *et al*., 1997) and the model was manually rebuilt in COOT (Emsley and Cowtan, 2004). ARP/wARP, version 7.0.1 was used to improve electron density maps (Cohen *et al*., 2008). Solvent molecules were added manually and the structure examined for the presence of intermediate complexes in the active site. The intermediate structure and the dictionary definition for MCP were built using JLigand, version 1.0.20 (A. Lebedev, http://www.ysbl.york.ac.uk/mxstat/JLigand/). Intermediate structures were built and fitted in accordance with both the F_o-F_c and $2F_o$-F_c electron density maps.

7.2.9.4 Structure validation

Post refinement, the quality of the model was checked using the program PROCHECK, version 3.3 (Laskowski *et al*., 1993).

7.3 Results

7.3.1 Bioinformatics

The L-haloacid dehalohenase from AQP5750 has 62% amino acid sequence identity to the L-haloacid dehalohenase from *Sulfitobacter* sp. EE-36. This protein has not been biochemically or structurally characterised. The L-haloacid dehalohenase from AQP5750 has 22% sequence identity to the L-haloacid dehalohenase from *B. cepacia*. The *B. cepacia* L-haloacid dehalohenase has been biochemically characterised and the crystal structure solved at a resolution of 1.93 Å (Schmidberger *et al*., 2007). Protein sequence alignments between the L-haloacid dehalohenases from AQP5750 and *B. cepacia* were performed using ClustalW, European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/clustalw2) (Figure 7.4). The catalytic Asp 11 in the L-haloacid dehalohenase from *B. cepacia* is conserved and proposed to be Asp 18 in the AQP5750 L-haloacid dehalohenase. Some of the equivalent catalytically important residues in the L-haloacid dehalohenase from *B. cepacia*

are different in the AQP5750 L-haloacid dehalohenase. Therefore, due to the differences observed in catalytic amino acid residues, the AQP5750 L-haloacid dehalohenase might have a slightly different mechanism in comparison to the *B. cepacia* L-haloacid dehalohenase. The crystal structure of the AQP5750 L-haloacid dehalohenase would confirm the orientation and positioning of the amino acids in the active site, so that the mechanisms can be compared and any potential differences highlighted.

```
sequence1       --MTPSHPARPSRSGILVFDVNETLLDLTSLSPLFERVFGDAKVLR--EWFPELILYSQT 56
sequence2       MDYKDDDDKLVDSLRACVFDAYGTLLDVHSAVMRNADEVGASAEALSMLWRQRQLEYSWT 60
                   . ..    .     ***.  ****: *        .* :       *  . : ** *

sequence1       LTLTGLYRPFGEIAAAVFEMVAANHQAKVTPDDIAELKTRLTSMPAYPDVAPALTRLQDA 116
sequence2       RTLMHQYADFWQLTDEALTFALRTYHLEDRKGLKDRLMSAYKELSAYPDAAETLEKLKSA 120
                 **   *  * :::  .: :.  .:: :   .   .* :  ..:.****.* :* :*:.*

sequence1       GFRLVTLTNSAPSPAPSPLEKAGIASFFEAHLTVHSSQRFKPHPSVYDSTAETLGAKPEE 176
sequence2       GYIVAILSNGNDEMLQAALKASKLDRVLDSCLSADDLKIYKPDPRIYQFACDRLGVNPNE 180
                *: :. *:*.  .   :.*: : :  .::: *:... : :**.* :*: :.: **.:*:*

sequence1       LCMIACHIWDTIGAQARGWRGGFVARPHNTPLTLAEVPQPDFIGRDMGELADQLIASLTA 236
sequence2       VCFVSSNAWDLGGAGKFGFNTVRINRQGNPPE--YEFAPLKHQVNSLSELWPLLAKNVTK 238
                :*:::.:. **  **   *:.   : *  *.*    *..  ..  ..:.**   *  .:*

sequence1       --
sequence2       AA 240
```

Figure 7.4- ClustalW amino acid sequence alignment of the L-haloacid dehalohenase from AQP5750 (sequence 1) and *B. cepacia* (sequence 2). The residues that are essential for catalysis in the L-haloacid dehalohenase from *B. cepacia* are highlighted in red and equivalent residues in the L-haloacid dehalohenase from AQP5750 are highlight in yellow. Residues in the active site of the AQP5750 L-haloacid dehalohenase which are different are highlight by the black boxes.

7.3.2 Cloning and over-expression of the L-haloacid dehalohenase from AQP5750

PCR was used to amplify the L-haloacid dehalohenase gene from AQP5750 and a band of the correct size was observed on a 1% agarose gel (Figure 7.5). The amplified gene was then cloned into the cloning vector pJET1.2/blunt and the expression vector pET-28a with the incorporation of an N-terminal His-tag. Over-expression trials of the L-haloacid dehalohenase in BL21 (DE3)-RIPL were conducted by varying the induction conditions. BL21 (DE3)-RIPL expression strain contains the T7 promoter and is induced using IPTG.

As discussed in section 6.3.2, proteins encoded by rare codons can limit the expression in *E. coli*. The expression cell line BL21-CodonPlus (DE3)-RIPL contains extra copies of the genes that code for the tRNAs that may slow down the translation of proteins. There are 10 rare proline codons (proL) in the AQP5750 L-haloacid dehalohenase (encoded by CCC). None of the rare proline codons are repeated next to each other. BL21-CodonPlus (DE3)-RIPL contains extra copies of the proL (CCC) tRNA gene, allowing higher levels of expression of heterologous proteins encoded by this rare codon. Expression of the L-haloacid dehalohenase in BL21-CodonPlus (DE3)-RIPL was very good with best expression observed with 1 mM IPTG at 28ºC, with an incubation time of 20 h post induction (Figure 7.6, lane 8).

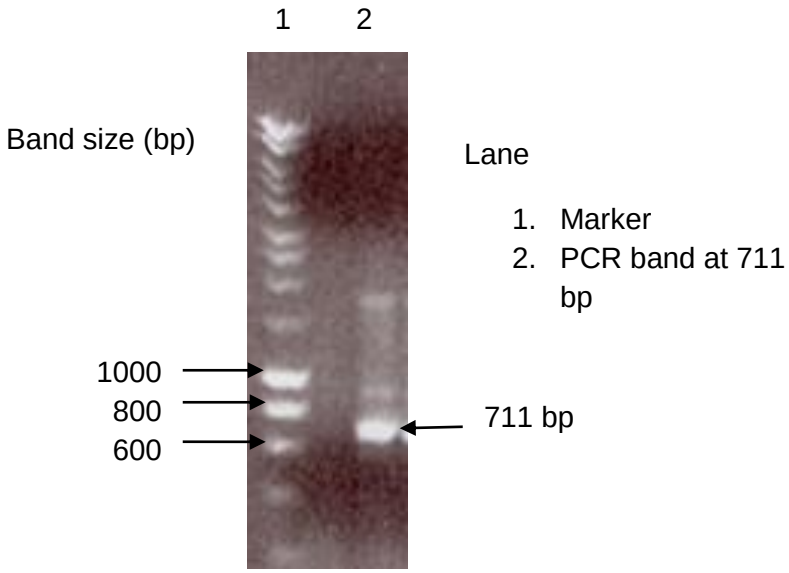


Figure 7.5- 1% agarose gel showing the amplified l-haloacid dehalogenasae gene from AQP5750 at 711 bp.

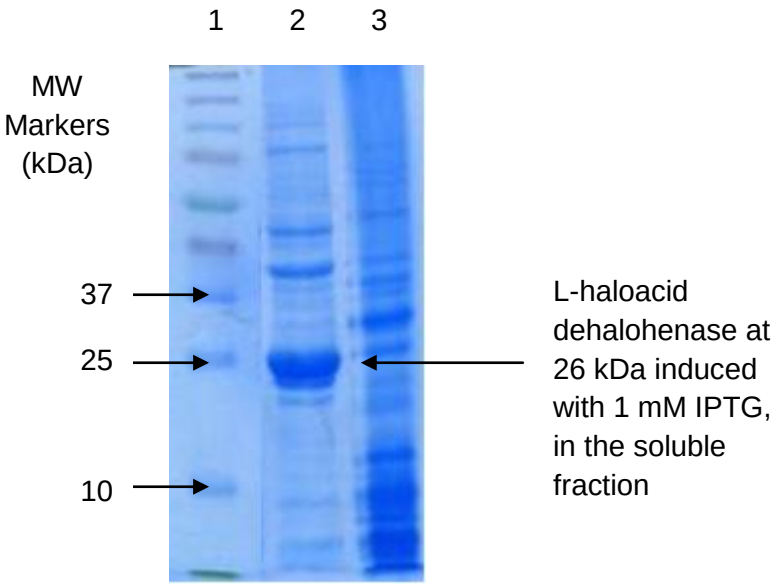


Figure 7.6- SDS-PAGE gel showing the over-expression of the AQP5750 L-haloacid dehalohenase. Lane 1- MW Marker, lane 2- 20 h post induction (soluble fraction), lane 3- 20 h post induction (insoluble fraction). The best over-expression was observed in the soluble fraction of the BL21-CodonPlus (DE3)-RIPL culture induced with 1 mM IPTG, after 20 h at 28ºC.

7.3.3 Purification of the L-haloacid dehalohenase from AQP5750

6 g of cell paste was harvested from 2 L of recombinant BL21-CodonPlus (DE3)-RIPL by centrifugation (section 2.1.6). Bacterial cells were lysed by sonication and the soluble protein purified by nickel affinity chromatography (Figure 7.7). Fractions 56-74 assayed positive for L-haloacid dehalohenase activity, the protein peak was pooled together and a 5 µl sample was analysed using SDS-PAGE (Figure 7.8, lane 3). The pooled fractions from the nickel affinity chromatography column were concentrated using a centrifugal concentrator with a 10 kDa membrane (Vivaspin 20; Viva science) at 3,000 x *g*, at 4ºC until the final volume reached 1 ml. The concentrated protein sample was purified on a Superdex 75 GF chromatography column (Figure 7.9).

Fractions eluting between 40-54 ml assayed positive for dehalogenase activity, suggesting that the protein eluted in the form of a dimer. The fractions from the protein peak were pooled together and a 5 µl sample analysed using SDS-PAGE (Figure 7.8, lane 4). 50 mg of the purified AQP5750 L-haloacid dehalohenase was obtained from a 2 L culture of recombinant BL21-CodonPlus (DE3)-RIPL.

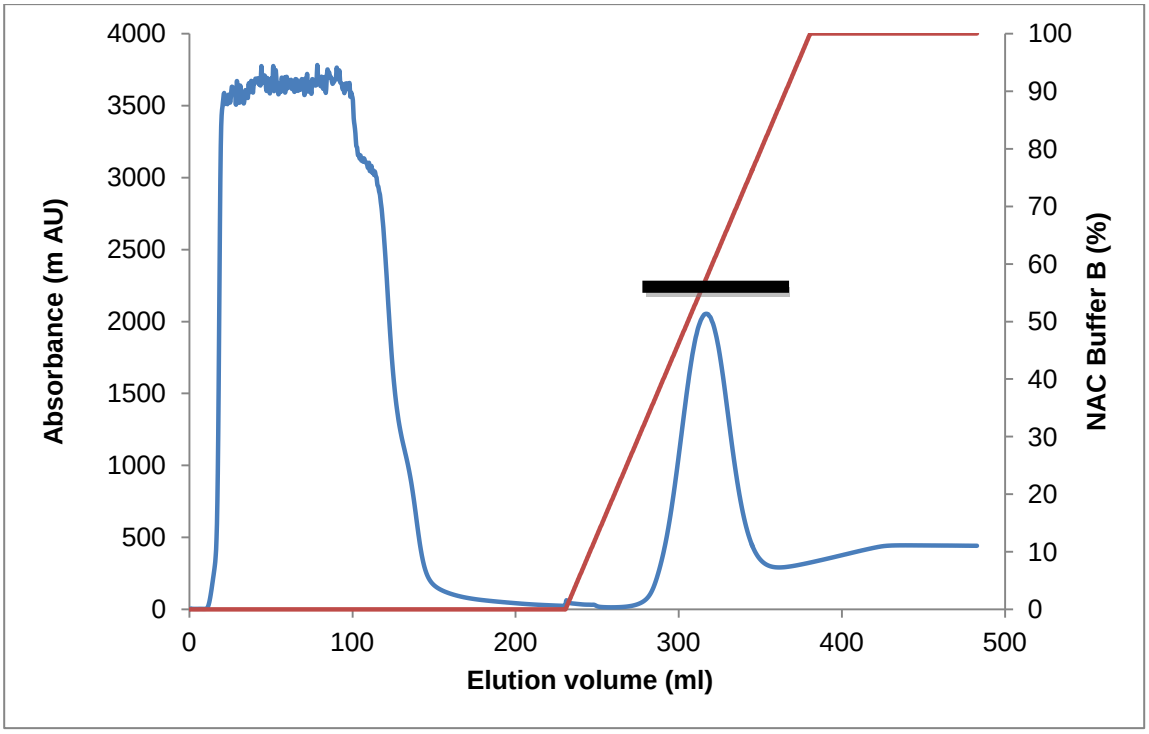


Figure 7.7- Elution profile from the purification of the AQP5750 L-haloacid dehalohenase from the nickel affinity chromatography column. The red line shows the concentration of NAC buffer B. Fractions 56-74 (280-370 ml) containing the L-haloacid dehalohenase are highlighted by the black bar.

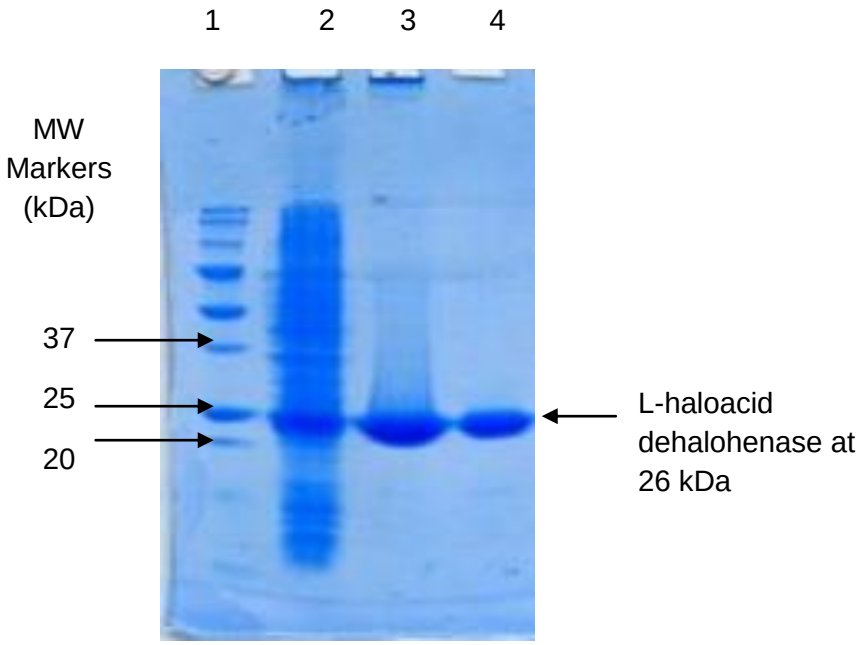


Figure 7.8- SDS-PAGE gel of the purified AQP5750 L-haloacid dehalohenase. Lane 1- MW Marker, Lane 2- crude protein extract, lane 3- protein peak of the L-haloacid dehalohenase from the nickel affinity chromatography column, lane 4- protein peak of the L-haloacid dehalohenase from the Superdex 75 GF chromatography column.

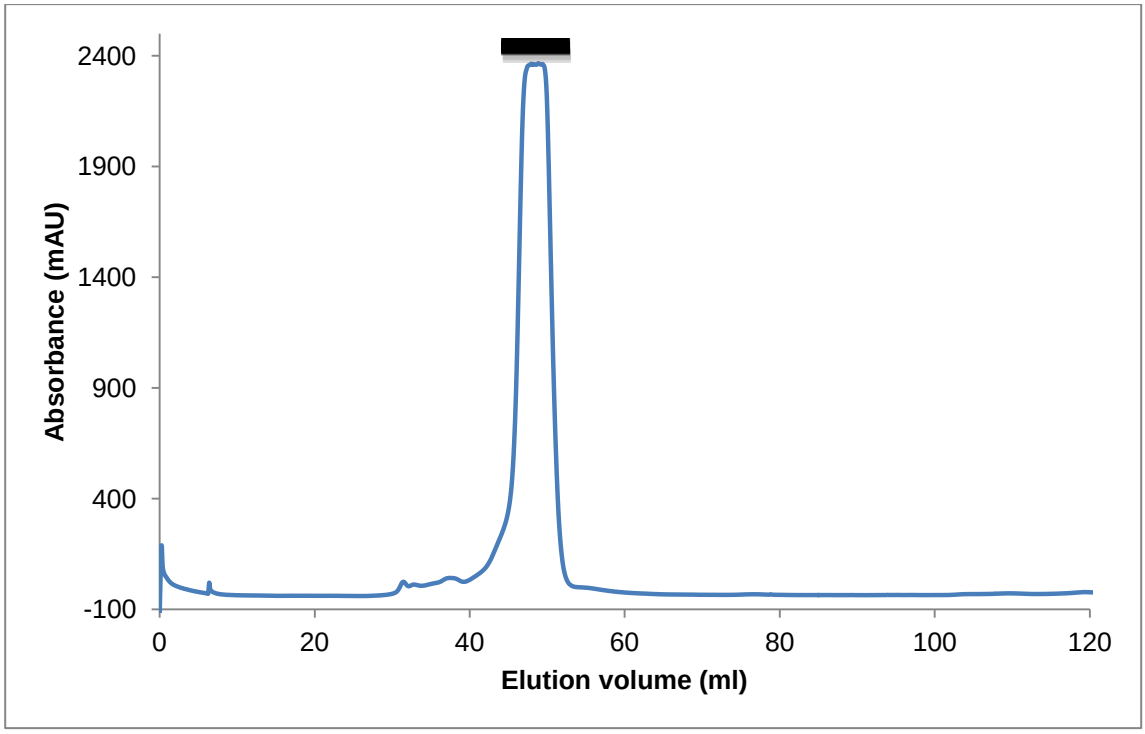


Figure 7.9- Elution profile from the purification of the AQP5750 L-haloacid dehalohenase from the GF chromatography column (Superdex 75). Fractions 40-54 containing the L-haloacid dehalohenase are highlighted by the black bar. 1 ml fractions were collected throughout GF chromatography.

7.3.4 Biochemical characterisation of the L-haloacid dehalohenase from AQP5750

The colorimetric dehalogenase assay is based on the method described by Holloway *et al.* (1998). It measures the decrease in pH as the dehalogenase turns over the substrate and a halide/proton product is formed. This can be visualised by eye as the assay solution changes from red (at a pH ≥8) to yellow (at a pH ≤4). The assay can also be followed spectrophotometrically at 540 nm. Enzyme activity is either shown as a percentage compared to a control reaction or in EU. 1 EU is defined as the production of 1 mM product/min/mg of enzyme.

7.3.4.1 Substrate specificity of the L-haloacid dehalohenase from AQP5750

The activity of the AQP5750 L-haloacid dehalohenase towards substrates (at a concentration of 10 mM) with varying halogens (chlorine, bromine and iodine), carbon chain lengths and enantiomers were tested (Figure 7.10) (structures shown in Appendix, 11.6). The L-haloacid dehalohenase from AQP5750 shows activity towards the l-enantiomer of the substrate only, with the halogen atom attached at the α-carbon.

The L-haloacid dehalohenase from AQP5750 shows highest activity towards monobromoacetic acid (this reading was taken as 100% in comparison to all other substrates). Following monobromoacetic acid, the AQP5750 L-haloacid dehalohenase showed decreasing activity towards monochloroacetic acid (71%), S-bromopropionic acid (71%), S-chloropropionic acid (10%) and dichloroacetic acid (9%). The L-haloacid dehalohenase from AQP5750 displayed no activity towards longer chain substrates such as 2-chlorobutyric acid. The enzyme has a higher activtity towards brominated substrates and displayed no activity towards 2-chlorobutyric acid, trichloroacetic acid, 3-bromopropionic acid, sodium fluoroacetate, R-chloropropionic acid, R-bromopropionic acid, 1,2-dichloropropane and 1-bromoheptane.

In comparison, the L-haloacid dehalohenase from *P. ingrahamii* (20% amino acid sequence identity) shows highest activity towards substrates with shorter carbon chain lengths, without preference towards a chlorine or bromine at the α-carbon position (Chapter 6, Figure 6.8). The L-haloacid dehalohenase from *B. cepacia* (22%) shows highest activity towards brominated substrates such as monobromoacetic acid and 2-bromopropionic acid (Tsang *et al.*, 1999). The L-haloacid dehalohenases from *Pseudomonas* sp. strain YL (18%) and *S. tokodaii* (19%) show highest activity towards long chain substrates such as 2-chlorobutryic acid (Liu *et al*., 1994, Rye *et al*., 2009). In contrast, the L-haloacid dehalohenase from *X. autotrophicus* (20%) prefers substrates with smaller carbon chain lengths, showing highest activity towards dibromoacetic acid and dichloroacetic acid (Van der Ploeg *et al*., 1991).

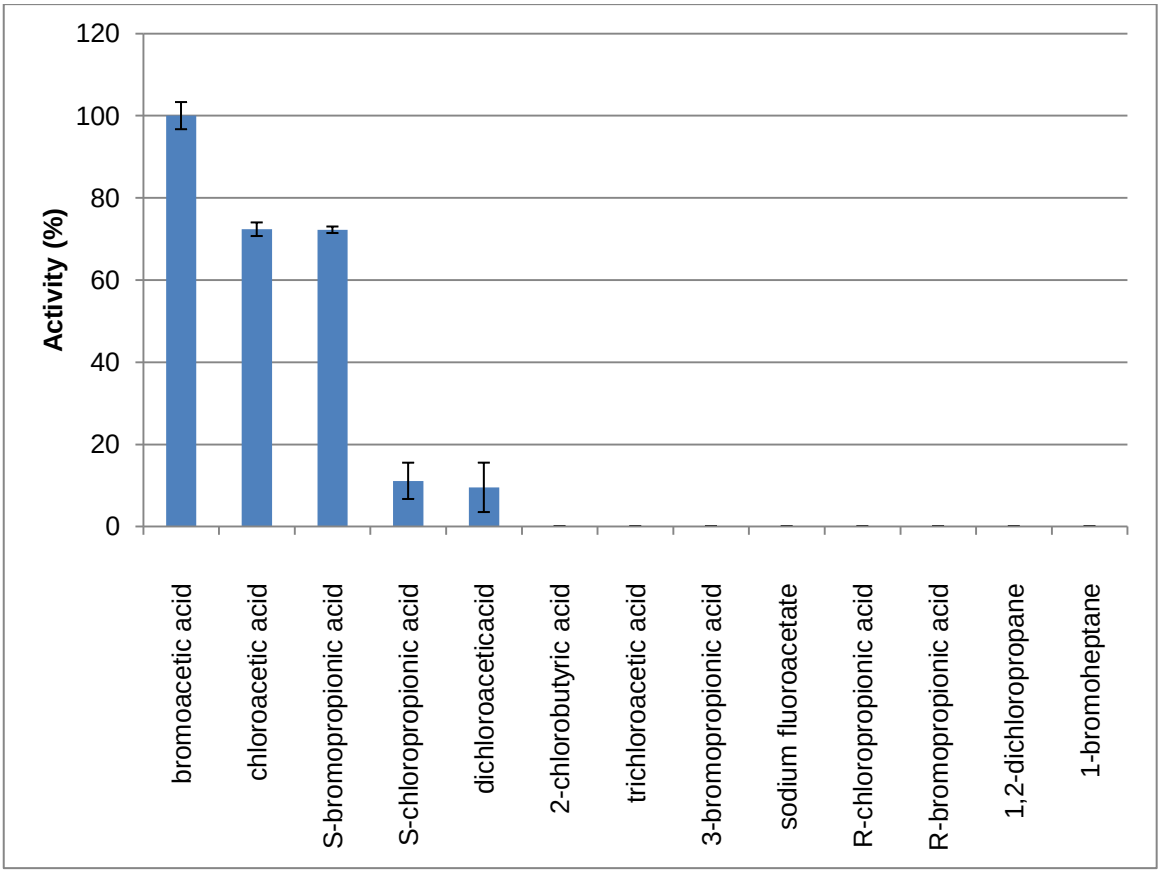


Figure 7.10- Graph showing the substrate specificity of the L-haloacid dehalohenase from AQP5750.

7.3.4.2 The thermostability of the L-haloacid dehalohenase from AQP5750

Thermal stability assays were carried out using pre-incubated assay solutions, with 10 mM of chloroacetic acid as the substrate. The thermostability of the L-haloacid dehalohenase from AQP5750 was determined by incubating the enzyme between 30-70ºC for 30, 60 and 90 min (Figure 7.11). Enzyme activity is displayed as a percentage in comparison to untreated protein.

The AQP5750 L-haloacid dehalohenase has high thermal stability, retaining 85% of activity when incubated at 50ºC for 90 min. The enzyme was less stable at 60ºC, showing a loss of 86% of activity after incubation for 90 min.

In comparison, the L-haloacid dehalohenase from the thermophilic archaeon *S. tokodaii* shows high thermal stability, retaining 92% of activity after incubation at 50ºC for 90 min (Rye *et al*., 2009). The L-haloacid dehalohenase from mesophillic *Pseudomonas* sp. YL is thermostable, retaining 100% activity when incubated at 60ºC for 30 min (Liu *et al*., 1994). The L-haloacid dehalohenase from the psychrophilic bacterium *P. ingrahamii* is also relatively thermostable, retaining 87% of activity when incubated at 50ºC for 90 min.

AQP5750 L-haloacid dehalohenase is a relatively thermostable enzyme from a mesophilic marine bacterium in the *Rhodobacteraceae* family. As with the L-haloacid dehalohenases from *P. ingrahamii* and *Pseudomonas* sp. YL, this provides evidence that enzymes from non-thermophilic microorganisms can be thermostable.

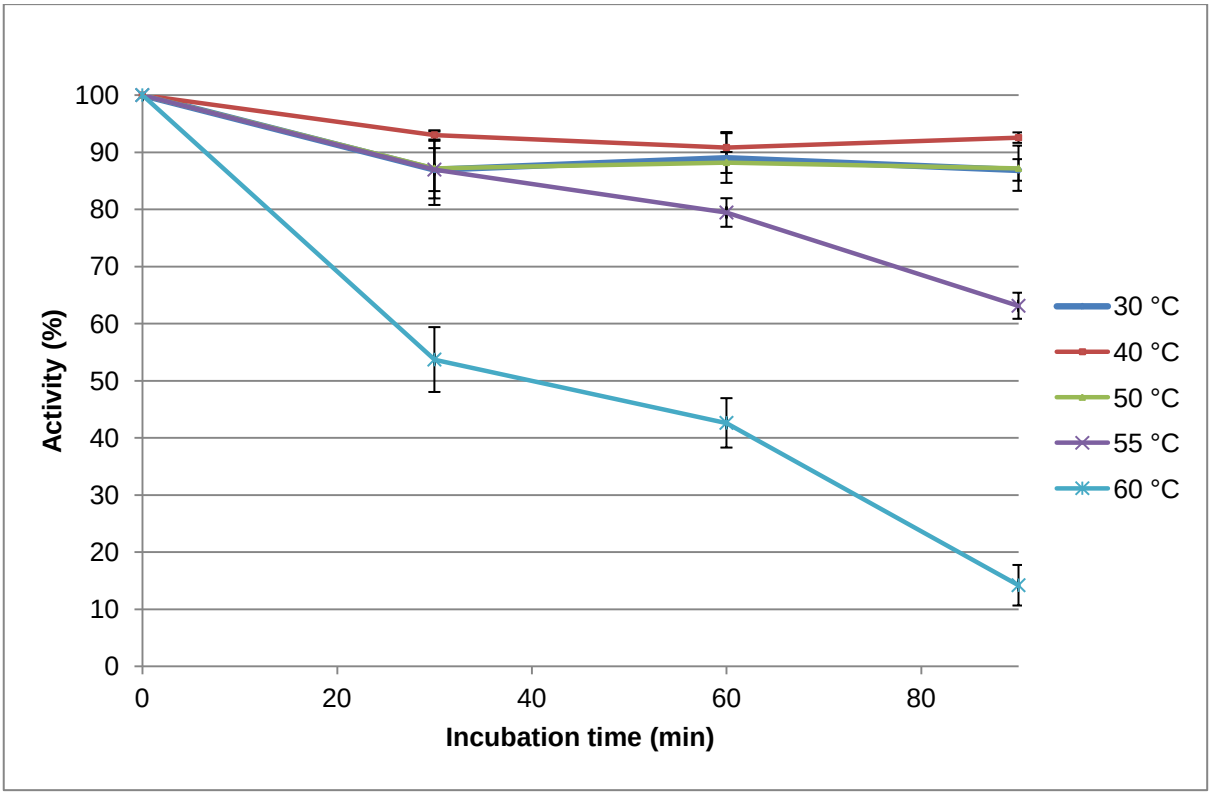


Figure 7.11- Graph showing the thermostability of the L-haloacid dehalohenase from AQP5750.

7.3.4.3 Solvent stability of the L-haloacid dehalohenase from AQP5750

To characterise the solvent stability of the L-haloacid dehalohenase from AQP5750 the enzyme was incubated in ethanol, methanol, DMSO and acetonitrile at varying concentrations between 10-70% for 1 h (Figure 7.12). Following incubation, the enzyme was assayed for L-haloacid dehalohenase activity with 10 mM chloroacetic acid and the results displayed as a percentage compared to untreated enzyme. The L-haloacid dehalohenase from AQP5750 showed increased activity of 150% in 40% DMSO and 123% in 30% methanol. These solvents may increase the availability of the substrate (without denaturing the enzyme) as chloroacetic acid is more soluble in DMSO and methanol in comparison to water. The enzyme was stable up to 20% ethanol and 10% acetonitrile.

In comparison, the L-haloacid dehalohenase from *P. ingrahamii* retains 95%, 75% and 55% of activity after incubation with 40% DMSO, methanol and ethanol for 1 h (chapter 6, Figure 6.10). The L-haloacid dehalohenase from *S. tokodaii* is less stable in organic solvents, showing a loss of 60% activity towards l-chloropropionic acid when incubated for 1 h in 10% ethanol, methanol and acetonitrile (Rye *et al*., 2009). The L-haloacid dehalohenase from *P. putida* shows higher activity towards substrates with longer carbon chain lengths in the presence of anhydrous DMSO (Hasan *et al*., 1991).

The AQP5750 L-haloacid dehalohenase is the first reported case of increased L-haloacid dehalohenase activity in solvent stability experiments.

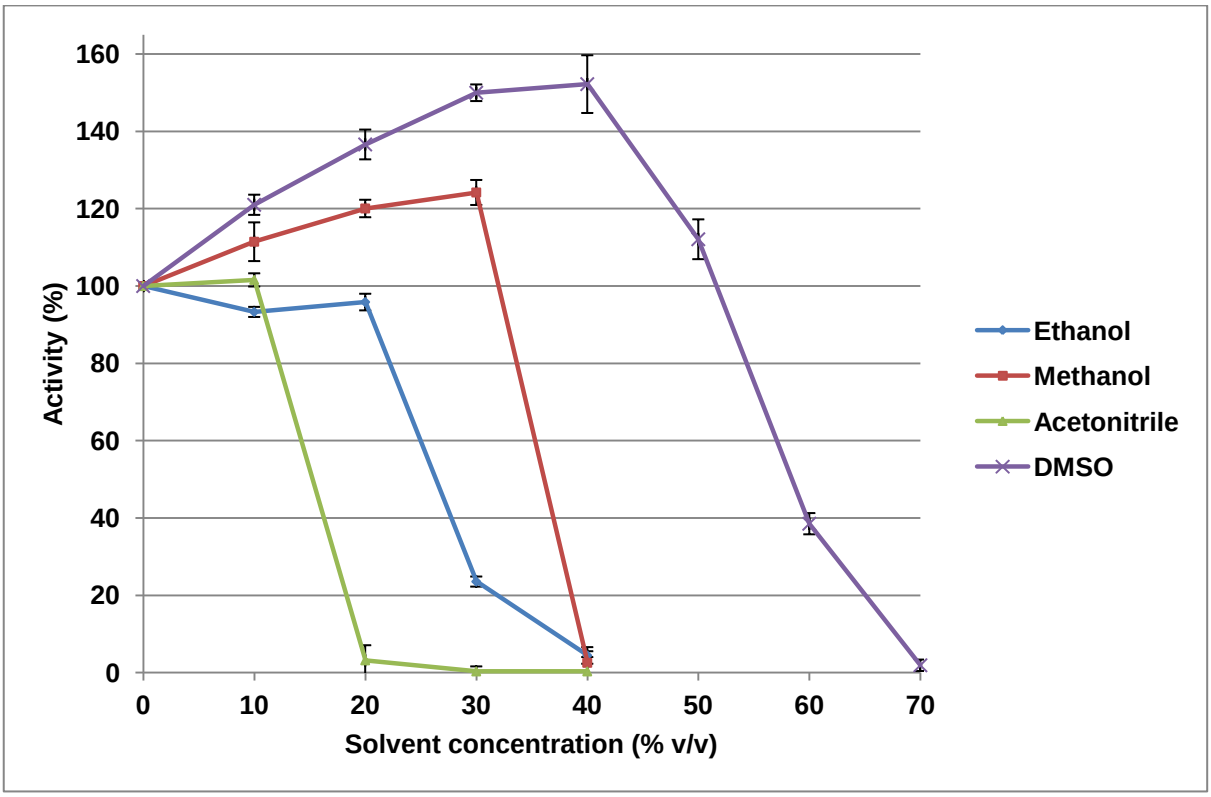


Figure 7.12- Graph showing the solvent stability of the L-haloacid dehalohenase from AQP5750.

7.3.4.4 Temperature optimum of the L-haloacid dehalohenase from AQP5750

The temperature optimum of the AQP5750 L-haloacid dehalohenase was determined by assaying the enzyme between 25-65ºC (Figure 7.13). Results are displayed in EU. 1 EU is the amount of enzyme required (mg) to produce 1 mM product/min.

The optimal temperature of the AQP5750 L-haloacid dehalohenase is 55ºC. In comparison, the thermostable L-haloacid dehalohenase from the thermophilic archaeon *S. tokodaii* has a temperature optimum of 60ºC (Rye *et al*., 2009). The L-haloacid dehalohenase from mesophilic *Pseudomonas* sp. YL has a temperature optimum of 65ºC (Liu *et al*., 1994) and the L-haloacid dehalohenase from the psychrophilic *P. ingrahamii* has a temperature optimum of 45ºC.

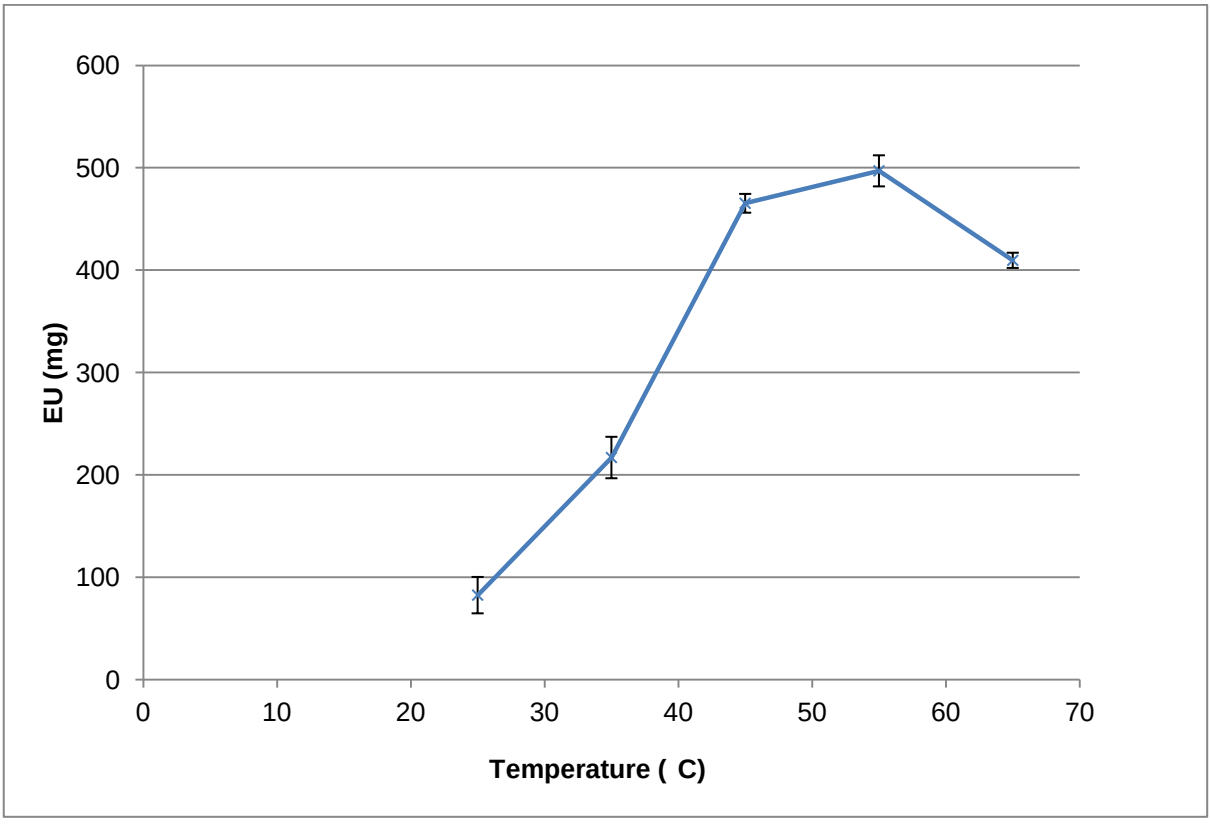


Figure 7.13- Graph showing the temperature optimum of the L-haloacid dehalohenase from AQP5750.

7.3.4.5 The Michaelis-Menten kinetics of the L-haloacid dehalohenase from AQP5750

The Michaelis-Menten kinetics of the AQP5750 L-haloacid dehalohenase was determined by assaying the enzyme at 55ºC, with bromoacetic acetic acid at concentrations between 1.2-10 mM (Figure 7.14). The V_{max} and K_m were calculated to be 1.75 µM min^{-1} mg^{-1} and 6.72 mM. The L-haloacid dehalohenase from *Pseudomonas* sp. YL, *B. cepacia, S. tokodaii* and have K_m values of 1.1 mM, 1.13 mM and 6.23 mM respectively (Liu *et al*., 1997, Tsang *et al*., 1999, Rye *et al*., 1999).

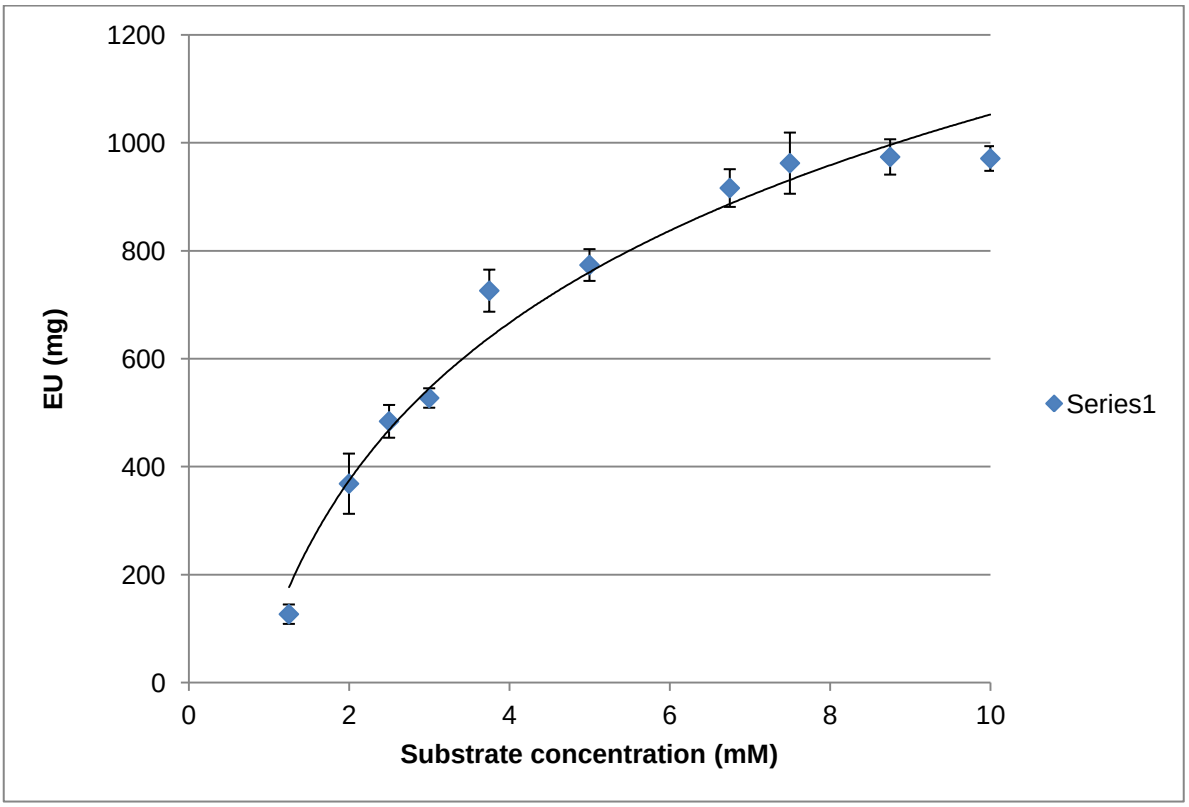


Figure 7.14- Graph showing the activity of the L-haloacid dehalohenase from AQP5750 with increasing bromoacetic acid concentrations (1.25-10 mM).

7.3.5 Crystallization

Purified protein in GF buffer (Table 2.7) was concentrated to 10 mg/ml. Microbatch crystallization trials were conducted using a number of commercial screens using a 1:1 ratio of protein to crystallization condition. Protein crystals were observed in many conditions with varying concentrations of salts, buffers and precipitants. Figure 7.15 shows the crystals that diffracted to the highest resolution grown in 0.2 M LiCl, 0.1 M Tris, 20% PEG 6000 at pH 8.

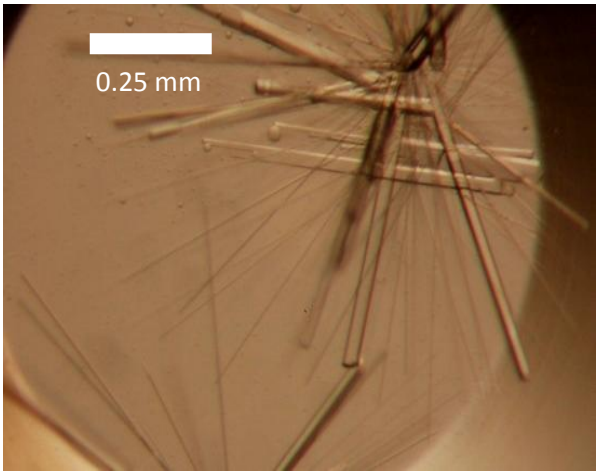


Figure 7.15- Crystals of the L-haloacid dehalohenase from AQP5750 obtained from 0.2 M LiCl, 0.1 M Tris, 20% PEG 6000 at pH 8.

7.3.5.1 Crystal Soaks

Crystal soaking experiments were performed with crystals from the same condition as described in section 7.2.7.3 (Figure 7.15). The composition of cryo-protectants containing 25 mM MCP are given in Table 7.2.

7.3.6 Structure determination of the L-haloacid dehalohenase from AQP5750

7.3.6.1 X-ray data collection

X-ray datasets were collected as described inTable 7.2. All crystals belong to the space group $P2_12_12_1$. Table 7.3 describes the data processing statistics for the native L-haloacid dehalohenase, the L-haloacid dehalohenase complex with a sulfate bound in the active site and the L-haloacid dehalohenase complex with the intermediate MCP covalently bound in the active site.

Crystal	Native L-haloacid dehalohenase	L-haloacid dehalohenase complex with sulfate	L-haloacid dehalohenase complex with MCP
Resolution (Å)	39.93-1.79 (1.83-1.79)	55.34-1.64 (1.68-1.64)	41.97-1.66 (1.71-1.66)
Cell dimension $P2_12_12_1$	a, b, c= 43.8 Å, 68.3 Å, 159.7 Å	a, b, c= 42.8 Å, 68.6 Å, 283.3 Å	a, b, c= 42.8 Å, 68.9 Å, 157.2 Å
No. of protomers	2	4	2
Solvent content (%) V_M ($Å^3$ Da^{-1})	45.4 (2.2)	37.1 (1.9)	45 (2.2)
No. of measured reflections	327159	1317980	531159
No. of unique reflections	46369	103191	50717
Completeness	99.8 (99.8)	99.9 (99.8)	99.9 (100)
(I)/ δ (I)	15.1 (2.8)	18.2 (3.4)	19.5 (3.7)
R_{merge} † (%)	0.085 (0.728)	0.075 (0.83)	0.073 (0.67)

Table 7.3- Data processing statistics for the L-haloacid dehalohenase from AQP5750. Values for the outer resolution shell are given in brackets. $R_{merge}=\Sigma_h\Sigma_J|<I_h>-I_J(h)|/\Sigma_h\Sigma_J I(h)$, where I(h) is the intensity of the reflections h, Σ_h is the sum over all the reflections and Σ_J is the sum over J measurements of the reflections.

7.3.6.2 Structure solution

A modified dimeric polyalanine L-haloacid dehalohenase from *B. cepacia* (PDB: 2NO4) model minus the N and C terminus (residues 1-8 and 202-225) was used for molecular replacement using the program MOLREP, version 10.2 (Vagin and Teplyakov, 1997). The rotation function was calculated using a radius of 21 Å at a resolution of 2.5 Å. A translation search at a resolution of 5 Å identified the correct rotation solution as 24, it had a correlation at 28.6%, with the background solution at 20.2%. Refinement of the structure solution using REFMAC version 5.6 (Murshudov *et al*., 1997) was attempted but was unsuccessful. Non-crystallographic symmetry averaging and rigid body refinement was also attempted to try to improve the structure solution but this was also unsuccessful.

The program MOLREP, version 10.2 (Vagin and Teplyakov 1997) was used to modify the 2NO4 monomer according to the sequence of the L-haloacid dehalohenase from AQP5750. The resulting align.pdb was superimposed onto the structure from molecular replacement and subsequently refined.

Further L-haloacid dehalohenase structure complexes were solved using the final structure of the L-haloacid dehalohenase from AQP5750 as a model in molecular replacement.

7.3.7 Model building and validation

7.3.7.1 Native structure of the L-haloacid dehalohenase

The structure solution was manually built in COOT (Emsley and Cowtan, 2004) and refined using REFMAC version 5.6 (Murshudov *et al*., 1997). The R-factor and R_{free} were 41.6% and 46.2% respectively. The solution was then submitted to automated refinement using ARP/wARP version 7.0.1 (Cohen *et al*., 2008). The resulting R-factor and R_{free} were 20.8% and 24.8% respectively. Further modelling and refinement was carried out resulting in a final R-factor and R_{free} value of 16.3% and 20.9% respectively. The G-factor was calculated at 0.1 (confirming that the structure has normal stereochemical properties) using PROCHECK version 3.3 (Laskowski *et al*., 1993).

No density was observed in the active site of native L-haloacid dehalohenase structure. The final refinement statistics for the native L-haloacid dehalohenase and validation results are shown in Table 7.4 and the Ramachandran plot in Appendix, 11.12.

7.3.7.2 Complex structure of the L-haloacid dehalohenase with sulfate

Modelling and refinement was carried out as described in the materials and methods section 7.2.9.3. Clear electron density was observed for a sulfate ion in the active site of the L-haloacid dehalohenase in all four chains (A, B, C and D). The final refinement statistics for the L-haloacid dehalohenase and validation results are shown in Table 7.4 and the Ramachandran plot in Appendix, 11.13. The G-factor was calculated at 0.2 (confirming that the structure has normal stereochemical properties) using PROCHECK version 3.3 (Laskowski *et al*., 1993).

7.3.7.3 Complex structure of the L-haloacid dehalohenase with MCP

Modelling and refinement was carried out as described in the materials and methods section 7.2.9.3. Clear electron density was observed for the intermediate MCP covalently bound to Asp 18 in the active site in both chains A and B in the AQP5750 L-haloacid dehalohenase structure. The final refinement statistics for the L-haloacid dehalohenase and validation results are shown in Table 7.4 and the Ramachandran plot in Appendix, 11.14. The G-factor was calculated at 0.1 (confirming that the structure has normal stereochemical properties) using PROCHECK version 3.3 (Laskowski *et al*., 1993).

The Ramachandran plot showed Glu 21 in the L-haloacid dehalohenase native and complex structures as an outlier. This residue is in the active site pocket of the L-haloacid dehalohenase and is clearly defined by the electron density map as being in the correct conformation.

Crystal	Native L-haloacid dehalohenase	L-haloacid dehalohenase complex with sulfate	L-haloacid dehalohenase complex with MCP
Resolution (Å)	1.8	1.64	1.66
Overall R-factor (%)	16.3	16.8	17.7
R_{free} (5.1 % total data)	20.9	20.7	20.2
No. of residues (specific residue name and number)	A (R8-T234) B (A11-T234)	A (S12-A236) B (A8-A236) C (S12-A236) D (S12-A236)	A (S12-A236) B (S12-A236)
No. of waters modelled	746	884	506
No. of sulfate ions modelled	0	4	0
No. of intermediates modelled	0	0	2
RMSD bond length (Å)	0.012 (0.022)	0.010 (0.022)	0.011 (0.022)
RMSD bond length (º)	1.32 (1.98)	1.3 (1.98)	1.3 (1.98)
Wilson B factor ($Å^2$)	29.8	24	22.1
Average B factor	22.6	24.1	22.3
Protein ($Å^2$)	20.2	22.7	21
Water ($Å^2$)	35.5	36.1	34.3
Sulfate ion ($Å^2$)	N/A	21.8	N/A
Intermediate ($Å^2$)	N/A	N/A	19.3
REFMAC RMS error ($Å^2$)	0.1	0.1	0.1
Ramachandran analysis (% of residues)			
Most favoured	94.1	94.8	95.8
Additionally allowed	5.4	4.3	3.6
Generously allowed	0	0.3	0.5
Disallowed	0.5	0.6	0.1
G-factor	0.1	0.2	0.1

Table 7.4 -The final refinement statistics for the L-haloacid dehalohenase AQP5750 structures. Target values are given in brackets.

7.3.8 Structure of the L-haloacid dehalohenase from AQP5750

7.3.8.1 Monomer structure

The L-haloacid dehalohenase from AQP5750 is a homodimer. The approximate dimensions of the monomer are 40 Å x 27 Å x 33 Å. There are two domains; the core domain has a typical Rossman fold which consists of six parallel open twisted β-sheets with a strand order of S6, S5, S4, S1, S2, S3 (with Richardson topology of +1,+1,-3,-1,-1 (Richardson, 1981)) surrounded by four helices and three 3_{10} helices formed by residues 14-17 and 119-232. The cap domain is made up of a 4 helix bundle formed by residues 26-115, which is a total of 49% of the protein. Figures 7.16 and 7.17 show the structure of the monomer with the α-carbon backbone displayed as a ribbon diagram and the tertiary structure of the monomer. The secondary structure composition consists of 36.8% α-helix, 11.4% β-sheets and 5.5% 3_{10} helix. The secondary structure assignments of the monomer are displayed in Figure 7.18. The core and cap domain of the AQP5750 L-haloacid dehalohenase is structurally similar to the L-haloacid dehalohenases from *B. cepacia*, *X. autotrophicus*, *S. tokodaii* and *Pseudomonas* sp. strain YL with RMSDs of 1.49 Å, 1.49 Å, 1.47 Å and 1.40 Å respectively. Figure 7.19 shows the superimposition of the α-carbon backbone displayed as a ribbon diagram of the L-haloacid dehalohenase from AQP5750 with the L-haloacid dehalohenases from *B. cepacia* (A), *X. autotrophicus* (B), *S. tokodaii* (C) and *Pseudomonas* sp. strain YL.

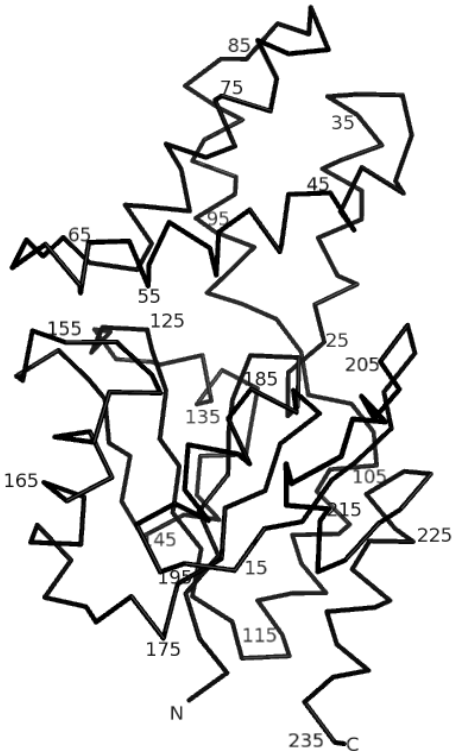


Figure 7.16- Diagram displaying the α-C trace of the AQP5750 L-haloacid dehalohenase monomer, with every 10th residue labelled. The Figure was produced using the program PyMOL (DeLano Scientific).

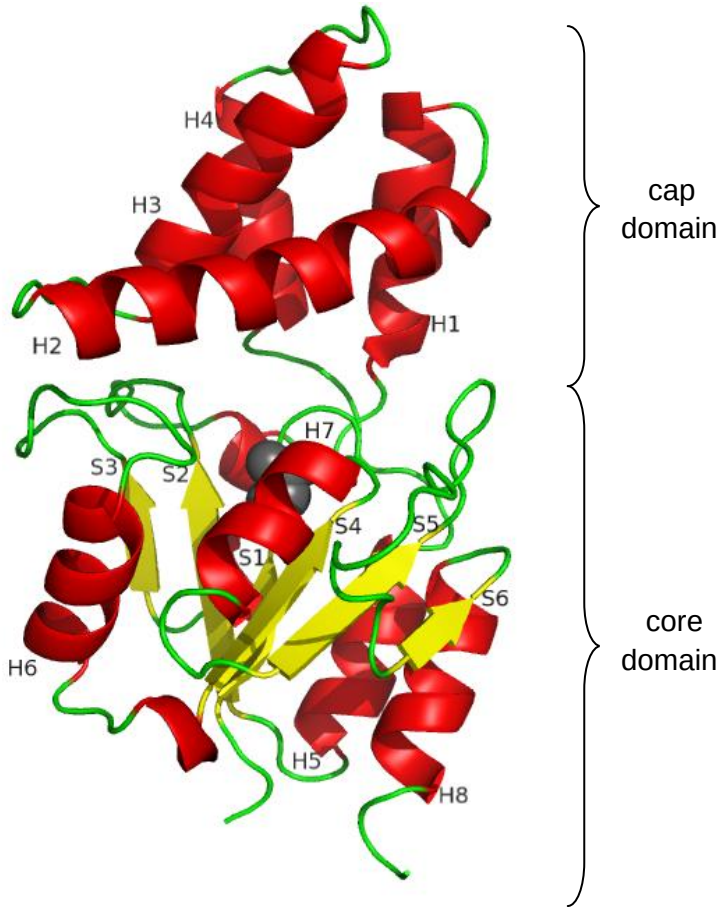


Figure 7.17- The tertiary structure of the L-haloacid dehalohenase monomer from AQP5750. The monomer is in the same orientation as in Figure 7.16. The α-helices (H) and β-sheets (S) are numerically labelled indicating their structural order. The catalytic Asp 18 residue is displayed as a sphere to highlight the location of the active site. The Figure was produced using the program PyMOL (DeLano Scientific).

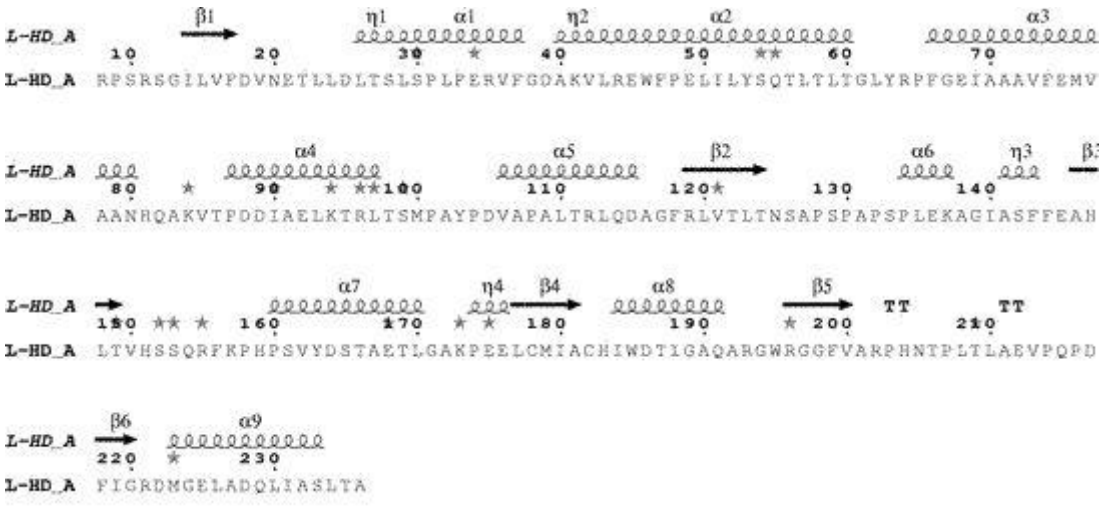


Figure 7.18- Secondary structure assignment of the AQP5750 L-haloacid dehalohenase. The Figure was produced using ESPript (Gouet *et al*., 1999).

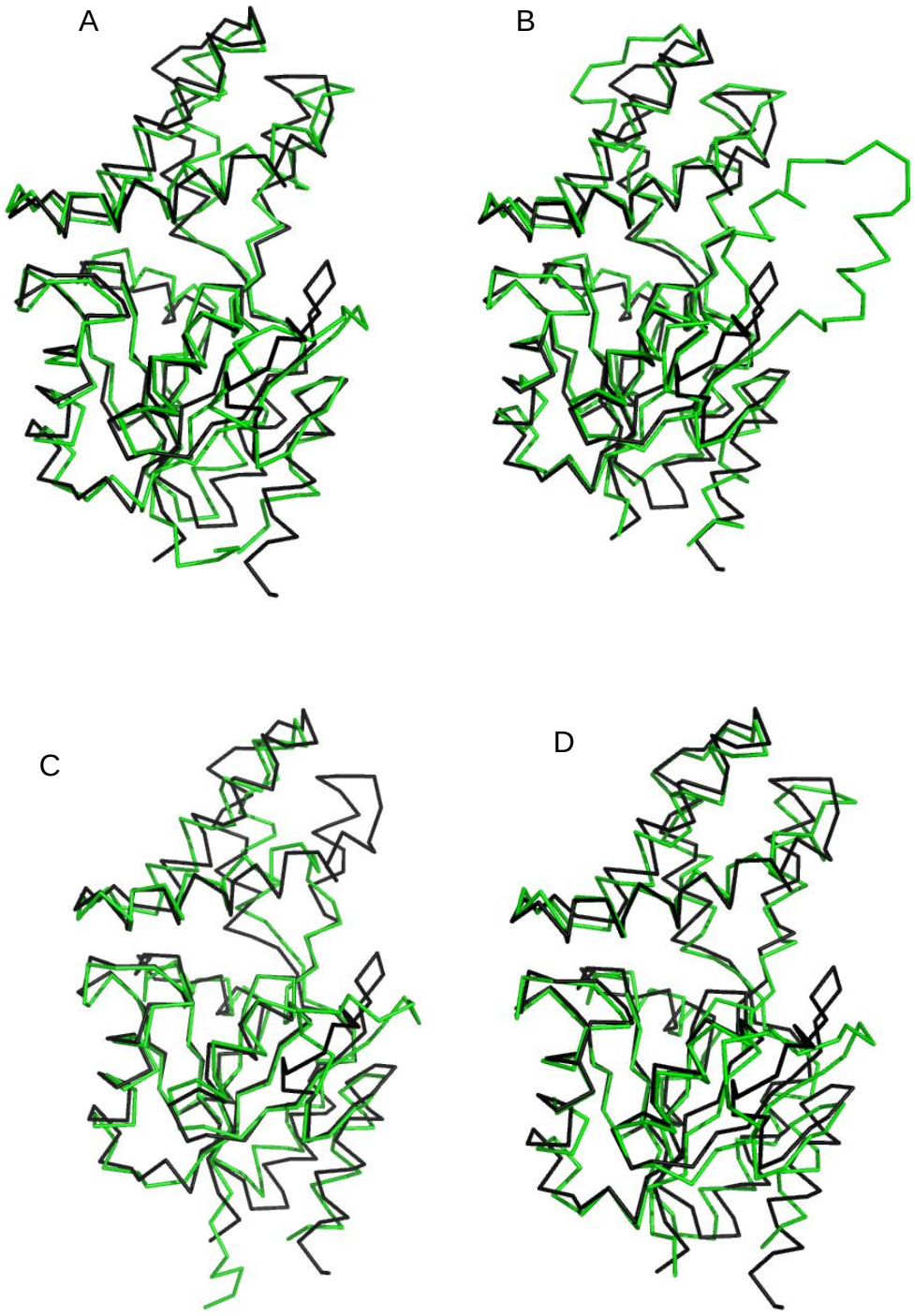


Figure 7.19- Diagram displaying the α-C trace of the L-haloacid dehalohenase monomer from AQP5750 (black) superimposed onto the α-C trace of the L-haloacid dehalohenases (green) from *B. cepacia* (A), *X. autotrophicus* (B), *S. tokodaii* (C) and *Pseudomonas* sp. strain YL (D). The Figure was produced using the program PyMOL (DeLano Scientific).

7.3.8.2 Oligomeric structure

The crystal structure of L-haloacid dehalohenase from AQP5750 exists as a homodimer (Figure 7.20) which correlates to the estimated size calculated when the protein was purified on a GF Superdex 75 chromatography column (section 7.2.5). The L-haloacid dehalohenase from *B. cepacia*, *X. autotrophicus* and *S. tokodaii* are also dimeric proteins. The two subunits are related by a 2-fold axis parallel to α-H2, which are in closer proximity to each other at the N-terminal end. The dimer has approximate dimensions of 70 Å x 40 Å x 43 Å. 18.7% of the surface of each monomer is buried on formation of the dimer. 13.4%, 13.5% and 19% of the monomer surface area is buried on dimer formation with the L-haloacid dehalohenases from *S. tokodaii, Pseudomonas* sp. YL and *X. autotrophicus*.

There are 3 main areas of contact between the two dimer subunits. The first area of contacts is observed between residues Glu 55-Thr 56 in α-H2. The second area of contact is observed between residues in α-H3 and the residues present between β-S5 and β-S3. The interacting residues are Arg 64-Glu 212, Glu 75-Thr 209, Asn 80-His 204 and His 81-His 204. The third area of contact is observed between Tyr 53 in H2 and Trp185 in H7.

The interactions involved in the formation of the dimerization domain in the L-haloacid dehalohenase from AQP5750 are significantly different to the dimerization interactions which have previously been described in the L-haloacid dehalohenase from *Pseudomonas* sp. YL and *S. tokodaii* (Hisano *et al*., 1996, Rye *et al*., 2009). Tyr 53 and equivalent residues in the L-haloacid dehalohenases from AQP5750, *B. cepacia* (Tyr 57), *S. tokodaii* (Tyr 31), *Pseudomonas* sp. YL (Tyr 57) and *X. autotrophicus* (Tyr 45) is conserved and involved in dimer stabilisation in all structures.

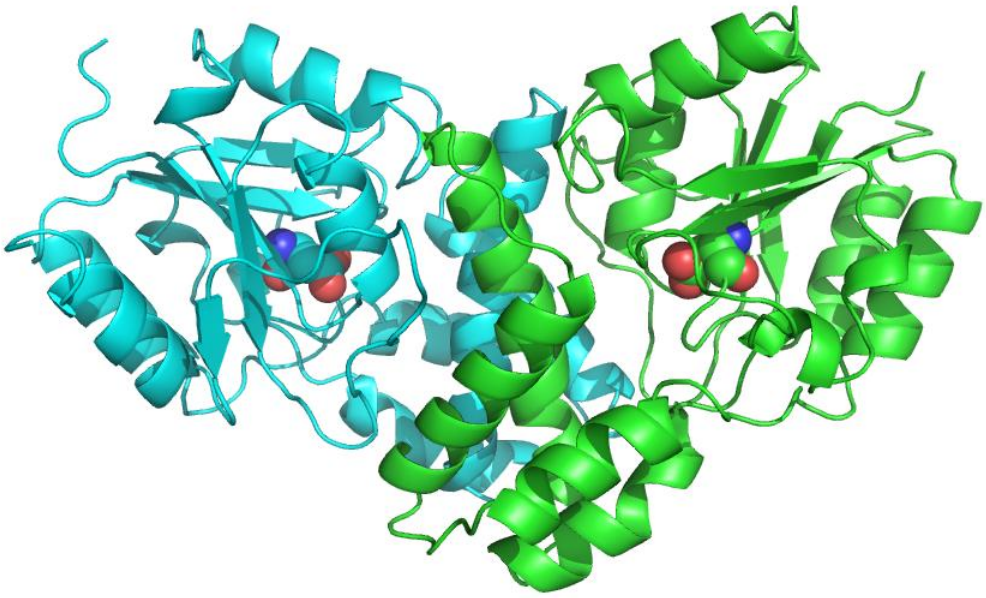


Figure 7.20- L-haloacid dehalohenase from AQP5750 shown as a dimer. The catalytic Asp 18 residues are displayed as spheres to highlight the location of the active site. The Figure was produced using the program PyMOL (DeLano Scientific).

7.3.8.3 Active site

The active site is located between the cap and the core domain following the S1 β-sheet. Residue side chains from both domains make up the active site cavity. There are a number of catalytically important residues which have been determined by point mutation experiments in the L-haloacid dehalohenase from *X. autotrophicus, Pseudomonas* sp. YL and *B. cepacia* (Kurihara *et al*., 1995, Pang *et al*., 2001, Schmidberger *et al*., 2007). Catalytically important residues in the L-haloacid dehalohenase from *B. cepacia* and equivalent residues in the L-haloacid dehalohenase from AQP5750 are displayed in Table 7.5.

Catalytically important residues in the L-haloacid dehalohenase from *B. cepacia*	Equivalent residues in the L-haloacid dehalohenase from AQP5750
Asp 11	Asp 18
Thr 15	Thr 22
Arg 52	**Phe 47**
Ser 119	**Thr 124**
Lys 152	Lys 157
Tyr 158	Tyr 163
Ser 176	**Ala 181**
Asn 178	**His 183**
Asp 181	Asp 186

Table 7.5- Catalytically important residues in the L-haloacid dehalohenases from *B. cepacia* and the equivalent residues in the AQP5750 L-haloacid dehalohenase. Residues which are different in the AQP5750 L-haloacid dehalohenase are highlighted in bold.

Point mutation experiments of the residues displayed in Table 7.5 in the L-haloacid dehalohenase from *B. cepacia* and equivalent residues in the L-haloacid dehalohenases *Pseudomonas* sp. YL and *X. autotrophicus* results in a loss of activity between 70-100% (Kurihara *et al*., 1995, Pang *et al*., 2001, Schmidberger *et al*., 2007). To date, these residues are conserved in all biochemically and structurally characterised L-haloacid dehalohenases. The changes of these residues in the AQP5750 L-haloacid dehalohenase indicate that they may have an interesting role in the catalytic mechanism which will be different to previously characterised L-haloacid dehalohenases. The catalytic site of the L-haloacid dehalohenase from AQP5750 is therefore of significant interest due to the differences observed compared to related enzymes.

Conserved residues Arg 42 and Asn 178 in the L-haloacid dehalohenase from *B. cepacia* have been proposed to be involved in a “lock down” mechanism of the substrate (Schmidberger *et al*., 2007). This was concluded when observing

different conformations of Arg 42 and the equivalent residues in the L-haloacid dehalohenase from *Pseudomonas* sp. YL (Arg 41) and *X. autotrophicus* (Arg 39) in an open position when the active site is empty, compared to a more closed position when an intermediate is bound to the catalytic residue Asp 11 (Asp 10 and Asp 8 in *Pseudomonas* sp. YL and *X. autotrophicus*). In the closed position Arg 42 is hydrogen bonded to Asn 178 and to the main chain carboxyl oxygen of Tyr 13. This has lead to the belief that these residues are highly conserved and may be involved in the substrate entering and binding in the active site. It was also proposed that Asp 180 activates a water molecule causing a dipole moment across the $N^{\delta 2}$ functional group of Asn 178. This would cause the $O^{\delta 2}$ of Asn 178 to possess a δ negative charge, making Asn 178 to bind more strongly to Arg 42. This "lock down" mechanism observed in the intermediate structure would be removed after the ester intermediate was hydrolysed causing the bond between Asn 178 and Arg 42 to be weakened. This would make Arg 42 adopt a more open conformation once the product is formed.

An intermediate structure of the *B. cepacia* L-haloacid dehalohenase with l-2-monochloropropanoate (ASl-11) covalently bound to the active site residue Asp 11 was obtained (Schmidberger *et al*., 2007). The equivalent Asp residue is completely conserved in all characterised L-haloacid dehalohenases. Asp 18 was proposed to be the catalytic residue in the AQP5750 L-haloacid dehalohenase. Figures 7.21 and 7.22 show the superimpositions of the active site residues of the L-haloacid dehalohenase from AQP5750 and *B. cepacia* (with ASl-11 bound).

With the aim to confirm Asp 18 as the catalytic residue and to determine the roles of Phe 47, Thr 124, Ala 181 and His 183 in the L-haloacid dehalohenase from AQP5750, substrate soaks were performed with a cryoprotectant containing 25 mM MCP at pH 4.5 (H_2SO_4) for 1 h before freezing under liquid nitrogen. This method of trapping an intermediate structure covalently bound to the catalytic Asp residue in the active site has previously been used to elucidate the intermediate structures of the L-haloacid dehalohenase from *Pseudomonas* sp. YL and *B. cepacia* (Ridder *et al*., 1999, Schmidberger *et al*., 2007). The method

works on the basis that the hydrolysis of the ester intermediate in the S_N2 dehalogenase mechanism cannot take place at a pH ≤5.

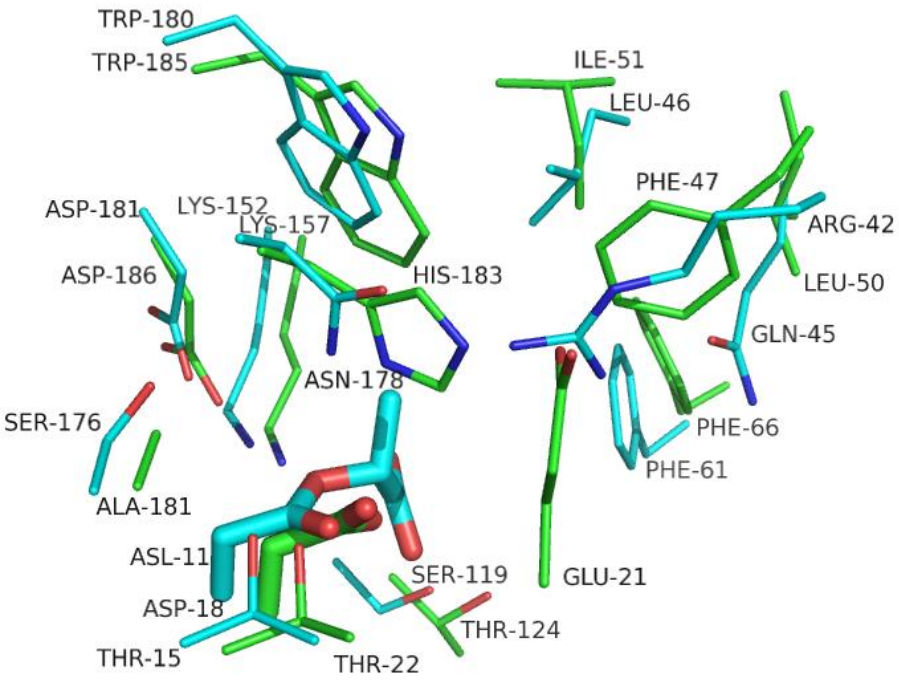


Figure 7.21- A superimposition of the active site residues in the L-haloacid dehalohenase from AQP5750 (green) and *B. cepacia* (cyan). The *B. cepacia* L-haloacid dehalohenase has the intermediate ASl-11 bound. The Figure was produced using the program PyMOL (DeLano Scientific).

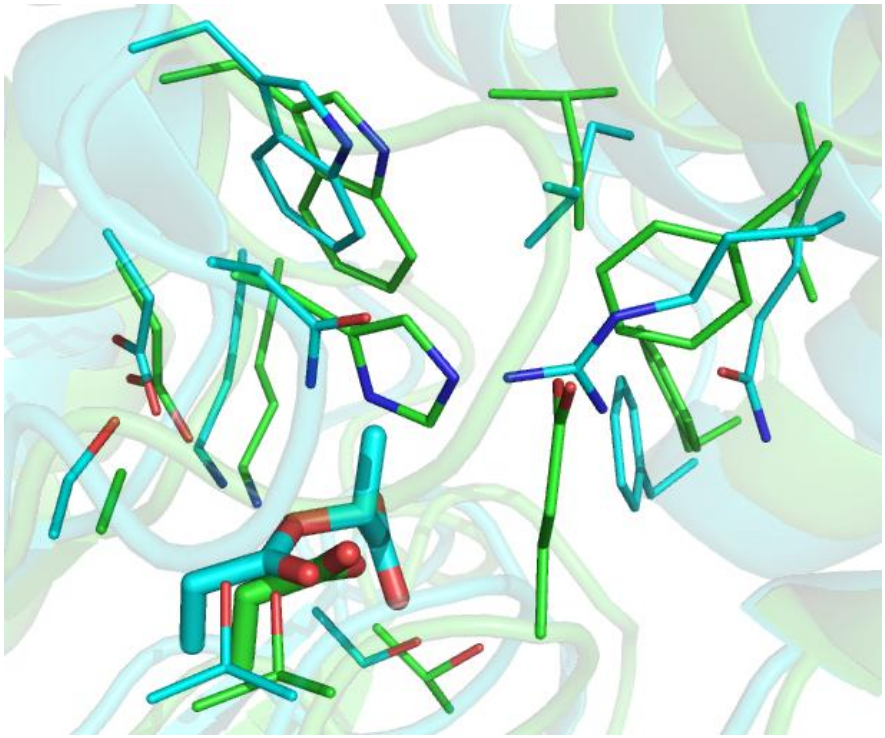


Figure 7.22- A superimposition of the active site residues in the L-haloacid dehalohenase from AQP5750 (green) and *B. cepacia* (cyan) with the intermediate (ASl-11) bound in relation to the tertiary structure. The Figure was produced using the program PyMOL (DeLano Scientific).

7.3.8.4 Sulfate binding

The structure for substrate soaked L-haloacid dehalohenase from AQP5750 was elucidated to 1.64 Å resolution. Clear electron density for a tetrahedral shaped sulfate ion was observed in the active site next to the catalytic Asp 18. Density for the sulfate ion was confirmed by calculating a F_o-F_c omit map. Figure 7.23 shows the $2F_o$-F_c electron density map of the sulfate ion in the active site. The sulfate ion forms hydrogen bonds with residues Asp 18, Thr 124, His 183 and Lys 157. The structure was very similar to the substrate free L-haloacid dehalohenase from AQP5750 with an RMSD of 0.42 Å.

The substrate free L-haloacid dehalohenase structure from *B. cepacia* also has a sulfate ion in the active site, as ammonium sulfate was at high concentration in the crystallization condition (Schmidberger *et al*., 2007).

Further crystal substrate soaking experiments with the AQP5750 L-haloacid dehalohenase were performed without the presence of H_2SO_4 in the cryo-protectant.

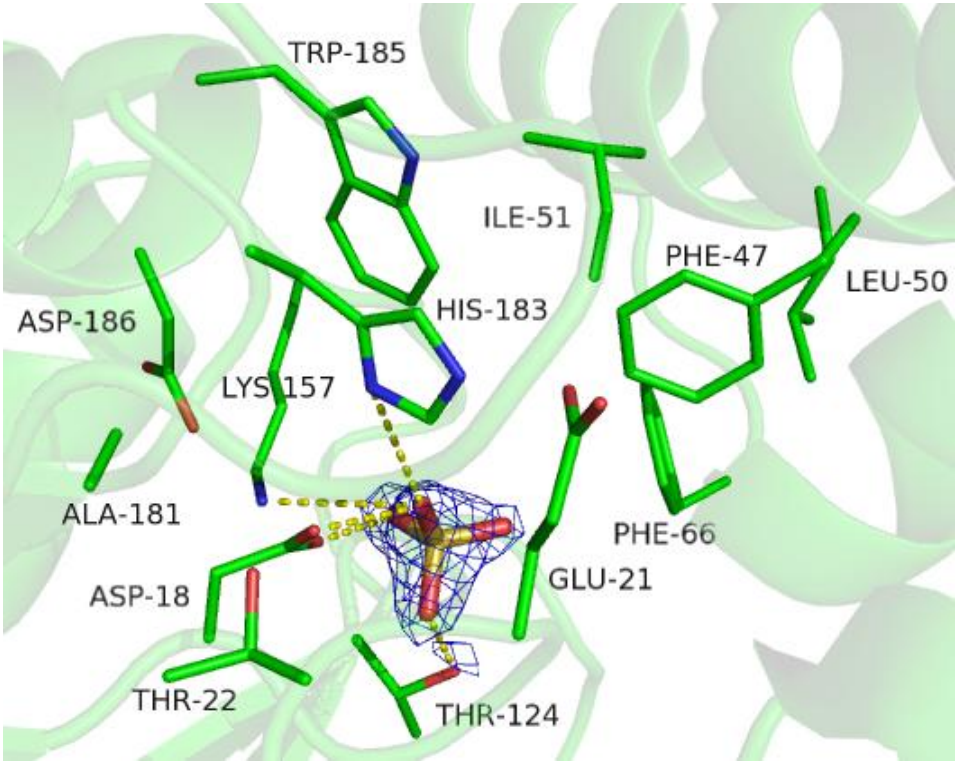


Figure 7.23- The active site residues in the L-haloacid dehalohenase from AQP5750 with a sulfate ion bound in the active site. The $2F_o$-F_c electron density map is displayed around the sulfate ion, which is hydrogen bonded to residues His 183, Lys 157, Asp 18 and Thr 124. The Figure was produced using the program PyMOL (DeLano Scientific).

7.3.8.5 Structure of the intermediate AQP5750 L-haloacid dehalohenase complex bound to MCP

The intermediate structure for substrate soaked L-haloacid dehalohenase from AQP5750 was elucidated to 1.7 Å resolution. Clear electron density connecting the catalytic Asp 18 to the intermediate structure of MCP is observed. Density for the MCP intermediate was confirmed by calculating a F_o-F_c omit map. Figure 7.24 shows the $2F_o$-F_c electron density map of the active site of the AQP5750 L-haloacid dehalohenase. No chlorine molecules were observed in the active site,

confirming that the halogen atom had been cleaved from the substrate and was no longer present in the active site. The structure was very similar to the native structure with an RMSD of 0.55 Å.

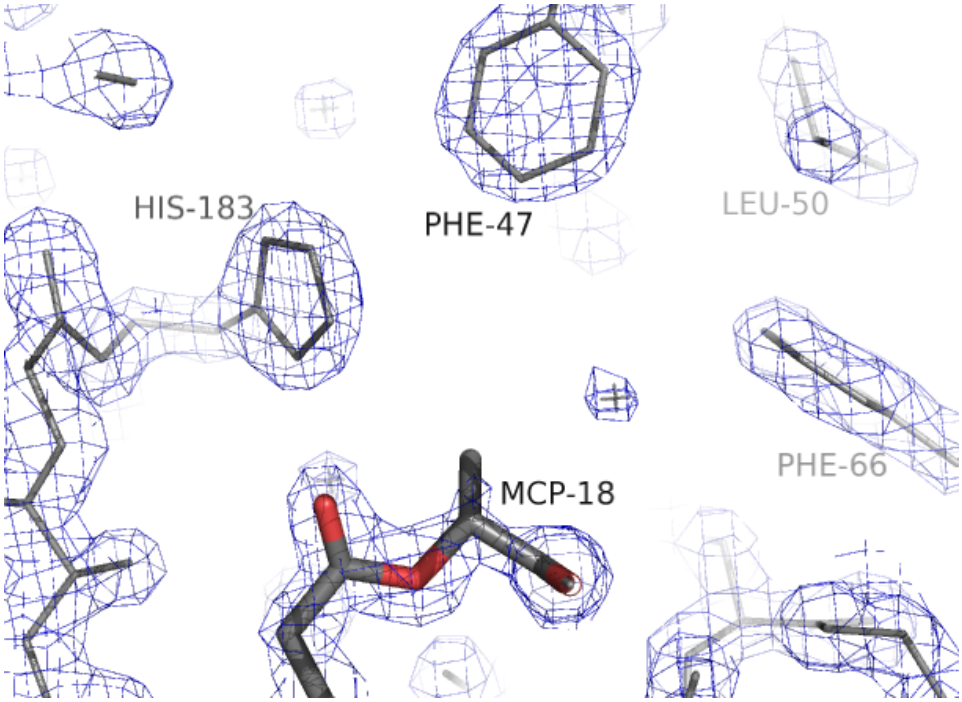


Figure 7.24- The intermediate structure of MCP covalently bound to Asp 18 in the AQP5750 L-haloacid dehalohenase. The $2F_o$-F_c electron density map is displayed for the active site residues. The Figure was produced using the program PyMOL (DeLano Scientific).

7.3.8.6 Active site of the L-haloacid dehalohenase

Figure 7.25 shows the superimposition of the active site residues from the intermediate structures of the L-haloacid dehalohenases from AQP5750 and *B. cepacia* with MCP bound to the catalytic Asp residues. The active site residues which are displayed in the L-haloacid dehalohenase from AQP5750 are Asp 18 (covalently bound to MCP), Glu 21, Thr 22, Phe 47, Leu 50, Ile 57, Phe 66, Thr 124, Lys 157, Ala 181, His 183, Trp 185, Asp 186. As previously observed, highly conserved residues Arg 42, Ser 119, Ser 176 and Asn 178 in the *B. cepacia* L-haloacid dehalohenase are replaced by Phe 47, Thr 124, Ala 181 and His 183 in the AQP5750 L-haloacid dehalohenase respectively.

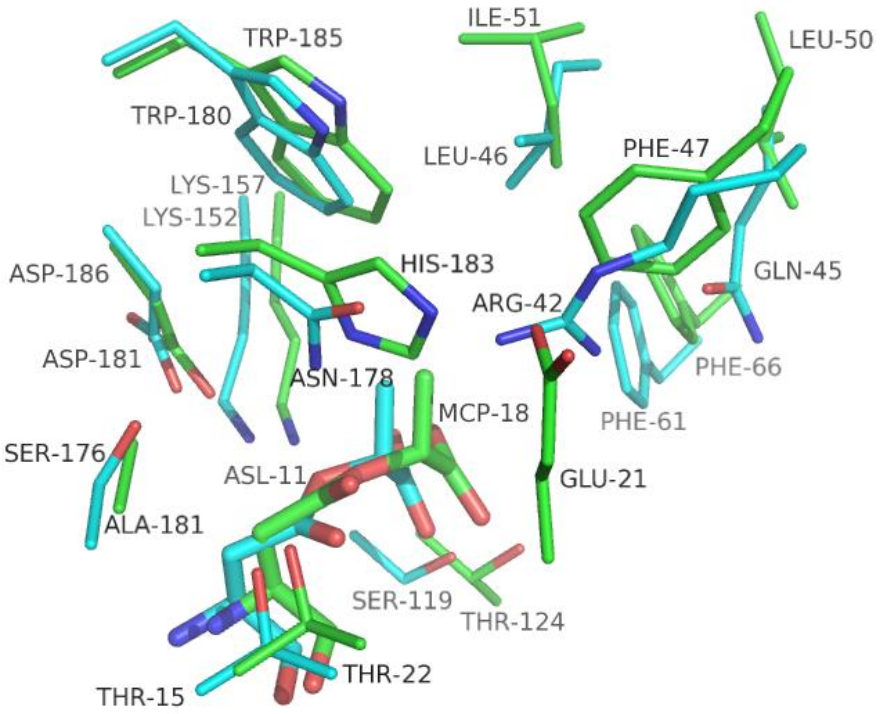


Figure 7.25- A superimposition of the active site residues of the L-haloacid dehalohenase from AQP5750 (green) and *B. cepacia* (cyan) with the intermediate MCP bound to the catalytic Asp. The Figure was produced using the program PyMOL (DeLano Scientific).

7.5.8.7- Residues contributing to substrate binding in the active site

Residues making up the hydrophobic pocket surrounding the active site in L-haloacid dehalohenases help stabilise the alkyl groups of the substrate. Residues which make up the hydrophobic pocket in the AQP5750 L-haloacid dehalohenase are Phe 47, Leu 50, Ile 51, Phe 66, Lys 157, Tyr 185.

The role of Arg 42 and Asn 178 in the L-haloacid dehalohenase from *B. cepacia* in substrate trafficking and binding in the active site has previously been discussed in section 7.3.7.3. Phe 47 and His 183 in the AQP5750 L-haloacid dehalohenase replace these conserved residues. Due to the functional differences in these residues Phe 47 and His 183 will not have a similar role to Arg 42 and Asn 178 in substrate entering and binding in the active site of the enzyme.

7.3.8.7 The Roles of His 183, Asp 186 and Glu 21 in the L-haloacid dehalohenase from AQP5750

Asp 180 and Asp 181 in the L-haloacid dehalohenase from *Pseudomonas* sp. YL and *B. cepacia* have been proposed to act as the base which activates the catalytic water molecule by removing a proton (Schmidberger *et al*., 2007, Nakamura *et al*., 2009). The equivalent residue (Asp 186) in the L-haloacid dehalohenase from AQP5750 is present and in a similar orientation (Figure 7.26).

Although the equivalent aspartic acid residue is present (Asp 186), the MCP bound structure suggests that it is likely that His 183 and Glu 21 also play an important role in activating the catalytic water molecule in the AQP5750 L-haloacid dehalohenase. This water molecule attacks the ester intermediate in the second part of the S_N2 dehalogenase mechanism.

The L-haloacid dehalohenases from *B. cepacia, Pseudomonas* sp. YL, *S. tokodaii* and *X. autotrophicus* do not contain the equivalent Glu or His in their active sites. Figure 7.26 shows the catalytic water molecule (H_2O 41) hydrogen bonded to the intermediate MCP, His 183 and Asp 186. Therefore, as predicited in the original bioinformatics (section 7.3.1), the crystal structures confirm that the AQP5750 L-haloacid dehalohenase reaction mechanism proceeds slightly differently to the related *Pseudomonas* sp. YL and *B. cepacia* enzymes.

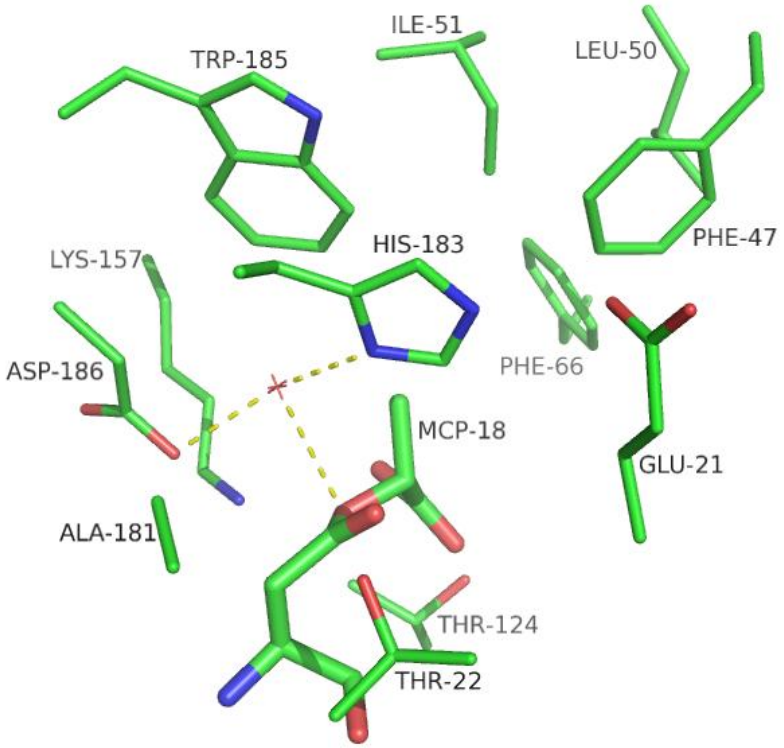


Figure 7.26- The active site residues in the L-haloacid dehalohenase from AQP5750 with MCP covalently bound to the catalytic Asp 18.The catalytic H_2O is displayed and the hydrogen bonds (shown as a dashed line) to MCP, His 183 and Asp 186 are shown. The Figure was produced using the program PyMOL (DeLano Scientific).

7.4 Summary

The AQP5750 L-haloacid dehalohenase shows highest activity towards brominated substrates with short carbon chains. The enzyme showed increased activity up to 150% in 40% DMSO and 30% methanol. The enzyme is stable up to temperatures of 50ºC, but shows an optimal activity at 55ºC. The AQP5750 L-haloacid dehalohenase has a K_m value of 6.72 mM. The crystal complex structure with covalently bound MCP (determined by X-ray crystallography) confirmed Asp 18 as the main catalytic residue. Residues His 183, Asp 186 and Glu 21 in the active site of the AQP5750 L-haloacid dehalohenase are proposed to be involved in activation of the catalytic water which attacks the ester intermediate in the second part of the S_N2 dehalogenase mechanism.

Chapter 8- Haloalkane dehalogenase from AQP5750

8.1 Introduction

Aquapharm Biodiscovery Ltd determined that AQP5750 had 99.0% sequence identity to *Rhodobacteraceae* family bacteria. Chapter 5 describes the genome sequencing and the identification of a haloalkane dehalogenase gene in AQP5750.

This chapter describes the cloning, over-expression, purification, biochemical characterisation, crystallization and structural analysis of the haloalkane dehalogenase from AQP5750.

8.2 Materials and Methods

8.2.1 Bioinformatics

To confirm the presence of catalytic amino acids, protein sequence alignments of the AQP5750 haloalkane dehalogenase against the biochemically and structurally characterised haloalkane dehalogenase from *S. paucimobilis* were performed using ClustalW, European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/clustalw2).

8.2.2 AQP5750 growth

Section 2.1.3 describes the bacterial growth of AQP5750. The bacterial paste was harvested (section 2.1.6) and the genomic DNA extracted (section 2.2.3).

8.2.3 Cloning of the haloalkane dehalogenase from AQP5750

PCR primers were designed (section 2.2.4.1) to amplify the haloalkane dehalogenase gene (see Appendix, 11.15), with the incorporation of the restriction sites *Not*I and *Nde*I for cloning into the pET-28a vector (Table 8.1). PCR reactions were performed using PCR composition 2 (section 2.2.4.4) with PCR program 1 (Table 2.2.4.2). The PCR product was run on a 1% agarose gel to visualise the amplified DNA (section 2.2.2). The PCR band was gel extracted (section 2.2.6) and ligated into the cloning vector pJET1.2/blunt (section 2.2.8.1). The DNA from the ligation reactions were transformed (section 2.2.8.3) into *E.*

coli XL 10-Gold® chemically competent cells (section 2.2.7). The pJET1.2/blunt construct and the pET-28a vector were double digested with *Not*I and *Nde*I restriction enzymes (section 2.2.8.5 and 2.2.8.6) and the digested pET-28a was SAP treated (section 2.2.7.7). The ligation reaction into the expression vector pET-28a was performed (section 2.2.8.6) and the reaction transformed (section 2.2.8.3) into *E. coli* XL 10-Gold® competent cells (section 2.2.7). Gel extracted PCR bands (section 2.2.6) and plasmid constructs (section 2.2.8.4) were sent for Sanger sequencing (section 2.2.8.8).

Forward primer	*Nde*I 5'- **CATATG**ATGACTACGTCTTTCCGCGA -3'
Reverse primer	*Not*I 5'- **GCGGCCGC**TTACAAGCTGGCGTGCCA -3'

Table 8.1- Forward and reverse primer sequences with incorporated restriction sites *Nde*I and *Not*I for the amplification of the haloalkane dehalogenase in AQP5750.

8.2.4 Over-expression of the haloalkane dehalogenase from AQP5750

The extracted pET-28a construct (section 2.2.8.4) from *E. coli* XL 10-Gold® was transformed into the competent *E. coli* cell line BL21-CodonPlus (DE3)-*Rosetta*2. The plasmid was extracted (section 2.2.8.4) and sent for Sanger sequencing (section 2.2.8.8). Positively transformed strains of *E. coli* were preserved as glycerol stocks (section 2.1.5). Protein over-expression conditions were optimised (section 2.3.1) and the resultant bacterial cell pellets were lysed with BugBuster (1/10 dilution with Tris-H_2SO_4 pH 8.2) as described in section 2.3.2.2. The soluble and insoluble protein fractions were analysed using SDS-PAGE (section 2.3.6).

8.2.5 Purification of the haloalkane dehalogenase from AQP5750

A 2 L culture of transformed BL21-CodonPlus (DE3)-*Rosetta*2 was grown (section 2.1.4). The protein expression was induced (section 2.3.1) and the bacterial paste harvested (section 2.1.6). The bacterial cell pellet was resuspended in NAC buffer A (Table 2.7) to a final concentration of 10% (w/v) and the cells lysed by sonication (section 2.3.2.1). The soluble protein was purified by nickel affinity chromatography and the fractions corresponding to the haloalkane dehalogenase were concentrated (section 2.3.3.8). The concentrated protein was further purified using a Superdex 75 GF chromatography column (section 2.3.3.7).

8.2.6 Substrate specificity of the haloalkane dehalogenase from AQP5750

Protein parameters including amino acid composition, theoretical pI and extinction coefficient were calculated (section 2.3.4) based on the haloalkane dehalogenase protein sequence (see Appendix, 11.16). The purified protein concentration was determined (section 2.3.5) and the protein diluted with GF buffer (Table 2.7) to 0.2 mg/ml. This stock solution was stored at 4ºC for no longer than 72 h and was used in all biochemical characterisation reactions. Section 2.3.9.4 describes the methods used to determine the substrate specificity of the AQP5750 haloalkane dehalogenase. Substrates used to test for dehalogenase activity were 1-bromooctane, 1,3-dibromopropane, 1-bromohexane, cyclohexyl bromide, cyclohexyl chloride, 1,4-dichlorobutane, 2-chloropropane, dibromoethane, 2-chlorobutane, 1,2-dichloropropane, 1,2-dichloroethane, 1-bromoheptane, 1-chlorohexane and 1,3-dichloro-2-propanol (structures shown in Appendix, 11.6).

8.2.7 Crystallization

8.2.7.1 Crystal screening and X-ray data collection

Methods for crystal screening and the preparation of crystals for data collection are described in chapter 7, section 7.2.7.

8.2.7.2 Crystal soaks

Crystals were removed from the droplet and placed in an appropriate cryo-protectant containing 25 mM of 1-bromohexane (BHX) for 1 h (Table 8.2) before being directly frozen in liquid nitrogen using a nylon loop.

8.2.8 X-ray Data Collection

Data were collected at beamlines I02 and I03 at the Diamond Light Synchrotron Radiation Source (www.diamond.ac.uk). The I02 beamline has an ADSC Q315r detector (charged couple detector) and the I03 beamline has a Pilatus 6M-F detector. Table 8.2 displays the X-ray data collection conditions for the haloalkane dehalogenase crystals.

Crystal	Native haloalkane dehalogenase	Haloalkane dehalogenase complex with 1-hexanol (1HO)
Cryo-protectant	0.1 M $MgCl_2$ 200 mM Tris-HCl 20% PEG 4000 25% PEG 400 pH 8.0	200 mM Tris-HCl 20% PEG 3350 30% PEG 400 25 mM BHX pH 4.0
Soaking time	N/A	1 h
Beam	Diamond I02	Diamond I03
Wavelength (Å)	0.98	0.98
No. Frames	360	2400
Oscillation in degrees	0.5	0.15

Table 8.2- X-ray data collection conditions for the AQP5750 haloalkane dehalogenase crystals.

8.2.9 Structure determination

8.2.9.1 Data processing

Data were processed using the program Xia2 version 0.3.3.0 (Winter *et al*., 2010), which uses the data reduction software programs XDS (Kabsch, 1993), MOSFLM (Leslie, 1992), Labelit (Sauter *et al*., 2004), Pointless (Evans, 2005) and CCP4 (Collaborative Computational Project, Number 4. 1994).

8.2.9.2 Phase determination, model building, refinement and structure validation

Phase determination, model building, refinement and structure validation were carried out as described in section 7.2.9.

8.3 Results

8.3.1 Bioinformatics

The haloalkane dehalogenase from AQP5750 shares relatively high amino acid sequence identity to the biochemically and structurally characterised haloalkane dehalogenase from *S. paucimobilis* (64%), *M. tuberculosis* (59%) *R. rhodochrous* (47%) and *Bradyrhizobium japonicum* (44%) respectively. Figure 8.1 shows an amino acid sequence alignment of the haloalkane dehalogenase from AQP5750 and *S. paucimobilis*.

```
sequence1       MTTSFR--DKKKFATVHGKQMAYIEEGTGDPIVFLHGNPMSSYLWRNIMPHLAGKGRLIA 58
sequence2       MSLGAKPFGEKKFIEIKGRRMAYIDEGTGDPILFQHGNPTSSYLWRNIMPHCAGLGRLIA 60
                *: . :  .:***  ::*::****:*******:* **** ********** ** *****

sequence1       PDLIGMGDSDKLDNSGPDSYTFAEHCTYLFALLEQLGVTENVTLVIHDWGSGLGFHWAHT 118
sequence2       CDLIGMGDSDKLDPSGPERYAYAEHRDYLDALWEALDLGDRVVLVVHDWGSALGFDWARR 120
                 ************ ***: *::***  ** ** * *.: :.*.**:*****.***.**:

sequence1       HSDAVKGIAFMEAIVETRESWDAFPERAREMFQALRSPAGEEMVLEKNLFVEALVPGSIL 178
sequence2       HRERVQGIAYMEAIAMPIE-WADFPEQDRDLFQAFRSQAGEELVLQDNVFVEQVLPGLIL 179
                * : *:***:****. . * *  ***: *::***:** ****:**:.*:*** ::** **

sequence1       RDLTEEEMNEYRRPFANAGEDRRPTLTFPRQVPIEGQPKDVTELVDAYVDWLGQTSIPKL 238
sequence2       RPLSEAEMAAYREPFLAAGEARRPTLSWPRQIPIAGTPADVVAIARDYAGWLSESPIPKL 239
                * *:* **  **.**  *** *****::***:** * * **. :.  *..**.::.****

sequence1       FINADPGVLITGEVRDRVRSWPNLTEVTVAGLHFIQEDSPDEIGAAVRDWHAl---- 291
sequence2       FINAEPGALTTGRMRDFCRTWPNQTEITVAGAHFIQEDSPDEIGAAIAAFVRRLRPA 296
                ****:**.* **.:**  *:*** **:**** **************:  :
```

Figure 8.1- ClustalW amino acid sequence alignment of the haloalkane dehalogenase from AQP5750 (sequence 1) and *S. paucimobilis* (sequence 2). The residues that are essential for catalysis in the *S. paucimobilis* haloalkane dehalogenase are highlighted in red. Equivalent residues in the AQP5750 haloalkane dehalogenase are highlighted in yellow.

8.3.2 Cloning and over-expression of the haloalkane dehalogenase from AQP5750

PCR was used to amplify the haloalkane dehalogenase gene from AQP5750 and a band of the correct size was observed on a 1% agarose gel (Figure 8.2). The amplified gene was then cloned into the cloning vector pJET1.2/blunt and the expression vector pET-28a with the incorporation of an N-terminal His-tag. Over-expression trials of the haloalkane dehalogenase in BL21-CodonPlus (DE3)-*Rosetta*2 were conducted by varying the induction conditions. BL21-CodonPlus (DE3)-*Rosetta*2 expression strain contains the T7 promoter and is induced using IPTG. As discussed in section 6.3.2, proteins encoded by rare codons can limit the expression in *E. coli*. The expression cell line BL21-CodonPlus (DE3)-*Rosetta*2, contains extra copies of the genes that code for the tRNAs that may slow down the translation of proteins. There is 1 rare leucine codon which is encoded by CTA and 6 rare proline codons encoded by CCC in the haloalkane dehalogenase from AQP5750. None of the rare codons are repeated next to each other. BL21-CodonPlus (DE3)-*Rosetta*2 contains extra copies of the proL (CCC) tRNA gene, allowing higher levels of expression of heterologous proteins encoded by this rare codon. Expression of the haloalkane dehalogenase in BL21-CodonPlus (DE3)-*Rosetta*2 was good. The best over-expression was observed with 1 mM IPTG at 28ºC, with incubation for 20 h post induction (Figure 8.3, lane 3).

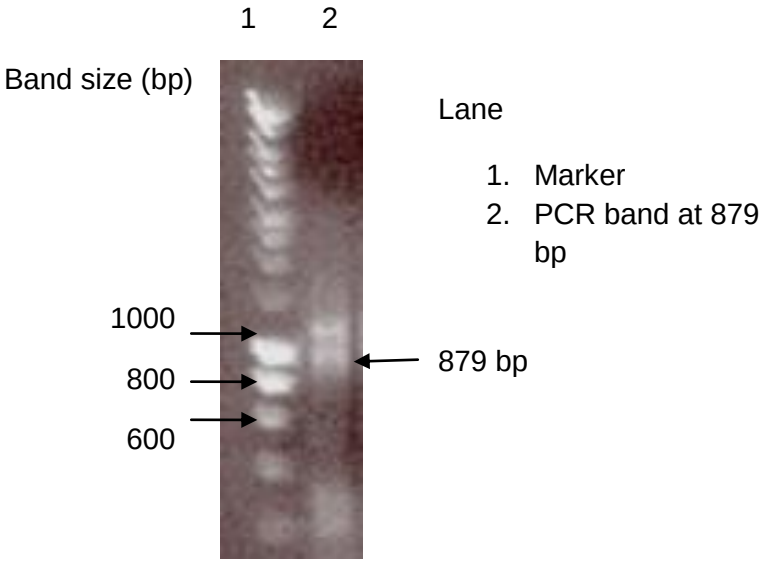


Figure 8.2- 1% agarose gel showing the amplified haloalkane dehalogenase gene from AQP5750.

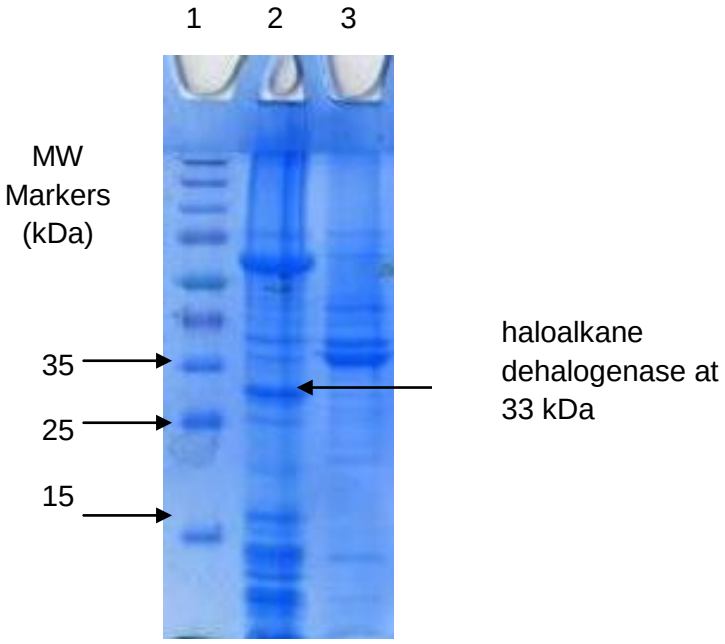


Figure 8.3- SDS-PAGE gel showing the over-expressed AQP5750 haloalkane dehalogenase in BL21-CodonPlus (DE3)-*Rosetta*2 induced with 1 mM IPTG for 20 h. Lane 1- MW Marker, lane 2- soluble fraction and Lane 3- insoluble fraction.

8.3.3 Purification of the haloalkane dehalogenase from AQP5750

5 g of cell paste was harvested from 2 L of recombinant BL21-CodonPlus (DE3)-*Rosetta*2 by centrifugation (section 2.1.6). Bacterial cells were lysed by sonication and the soluble protein purified by nickel affinity chromatography (Figure 8.4). Fractions 44-62 assayed positive for haloalkane dehalogenase activity, the protein peak was pooled together and a 5 µl sample was analysed using SDS-PAGE (Figure 8.5, lane 3). The pooled fractions from the nickel affinity chromatography column were concentrated using a centrifugal concentrator with a 10 kDa membrane (Vivaspin 20; Viva science) at 3,000 x *g*, at 4ºC until the final volume reached 1 ml. The concentrated protein sample was purified on a Superdex 75 GF chromatography column (Figure 8.6). Fractions eluting between 53-68 ml assayed positive for activity, suggesting that the protein eluted in the form of a monomer. The fractions from the protein peak were pooled together and a 5 µl sample analysed using SDS-PAGE (Figure 8.5, lane 4). 40 mg of haloalkane dehalogenase from AQP5750 was obtained from a 2 L culture of recombinant BL21-CodonPlus (DE3)-*Rosetta*2**.**

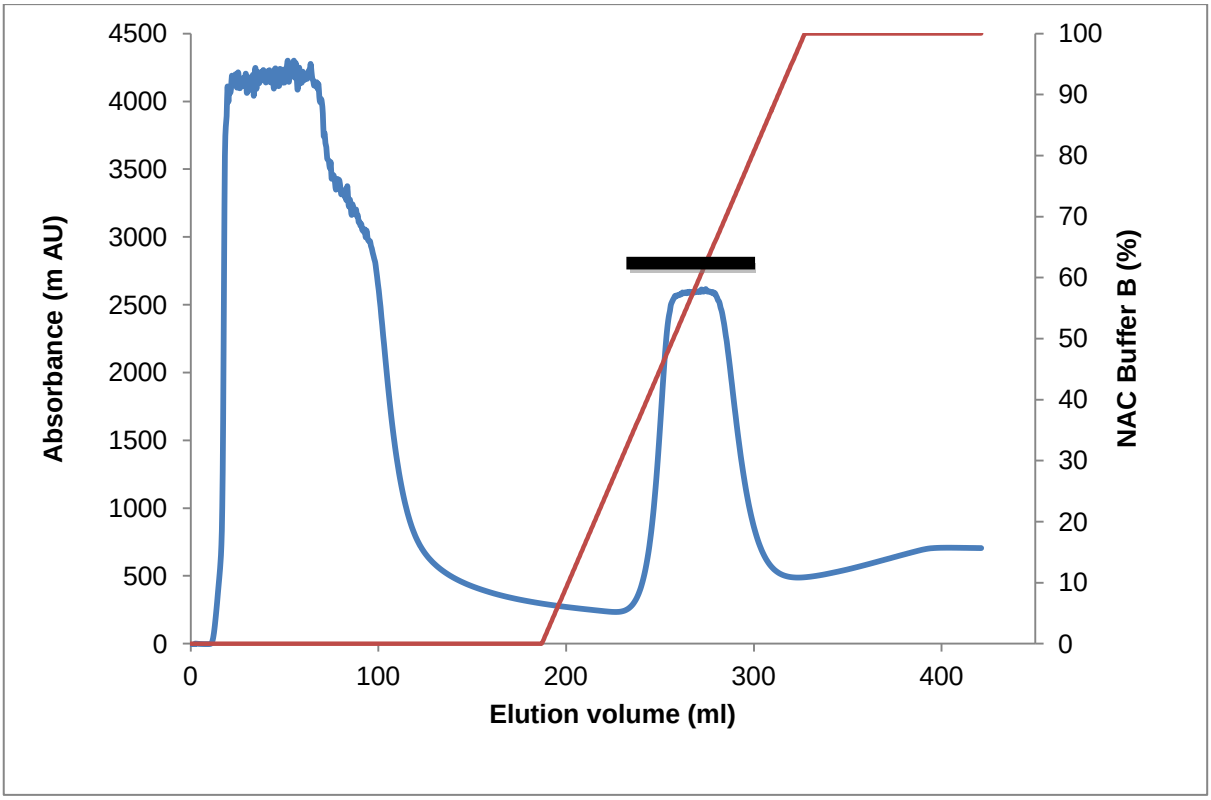


Figure 8.4- Elution profile from the purification of the AQP5750 haloalkane dehalogenase from the nickel affinity chromatography column. The red line shows the concentration of NAC buffer B. Fractions 44-62 (220-310 ml) containing the haloalkane dehalogenase are highlighted by the black bar.

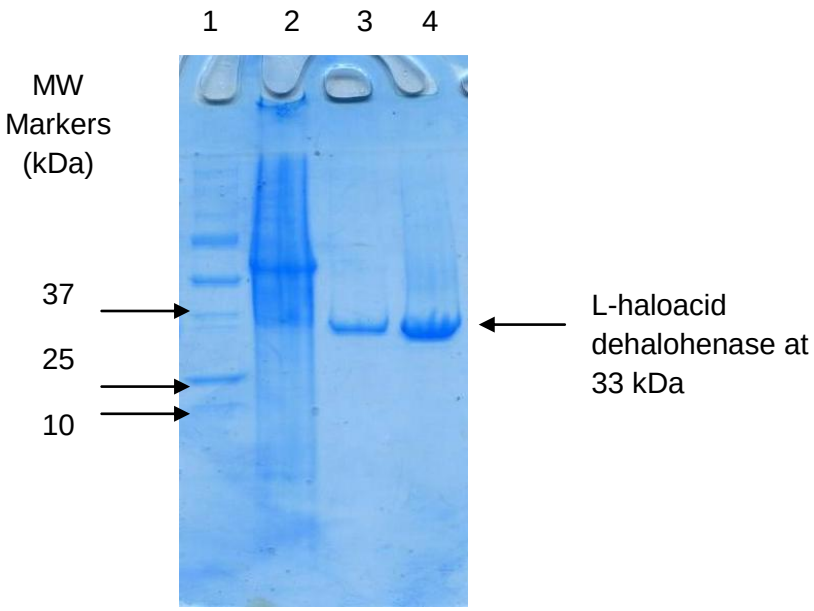


Figure 8.5- SDS-PAGE gel of the purified haloalkane dehalogenase from AQP5750. Lane 1- MW Marker, Lane 2- crude protein extract, lane 3- protein peak of the haloalkane dehalogenase from the nickel affinity chromatography column, lane 4- protein peak of the haloalkane dehalogenase from the Superdex 75 GF chromatography column.

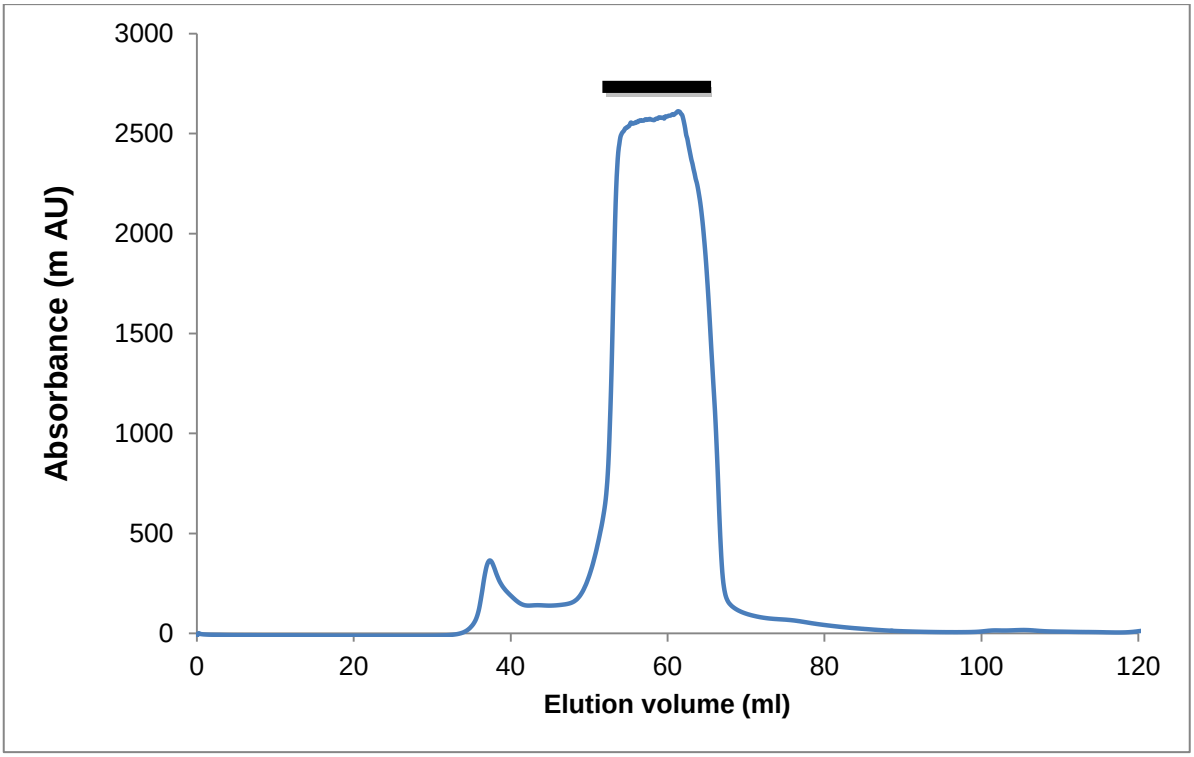


Figure 8.6- Elution profile from the purification of the haloalkane dehalogenase from the GF chromatography column (Superdex 75). Fractions 53-68 containing the haloalkane dehalogenase are highlighted by the black bar. 1 ml fractions were collected throughout GF chromatography.

8.3.4 Biochemical characterisation of the haloalkane dehalogenase from AQP5750

The colorimetric dehalogenase assay is based on the method described by Holloway *et al.* (1998). It measures the decrease in pH as the dehalogenase turns over the substrate and a halide/proton product is formed. This can be visualised by eye as the assay solution changes from red (at pH ≥8) to yellow (at pH ≤4). The assay can also be followed spectrophotometrically at 540 nm.

8.3.5 Substrate specificity of the haloalkane dehalogenase from AQP5750

The activity of the haloalkane dehalogenase from AQP5750 towards substrates utilised by haloalkane dehalogenases with high amino acid sequence identity were tested. The substrates had varying linear and cyclic structures, carbon chain lengths and the positioning and number of halogens (chlorine or bromine).

Reaction rate experiments were attempted but quantitative biochemical analysis and kinetic studies could not be carried out due to low levels of activity observed with these substrates and large variations in repeat experiments. Small amounts of activity towards 1,6-dichlorohexane, 1-bromooctane, 1,3-dibromopropane and 1-bromohexane were visualised (over 1 h) by observing a colour change from red to yellow (substrate structures shown in Appendix, 11.6) .

Wild type microorganism AQP5750 was isolated from a tube worm. Some species of tube worms (*Capietellidae, Spionidae* and *Cirratulidae*) are known to produce a range of structurally diverse halogenated compounds (Fielman *et al*., 2001). Many of these are complex aromatic compounds such as 2,4,6-tribromophenol and bromooctylpyrrole. Microorganisms living in symbiosis with the tube worm may express dehalogenase enzymes capable of degrading some of these toxic compounds. The complexity of the metabolites produced by the worm can make identifying the dehalogenase substrates difficult.

The haloalkane dehalogenase from *S. paucimobilis, M. tuberculosis, R. rhodochrous* and *B. japonicum* have high amino acid sequence identity (> 44%) in comparison to the AQP5750 haloalkane dehalogenase. However, they show varying activities towards a range of substrates. The haloalkane dehalogenase from *S. paucimobilis* (sequence identity of 64%) shows highest activity towards multiple halogenated alkanes including 1,2-dibromoethane, 1-bromoethane, 1,2-dibromopropane and 1-bromobutane (Nagata *et al*., 1997, Nagata *et al*., 1999). The haloalkane dehalogenase from *M. tuberculosis* (59%) shows highest activity towards 1-chlorobutane, 1-chlorodecane, 1-bromobutane and 1,2-dibromoethane (Jesenska, 2000). The haloalkane dehalogenase from *R. rhodochrous* (47%) shows highest activity towards 1,2-dibromoethane, 1-bromobutane, 1-chlorobutane, 2-bromobutane and 2-chlorobutane (Newman *et al*., 1999) and the haloalkane dehalogenase from *B. japonicum* (44%) shows highest activity towards 1,2 dibromoethane and 1,2-dibromopropane (Sfetsas *et al*., 2009). These haloalkane dehalogenases with high sequence similarity show a diverse range of activity.

1,2-dibromoethane is a good substrate for the haloalkane dehalogenases from *S. paucimobilis* (K_m 1.9 mM)*, M. tuberculosis, R. rhodochrous* (K_m 1.2 mM) and *B. japonicum* (K_m 0.1 mM) (Nagata *et al*., 1999, Newman *et al*., 1999, Sfetsas *et al*., 2009). Despite the high amino acid sequence identity of these haloalkane dehalogenases, the enzyme from AQP5750 did not show activity towards 1,2-dibromoethane. Substrate variations between the haloalkane dehalogenases are thought to arise from the structural difference in the cap domain of the protein which alters the size and shape of the binding pocket in the active site (Chovancová *et al*., 2007). Differences in residues in the hydrophobic active site pockets can affect substrate specificity and activity. The haloalkane dehalogenases from *S. paucimobilis* and AQP5750 have high amino acid sequence identity (64%). The haloalkane dehalogenase from *S. paucimobilis* shows a preference towards long chain haloalkanes and displays low activity to 1,2-dichloroethane and 1,2-dichloropropane. 1,2-dichloroethane is a xenobiotic compound which is produced as a by-product in numerous industrial processes. It is therefore of great interest to find haloalkane dehalogenase which can degrade this molecule (Nagata *et al*., 1997, Oakley *et al*., 2002).

With the aim to gain a further insight into the substrate specificity observed with the AQP5750 haloalkane dehalogenase in comparison to haloalkane dehalogenases with high sequence identity, the crystal structure of the enzyme was determined.

8.3.6 Crystallization

Purified protein in GF buffer (Table 2.7) was concentrated to 10 mg/ml. Microbatch crystallization trials were conducted using a number of commercial screens using a 1:1 ratio of protein to crystallization condition. Protein crystals were observed in many conditions with varying concentrations of salts, buffers and precipitants. Figure 8.7 shows the crystals that diffracted to the highest resolution. The crystallization condition was 0.3 M $MgCl_2$, 0.1 M Tris-HCl and 30% PEG-4000.

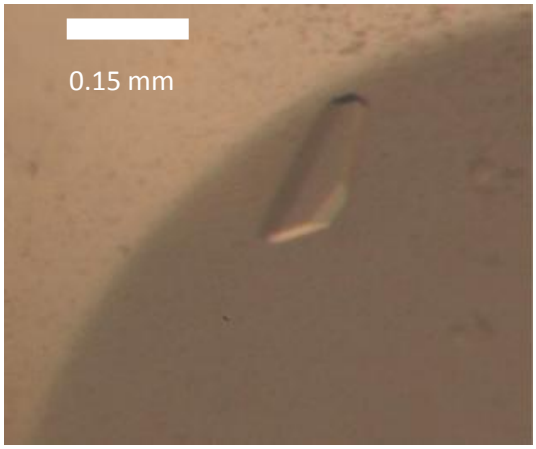


Figure 8.7- A crystal of the haloalkane dehalogenase from AQP5750 obtained from 0.3 M $MgCl_2$, 0.1 M Tris-HCl and 30% PEG-4000.

8.3.6.1 Crystal Soaks

Figures 8.8 shows the crystals used for the substrate soaking experiment. The crystallization condition contained 200 mM sodium thiocyanate with 20% PEG-3350. Crystal soaks were performed as described in section 8.4.7.2 and the cryo-protectants used are given in Table 8.2.

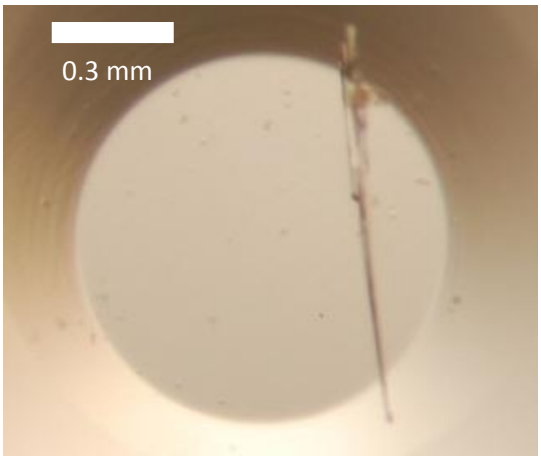


Figure 8.8- Crystals of the haloalkane dehalogenase from AQP5750 obtained from 200 mM sodium thiocyanate with 20% PEG-3350.

8.3.7 Structure determination of the haloalkane dehalogenase from AQP5750

8.3.7.1 X-ray data collection

X-ray datasets were collected as described in Table 8.2. The haloalkane dehalogenase crystallized in space group $P2_1$ and the haloalkane dehalogenase used for crystal soaks with product 1-hexanol (1HO) bound, crystallized in space group $P2_12_12$.

Table 8.3 describes the data processing statistics for the native haloalkane dehalogenase and haloalkane dehalogenase complex with two molecules of 1HO bound.

Crystal	Native haloalkane dehalogenase	Haloalkane dehalogenase complex with 1HO
Resolution (Å)	49-1.79 (1.89-1.79)	47.84-1.61 (1.66-1.61)
Cell dimension	a, b, c= 55.18 Å, 75.18 Å, 64.7 Å α, β, γ=90º, 93.1º, 90º	a, b, c= 64.5 Å, 71.41 Å, 59.75 Å α, β, γ=90º
Space group	$P2_1$	$P2_12_12$
No. of protomers	2	1
Solvent content (%) V_M ($Å^3$ Da^{-1})	40.14 (2.1)	41.1(2.1)
No. of measured reflections	178669	465006
No. of unique reflections	49155	36192
Completeness	99.6 (99.0)	99.9 (100)
(I)/ δ (I)	9.6 (2.7)	26 (3.6)
R_{merge} † (%)	0.086 (0.47)	0.06 (0.20)

Table 8.3 -Data processing statistics for the haloalkane dehalogenase from AQP5750. Values for the outer resolution shell are given in brackets. $R_{merge}=\Sigma_h\Sigma_J|<I_h>-I_J(h)|/\Sigma_h\Sigma_J I(h)$, where I(h) is the intensity of the reflections h, Σ_h is the sum over all the reflections and Σ_J is the sum over J measurements of the reflections.

8.3.8 Structure solution

The haloalkane dehalogenase from *S. paucimobilis* (PDB: 1CV2) was used as a model for the molecular replacement using the program MOLREP, version 10.2 (Vagin and Teplyakov, 1997). The rotation function was calculated using a radius of 28 Å at a resolution of 3.0 Å. This resulted in 2 strong rotation peaks at 16.32 and 14.03 with the wrong solution being no greater than 4.63. A translation search at a resolution of 3 Å identified the correct rotation solution as 14.03, with a peak score of 0.49 and the wrong solution being no greater than 0.24.

The haloalkane dehalogenase structure complex was solved using the final structure of the native AQP5750 haloalkane dehalogenase as a model for the molecular replacement.

8.3.9 Model building and validation

Refinement of the structure solution using REFMAC version 5.6 (Murshudov *et al*., 1997) was performed giving an R-factor of 27.4 and an R_{free} of 33.6. The model was then submitted to automated refinement using ARP/wARP version 7.0.1 (Cohen *et al*., 2008). The resulting R-factor and R_{free} were 25.0% and 32.3%. Further manual modelling and refinement was carried out as described in section 8.2.9.2, resulting in a final R-factor and R_{free} value of 20.9% and 27.3% respectively.

The G-factor was calculated at 0 (confirming that the structure has normal stereochemical properties) using PROCHECK version 3.3 (Laskowski *et al*., 1993). The final refinement statistics for the haloalkane dehalogenase structure and validation results are shown in Table 8.4 and the Ramachandran plot in Appendix, 11.17.

8.3.9.1 Haloalkane dehalogenase complex with 1HO

Modelling and refinement were carried out as described in section 8.2.9.2. Electron density was observed for the product 1HO in the AQP5750 haloalkane dehalogenase structure. The final refinement statistics for the haloalkane dehalogenase and validation results are shown in Table 8.4 and the Ramachandran plot in Appendix, 11.18. The G-factor was calculated at 0

(confirming that the structure has normal stereochemical properties) using PROCHECK version 3.3 (Laskowski *et al*., 1993).

A bound chlorine atom was found in the active site of the native and complex haloalkane dehalogenase structures.

Crystal	Native Haloalkane dehalogenase	Haloalkane dehalogenase complex with 1HO
Resolution (Å)	49-1.79	47.84-1.61
Overall R-factor (%)	20.9	11.7
R_{free} (5 % total data)	27.8	18.4
No. of residues modelled (specific residue name and number)	A (M1-L292) B (M1-L292)	A (T2-L292)
No. of waters modelled	243	352
No. of chlorines modelled	1	1
No. of intermediates modelled	0	2
RMSD bond length (Å)	0.017 (0.022)	0.013 (0.022)
RMSD bond length (º)	1.61 (1.95)	1.35 (1.95)
Wilson B factor ($Å^2$)	31.0	27.4
Average B factor		
Protein ($Å^2$)	27.6	22.8
Water ($Å^2$)	33.2	38.8
Chlorine($Å^2$)	27.24	16.23
Intermediate ($Å^2$)	N/A	33.71
REFMAC RMS error ($Å^2$)	0.17	0.09
Ramachandran analysis (% of residues)		
Most favoured	88.6	91.2
Additionally allowed	10.6	8.4
Generously allowed	0.8	0.4
Disallowed	0	0
G-factor	0	0

Table 8.4- The final refinement statistics for the haloalkane dehalogenase AQP5750 structures. Target values are given in brackets.

8.3.10 Structure of the haloalkane dehalogenase from AQP5750

8.3.10.1 Monomer structure

The haloalkane dehalogenase from AQP5750 adopts the typical α/β hydrolase fold. The protein has two domains, the core and the cap domain which are highlighted in Figure 8.9. The core domain is conserved in all of the α/β hydrolase fold superfamily proteins and is composed of 8 β-strands surrounded by 6 α-helices, two on one side and 4 on the other. The cap domain is composed of 6 α-helices and is the most variable section of haloalkane dehalogenases. Changes to the cap domain are thought to give rise to the different substrate specificities of biochemically characterised haloalkane dehalogenases (Janssen *et al*., 2004).

The secondary structure composition of haloalkane dehalogenase from AQP5750 consists of 32.1% α-helix, 17.4% β-strands and 8.5% 3_{10} helix. The secondary structure assignments of the monomer are displayed in Figure 8.10 and Figure 8.11 shows the structure of the monomer with the α-carbon backbone displayed as a ribbon diagram. The approximate dimensions of the monomer are 39 Å x 40 Å x 50 Å.

The overall structure of the haloalkane dehalogenase from AQP5750 is very similar to the haloalkane dehalogenases from *S. paucimobilis, M. tuberculosis, R. rhodochrous* and *B. japonicum* with RMSDs of 0.75, 0.71, 1.03 and 0.99 respectively. Figure 8.12 shows the superimposition of the α-carbon backbone of the haloalkane dehalogenase from AQP5750 with the haloalkane dehalogenase from *S. paucimobilis, M. tuberculosis, R. rhodochrous* and *B. japonicum*.

The native structure revealed a dimer in the asymmetric unit. 8.7% of the surface of each monomer is buried on formation of the dimer. The protein was estimated to elute as a monomer from a GF 75 chromatography column and is believed to be catalytically active as a monomer. The substrates soaked crystal (in BHX) has one monomer in the asymmetric unit which forms a crystallographic dimer within the unit cell. This dimer is similar to the one observed in the asymmetric unit of the native structure.

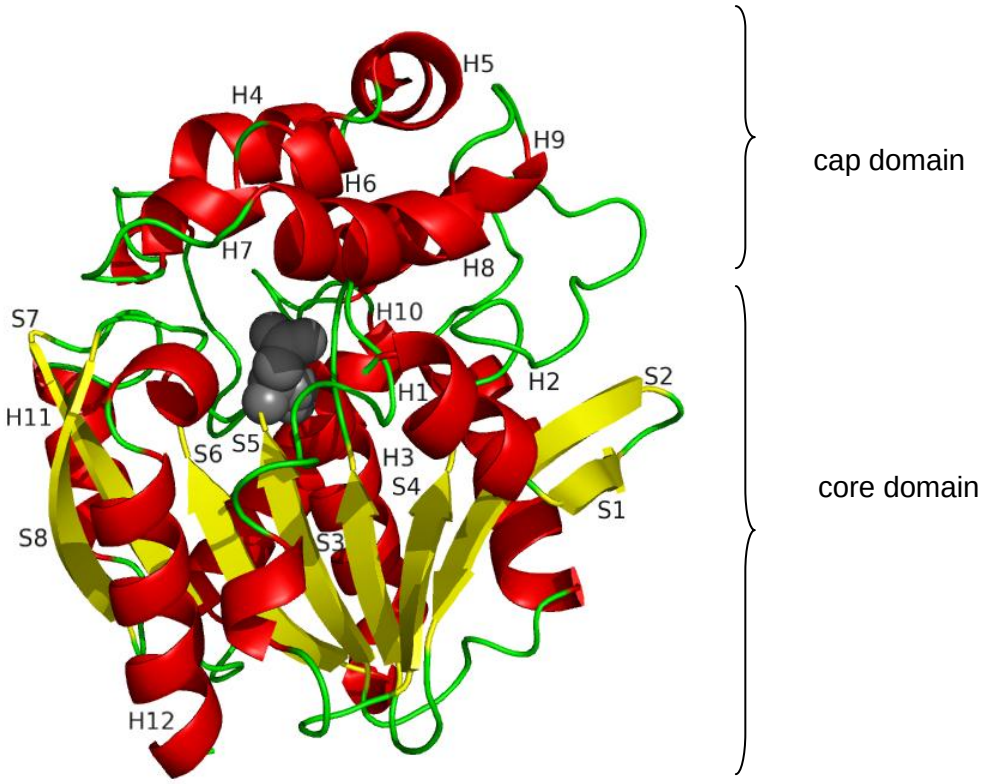


Figure 8.9- The tertiary structure of the haloalkane dehalogenase monomer from AQP5750. The monomer is in the same orientation as in Figure 8.11 and the α-helices and β-strands are labelled with an S or H followed by the number which indicates their structural order. The catalytic Asp 106 is displayed as spheres to highlight the location of the active site. The Figure was produced using the program PyMOL (DeLano Scientific).

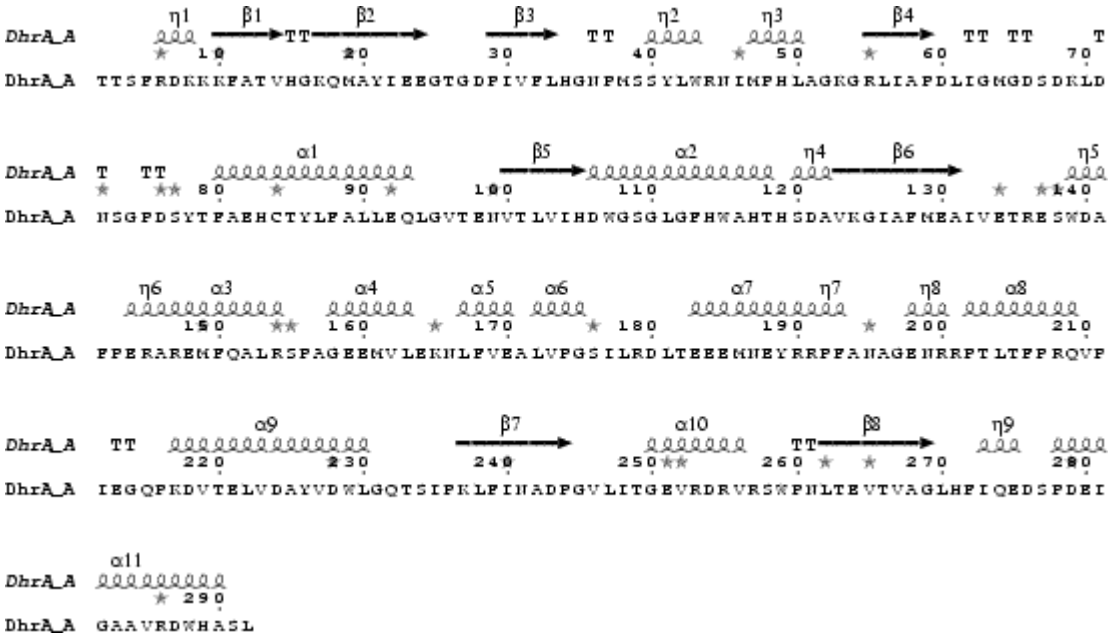


Figure 8.10- Secondary structure assignment of the haloalkane dehalogenase from AQP5750. The Figure was produced using ESPript (Gouet *et al*., 1999).

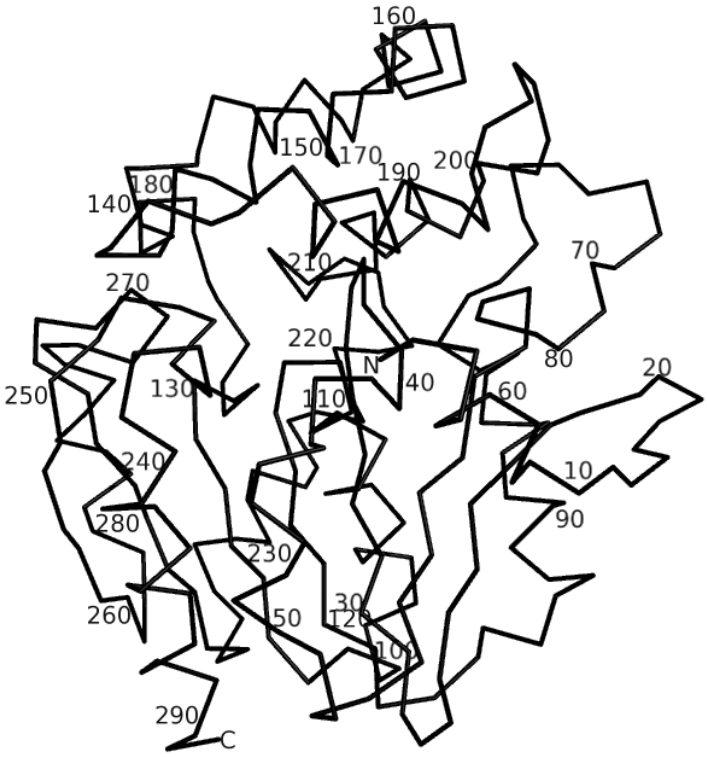


Figure 8.11- Diagram displaying the α-C trace of the AQP5750 haloalkane dehalogenase monomer, with every 10th residue labelled. The Figure was produced using the program PyMOL (DeLano Scientific).

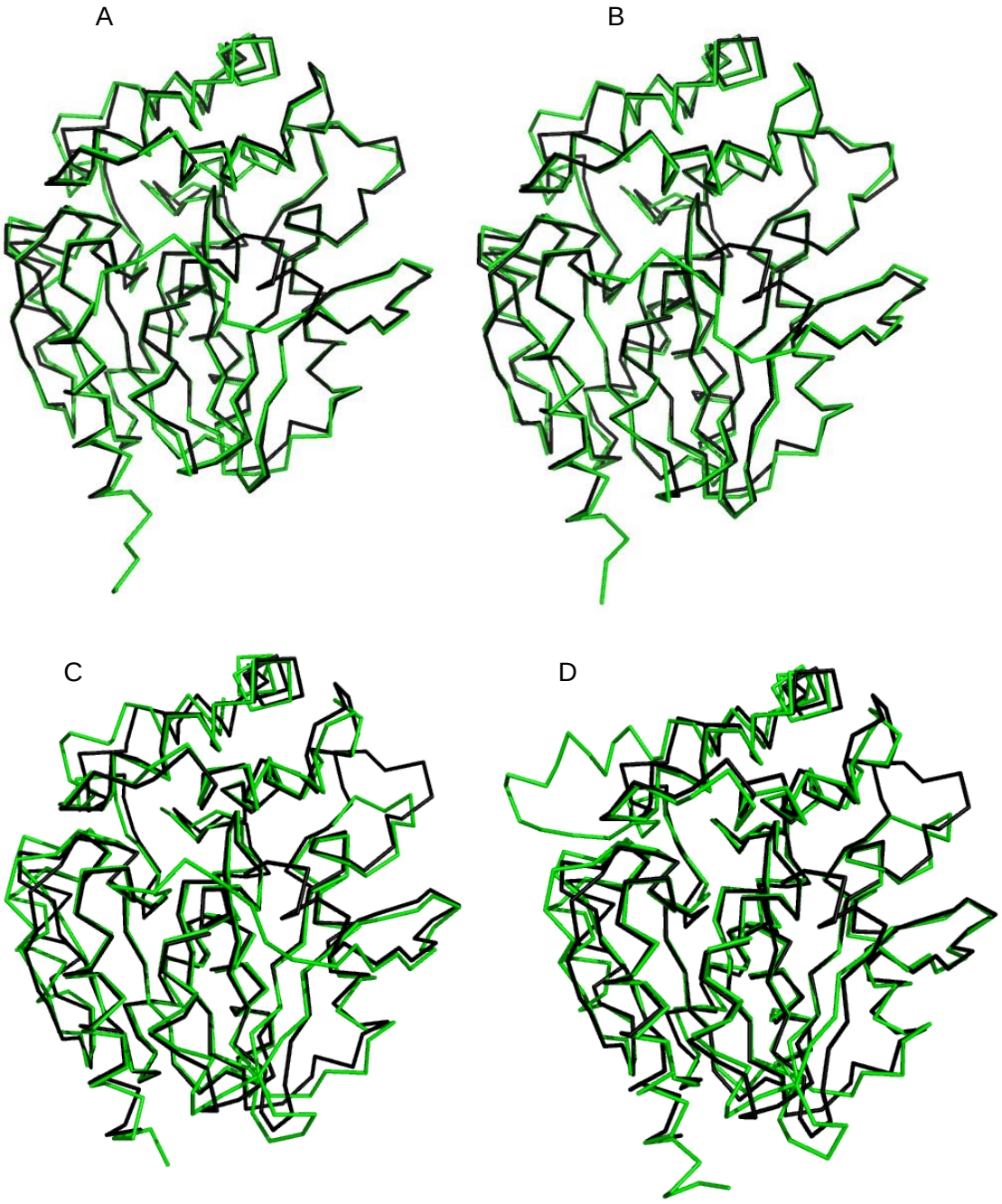


Figure 8.12- Diagram displaying the α-C trace of the haloalkane dehalogenase monomers from AQP5750 (black) superimposed onto the α-C trace of the haloalkane dehalogenases (green) from *S. paucimobilis* (A), *M. tuberculosis* (B), *R. rhodochrous* (C), *B. japonicum* (D). The Figure was produced using the program PyMOL (DeLano Scientific).

8.3.11 Active site

By using the structure of the haloalkane dehalogenase from *S. paucimobilis*, the active site in the haloalkane dehalogenase from AQP5750 was located. The active site is situated between the core and the cap domain with residues from both domains forming the active site cavity. The catalytic triad consists of Asp 106 (located between β-strand 5 and α-helix 2), His 271 (located between β-strand 8 and 3_{10} helix 9) and Glu 130 (located in β-strand 6). These residues in the haloalkane dehalogenase from AQP5750 are identical to the catalytic triad in the haloalkane dehalogenase from S. *paucimobilis* (Asp 108, His 272 and Glu 132). Figure 8.13 shows the superimposition of the amino acids making up the hydrophobic active site cavities in the haloalkane dehalogenase from AQP5750 and *S. paucimobilis*.

A chlorine atom is found to be present in the active site; this is not surprising as NaCl was a component of the purification buffers. A chlorine atom in an identical position to that in the active site of the haloalkane dehalogenase from *S. paucimobilis* was observed (Oakley *et al*., 2002). Figure 8.14 shows the superimposition of the active site residues of the haloalkane dehalogenases from AQP5750 and *S. paucimobilis* in relation to the three-dimensional structure of the proteins.

The difference in the active site residues in the haloalkane dehalogenases from *S. paucimobilis* and AQP5750 are shown in Table 8.5. Additionally, Phe 150 is shifted away from the active site cavity in the haloalkane dehalogenase from AQP5750 in comparison to the equivalent Phe 151 in the *S. paucimobilis* haloalkane dehalogenase.

Although many haloalkane dehalogenases have high sequence homology, small changes to the active site cavities and changes to the residues that make up the hydrophobic access tunnels can alter substrate specificity (Janssen *et al*., 2004). The haloalkane dehalogenase from AQP5750 displayed low activity towards the substrates tested. To gain an insight into the mechanism and potential structural

changes to the active site cavity upon binding of a substrate, crystal soaking experiments with BHX were performed.

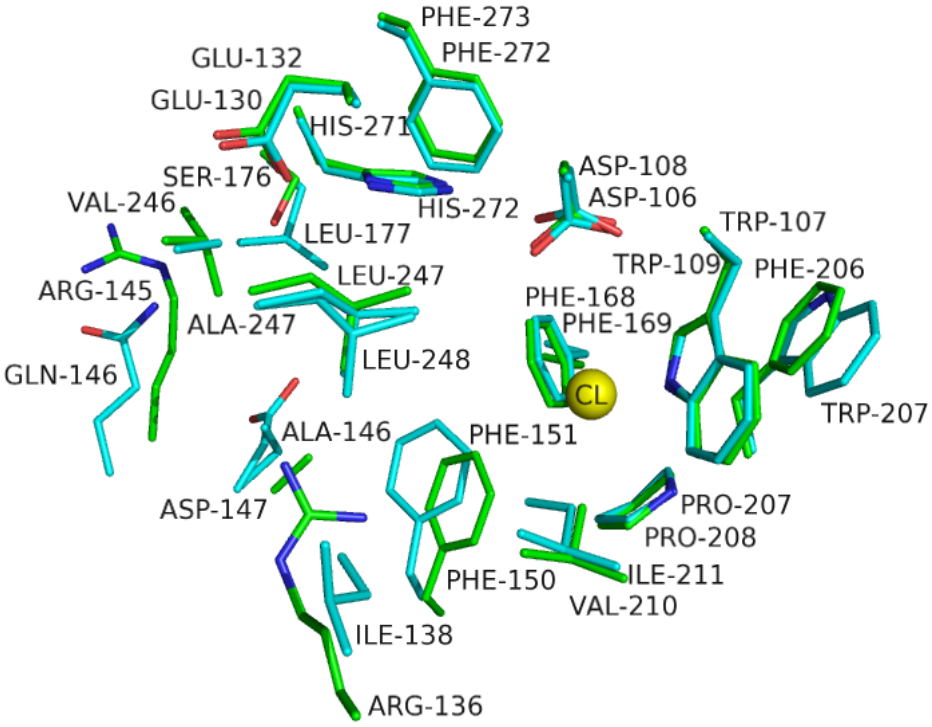


Figure 8.13- A superimposition of the active site residues in the haloalkane dehalogenase from AQP5750 (green) and *S. paucimobilis* (cyan). The chlorine in the active site is shown as a yellow sphere. The Figure was produced using the program PyMOL (DeLano Scientific).

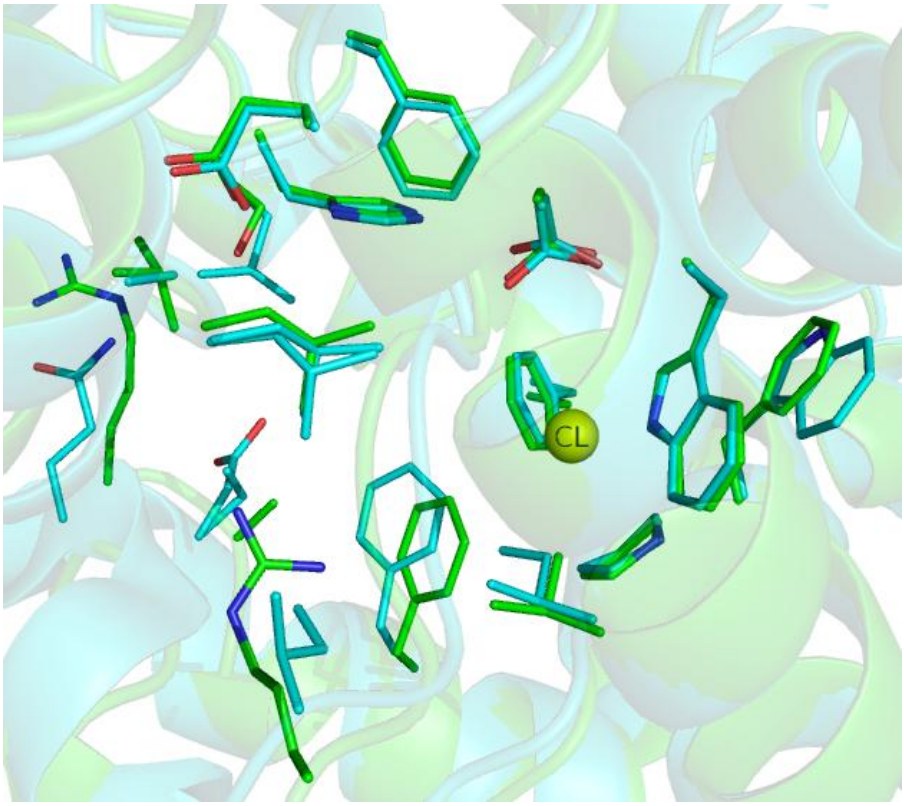


Figure 8.14- A superimposition of the active site residues in the haloalkane dehalogenase from AQP5750 (green) and *S. paucimobilis* (cyan) in relation to the tertiary structure. The chlorine in the active site of the AQP5750 haloalkane dehalogenase is displayed as a yellow sphere. The Figure was produced using the program PyMOL (DeLano Scientific).

Haloalkane dehalogenase from *S. paucimobilis*	Haloalkane dehalogenase from AQP5750
Ala 247	Val 246
Gln 146	Arg 145
Asp 147	Ala 146
Leu 177	Ser 176
Ile 138	Arg 136
Trp 207	Phe 206

Table 8.5- The difference in the active site residues in the haloalkane dehalogenases from *S. paucimobilis* and AQP5750.

8.3.12 Substrate binding

The structure for substrate soaked AQP5750 haloalkane dehalogenase was elucidated to a 1.61 Å resolution. Electron density for the fully dehalogenated product 1HO at a partial occupancy of 0.5 was observed in the active site. Electron density for a chlorine atom was also observed in active site in the same position as in the native haloalkane dehalogenase structure.

Both the $2F_o$-F_c and the F_o-F_c maps confirmed the presence of the 1HO in the active site. The 1HO product sits in a pocket formed by residues Leu 247, Val 210, Phe 150 and Phe 142 in the active site. The $2F_o$-F_c and the F_o-F_c maps also confirmed the presence of a second 1HO product molecule at an occupancy of 0.5 on the surface of the protein near the entrance of the active site. This 1HO product is in close proximity to Met 149 and Ala 152. Figure 8.15 shows the $2F_o$-F_c electron density map of the 1HO product in the active site.

The overall structure of substrate soaked haloalkane dehalogenase from AQP5750 is very similar to the native structure, with an RMSD of 0.35 Å. Figure 8.16 shows the superimposition of the active site residues in the native and complex haloalkane dehalogenase structures, showing that most residues are in similar positions. In the complex structure, Arg 145 at the entrance of the active site tunnel has shifted 2.7 Å away from the hydrophobic active site cavity into a

more open conformation. B-factors of Arg 145 in the native and complex structure of the AQP5750 haloalkane dehalogenase are similar to surrounding residues, indicating full occupancy in their positions. In one direction, the opening of the active site cavity measured 9 Å (Leu 172 to Arg 145), perpendicular to this measurement the active site cavity measured 10 Å (Met 149 to Ser 178).

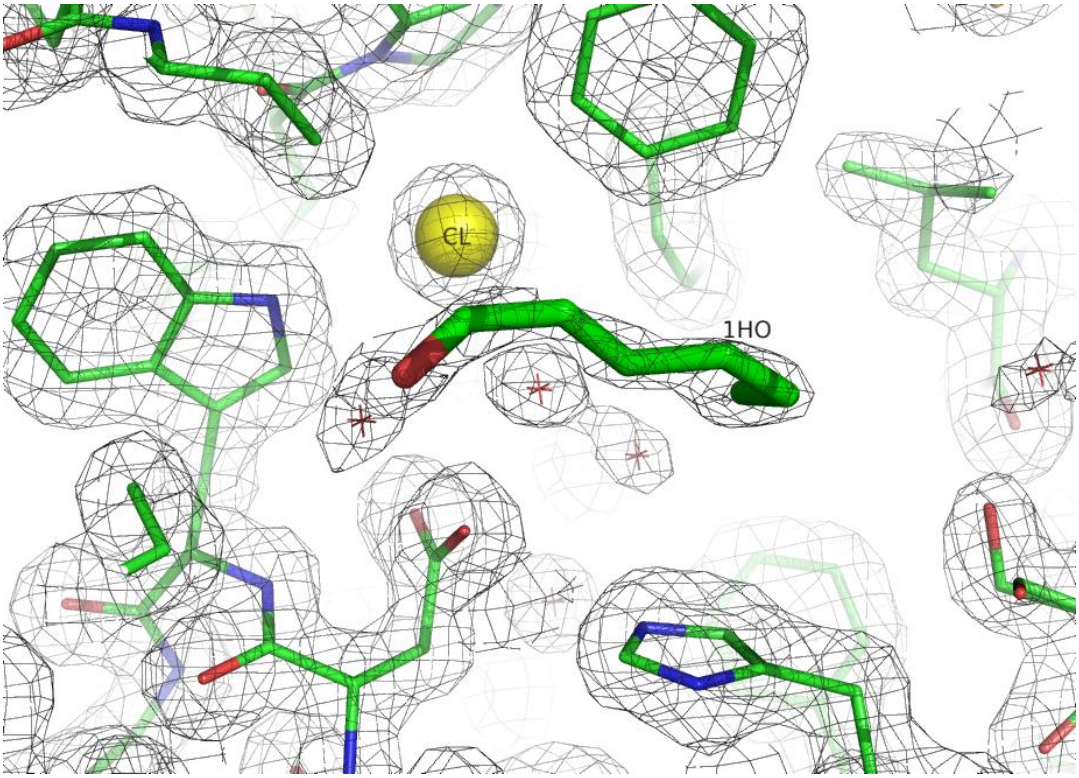


Figure 8.15- The $2F_o$-F_c electron density map for the active site cavity of the haloalkane dehalogenase from AQP5750 with the product 1HO displayed in the centre of the cavity and the chlorine shown as a sphere. The Figure was produced using the program PyMOL (DeLano Scientific).

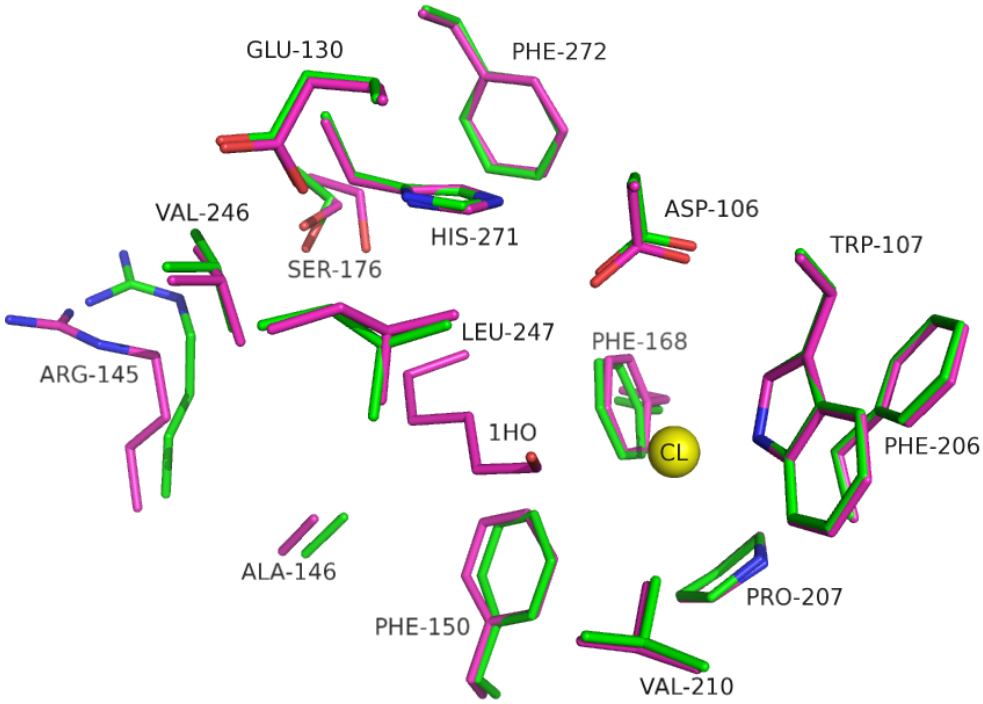


Figure 8.16- A superimposition of the active site residues in the native (green) and complex (magenta) AQP5750 haloalkane dehalogenase structures. The chlorine present in the active site is shown as a sphere. The Figure was produced using the program PyMOL (DeLano Scientific).

8.3.13 Structural comparisons

Figure 8.17 shows the superimposition of the active site residues in the haloalkane dehalogenases from AQP5750 and *S. paucimobilis* with their bound products 1HO and 1-butanol (1BO). The products 1BO and 1HO are in similar positions in the active sites.

Differences in the size and shape of the active site cavities in characterised haloalkane dehalogenases can alter their substrate specificity (Janssen, 2004). Mutagensis experiments of Leu 177 (which partially blocks the active site tunnel) in the *S. paucimobilis* haloalkane dehalogenase altered activity towards a range of substrates. Mutagenesis of Leu 177 to smaller, non-polar amino acids (Ala and Gly) generally increased the activity of the enzyme, while mutagenesis to negative, positively or polar residues (Asp, Arg, Ser and Thr) reduced activity (Chaloupkova *et al.,* 2003). The authors concluded that this could be due to

electrostatic interactions between polar residues at the entrance to the active site cavity and the dipole moment of the halogenated substrate binding the active site of the enzyme.

Hydrophobic residues Leu 177 and Ile 138 are present at opposite ends of the bound product (1BO) in the *S. paucimobilis* haloalkane dehalogenase. The corresponding residues in the AQP5750 haloalkane dehalogenase are Ser 176 and Arg 136. These two differences in amino acids may significantly change the environment of the active site cavity in comparison to the *S. paucimobilis* haloalkane dehalogenase. This could account for the poor substrate specificity observed with the AQP5750 haloalkane dehalogenase.

In contrast to the *S. paucimobilis* haloalkane dehalogenase, the haloalkane dehalogenase from *X. autotrophicus* shows high activity towards 1,2-dichloroethane (Keuning *et al*., 1985). There are large differences between the cap domains of the haloalkane dehalogenases from *S. paucimobilis* and *X. autotrophicus.* These differences give rise to a much larger active site cavity in the *S. paucimobilis* haloalkane dehalogenase. The larger active site cavity in this enzyme is thought to limit the specific binding of smaller substrates such as 1,2-dichloroethane. A complex structure with the haloalkane dehalogenase from *S. paucimobilis* bound to 1,2-dichloroethane was elucidated to 1.8 Å (Oakley *et al*., 2002). 1,2-dichloroethane is modelled in two positions in the active site cavity. Both molecules are relatively disordered and form contacts with Leu 177, Asp 108 and His 272. The active site cavities in the haloalkane dehalogenases from *S. paucimobilis* and AQP5750 are similar in size. The large active site cavity of the AQP5750 haloalkane dehalogenase may explain the low activity displayed towards small substrates such as 1,2-dichloroethane.

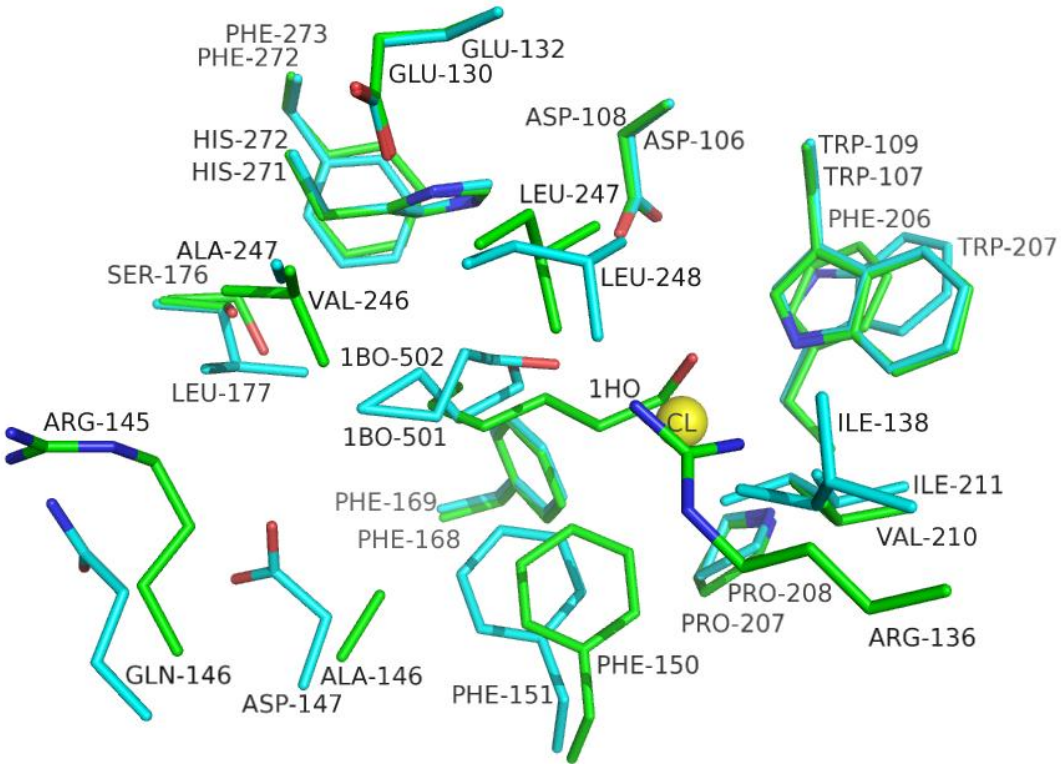


Figure 8.17- A superimposition of the active site residues in the haloalkane dehalogenase from AQP5750 with the product 1HO (green) and the haloalkane dehalogenase from *S. paucimobilis* with the product 1BO (cyan). The chlorine ion is shown as a sphere. The Figure was produced using the program PyMOL (DeLano Scientific).

8.3.14 Halide stabilisation

The *X. autotrophicus* haloalkane dehalogenase has two tryptophan halide stabilising residues (Trp 125 and Trp 175). The second halide stabilisation residues in the haloalkane dehalogenases from *S. paucimobilis* (Trp 109 and Asn 41) and AQP5750 (Trp 109 and Asn 41) is an asparagine residue.

Kinetic experiments show that the two Trp residues in the haloalkane dehalogenase from *X. autotrophicus* show a higher degree of halide stabilisation in comparison to the Trp and Asn in the haloalkane dehalogenase from *S. paucimobilis* (Prokop *et al*., 2003). The presence of a Trp and Asn as halide stabilising residues in the AQP5750 haloalkane dehalogenase may contribute to low activity observed in comparison to the *X. autotrophicus* haloalkane dehalogenase.

8.4 Summary

Substrate specificity experiments showed that the AQP5750 haloalkane dehalogenase does not show high activity towards substrates used by haloalkane dehalogenases sharing high amino acid sequence identity. The crystal structure of the AQP5750 haloalkane dehalogenase in the native state and product (1HO) bound forms were determined. The large active site cavity and the presence of Ser 176 and Arg 136 in the hydrophobic binding pocket may alter the binding of substrates, which could account for the low activity observed.

Summary and future work

9.1 Summary

Dehalogenase enzymes have the potential to be used in industrial and pharmaceutical applications and in bioremediation processes. A range of microbes from the surface of algae, sediments and crustaceans were isolated (Exeter microbial library) and assayed for bromoperoxidase and dehalogenase activity. A total of 56 different bacterial species were identified using 16S rRNA PCR. No dehalogenase or bromoperoxidase activity was detected in any of the microorganisms in the Exeter microbial library. All microorganisms in the Exeter microbial library have been incorporated into the proprietary marine microorganism collection at Aquapharm Biodiscovery Ltd. This collection exceeds over 7,000 microbes and is screened for anti-cancer, anti-oxidant, anti-inflammatory and anti-microbial compounds.

Strains AQP5750 and AQP4626 (from the Aquapharm Biodiscovery Ltd Library) tested positive for L-haloacid dehalohenase activity. Strain AQP5750 was isolated from a Polychaeta (tube worm) from Tralee beach, Argyll, UK and strain AQP4626 was isolated from a sponge of unknown species from Linnhe Marine, Argyll, UK. Both AQP5750 and AQP4626 strains were identified as having 99.7% and 99.0% sequence identity to the *Rhodobacteraceae* family of bacteria and *P. arctica* respectively. The L-haloacid dehalohenases from these wild type microorganisms were partially purified with the aim of identifying the proteins using peptide mass spectrometry. The mass spectrometry results indicated that there was no amino acid sequence identity of the partially purified proteins towards L-haloacid dehalohenases. L-haloacid dehalohenases from related species to AQP4626 and AQP5750 were used to design degenerate oligonucleotide primers. Amplification of the L-haloacid dehalohenase genes by degenerate PCR was unsuccessful.

The AQP5750 genome was sequenced using the 'in house' Illumina GA2 sequencer. *De novo* assembly was used to arrange the 72 bp paired-end reads into 1082 contigs. Using the NCBI BLAST tool (Altschul *et al*., 1990), a L-haloacid dehalohenase and a haloalkane dehalogenase were identified within the 4.5 Mbp genome.

To date, no L-haloacid dehalohenases from psychrophilic microorganisms have been characterised. The L-haloacid dehalohenase from psychrophilic *P. ingrahamii* was cloned into pET-28a with an N-terminal His-tag and over-expressed in *E. coli*. The enzyme was purified by nickel affinity and GF chromatography and biochemically characterised. The recombinant L-haloacid dehalohenase from *P. ingrahamii* shows activity towards only the l-enantiomer of the substrate, with the halogen atom attached at the α-carbon. The enzyme shows activity towards monobromoacetic (100%), monochloroacetic acid (62%), S-chloropropionic acid (42%), S-bromopropionic acid (31%), dichloroacetic acid (28%) and 2-chlorobutyric acid (10%). This concluded that the *P. ingrahamii* L-haloacid dehalohenase shows highest activity towards substrates with short carbon chains (≤C3), without preference towards a chlorine or bromine at the α-carbon position. The optimal temperature for activity of the L-haloacid dehalohenase from *P. ingrahamii* is 45ºC. The enzyme displays high thermal stability retaining 87% of its activity when incubated at 50°C for 90 min. The high thermostability of this enzyme provides evidence that proteins from extreme psychrophiles may be thermostable. The enzyme is relatively stable in organic solvents retaining 95%, 75% and 55% of activity in up to 40% methanol, DMSO and ethanol respectively. The V_{max} and K_m are 0.6 µM min^{-1} mg^{-1} 1.36 mM respectively.

The L-haloacid dehalohenase from AQP5750 was cloned into pET-28a with an N-terminal His-tag and over-expressed in *E. coli*. The enzyme was purified by nickel affinity and GF chromatography. The enzyme was subsequently biochemically characterised and the crystal structure determined.

The recombinant AQP5750 L-haloacid dehalohenases shows activity towards only the l-enantiomer of the substrate, with the halogen atom attached at the α-carbon. The enzyme shows activity towards monobromoacetic acid (100%), monochloroacetic acid (71%), S-bromopropionic acid (71%), S-chloropropionic acid (10%) and dichloroacetic acid (9%). This concluded that the AQP5750 L-haloacid dehalohenase shows higher activity towards brominated substrates and shows no activity towards longer chain substrates such as 2-chlorobutyric acid. The optimal temperature for activity of the enzyme is 55ºC and the V_{max} and K_m are 1.75 µM min^{-1} mg^{-1} and 6.72 mM respectively. The AQP5750 L-haloacid dehalohenase shows high thermal stability, retaining 85% of activity when incubated at 50ºC for 90 min and shows increased activity of up to 150% in 40% DMSO and 30% methanol. The AQP5750 L-haloacid dehalohenase is the first reported case of increased activity in solvent stability experiments for this class of enzyme. Solvent stable enzymes are attractive for the use in biotransformation reactions because of their robust nature.

The AQP5750 L-haloacid dehalohenase crystal structure was solved using molecular replacement using the L-haloacid dehalohenase from *B. cepacia* (PDB: 2NO4) as a model. Crystal soaks were performed with MCP and an intermediate crystal complex with the substrate covalently bound to the main catalytic residue Asp 18 was obtained. Superimpositions of the AQP5750 L-haloacid dehalohenase with previously characterised L-haloacid dehalohenase crystal structures revealed some major differences in active site cavity. Residues His 183, Asp 186 and Glu 21 in the active site of the AQP5750 L-haloacid dehalohenase are proposed to be involved in activation of the catalytic water which attacks the ester intermediate in the second part of the S_N2 dehalogenase mechanism.

The haloalkane dehalogenase from AQP5750 was cloned into pET-28a with an N-terminal His-tag and over-expressed in *E. coli*. The enzyme was purified by nickel affinity and GF chromatography. The enzyme was subsequently biochemically characterised and the crystal structure determined. Small amounts of activity towards 1,6-dichlorohexane, 1-bromooctane, 1,3-dibromopropane and

1-bromohexane were visualised (over 1 h) by observing a colour change from red to yellow in the colorimetric assay used. The crystal structure was solved using molecular replacement using the *S. paucimobilis* haloalkane dehalogenase (PDB: 1CV2) as a model. Crystal soaks were performed with BHX and an intermediate crystal complex with the product 1HO bound was obtained. Superimpoistions of the AQP5750 haloalkane dehalogenase with the previously characterised *S. paucimobilis* haloalkane dehalogenase structure revealed some differences in the hydrophobic active site cavity. Ser 176 and Arg 136 in the AQP5750 haloalkane dehalogenase replaced Leu 177 and Ile 138 in the *S. paucimobilis* haloalkane dehalogenase. These differences may significantly change the environment of the active site cavity and could account for the poor substrate specificity observed for the substrates tested.

9.2 Future work

There a number of ways in which this project can be developed.

Chapter 5 describes the biochemical characterisation of the *P. ingrahamii* L-haloacid dehalohenase. The recombinant protein was over-expressed at concentrations high enough for biochemical characterisation but crystallization trials could not be attempted. Protein over-expression and stability could be optimised so that crystallization trials can be attempted. The crystal structure of this enzyme could help provide an insight into why the protein shows high thermal stability.

The AQP4626 genome could be sequenced with the aim to identify the L-haloacid dehalohenase gene. This enzyme could then be cloned, over-expressed in a bacterial expression system and biochemically characterised. This would allow the direct comparison of two L-haloacid dehalohenase proteins from the same genus (*P. ingrahammi* and *P. arctica*).

The genome of AQP5750 was sequenced with the aim of identifying bromoperoxidase and dehalogenase genes. The genome sequence gives a wealth of information about the wild type bacterium which has not been investigated. Further analysis of the AQP5750 proteome could result in the identification of many other novel proteins of industrial importance.

The crystal structure of the native and complex (with MCP) AQP5750 L-haloacid dehalohenase have given an insight into the catalytic mechanism. Point mutations of key catalytic residues His 183, Asp 186 and Glu 21 which are proposed to be involved in activation of the catalytic water, followed by biochemical characterisation and structural elucidation would allow a better understanding of the second part of the S_N2 dehalogenase mechanism.

The crystal structure of the native and complex (with 1HO) AQP5750 haloalkane dehalogenase have given an insight into the low activity and poor substrate specificity observed. Point mutagenesis of residues Ser 176 and Arg 136 to hydrophobic residues followed by biochemical characterisation of the mutant

enzymes could give a better understanding of the substrate specificity and mechanism of the enzyme. Molecular modelling could be performed with the aim of determining substrates which the enzyme could have high activity towards.

A better understanding of the AQP5750 dehalogenase mechanisms could lead to protein engineering experiments with the aim of improving activity and changing substrate specificity for a specific industrial application.

References

Al Khudary, R., R. Venkatachalam, *et al.* (2010). "A cold-adapted esterase of a novel marine isolate, *Pseudoalteromonas arctica*: gene cloning, enzyme purification and characterization." Extremophiles **14**(3): 273-285.

Allpress, J. D. and P. C. Gowland (1998). "Dehalogenases: environmental defence mechanism and model of enzyme evolution." Biochemical Education **26**(4): 267-276.

Al-Sulaimani, H., S. Joshi, *et al.* (2011). "Microbial biotechnology for enhancing oil recovery: Current developments and future prospects." Biotechnology and Bioengineering **1**(2): 147-158.

Altschul, S. F., W. Gish, *et al.* (1990). "Basic local alignment search tool." Journal of Molecular Biology **215**(3): 403-410.

Angibaud, P., J. Chaumette, *et al.* (1995). "Asymmetric Synthesis of 2-chloro and 2-bromo-alkanoic acids by halogenation of α-D-glucofuranose-derived silyl ketene acetals" Tetrahedron: Asymmetry **6**(8): 1919-1932.

Asherie, N. (2004). "Protein crystallization and phase diagrams." Methods **34**(3): 266-272.

Assis, H., P. Sallis, *et al.* (1998). "Biochemical characterization of a haloalcohol dehalogenase from *Arthrobacter erithii* H10a." Enzyme and Microbial Technology **22**(7): 568-574.

Auman, A. J., J. L. Breezee, *et al.* (2006). "*Psychromonas ingrahamii* sp. nov., a novel gas vacuolate, psychrophilic bacterium isolated from Arctic polar sea ice." International Journal of Systematic and Evolutionary Microbiology **56**(5): 1001-1007.

Beloqui, A., P. D. de María, *et al.* (2008). "Recent trends in industrial microbiology." Current Opinion in Microbiology **11**(3): 240-248.

Breezee, J., N. Cady, *et al.* (2004). "Subfreezing growth of the sea ice bacterium "*Psychromonas ingrahamii*"." Microbial Ecology **47**(3): 300-304.

Brindley, A., A. Dalby, *et al.* (1998). "Preliminary X-ray analysis of a new crystal form of the vanadium-dependent bromoperoxidase from *Corallina officinalis*." Acta Crystallographica Section D: Biological Crystallography **54**(3): 454-457.

Buckland, B. C., D. K. Robinson, *et al.* (2000). "Biocatalysis for pharmaceuticals-status and prospects for a key technology." Metabolic Engineering **2**(1): 42-48.

Burroughs, A. M., K. N. Allen, *et al.* (2006). "Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes." Journal of Molecular Biology **361**(5): 1003-1034.

Burton, S., D. A. Cowan, *et al.* (2002). "The search for the ideal biocatalyst." Nature Biotechnology **20**: 35-46.

Burton, S. G. (2003). "Oxidizing enzymes as biocatalysts." Trends in Biotechnology **21**(12): 543-549.

Butler, A. (2005). "Molecular Approaches in Marine Pharmacology." Department of Chemistry & Biochemistry, UC Santa Barbara.

Butler, A. and J. N. Carter-Franklin (2004). "The role of vanadium bromoperoxidase in the biosynthesis of halogenated marine natural products." Natural Product Reports **21**(1): 180-188.

Butler, A. and J. Walker (1993). "Marine haloperoxidases." Chemical Reviews **93**(5): 1937-1944.

Cabrita, M. T., C. Vale, *et al.* (2010). "Halogenated compounds from marine algae." Marine Drugs **8**(8): 2301-2317.

Cavicchioli, R., K. S. Siddiqui, *et al.* (2002). "Low-temperature extremophiles and their applications." Current Opinion in Biotechnology **13**(3): 253-261.

Chaloupková, R., J. Sýkorová, *et al.* (2003). "Modification of activity and specificity of haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26 by engineering of its entrance tunnel." Journal of Biological Chemistry **278**(52): 52622.

Champdoré, M., M. Staiano, *et al.* (2007). "Proteins from extremophiles as stable tools for advanced biotechnological applications of high social interest." Journal of the Royal Society Interface **4**(13): 183.

Chayen, N. E., P. D. Shaw Stewart, *et al.* (1992). "Microbatch crystallization under oil--a new technique allowing many small-volume crystallization trials." Journal of Crystal Growth **122**(1-4): 176-180.

Cherry, J. R. and A. L. Fidantsef (2003). "Directed evolution of industrial enzymes: an update." Current Opinion in Biotechnology **14**(4): 438-443.

Chovancová, E., J. Kosinski, *et al.* (2007). "Phylogenetic analysis of haloalkane dehalogenases." PROTEINS: Structure, Function, and Bioinformatics **67**(2): 305-316.

Cohen, S. X., B. Jelloul, *et al.* (2007). "ARP/wARP and molecular replacement: the next generation." Acta Crystallographica Section D: Biological Crystallography **64**(1): 49-60.

Collaborative (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallographica Section D **50**(5): 760-763.

Colonna, S., N. Gaggero, *et al.* (1999). "Recent biotechnological developments in the use of peroxidases." Trends in Biotechnology **17**(4): 163-168.

Crowther, R. A. and D. Blow (1967). "A method of positioning a known molecule in an unknown crystal structure." Acta Crystallographica **23**(4): 544-548.

Cubitt, A. B., R. Heim, et al. (1995). "Understanding, improving and using green fluorescent proteins." TRENDS in biochemical Sciences **20**(11): 448-455.

D'Amico, S., T. Collins, *et al.* (2006). "Psychrophilic microorganisms: challenges for life." EMBO reports **7**(4): 385-389.

Dauter, Z. (1999). "Data-collection strategies." Acta Crystallographica Section D: Biological Crystallography **55**(10): 1703-1717.

Davis, B. G. and V. Boyer (2001). "Biocatalysis and enzymes in organic synthesis." Natural Product Reports **18**(6): 618-640.

De Backer, M., S. McSweeney, et al. (2002). "The 1.9 Å crystal structure of heat-labile shrimp alkaline phosphatase." Journal of Molecular Biology **318**(5): 1265-1274.

De Boer, E., H. Plat, *et al.* (1987). "Vanadium containing bromoperoxidase: an example of an oxidoreductase with high operational stability in aqueous and organic media." Biotechnology and Bioengineering **30**(5): 607-610.

De Jong, R. M., K. H. Kalk, *et al.* (2006). "The X-ray structure of the haloalcohol dehalogenase HheA from Arthrobacter sp. strain AD2: insight into

enantioselectivity and halide binding in the haloalcohol dehalogenase family." Journal of Bacteriology **188**(11): 4051.

de Jong, R. M., J. J. W. Tiesinga, *et al.* (2003). "Structure and mechanism of a bacterial haloalcohol dehalogenase: a new variation of the short-chain dehydrogenase/reductase fold without an NAD(P)H binding site." The EMBO Journal **22**(19): 4933-4944.

Eijsink, V. G. H., S. Gaseidnes, *et al.* (2005). "Directed evolution of enzyme stability." Biomolecular Engineering **22**(1-3): 21-30.

Eisenthal, R., M. E. Peterson, *et al.* (2006). "The thermal behaviour of enzyme activity: implications for Biotechnology." Trends in Biotechnology **24**(7): 289-292.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallographica Section D: Biological Crystallography **60**(12): 2126-2132.

Eustáquio, A. S., F. Pojer, *et al.* (2008). "Discovery and characterization of a marine bacterial SAM-dependent chlorinase." Nature Chemical Biology **4**(1): 69.

Evans, P. (2005). "Scaling and assessment of data quality." Acta Crystallographica Section D: Biological Crystallography **62**(1): 72-82.

Evans, P. and A. McCoy (2008). "An introduction to molecular replacement." Acta Crystallographica Section D **64**(1): 1-10.

Fang, Y., Z. Lu, *et al.* (2006). "A newly isolated organic solvent tolerant Staphylococcus saprophyticus M36 produced organic solvent-stable lipase." Current Microbiology **53**(6): 510-515.

Faulkner, D. J. (2001). "Marine natural products." Natural Product Reports **19**(1): 1-49.

Feling, R. H., G. O. Buchanan, *et al.* (2003). "Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*." Angewandte Chemie International Edition **42**(3): 355-357.

Feller, G. (2003). "Molecular adaptations to cold in psychrophilic enzymes." Cellular and Molecular Life Sciences **60**(4): 648-662.

Field, J. A., F. J. M. Verhagen, *et al.* (1995). "Natural organohalogen production by *Basidiomycetes*." Trends in Biotechnology **13**(11): 451-456.

Fielman, K. T., S. A. Woodin, *et al.* (2001). "Polychaete indicator species as a source of natural halogenated organic compounds in marine sediments." Environmental Toxicology and Chemistry **20**(4): 738-747.

Georlette, D., V. Blaise, *et al.* (2004). "Some like it cold: biocatalysis at low temperatures." FEMS Microbiology Reviews **28**(1): 25-42.

Gerday, C., M. Aittaleb, *et al.* (2000). "Cold-adapted enzymes: from fundamentals to biotechnology." Trends in Biotechnology **18**(3): 103-107.

Gerebtzoff, G., X. Li Blatter, *et al.* (2004). "Halogenation of drugs enhances membrane binding and permeation." ChemBioChem **5**(5): 676-684.

Gerike, U., M. J. Danson, *et al.* (1997). "Sequencing and Expression of the Gene Encoding a Cold Active Citrate Synthase from an Antarctic Bacterium, Strain DS2 3R." European Journal of Biochemistry **248**(1): 49-57.

Gomes, J. and W. Steiner (2004). "The biocatalytic potential of extremophiles and extremozymes." Food Technology and Biotechnology **42**(4): 223-235.

Gosink, J. J., R. L. Irgens, *et al.* (1993). "Vertical distribution of bacteria in Arctic sea ice." FEMS Microbiology Letters **102**(2): 85-90.

Gouet, P., E. Courcelle, *et al.* (1999). "ESPript: analysis of multiple sequence alignments in PostScript." Bioinformatics **15**(4): 305.

Gregg, K., B. Hamdorf, *et al.* (1998). "Genetically modified ruminal bacteria protect sheep from fluoroacetate poisoning." Applied and Environmental Microbiology **64**(9): 3496.

Gribble, G. W. (1973). "Fluoroacetate toxicity." Journal of Chemical Education **50**(7): 460.

Gribble, G. W. (1996). "The diversity of natural organochlorines in living organisms." Pure and Applied Chemistry **68**(9): 1699-1712.

Gribble, G. W. (1999). "The diversity of naturally occurring organobromine compounds." Chemical Society Reviews **28**(5): 335-346.

Gribble, G. W. (2003). "The diversity of naturally produced organohalogens." Chemosphere **52**(2): 289-297.

Groudieva, T., R. Grote, *et al.* (2003). "*Psychromonas arctica* sp. nov., a novel psychrotolerant, biofilm-forming bacterium isolated from Spitzbergen." International Journal of Systematic and Evolutionary Microbiology **53**(2): 539.

Häggblom, M. M. and I. D. Bossert (2004). "Halogenated organic compounds- A global perspective." Dehalogenation: 3-29.

Haki, G. and S. Rakshit (2003). "Developments in industrially important thermostable enzymes: a review." Bioresource Technology **89**(1): 17-34.

Hall-Stoodley, L., J. W. Costerton, *et al.* (2004). "Bacterial biofilms: from the natural environment to infectious diseases." Nature Reviews Microbiology **2**(2): 95-108.

Hanahan, D. (1983). "Studies on transformation of *Escherichia coli* with plasmids." Journal of Molecular Biology **166**(4): 557-580.

Harrington, R., C. Hamilton, *et al.* (1983). "Metoclopramide. An updated review of its pharmacological properties and clinical use." Drugs **25**(5): 451.

Hasan, A., H. Takada, *et al.* (1991). "Catalytic action of l-2-haloacid dehalogenase on long chain l-2-haloalkanoic acids in organic solvents." Biotechnology and Bioengineering **38**(9): 1114-1117.

Hell, C. (1881). "Ueber eine neue Bromirungsmethode organischer Säuren." Berichte der deutschen chemischen Gesellschaft **14**(1): 891-893.

Henzel, W. J. "Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases." Proceedings of the National Academy of Science of the United States of America: Biochemistry **90**: 5011-5015.

Hisano, T., Y. Hata, *et al.* (1996). "Crystal structure of l-2-haloacid dehalogenase from Pseudomonas sp. YL." Journal of Biological Chemistry **271**(34): 20322.

Holloway, P. and J. T. Trevors (1998). "A colorimetric assay for detecting haloalkane dehalogenase activity." Journal of Microbiological Methods **32**(1): 31-36.

Holscher, T., R. Krajmalnik-Brown, *et al.* (2004). "Multiple nonidentical reductive-dehalogenase-homologous genes are common in *Dehalococcoides*." Applied and Environmental Microbiology **70**(9): 5290.

Hossain, A., A. Salleh, *et al.* (2008). "Biodiesel fuel production from algae as renewable energy." American Journal of Biochemistry and Biotechnology **4**(3): 250-254.

Huang, J., Y. Xin, *et al.* (2010). "Phylogenetic diversity and characterization of 2-haloacid degrading bacteria from the marine sponge *Hymeniacidon perlevis*." World Journal of Microbiology and Biotechnology: **27** (8) 1-8.

Ichiyama, S., T. Kurihara, *et al.* (2004). "Reactivity of asparagine residue at the active site of the D105N mutant of fluoroacetate dehalogenase from Moraxella sp. B." Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics **1698**(1): 27-36.

Inatomi, K. I., M. Ohba, *et al.* (1983). "Chemical properties of proteinaceus cell wall from an acido-thermophile, *Sulfolobus acidocaldarius*." Chemistry Letters (8): 1191-1194.

Isupov, M. N., A. R. Dalby, *et al.* (2000). "Crystal structure of dodecameric vanadium-dependent bromoperoxidase from the red algae *Corallina officinalis*." Journal of Molecular Biology **299**(4): 1035-1049.

Izbicka, E., R. Lawrence, *et al.* (1998). "*In vitro* antitumor activity of the novel marine agent, ecteinascidin-743 (ET-743, NSC-648766) against human tumors explanted from patients." Annals of Oncology **9**(9): 981.

Jacquel, N., C. W. Lo, *et al.* (2008). "Isolation and purification of bacterial poly (3-hydroxyalkanoates)." Biochemical Engineering Journal **39**(1): 15-27.

Janssen, D. B. (2004). "Evolving haloalkane dehalogenases." Current Opinion in Chemical Biology **8**(2): 150-159.

Janssen, D. B., F. Pries, *et al.* (1994). "Genetics and biochemistry of dehalogenating enzymes." Annual Reviews in Microbiology **48**(1): 163-191.

Jensen, P. and W. Fenical (1996). "Marine bacterial diversity as a resource for novel microbial products." Journal of Industrial Microbiology and Biotechnology **17**(5): 346-351.

Jesenska, A. (2000). "Dehalogenation of haloalkanes by Mycobacterium tuberculosis H37Rv and other mycobacteria." Applied and Environmental Microbiology **66**(1): 219.

Jia, Z. and P. L. Davies (2002). "Antifreeze proteins: an unusual receptor-ligand interaction." TRENDS in Biochemical Sciences **27**(2): 101-106.

Jitsumori, K., R. Omi, *et al.* (2009). "X-Ray crystallographic and mutational studies of fluoroacetate dehalogenase from Burkholderia sp. strain FA1." Journal of Bacteriology **191**(8): 2630.

Jung, Y. H., J. Y. Yi, *et al.* (2010). "Overexpression of Cold Shock Protein A of *Psychromonas arctica* KOPRI 22215 Confers Cold-Resistance." The Protein Journal **29**(2): 136-142.

Kabsch, W. (1993). "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants." Journal of Applied Crystallography **26**(6): 795-800.

Kennedy, J., J. R. Marchesi, *et al.* (2008). "Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments." Microbial Cell Factories **7**: 27.

Keuning, S., D. B. Janssen, *et al.* (1985). "Purification and characterization of hydrolytic haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10." Journal of Bacteriology **163**(2): 635.

Kottmeier, S. T. and C. W. Sullivan (1988). "Sea ice microbial communities (SIMCO)." Polar Biology **8**(4): 293-304.

Koudelakova, T., E. Chovancova, *et al.* (2011). "Substrate specificity of haloalkane dehalogenases." Biochemical Journal **435**(2): 345-354.

Kurata, K., K. Taniguchi, *et al.* (1998). "Diterpenoid feeding-deterrents from *Laurencia saitoi*." Phytochemistry **47**(3): 363-369.

Kurihara, T. and N. Esaki (2008). "Bacterial hydrolytic dehalogenases and related enzymes: Occurrences, reaction mechanisms, and applications." The Chemical Record **8**(2): 67-74.

Kurihara, T., J. Q. Liu, *et al.* (1995). "Comprehensive site-directed mutagenesis of l-2-halo acid dehalogenase to probe catalytic amino acid residues." Journal of Biochemistry **117**(6): 1317.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-685.

Lahiri, S. D., G. Zhang, *et al.* (2004). "Analysis of the substrate specificity loop of the HAD superfamily cap domain." Biochemistry **43**(10): 2812-2820.

Laskowski, R. A., M. W. MacArthur, *et al.* (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." Journal of Applied Crystallography **26**(2): 283-291.

Lee, K. B. and G. S. Burrill (1995). Biotechnology: Reform, Restructure, Renewal, Ernst and Young, san Francisco, CA.

Lehmacher, A. and H. Bisswanger (1990). "Isolation and characterization of an extremely thermostable D-xylose isomerase from *Thermus aquaticus* HB 8." Journal of General Microbiology **136**(4): 679.

Leigh, J., A. Skinner, *et al.* (1988). "Partial purification, stereospecificity and stoichiometry of three dehalogenases from a Rhizobium species." FEMS Microbiology Letters **49**(3): 353-356.

Leslie, A. (1992). "Recent changes to the MOSFLM package for processing film and image plate data." Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography, No. 26.

Li, Y. (1999). "Global technical hexachlorocyclohexane usage and its contamination consequences in the environment: from 1948 to 1997." The Science of the Total Environment **232**(3): 121-158.

Li, Y. F., Y. Hata, *et al.* (1998). "Crystal Structures of Reaction Intermediates of l-2-Haloacid Dehalogenase and Implications for the Reaction Mechanism." Journal of Biological Chemistry **273**(24): 15035.

Liu, J. Q., T. Kurihara, *et al.* (1994). "Purification and characterization of thermostable and nonthermostable 2-haloacid dehalogenases with different stereospecificities from Pseudomonas sp. strain YL." Applied and Environmental Microbiology **60**(7): 2389.

Liu, J. Q., T. Kurihara, *et al.* (1998). "Reaction mechanism of fluoroacetate dehalogenase from Moraxella sp. B." Journal of Biological Chemistry **273**(47): 30897.

Lutje Spelberg, J. H., J. E. T. van Hylckama Vlieg, *et al.* (1999). "A tandem enzyme reaction to produce optically active halohydrins, epoxides and diols." Tetrahedron: Asymmetry **10**(15): 2863-2870.

Marcotte, E. M. (2007). "How do shotgun proteomics algorithms identify proteins?" Nature Biotechnology **25**(7): 755-757.

Marek, J., J. Vévodová, *et al.* (2000). "Crystal structure of the haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26." Biochemistry **39**(46): 14082-14086.

Marhuenda-Egea, F. C. and M. J. Bonete (2002). "Extreme halophilic enzymes in organic solvents." Current Opinion in Biotechnology **13**(4): 385-389.

Marshall, R. O., E. R. Kooi, *et al.* (1957). "Enzymatic conversion of D-glucose to D-fructose." Science **125**(3249): 648.

Mateo, C., J. M. Palomo, *et al.* (2007). "Improvement of enzyme activity, stability and selectivity via immobilization techniques." Enzyme and Microbial Technology **40**(6): 1451-1463.

Mazumdar, P. A., J. C. Hulecki, *et al.* (2008). "X-ray crystal structure of *Mycobacterium tuberculosis* haloalkane dehalogenase Rv2579." Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics **1784**(2): 351-362.

Messerschmidt, A. and R. Wever (1996). "X-ray structure of a vanadium-containing enzyme: chloroperoxidase from the fungus *Curvularia inaequalis*." Proceedings of the National Academy of Sciences **93**(1): 392.

Morley, K. L. and R. J. Kazlauskas (2005). "Improving enzyme properties: when are closer mutations better?" Trends in Biotechnology **23**(5): 231-237.

Mulder, J. and K. Harrap (1975). "Cytosine arabinoside uptake by tumour cells *in vitro*." European Journal of Cancer (1965) **11**(5): 373-379.

Murshudov, G., A. A. Vagin, *et al.* (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallographica Section D: Biological Crystallography **53**(3): 240-255.

Nagata, Y., R. Endo, *et al.* (2007). "Aerobic degradation of lindane (hexachlorocyclohexane) in bacteria and its biochemical and molecular basis." Applied Microbiology and Biotechnology **76**(4): 741-752.

Nagata, Y., K. Hynková, *et al.* (1999). "Construction and characterization of histidine-tagged haloalkane dehalogenase (LinB) of a new substrate class from a [gamma]-hexachlorocyclohexane-degrading bacterium, *Sphingomonas paucimobilis* UT26." Protein Expression and Purification **17**(2): 299-304.

Nagata, Y., K. Miyauchi, *et al.* (1997). "Purification and characterization of a haloalkane dehalogenase of a new substrate class from a gamma-hexachlorocyclohexane-degrading bacterium, *Sphingomonas paucimobilis* UT26." Applied and Environmental Microbiology **63**(9): 3707.

Nagata, Y., Z. Prokop, *et al.* (2005). "Degradation of hexachlorocyclohexane by haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* UT26." Applied and Environmental Microbiology **71**(4): 2183.

Nakamura, T., A. Yamaguchi, *et al.* (2009). "Roles of K151 and D180 in l-2-haloacid dehalogenase from Pseudomonas sp. YL: Analysis by molecular dynamics and ab initio fragment molecular orbital calculations." Journal of Computational Chemistry **30**(16): 2625-2634.

Nardi-Dei, V., T. Kurihara, *et al.* (1997). "Bacterial Dl-2-haloacid dehalogenase from Pseudomonas sp. strain 113: gene cloning and structural comparison with D-and l-2-haloacid dehalogenases." Journal of Bacteriology **179**(13): 4232.

Natarajan, R., R. Azerad, *et al.* (2005). "Microbial cleavage of C-F bond." Journal of Fluorine Chemistry **126**(4): 424-435.

Newman, J., S. Thomas, *et al.* (1999). "Haloalkane dehalogenases: structure of a Rhodococcus enzyme." Biochemistry **38**(49): 16105-16114.

Nichols, D., K. Lewis, *et al.* (2008). "Short peptide induces an" uncultivable" microorganism to grow *in vitro*." Applied and Environmental Microbiology **74**(15): 4889.

Niehaus, F., B. Frey, et al. (1997). "Cloning and characterisation of a thermostable [alpha]-DNA polymerase from the hyperthermophilic archaeon Thermococcus sp. TY." Gene **204**(1-2): 153-158.

Nightingale, P., G. Malin, *et al.* (1995). "Production of chloroform and other low-molecular-weight halocarbons by some species of macroalgae." 40: (4) Limnology and Oceanography: 680-689.

Oakley, A. J., Z. Prokop, *et al.* (2002). "Exploring the structure and activity of haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26: evidence for product-and water-mediated inhibition." Biochemistry **41**(15): 4847-4855.

Ogorzalek Loo, R. R., R. Hayes, *et al.* (2005). "Top-down, bottom-up, and side-to-side proteomics with virtual 2-D gels." International Journal of Mass Spectrometry **240**(3): 317-325.

Oikawa, T., T. Kazuoka, *et al.* (2003). "Paradoxical thermostable enzymes from psychrophile: molecular characterization and potentiality for biotechnological application." Journal of Molecular Catalysis B: Enzymatic **23**(2-6): 65-70.

Pang, B. and J. S. H. Tsang (2001). "Mutagenic analysis of the conserved residues in dehalogenase IVa of *Burkholderia cepacia* MBA4." FEMS Microbiology Letters **204**(1): 135-140.

Panova, A., L. J. Mersinger, *et al.* (2007). "Chemoenzymatic synthesis of glycolic acid." Advanced Synthesis & Catalysis **349**(8 9): 1462-1474.

Parsons, J. R., M. Sáez, *et al.* (2009). "Biodegradation of perfluorinated compounds." Reviews of Environmental Contamination and Toxicology **196**: 1-19.

Paul, N., R. De Nys, *et al.* (2006). "Chemical defence against bacteria in the red alga *Asparagopsis armata*: linking structure with function." Marine Ecology Progress Series **306**: 87-101.

Pelletier, I. and J. Altenbuchner (1995). "A bacterial esterase is homologous with nonhaem haloperoxidases and displays brominating activity." Microbiology **141**(2): 459.

Perelo, L. W. (2010). "Review: In situ and bioremediation of organic pollutants in aquatic sediments." Journal of Hazardous Materials **177**(1-3): 81-89.

Phadtare, S., J. Alsina, *et al.* (1999). "Cold-shock response and cold-shock proteins." Current Opinion in Microbiology **2**(2): 175-180.

Polizzi, K. M., A. S. Bommarius, *et al.* (2007). "Stability of biocatalysts." Current Opinion in Chemical Biology **11**(2): 220-225.

Pollard, D. J. and J. M. Woodley (2007). "Biocatalysis for pharmaceutical intermediates: the future is now." Trends in Biotechnology **25**(2): 66-73.

Prasher, D. C., V. K. Eckenrode, et al. (1992). "Primary structure of the Aequorea victoria green-fluorescent protein." Gene **111**(2): 229-233.

Prokop, Z., M. Monincová, *et al.* (2003). "Catalytic mechanism of the haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* UT26." Journal of Biological Chemistry **278**(46): 45094.

Querellou, J., T. Børresen, *et al.* (2010). "Marine Biotechnology: A New Vision and Strategy for Europe." A New Vision and Strategy for Europe. 15.

Ran, N., L. Zhao, *et al.* (2007). "Recent applications of biocatalysis in developing green chemistry for chemical synthesis at the industrial scale." Green Chemistry. **10**(4): 361-372.

Ridder, I. S., H. J. Rozeboom, *et al.* (1999). "Crystal structures of intermediates in the dehalogenation of haloalkanoates by l-2-haloacid dehalogenase." Journal of Biological Chemistry **274**(43): 30672.

Ridder, I. S., H. J. Rozeboom, *et al.* (1997). "Three-dimensional structure of l-2-haloacid dehalogenase from *Xanthobacter autotrophicus* GJ10 complexed with the substrate-analogue formate." Journal of Biological Chemistry **272**(52): 33015.

Riesenfeld, C. S., P. D. Schloss, *et al.* (2004). "Metagenomics: genomic analysis of microbial communities." Annual Review Genetics. **38**: 525-552.

Riley, M., J. Staley, *et al.* (2008). "Genomics of an extreme psychrophile, *Psychromonas ingrahamii*." BMC Genomics **9**(1): 210.

Roach, M. P., Y. P. Chen, *et al.* (1997). "*Notomastus lobatus* chloroperoxidase and *Amphitrite ornata* dehaloperoxidase both contain histidine as their proximal heme iron ligand." Biochemistry **36**(8): 2197-2202.

Rossmann, M. G., D. Blow, *et al.* (1964). "The relative positions of independent molecules within the same asymmetric unit." Acta Crystallographica **17**(4): 338-342.

Rye, C. A., M. N. Isupov, *et al.* (2007). "An order-disorder twin crystal of l-2-haloacid dehalogenase from *Sulfolobus tokodaii*." Acta Crystallographica Section D: Biological Crystallography **63**(8): 926-930.

Rye, C. A., M. N. Isupov, *et al.* (2009). "Biochemical and structural studies of a L-haloacid dehalohenase from the thermophilic archaeon *Sulfolobus tokodaii*." Extremophiles **13**(1): 179-190.

Sauter, N. K., R. W. Grosse-Kunstleve, *et al.* (2004). "Robust indexing for automatic data collection." Journal of Applied Crystallography **37**(3): 399-409.

Schaeffer, D. J. and V. S. Krylov (2000). "Anti-HIV activity of extracts and compounds from algae and cyanobacteria." Ecotoxicology and Environmental Safety **45**(3): 208-227.

Schanstra, J. P. and D. B. Janssen (1996). "Kinetics of halide release of haloalkane dehalogenase: evidence for a slow conformational change." Biochemistry **35**(18): 5624-5632.

Schmid, A., J. Dordick, *et al.* (2001). "Industrial biocatalysis today and tomorrow." Nature **409**(6817): 258-268.

Schmidberger, J. W., J. A. Wilce, *et al.* (2007). "Crystal structures of the substrate free-enzyme, and reaction intermediate of the HAD superfamily member, haloacid dehalogenase DehIVa from *Burkholderia cepacia* MBA4." Journal of Molecular Biology **368**(3): 706-717.

Schoemaker, H. E., D. Mink, *et al.* (2003). "Dispelling the myths-biocatalysis in industrial synthesis." Science **299**(5613): 1694.

Scott, C., G. Pandey, *et al.* (2008). "The enzymatic basis for pesticide bioremediation." Indian Journal of Microbiology **48**(1): 65-79.

Sen, R. (2008). "Biotechnology in petroleum recovery: the microbial EOR." Progress in Energy and Combustion Science **34**(6): 714-724.

Sfetsas, C. C., L. Milios, *et al.* (2009). "Characterization of 1,2-dibromoethane-degrading haloalkane dehalogenase from *Bradyrhizobium japonicum* USDA110." Enzyme and Microbial Technology **45**(5): 397-404.

Sharma, R., R. Ranjan, *et al.* (2005). "Unculturable' bacterial diversity: An untapped resource." Current Science **89**(1): 72-77.
Siddiqui, K. S. and R. Cavicchioli (2006). "Cold-adapted enzymes." Annual Review of Biochemistry. **75**: 403-433.

Simons, B. H., P. Barnett, *et al.* (1995). "Primary Structure and Characterization of the Vanadium Chloroperoxidase from the Fungus *Curvularia inaequails*." European Journal of Biochemistry **229**(2): 566-574.

Smith, J. M., K. Harrison, *et al.* (1990). "Purification and characterization of D-2-haloacid dehalogenase from *Pseudomonas putida* strain AJ1/23." Journal of General Microbiology **136**(5): 881.

Staley, J. T. and J. J. Gosink (1999). "Poles apart: biodiversity and biogeography of sea ice bacteria." Annual Reviews in Microbiology **53**(1): 189-215.

Sundaramoorthy, M., J. Terner, *et al.* (1998). "Stereochemistry of the chloroperoxidase active site: crystallographic and molecular-modeling studies." Chemistry & Biology **5**(9): 461-473.

Synnes, M. (2007). "Bioprospecting of organisms from the deep sea: scientific and environmental aspects." Clean Technologies and Environmental Policy **9**(1): 53-59.

Tao, J. and J. H. Xu (2009). "Biocatalysis in development of green pharmaceutical processes." Current Opinion in Chemical Biology **13**(1): 43-50.

Targ, A. G., N. Yasuda, *et al.* (1989). "Halogenation and anesthetic potency." Anesthesia & Analgesia **68**(5): 599.

Taylor, G. (2003). "The phase problem." Acta Crystallographica Section D **59**(11): 1881-1890.

Torz, M., P. Wietzes, *et al.* (2007). "Metabolism of mono-and dihalogenated C1 and C2 compounds by *Xanthobacter autotrophicus* growing on 1,2-dichloroethane." Biodegradation **18**(2): 145-157.

Trincone, A. (2011) "Marine Biocatalysts: Enzymatic Features and Applications." Marine Drugs **9**(4): 478-499.

Trincone, A. (2010). "Potential biocatalysts originating from sea environments." Journal of Molecular Catalysis B: Enzymatic **66**(3-4): 241-256.

Tsang, J. S. H. and L. Sam (1999). "Cloning and characterization of a cryptic haloacid dehalogenase from *Burkholderia cepacia* MBA4." Journal of Bacteriology **181**(19): 6003.

Twigg, L. E. and D. R. King (1991). "The impact of fluoroacetate-bearing vegetation on native Australian fauna: a review." Oikos: 412-430.

Vagin, A. and A. Teplyakov (1997). "MOLREP: an automated program for molecular replacement." Journal of Applied Crystallography **30**(6): 1022-1025.

Valverde, C., A. Orozco, *et al.* (2004). "Halometabolites and cellular dehalogenase systems: an evolutionary perspective." International Review of Cytology **234**: 143-199.

Van der Ploeg, J., G. Van Hall, *et al.* (1991). "Characterization of the haloacid dehalogenase from *Xanthobacter autotrophicus* GJ10 and sequencing of the dhlB gene." Journal of Bacteriology **173**(24): 7925.

Van Pée, K. H. (2001). "Microbial biosynthesis of halometabolites." Archives of Microbiology **175**(4): 250-258.

Verschueren, K. H. G., S. M. Franken, *et al.* (1993). "Refined X-ray Structures of Haloalkane Dehalogenase at pH 6 2 and pH 8 2 and Implications for the Reaction Mechanism." Journal of Molecular Biology **232**(3): 856-872.

Volhard, J. (1887). "4) Ueber Darstellung bromirter Säuren." Justus Liebigs Annalen der Chemie **242**(1 2): 141-163.

Winter, G. (2010). "xia2: an expert system for macromolecular crystallography data reduction." Journal of Applied Crystallography **43**(1): 186-190.

Wohlfarth, G. and G. Diekert (1997). "Anaerobic dehalogenases." Current Opinion in Biotechnology **8**(3): 290-295.

Woodley, J. M. (2008). "New opportunities for biocatalysis: making pharmaceutical processes greener." Trends in Biotechnology **26**(6): 321-327.

Yarnell, A. (2006). "Nature's X-factors." Chemical & Engineering News **84**(21): 12-14.

Zerbino, D. R. and E. Birney (2008). "Velvet: algorithms for de novo short read assembly using de Bruijn graphs." Genome Research **18**(5): 821.

Zhang, L. H., J. Duan, *et al.* (1998). "A simple and efficient method of preparing [alpha]-bromo carboxylic acids." Tetrahedron Letters **39**(52): 9621-9622.

10 Appendix

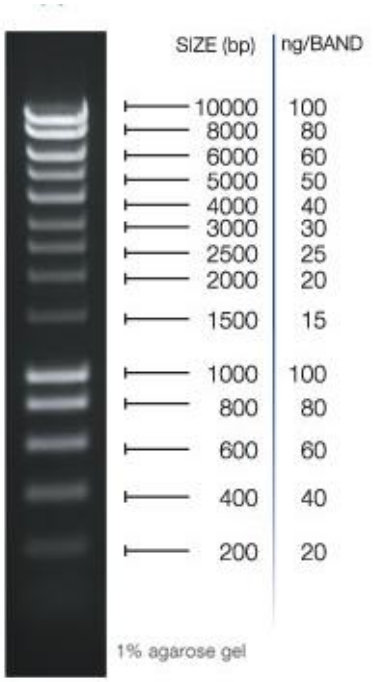


Figure 11.1- Agarose gel Hyperladder 1(Bioline).

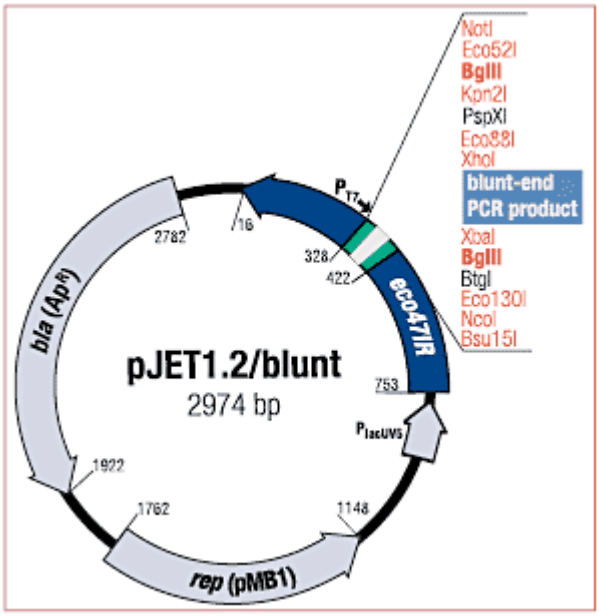


Figure 11.2- pJET1.2/blunt vector.

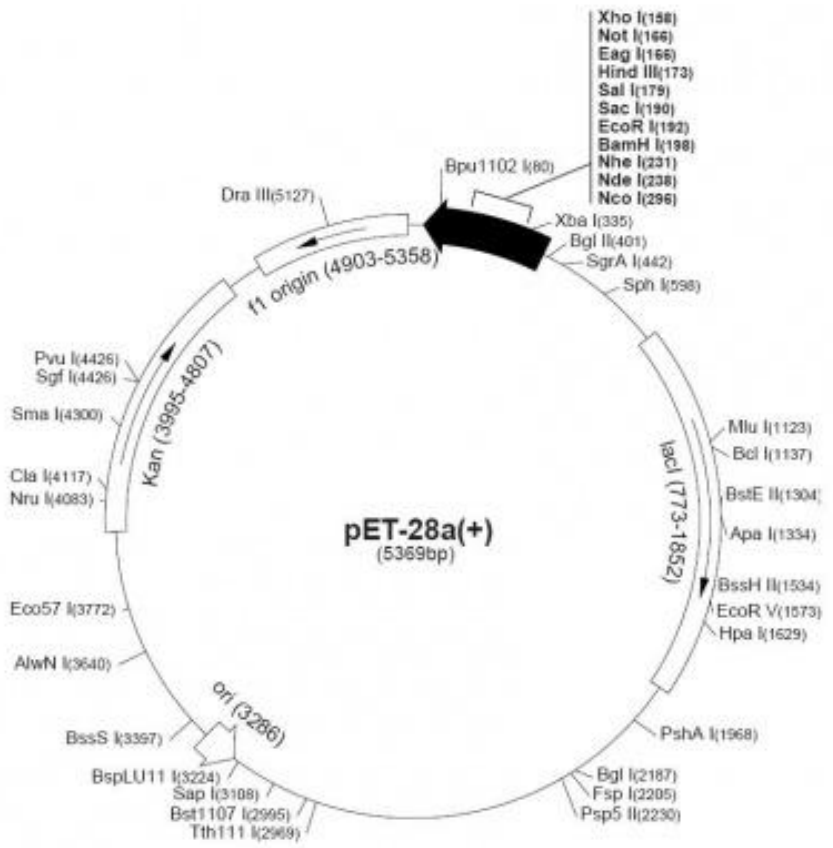


Figure 11.3- pET-28a vector.

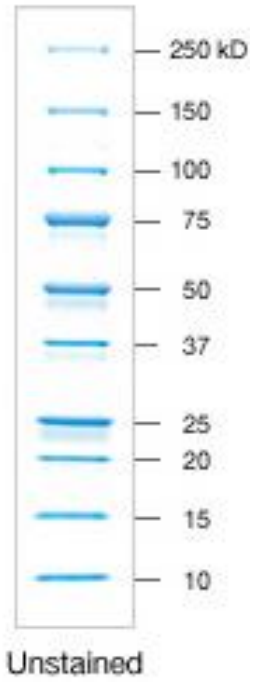


Figure 11.4- SDS-PAGE Precision plus ladder (Bio-Rad).

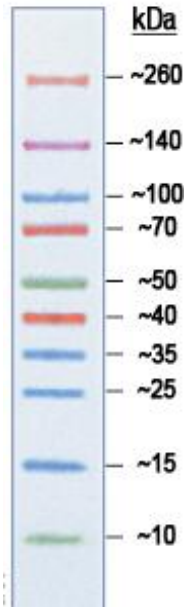
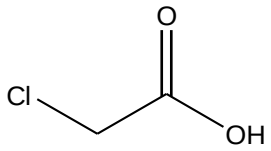
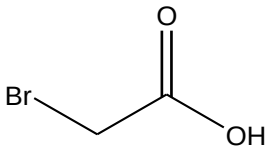
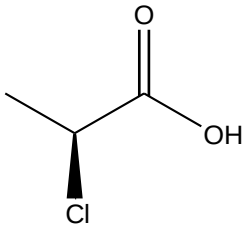
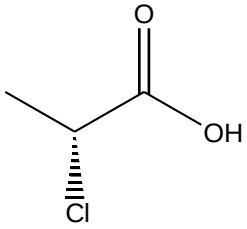
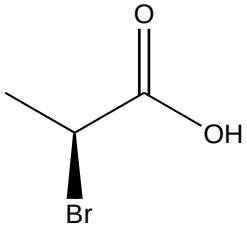
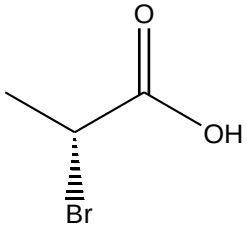
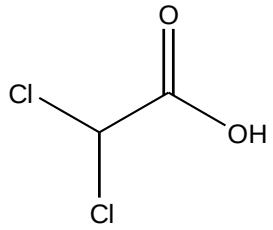
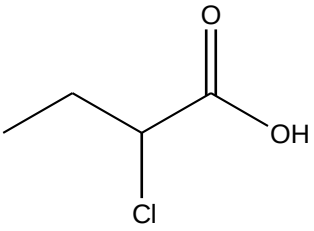
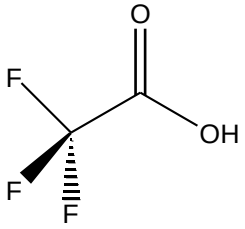
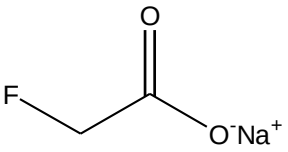
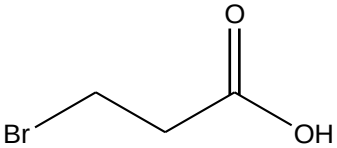
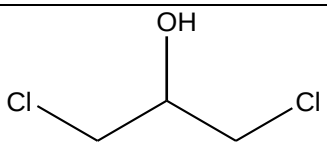
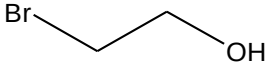
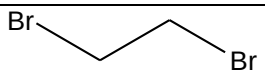
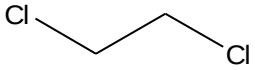
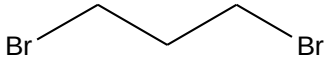
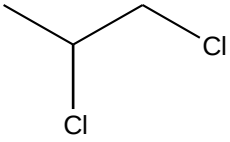
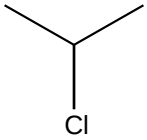
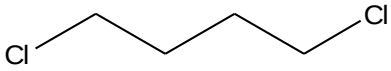
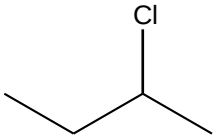
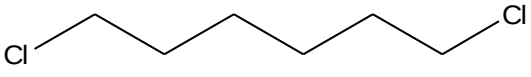
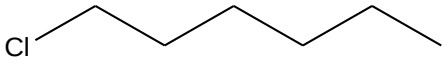


Figure 11.5- SDS-PAGE Spectra Mulicolor broad range protein ladder (Fermentas).

Name of compound	Structure
chloroacetic acid	
bromoacetic acid	
S-chloropropionic acid	
R-chloropropionic acid	
S-bromopropionic acid	
R-bromopropionic acid	

dichloroacetic acid	
2-chlorobutyric acid	
trifluoroacetic acid	
sodium fluoroacetate	
3-bromopropionic acid	
1,3-dichloro-2-propanol	
2-bromoethanol	

1,2-dibromoethane	
1,2-dichloroethane	
1,3-dibromopropane	
1,2-dichloropropane	
2-chloropropane	
1,4-dichlorobutane	
2-chlorobutane	
1,6-dichlorohexane	
1-chlorohexane	

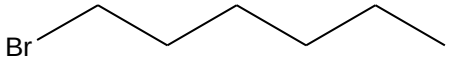
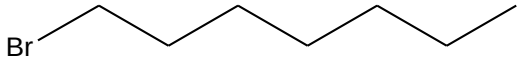
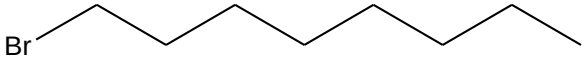
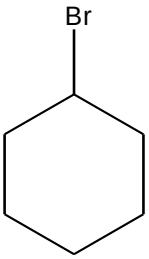
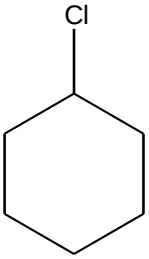
1-bromohexane	
1-bromoheptane	
1-bromooctane	
cyclohexyl bromide	
cyclohexyl chloride	

Table 11.6- The structures of the substrates used in the haloacid dehalogenase, haloalcohol dehalogenase and haloalkane dehalogenase activity assay.

M	R	W	S	Y	K	V	H	D	B	N
AC	AG	AT	GC	CT	GT	AGC	ACT	AGT	GCT	AGCT

Figure 11.7- A table displaying the wobble base codes.

ATGTCAGTGATACTTGCTTTTGATGTCTATGGAACATTAATTAATACACATGG
AATAGTGACTATTTTAGAAAAATGGTTAGGTAAAAATGCACAGGCTTTCTCTC
AGACTTGGCGAGATAAACAACTGGAATACTCTTTCCGCAGGGGATTAATGCA
GAATTACCAAAGCTTTGCAGTTTGCACCCGTCATGCTTTGGACTATACCTGT
AATTTCTATCGGGTTTCACTTTCAAAAGAGCAGAAGCAAGCGTTGATGGACG
GTTATAAAACCCTGCCAGCCTTTGATGATGCAGCAAATGCGTTAGCGCAATT
AAAGGGTGCCGGCTTTAAACTTTATGCCTTTTCTAATGGCACTAAAGAAGCC
GTTGAAGAGCTGCTTACCGGTGCAGGAATAAGGGCGTTTTTCGATGGAATT
GTCAGTTGCGATGATCTTAAATCATTTAAACCTAATCCAGCTGTCTACTGCCA
TTTTTTACGAGAATCAGGTGCTAAACACAACCAGGCATGGCTAATATCAAGC
AATCCCTTTGATGTAATTGGCGCAATATCAACAGGGATGCTGTCTGCATGGG
TACAGCGTTCTCAGGATGCCATTTTTGATCCCTGGGATGTTGATCCAACGGC
GATTATTACAAGCCTTAAAGAGCTTAAAGGCGTGCTACCCAAGCCGGACTAA

Figure 11.8- *P. ingrahammi* L-haloacid dehalohenase gene sequence.

MSVILAFDVYGTLINTHGIVTILEKWLGKNAQAFSQTWRDKQLEYSFRRGLQNY
QSFAVCTRHALDYTCNFYRVSLSKEQKQALMDGYKTLPAFDDAANALAQLKGA
GFKLYAFSNGTKEAVEELLTGAGIRAFFDGIVSCDDLKSFKPNPAVYCHFLRES
GAKHNQAWLSSNPFDVIGAISTGMLSAWVQRSQDAIFDPWDVDPTAIITSLKEL
KGVLPKPD-

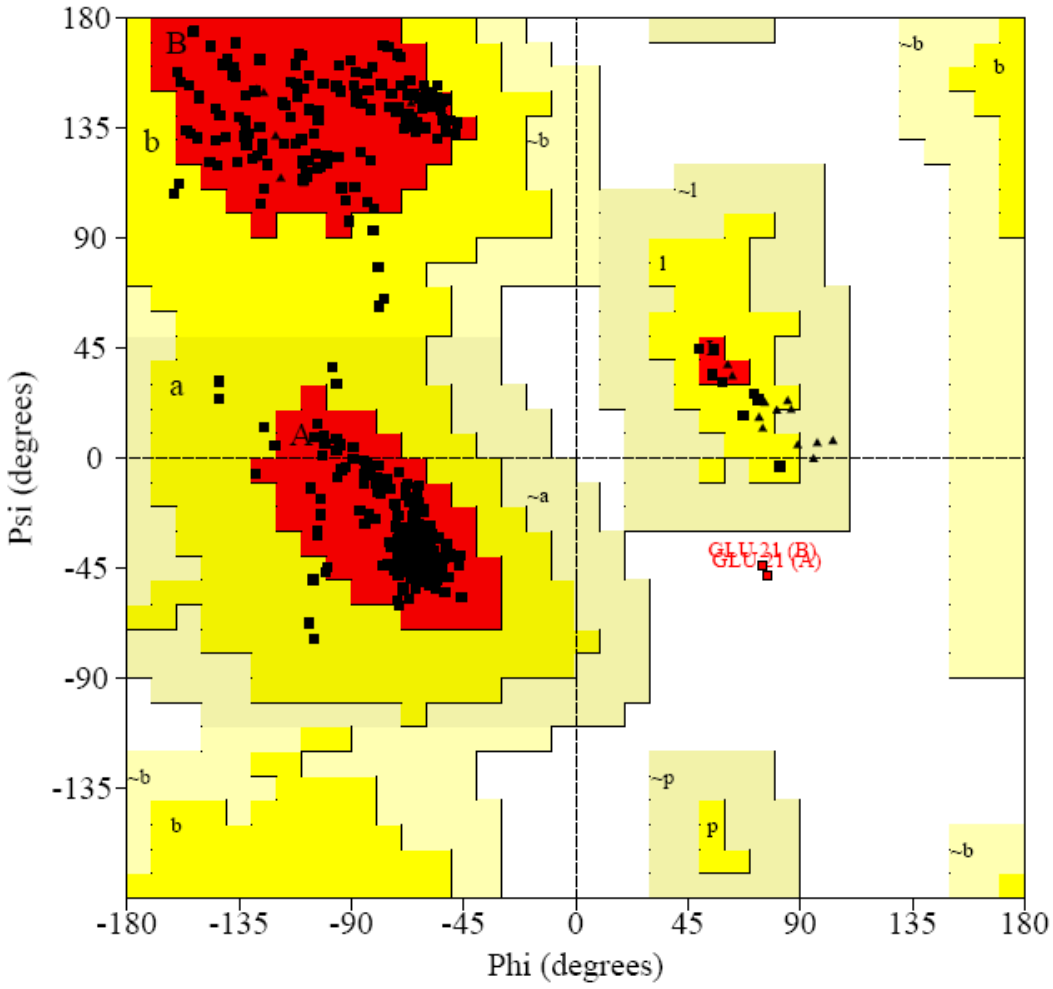
Figure 11.9- *P. ingrahammi* L-haloacid dehalohenase protein sequence.

ATGACCCCAAGCCACCCTGCCCGTCCGTCACGCTCCGGCATTCTGGTGTTT
GACGTCAATGAAACATTGCTGGACCTCACAAGCCTCTCCCCGCTCTTCGAG
CGCGTCTTCGGAGATGCGAAAGTCTTGCGCGAATGGTTCCCCGAACTGATC
CTCTATTCCCAGACCCTCACCCTGACCGGTCTCTATCGCCCCTTCGGAGAG
ATCGCCGCCGCCGTGTTTGAAATGGTCGCGGCAAACCATCAGGCAAAGGT
GACGCCCGACGACATCGCCGAATTGAAAACCCGCCTGACCTCCATGCCCG
CCTATCCCGATGTAGCCCCCGCCCTGACACGGTTGCAGGACGCGGGGTTC
CGGCTGGTCACCCTGACCAATTCCGCGCCCTCTCCCGCCCCCTCGCCTCT
GGAAAAGGCTGGCATTGCCTCTTTCTTCGAGGCGCATTTAACCGTTCATTCA
AGCCAGCGGTTCAAACCGCACCCTTCCGTCTATGACAGCACGGCAGAAACT
CTGGGCGCAAAGCCCGAGGAGCTTTGCATGATTGCCTGTCATATCTGGGAC
ACGATCGGCGCACAGGCGCGCGGCTGGCGTGGCGGATTTGTCGCCCGTC
CGCATAATACGCCGCTCACGCTGGCCGAGGTCCCGCAACCGGATTTCATCG
GGCGGGATATGGGCGAGCTGGCCGATCAACTGATCGCGAGCCTCACCGCC
TGA

Figure 11.10- AQP5750 L-haloacid dehalohenase gene sequence.

MTPSHPARPSRSGILVFDVNETLLDLTSLSPLFERVFGDAKVLREWFPELILYSQ
TLTLTGLYRPFGEIAAAVFEMVAANHQAKVTPDDIAELKTRLTSMPAYPDVAPAL
TRLQDAGFRLVTLTNSAPSPAPSPLEKAGIASFFEAHLTVHSSQRFKPHPSVYD
STAETLGAKPEELCMIACHIWDTIGAQARGWRGGFVARPHNTPLTLAEVPQPD
FIGRDMGELADQLASLTA-

Figure 11.11- AQP5750 L-haloacid dehalohenase protein sequence.



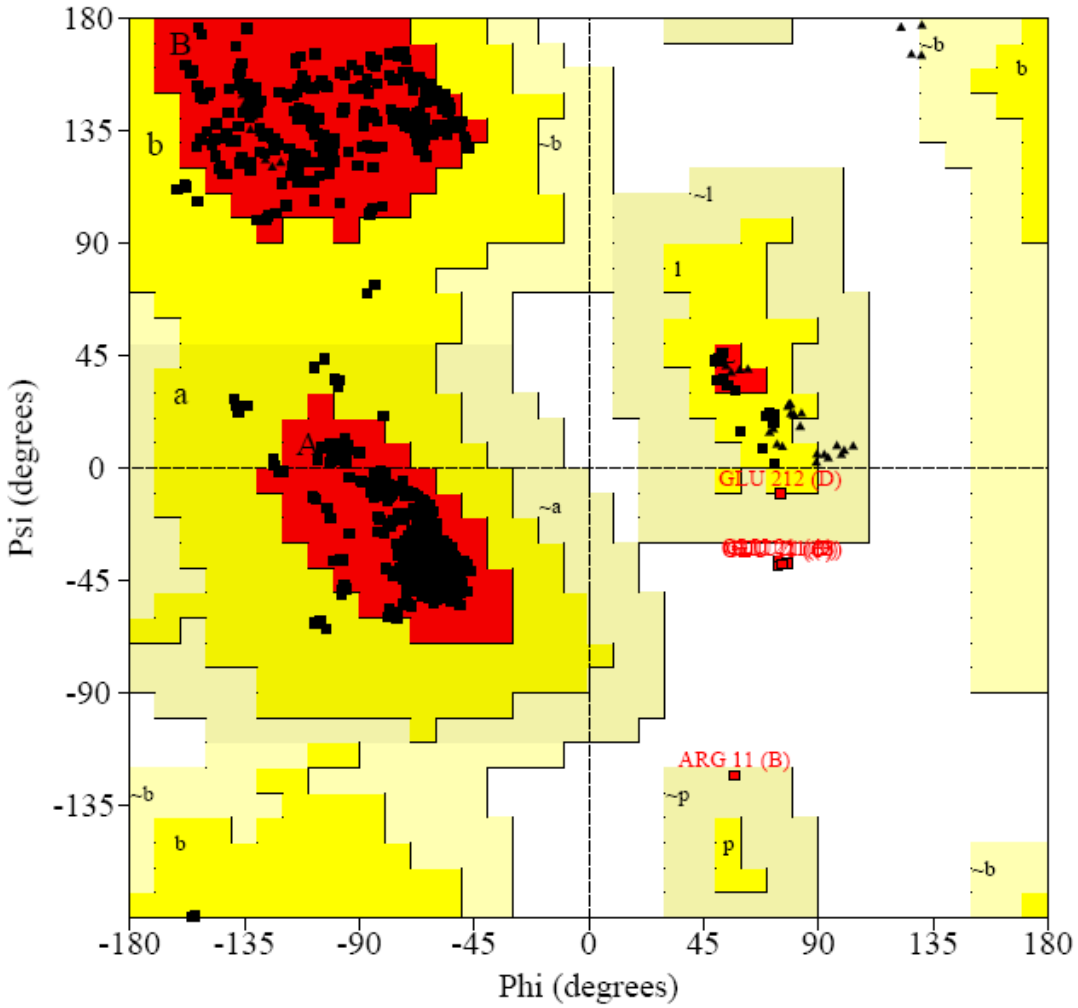
Plot statistics

Residues in most favoured regions [A,B,L]	363	94.0%
Residues in additional allowed regions [a,b,l,p]	21	5.4%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	2	0.5%
	----	------
Number of non-glycine and non-proline residues	386	100.0%
Number of end-residues (excl. Gly and Pro)	680	
Number of glycine residues (shown as triangles)	26	
Number of proline residues	37	

Total number of residues	1129	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 11.12- Ramachandran plot of the native AQP5750 L-haloacid dehalohenase.

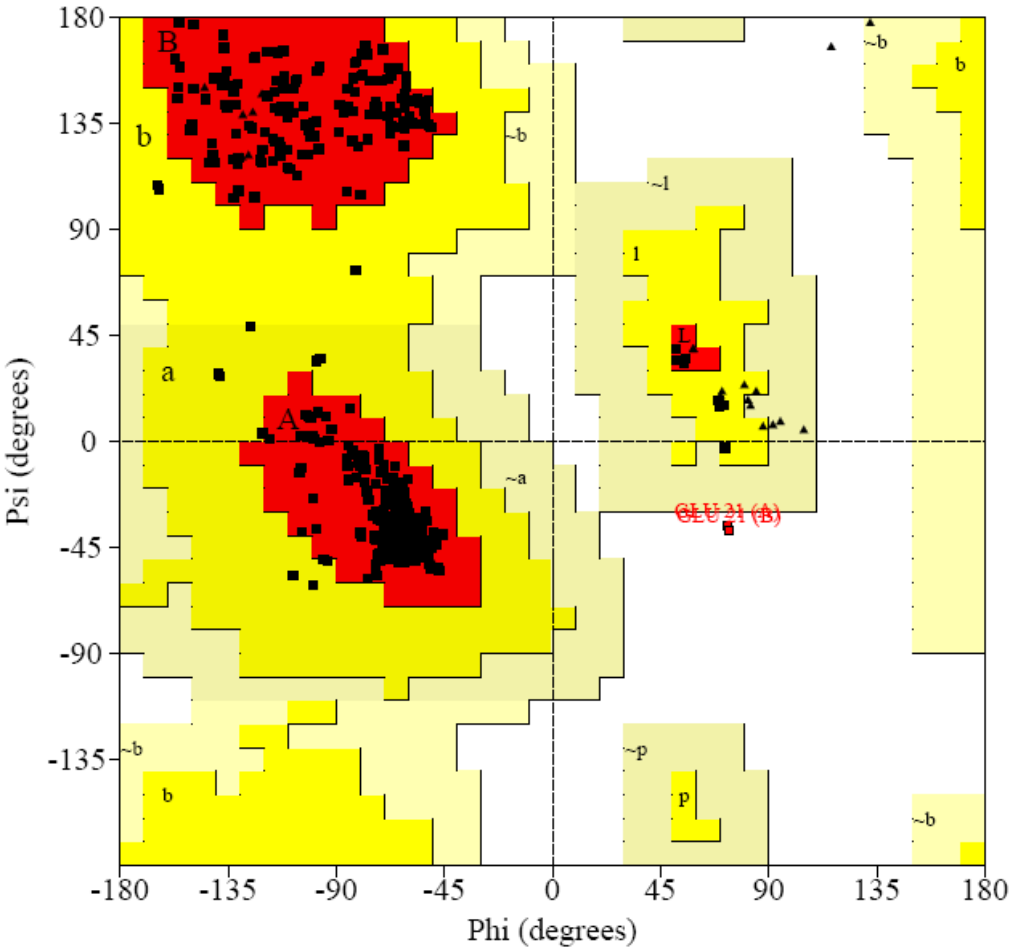


Plot statistics

Residues in most favoured regions [A,B,L]	733	95.1%
Residues in additional allowed regions [a,b,l,p]	32	4.2%
Residues in generously allowed regions [~a,~b,~l,~p]	2	0.3%
Residues in disallowed regions	4	0.5%
Number of non-glycine and non-proline residues	771	100.0%
Number of end-residues (excl. Gly and Pro)	833	
Number of glycine residues (shown as triangles)	52	
Number of proline residues	73	
Total number of residues	1729	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 11.13- Ramachandran plot of the sulfate complex AQP5750 L-haloacid dehalohenase.



Plot statistics

Residues in most favoured regions [A,B,L]	368	95.8%
Residues in additional allowed regions [a,b,l,p]	14	3.6%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	2	0.5%
Number of non-glycine and non-proline residues	384	100.0%
Number of end-residues (excl. Gly and Pro)	371	
Number of glycine residues (shown as triangles)	26	
Number of proline residues	36	
Total number of residues	817	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

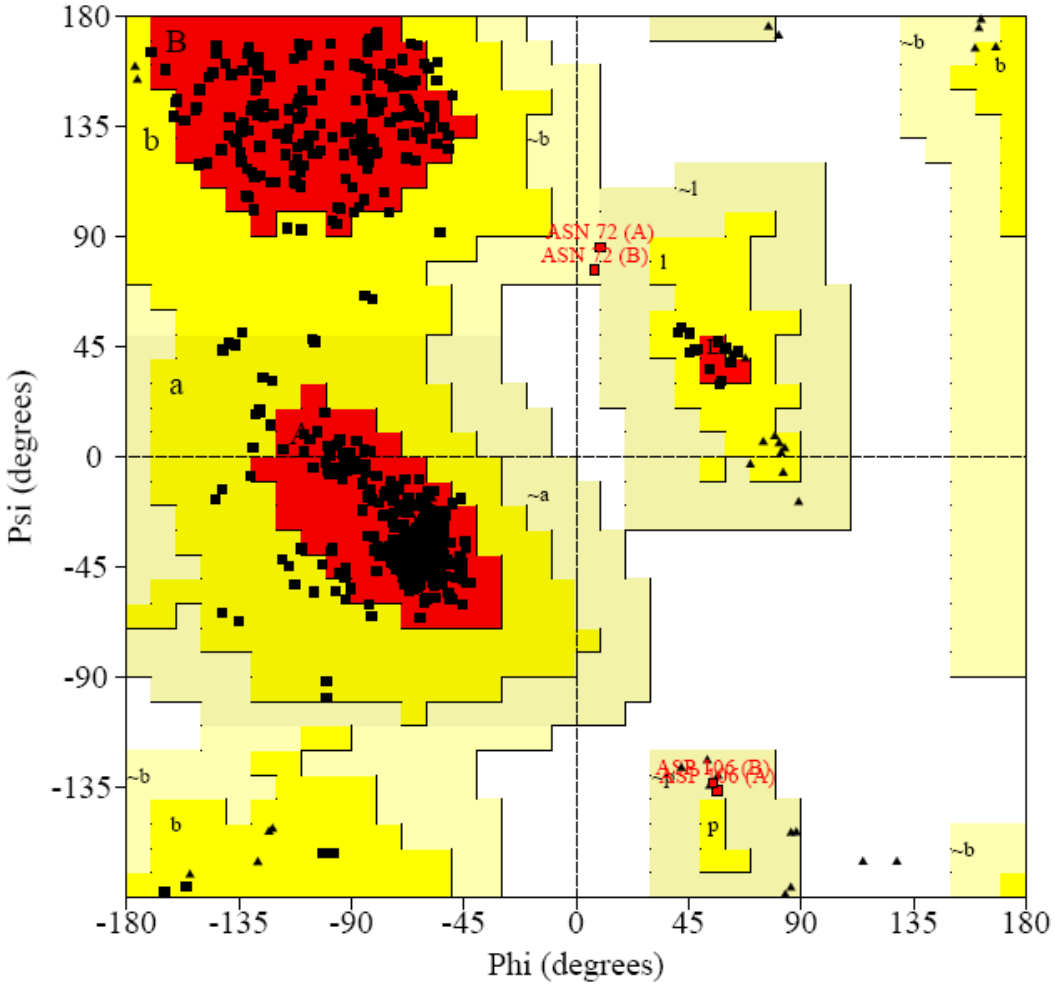
Figure 11.14- Ramachandran plot of the MCP complex AQP5750 L-haloacid dehalohenase.

ATGACTACGTCTTTCCGCGACAAAAAGAAATTTGCCACCGTCCACGGCAAAC
AGATGGCCTATATCGAAGAGGGGACCGGCGATCCAATTGTGTTCCTGCACG
GCAACCCGATGTCATCCTATCTGTGGCGCAACATCATGCCACATCTTGCGG
GCAAGGGGCGGTTGATCGCGCCAGACCTGATTGGCATGGGCGATAGTGAC
AAGCTCGATAATTCTGGCCCGGATAGCTACACCTTTGCAGAACACTGCACG
TATCTGTTTGCCTTGCTAGAACAATTGGGGGTCACCGAAAACGTCACGCTG
GTCATCCATGACTGGGGGTCCGGTTTAGGATTCCACTGGGCGCATACGCAC
TCGGATGCGGTCAAAGGCATTGCATTTATGGAAGCAATCGTCGAAACCCGC
GAGAGCTGGGATGCATTCCCCGAGCGCGCCCGCGAGATGTTCCAAGCGCT
GCGTTCGCCTGCTGGCGAAGAGATGGTTCTTGAAAAGAACCTCTTTGTCGA
GGCACTCGTGCCAGGTTCAATCCTGCGGGATTTGACCGAAGAGGAGATGAA
CGAATACCGCCGTCCCTTCGCCAACGCCGGCGAAGACCGTCGCCCCACAC
TGACATTCCCGCGTCAGGTCCCGATCGAAGGACAGCCCAAGGACGTGACC
GAACTCGTCGATGCCTATGTCGACTGGCTCGGCCAGACATCAATCCCCAAG
CTTTTCATCAATGCAGATCCCGGCGTGCTGATCACCGGCGAAGTGCGCGAT
CGCGTGCGCAGCTGGCCAAACCTCACCGAAGTGACGGTGGCCGGCTTGCA
CTTCATTCAGGAAGACTCCCCGGACGAGATCGGTGCAGCTGTGCGTGATTG
GCACGCCAGCTTGTAA

Figure 11.15- AQP5750 haloalkane dehalogenase gene sequence.

MTTSFRDKKKFATVHGKQMAYIEEGTGDPIVFLHGNPMSSYLWRNIMPHLAGK
GRLIAPDLIGMGDSDKLDNSGPDSYTFAEHCTYLFALLEQLGVTENVTLVIHDW
GSGLGFHWAHTHSDAVKGIAFMEAIVETRESWDAFPERAREMFQALRSPAGE
EMVLEKNLFVEALVPGSILRDLTEEEMNEYRRPFANAGEDRRPTLTFPRQVPIE
GQPKDVTELVDAYVDWLGQTSIPKLFINADPGVLITGEVRDRVRSWPNLTEVTV
AGLHFIQEDSPDEIGAAVRDWHAI-

Figure 11.16- AQP5750 haloalkane dehalogenase protein sequence.

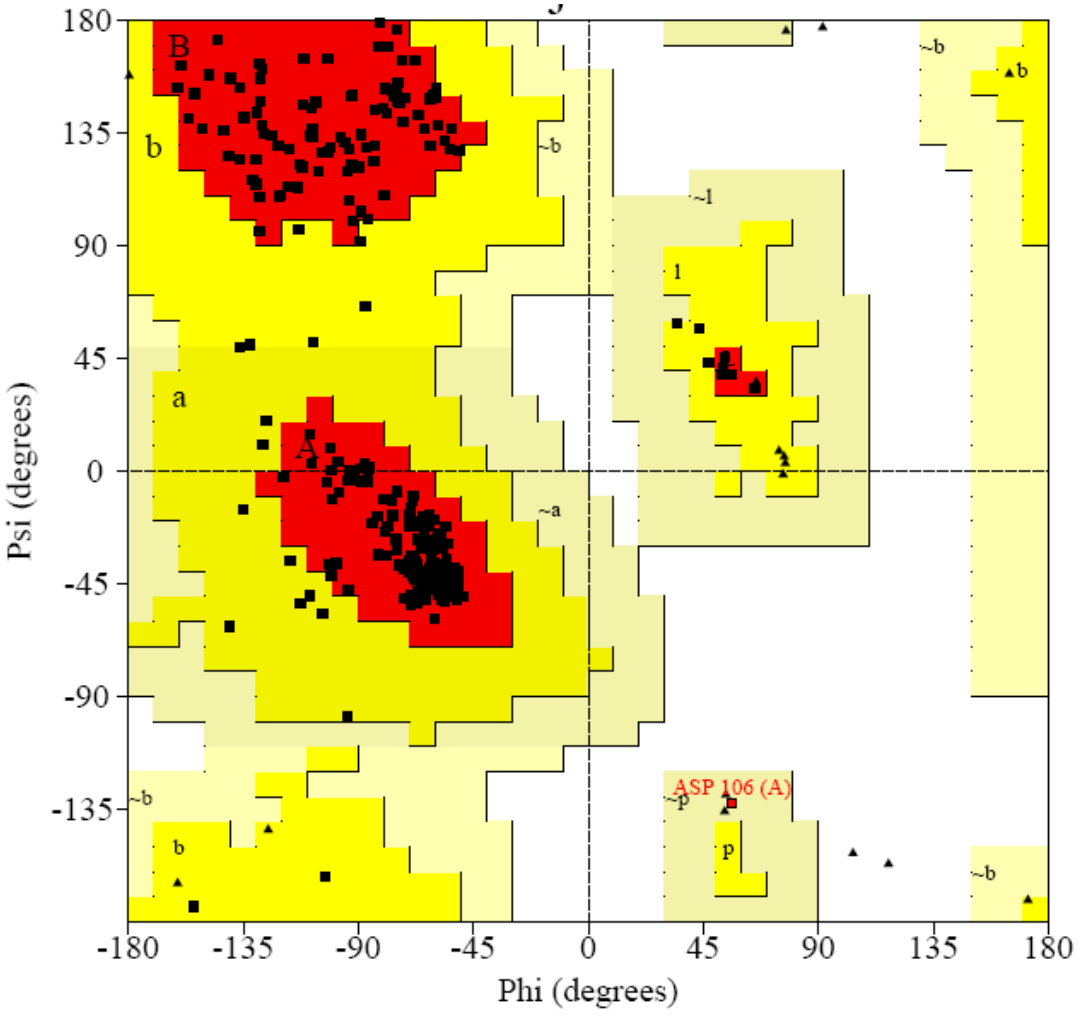


Plot statistics

Residues in most favoured regions [A,B,L]	443	88.6%
Residues in additional allowed regions [a,b,l,p]	53	10.6%
Residues in generously allowed regions [~a,~b,~l,~p]	4	0.8%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	500	100.0%
Number of end-residues (excl. Gly and Pro)	233	
Number of glycine residues (shown as triangles)	46	
Number of proline residues	34	
Total number of residues	813	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 11.17- Ramachandran plot of the native AQP5750 haloalkane dehalogenase.



Plot statistics

Residues in most favoured regions [A,B,L]	227	91.2%
Residues in additional allowed regions [a,b,l,p]	21	8.4%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.4%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	249	100.0%
Number of end-residues (excl. Gly and Pro)	317	
Number of glycine residues (shown as triangles)	23	
Number of proline residues	17	
Total number of residues	606	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 11.18- Ramachandran plot of the complex 1HO AQP5750 haloalkane dehalogenase.