Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

EXPRESSION, PURIFICATION AND MUTAGENESIS OF RECOMBINANT CLASS 1 ALDEHYDE DEHYDROGENASE

A thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Molecular Biology at Massey University

Jane Margaret Wyatt

1997

ABSTRACT

Aldehyde dehydrogenase (AlDH) catalyses the conversion of aldehydes, for example acetaldehyde and retinal, to carboxylic acids in an NAD^+-dependent reaction involved in the detoxification of aldehydes and alcohols. There are several isoenzymes, class 1 being the cytosolic form.

Three over-expression and purification systems have been tested, in order to gain a high yield of active, pure class 1 aldehyde dehydrogenase. The traditional method, using T7 polymerase-driven expression in *E. coli* followed by ion-exchange and *p*-hydroxyacetophenone-affinity chromatography, gave 3 mg/L human aldehyde dehydrogenase with a high specific activity of 1.2 units/mg. Human Class 1 AlDH has been over-expressed and purified using the GST Gene Fusion System (Pharmacia Biotech), avoiding the need for AlDH-affinity chromatography and therefore allowing straight-forward purification of mutated enzymes. The GST fusion system produced 2.6 mg/L pure AlDH with a specific activity of 0.39 units/mg. The methylotrophic yeast *Pichia pastoris* was chosen for its high yields, in the region of grams/litre. Preparatory work was carried out with the construction of the expression plasmids and screening of the *Pichia* transformants.

A highly conserved lysine residue (Lys-272) may be involved in acid-base catalysis of aldehyde oxidation, as well as of the esterase reaction also catalysed by AlDH. Preliminary work has been carried out on the generation of the K272A, K272R, K272H and K272L altered enzymes. The resultant activity level, kinetic behaviour and active site structure of these modified enzymes should help to elucidate further the mechanism of action of AlDH.

ACKNOWLEDGEMENTS:

I would like to offer a very large thank you to my supervisors, Dr Mike Hardman and Dr Mark Patchett. Thank you for your time and patience, and for allowing me my independence in order to make my own mistakes. In addition, the writing of this thesis has been greatly helped by Mike, who would read chapter drafts in a day (the headings did renumber themselves!), and my time in the lab would have been far more stressful without Mark's extensive knowledge of the best way to do things and his ability to think around a problem to find solutions.

My thanks are due to two of our Twilight Zone postdocs, Dr Shaun Lott and Dr Catherine Day. Shaun coped well with a never-ending barrage of questions and often managed to make me see the overall picture when I had been incapable of seeing the wood for the trees. Mrs Carole Flyger deserves a special mention for her perseverance in keeping the lab running when we all needed everything yesterday. My thanks also to everyone who has worked in the Twilight Zone over the last two years, for making the time an enjoyable one.

I would like to thank my friends and flatmates over the last two years, especially Kathryn Frith who has put up with my stresses and moans and then offered me a place to stay for my last three months of this thesis - thank you. Nachos will never be the same again.

My parents have always supported me in everything that I have strived to do. Your love, encouragement and friendship have meant so much to me. Thank you for believing in me; this thesis is dedicated to you.

Finally my thanks to Andrew Milne. We have survived both my thesis and you in Gisborne; Christchurch here we come!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION AND LITERATURE REVIEW.

1.1 ALDEHYDE DEHYDROGENASE.

Aldehyde dehydrogenase (AlDH) catalyses the oxidation of a broad variety of aldehyde substrates to their corresponding carboxylic acids, in an irreversible, NAD^+- dependant reaction, as shown in Figure 1-1.

In vivo, aldehydes are usually intermediates derived from the metabolism of compounds such as amino acids, biogenic amines, carbohydrates, vitamins, steroids and lipids (Lindahl, 1992). Aldehydes are highly reactive due to the electrophillic carbonyl group and react with cellular nucleophiles, for example nucleic acids and proteins (Schauenstein *et al.*, 1977). Due to this reactivity, and their long-lived nature, aldehydes have been suggested to have roles in cytotoxity, mutagenicity, genotoxicity and carcinogenicity (Lindahl, 1992).

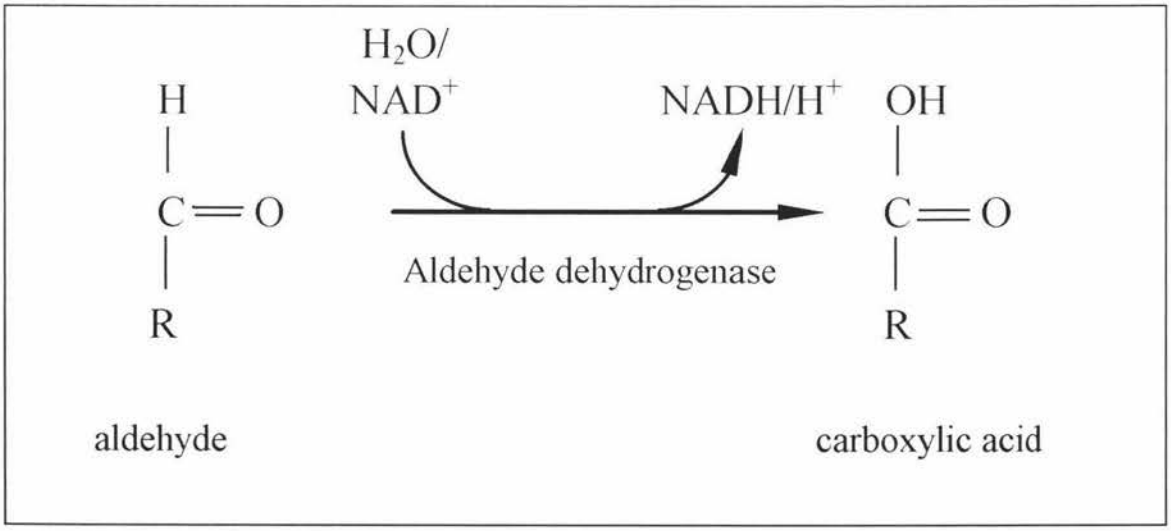

Figure 1-1: Schematic diagram of the oxidation catalysed by aldehyde dehydrogenase. The R group may be aliphatic or aromatic.

Aldehydes are metabolised by aldehyde oxidases, aldo-keto reductases and aldehyde dehydrogenases. All three have broad specificities yet each

appears to have a different physiological role (Lindahl, 1992); aldehyde oxidases are involved in the oxidation of purines and pyrimidines, aldo-keto reductases metabolise sugar aldehydes, ketones and quinones; and aldehyde dehydrogenases have a broad substrate specificity including aliphatic and aromatic aldehydes.

Aldehyde dehydrogenases have an active role in the detoxification system of the cell. As an example, lipid peroxidation is a process that occurs when free oxygen radicals attack the cell membrane lipids, producing a variety of aldehydes (Lindahl, 1992). The potentially harmful effects of these aldehydes are neutralised by aldehyde dehydrogenases.

Aldehyde dehydrogenases are also involved in the metabolism of ethanol, which is metabolised by two irreversible oxidation steps, mainly in the liver. Alcohol dehydrogenase oxidises the ethanol to acetaldehyde, and aldehyde dehydrogenase oxidises the acetaldehyde to acetate. Disulfiram is an anti-alcohol drug that specifically inhibits the action of aldehyde dehydrogenase, if disulfiram is ingested and alcohol is then taken acetaldehyde builds up in the liver and causes unpleasant symptoms for the person (Kitson, 1989).

Retinoic acid, the end product of retinol (vitamin A) metabolism, has been shown to regulate growth and development at a transcriptional level. Retinoic acid acts as a ligand that binds to receptor proteins (for example the retinoic acid receptor, RAR, and the retiniod X receptor, RXR) that then act as transcription factors (Mangelsdorf and Evans, 1995). The targets of these transcription factors appear to be genes involved in embryonal development and cell differentiation (Duester, 1996). While the conversion of retinol to retinal is the rate-limiting step, it is reversible, unlike the oxidation of retinal to retinoic acid, catalysed be aldehyde dehydrogenase and cytochrome P450 (Duester, 1996).

Aldehyde dehydrogenase also has a substantial esterase activity, the hydrolysis of nitrophenyl esters (Feldman and Weiner, 1972).

1.2 Classes of Aldehyde Dehydrogenase.

Aldehyde dehydrogenases are a family of related enzymes and have been classified into several isozyme forms using physical and kinetic properties, including primary sequence information. The nomenclature for the classification of the isozymes is: class 1, the cytosolic isozymes; class 2, the mitochondrial enzymes; and class 3, the tumour forms (a comprehensive list of the enzymes falling into each of these categories is published in Lindahl, 1992). More forms have since been characterised and classified and at least five forms have been purified from human tissue (Yin *et al.*, 1995). There are constitutive and inducible forms of aldehyde dehydrogenase in classes 1 and 3, but class 2 enzymes are constitutive only (Lindahl, 1992).

The class 1 and 2 forms are tetramers, are similar in terms of molecular weight (M_r 210,000 - 220,000) and have similarly low K_m values for acetaldehyde. The sequence comparison between the human cytosolic AlDH and the human mitochondrial AlDH shows a 68% identity at the amino acid level (Hempel *et al.* 1993).

A variety of kinetic and chemical modification studies have characterised the class 1 and 2 enzymes, including NAD^+ and NADH binding, steady state kinetics and stopped-flow kinetics (Blackwell *et al.*, 1989), discussed in greater depth in Sections 1.3.1 and 1.3.2.

The class 3 isozymes tend to be a dimer with an M_r of 110,000 and have only 30% identity in protein sequence with enzymes in the other classes (Hempel *et al.*, 1989; Yoshida *et al.*, 1991).

1.2.1 Class 1 (Cytosolic) Aldehyde Dehydrogenase.

The gene for the cytosolic aldehyde dehydrogenase has been isolated from a human genomic library (Hsu *et al.*, 1989) and the cDNA has been cloned and sequenced for both the human and sheep iso-forms (human, Zheng *et al.*, 1993); sheep, Staynor and Tweedie, 1995). The sheep and human cDNAs have a 90% identity and the amino acid sequence identity is 91%, and the sheep class 1 cDNA has a 65 % identity with human class 2 cDNA (Staynor and Tweedie, 1995).

The human and sheep class 1 aldehyde dehydrogenases have been purified from liver to homogeneity (human, Greenfield and Pietruszko, 1977; sheep, MacGibbon *et al.*, 1979). The human and sheep recombinant class 1 aldehyde dehydrogenases have also been overexpressed in *E. coli* and purified to homogeneity (human, Zheng *et al.*, 1993; sheep, Jones *et al.*, 1995).

1.3 Mechanism of Catalysis.

While no three-dimensional X-ray structure has yet been described, the structure of the active site and the catalytic mechanism of aldehyde dehydrogenase have been analysed. A proposed reaction pathway is shown in Figure 1-2 (Weiner *et al.*, 1995), which is based on the results of several studies using the biochemical approaches discussed in Sections 1.3.1-1.3.4.

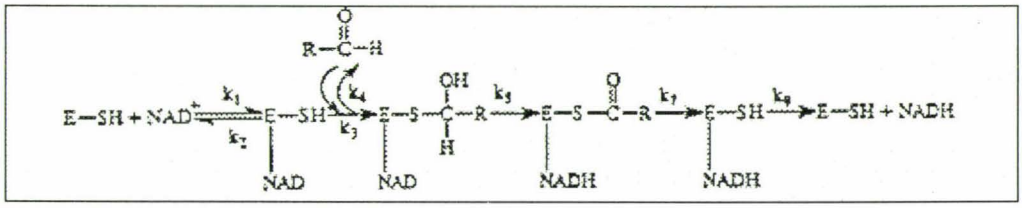

Figure 1-2: Model showing the reaction pathway for the aldehyde dehydrogenase catalysed oxidation of an aldehyde. From Weiner *et al.*, 1995.

In this model, an active site nucleophile attacks the aldehyde to form a thiohemiacetal intermediate, which is then oxidised to form an acyl enzyme. Deacylation of the acyl enzyme is probably catalysed by a general base. The active site nucleophile has been shown to be Cys-302 (Weiner *et al.*, 1995) and the general base has not yet been identified, but could be a histidine, lysine or arginine residue.

Investigations into the roles of the putative active site residues have been conducted using kinetic studies, chemical modification experiments, analysis of conserved residues within the families of enzymes, and site-directed mutagenesis. The contributions from each of these studies will be looked at in turn.

1.3.1 Kinetic Experiments.

Kinetic experiments using NAD^+ and various aldehydes have shown that the oxidation of aldehyde to acid occurs by an ordered bi-bi mechanism. NAD^+ binds to the enzyme first, then the aldehyde binds, catalysis occurs, the release of acid and then NADH (Vallari and Pietrusko, 1981; Blackwell *et al.*, 1989). This model of an ordered mechanism is supported by the tight binding of both NAD^+ and NADH to free AlDH (MacGibbon *et al.*, 1979), as the co-factor is able to be the first molecule to bind to the enzyme and the last to be released. NADH is a competitive inhibitor with respect to NAD^+, also consistent with this mechanism of catalysis (Blackwell *et al.*, 1989).

Stopped-flow experiments on class 1 aldehyde dehydrogenase showed a burst of NADH production that suggested that the hydride transfer step (k_5 in Figure 1-2) comes before the rate-limiting step. This supports the data showing that the binding of NAD^+ was not rate-limiting (Blackwell *et al.*, 1989). NADH release appears to be a biphasic reaction, where the enzyme-NADH complex undergoes conformational change before the release of NADH can occur (MacGibbon *et al.*, 1977). The conformational change and NADH release are the rate-limiting steps (Hill *et al.*, 1991) as is typically

the case for dehydrogenases, including glyceraldehyde 3-phosphate dehydrogenase (Blackwell *et al.*, 1989).

The number of NADH binding sites per tetramer has been shown to range between 1 and 3 (MacGibbon *et al.*, 1979) and, in contrast to the mitochondrial form, Mg^{2+} does not increase the number of functioning subunits (Dickenson and Hart, 1982). There are no cooperative subunit effects (Blackwell *et al.*, 1989).

1.3.2 Chemical Modification.

By allowing the enzyme to react with chemicals that 'tag' the amino acid residues involved in catalysis, the importance and function of these residues can be established. Properties affected by these modifications may include subunit association/dissociation, cofactor binding, substrate binding and product release, as well as alterations in specificity.

The Cys-302 residue has been shown to be the active site nucleophile through many chemical modification studies including reacting the enzyme with iodoacetamide (Hempel *et al.*, 1982), a brominated coenzyme analogue (von Bahr-Lindström *et al.*, 1985), substrate analogues (Blatter *et al.*, 1992) and a chromophoric reporter group (Kitson and Kitson, 1995). In each of these cases the modified aldehyde dehydrogenase was found to be inactive.

Glu-268 has also been implicated in the catalytic mechanism through chemical modification using the active site-directed reagent, bromoacetophenone; the labelled enzyme is irreversibly inactivated (Abriola *et al.*, 1990).

Some evidence has also been presented that Ser-74 could be the active site nucleophile. AlDH can be modified by reaction with *trans*-4-(*N*,*N*-dimethylamino) cinnamaldehyde (DACA), a chromophoric aldehyde which is a poor substrate. A tryptic digest of the modified-enzyme showed that the

labelled peptide contained Ser-74, but no cysteine residues (Loomes *et al.*, 1990).

1.3.3 Residue Conservation Analyses.

An alignment of the amino acid sequences of 16 AlDH proteins was carried out using the Wisconsin GCG Pileup program (Hempel *et al.*, 1993). Twenty-three invariant residues were found with 14 of these being glycine or proline. The glycine and proline residues are likely to be involved with the secondary and tertiary structure of the enzyme rather than the active site, as they allow flexibility or rigidity respectively, allowing the protein to adopt a certain conformation.

Cys-302 was completely conserved, as was Glu-268. No histidine residues were strictly or closely conserved. Lys-272 was found to be conserved in 15 of the 16 AlDHs, the exception being the replacement by an asparagine in succinic semi-aldehyde dehydrogenase from *E. coli* (Hempel *et al.*, 1993).

1.3.4 Mutagenesis Experiments.

Site-directed mutagenesis is a molecular biological tool which enables the DNA sequence to be altered in a predetermined way, causing a change in amino acid sequence for the resulting expressed protein. The effect of the alteration can be determined by crystallographic and kinetic experiments.

There needs to be caution in attributing the loss of catalytic activity to the substitution of an active site residue. The loss of activity could also be due to indirect conformational changes affecting, for example, subunit binding, cofactor binding, stability, and the folding of mobile loops (Knowles, 1987).

The putative active site nucleophile, Cys-302, in rat class 2 AlDH has been mutated to alanine (C302A), and the altered enzyme was found to be inactive (Farrés *et al.*, 1995). This residue was also mutated to a serine, a

poor nucleophile (C302S). In the latter case, enzyme activity was observed, but as the k_{cat}/K_m was 7 x 10^5 times lower than the wild-type the modified enzyme was a much poorer catalyst than the wild-type enzyme. These results are further evidence that the nucleophile required for catalytic activity is Cys-302.

Glu-268 has been altered to an aspartate, a glutamine and a lysine in human liver class 2 AlDH (Wang and Weiner, 1995). Less than 0.4% of the wild-type specific activity was present for each of the mutants. These substitutions did not affect NAD^+ binding, as the K_m value for NAD^+ was unchanged. Stopped-flow experiments showed an absence of the burst of NADH formation that occurs in the wildtype enzyme (Blackwell *et al.*, 1989); the rate-limiting step is up to and including the hydride transfer step in the E268Q mutant. This alteration in the rate-limiting step may indicate that the glutamate residue is involved in the activation of Cys-302 ready for catalysis, perhaps as part of a charged relay triad (Wang and Weiner, 1995).

The Ser-74 residue of class 2 AlDH has been converted into an alanine in order to investigate its role in the catalytic mechanism (Rout and Weiner, 1994). The mutant enzyme (S74A) still had some activity, although lowered to 10 % of the activity of the wild type. This shows that Ser-74 is not the active site nucleophile. However, it may be involved in the cofactor binding site, as the NAD^+ and NADH binding properties in the mutant are altered.

Highly conserved histidine residues have been mutated to alanine and other amino acids (Zheng and Weiner, 1993b). The results showed that none of the conserved histidines appeared to be involved in the active site, but two of the residues, H29 and H235, were likely to be important with the maintenance of the secondary or tertiary structure of the enzyme.

1.4 Over-expression Studies.

Further work into understanding the mechanism of action of AlDH is limited by two factors; firstly, the lack of a fully determined X-ray structure, and secondly, the lack of sufficient expressed and purified recombinant enzyme - both wild-type and altered - to carry out crystallographic and kinetic experiments.

Wild type classes 1 and 2 aldehyde dehydrogenases were first purified from human autopsy liver in 1977 by Greenfield and Pietruszko, and are routinely purified from the native source. However, more information about the structure and function of aldehyde dehydrogenase can be gained from the application of molecular biology and protein engineering techniques. Central to this change in approach was the cloning of many AlDH cDNAs and the subsequent heterologous expression of the various forms of the enzyme, as discussed in Section 1.3.4.

Three expression systems were tested within this work; pT7-7 expression, the GST fusion-protein system (both in *E. coli*), and expression using the yeast *Pichia pastoris*. The reasons for testing these systems will be discussed in turn.

1.4.1 The *E. coli* pT7-7 Expression System.

Expression of human class 1 and class 2 AlDH in *E. coli* has been achieved by cloning the cDNA into the pT7-7 expression system (Weiner, 1991; Zheng and Weiner, 1993, Jones *et al.*, 1995). The class 1 human AlDH cDNA has been expressed and purified using the following strategy giving homogenous aldehyde dehydrogenase with a specific activity of 0.25 μmol/min/mg (Zheng and Weiner, 1993) and 0.34 μmol/min/mg (Jones *et al.*, 1995).

The RNA polymerase from the bacteriophage T7 is a single polypeptide with a molecular weight of 99 kDa, which recognises a specific 23 base pair promoter. By placing the gene for T7 RNA polymerase under the control of an inducible promoter, and then placing the heterologous cDNA to be expressed under the control of the T7 promoter, the specific induction of the heterologous protein can be brought about (Tabor and Richardson, 1985). Induction will switch on expression of T7 RNA polymerase, which will recognise the T7 promoter, and in turn switch on the expression of the heterologous protein. This system is explained further in Section 3.3 and in Figure 3-1.

Induction can be achieved using heatshock where the recombinant pT7-7 plasmid is transformed into *E. coli* SRP84/pGP1-2 cells. The pGP1-2 plasmid carries the T7 RNA polymerase gene under a heat-controlled promoter system. Alternatively, by transforming the pT7-7 plasmid with the cDNA insert into *E. coli* BL21 (DE3) cells, induction can be attained by introducing IPTG, a non-hydrolysable lactose analog, to the growth medium. In this case, the T7 polymerase gene is integrated into the *E. coli* chromosome and under the control of an IPTG-induced lac^{kv5} promoter.

Purification of the over-expressed recombinant aldehyde dehydrogenase has been achieved by the following combination of methods (Jones *et al.*, 1995). The disrupted cells (typically using a French press) are treated with protamine sulphate, to remove nucleic acids. The protein solution is then passed through an ion-exchange column, where either the AlDH passes through a CM-Sephadex column immediately after the void column, or is bound to a DEAE-Iontosorb column and eluted with NaCl. The partially purified AlDH solution is then passed down an affinity column, where *p*-hydroxyacetophenone (HO  CO - CH3) is linked to the Sepharose 6B solid support through the *p*-hydroxy group.

This *p*-hydroxyacetophenone-Sepharose affinity column was first utilised by Ghenbot and Weiner (1992), on the basis that the acetophenone derivative was a good competitive inhibitor. However it was unexpected that *p*-hydroxyacetophenone would make an effective and specific ligand when NAD^+ was absent, given the ordered mechanism of binding of the co-factor and substrate (Ghenbot and Weiner, 1992). Although data was presented to suggest that the binding of enzymes with an altered active site to this affinity-based column would occur, most notably the C302A enzyme, this can not be held to be accurate for all mutated enzymes in the future, given that the mechanism of binding of the enzyme to the column has not been established.

The T7-7 system was tested as part of this work, with the expression and purification of human class 1 aldehyde dehydrogenase to be used as a comparison with the other expression systems tested. The results are detailed in Chapter 3.

1.4.2 The *E. coli* GST-fusion Expression System.

Glutathione-S-transferase (GST) is a 26 kDa enzyme which catalyses the reaction of alkylating agents with the -SH group of glutathione (Habig *et al.*, 1974). The enzyme has a specific affinity for glutathione and this has been utilised within a heterologous expression system (Pharmacia Biotech). A fusion protein can be expressed with the GST 'tag' on the C-terminus of the heterologous protein by placing the cDNA of the protein to be expressed beside the cDNA for GST in an expression plasmid (pGEX). The fusion protein can be easily purified by passing the bacterial lysate through a glutathione-linked Sepharose column (Smith and Johnson, 1988), the fusion protein will bind while contaminating proteins will wash through.

The pGEX plasmid (Figure 3-3) has a multiple cloning site so that the cDNA of the heterologous protein can be cloned in frame with the GST coding region (Smith and Johnson, 1988). Between the GST cDNA and the

multiple cloning site is a target sequence for a protease, for example thrombin or Factor Xa. Once expressed, this allows the fusion protein to be cleaved and the heterologous protein to be released from the GST protein.

The pGEX plasmid also contains: a tac promoter so that expression can be induced with IPTG; β-lactamase, the ampicillin resistance gene for the screening of recombinant *E. coli* colonies; and the *lac*I^q repressor protein to reduce leaky promoter action. Once cloned in frame, and transformed into a suitable *E. coli* host, expression of the fusion protein is induced using IPTG.

The cells are harvested, lysed, and the cell-free supernatant is loaded onto the glutathione-Sepharose column. The fusion protein will bind to the column and most other endogenous *E. coli* proteins will wash through. The fusion protein can either be cleaved while still bound to the column, releasing the heterologous protein, or it can be competitively eluted using free glutathione and then cleaved.

This system has been used to produce and purify many proteins, including a species-specific epitope from *Chlamydia trachomatis* at 5-10 mg per litre culture (Toye *et al.*, 1990), a *Taenia ovis* antigen (Johnson *et al.*, 1989), the SV40 T antigen-binding domain of the retinoblastoma protein (Kaelin *et al.*, 1991) and a cDNA encoding a retinoblastoma-binding protein (Kaelin *et al.*, 1992).

This expression system has the advantage in that aldehyde dehydrogenase can be purified without using the *p*-hydroxyacetophenone-Sepharose affinity column. A protein in which an amino acid is specifically altered, supposedly at the active site, may not bind to the affinity column which may rely on the integrity of the active site. By using the GST fusion protein system the expression and purification of the mutant enzyme should not be different to the wild-type.

The cDNA for human class 1 aldehyde dehydrogenase had previously been cloned into pGEX 4T-3 (E. Loughnane, work done in this lab), but no expression work had been carried out. As part of this work, pGEX-hAlDH was used to express human class 1 AlDH in *E. coli* XL-1 Blue cells, and trials were performed to find the most efficient way of purifying AlDH, using the glutathione-Sepharose. Results are detailed in Chapter 3.

1.4.3 The *Pichia pastoris* Yeast Expression System.

The methylotropic yeast *Pichia pastoris* offered a practical alternative to the *E. coli*-based expression systems. With the added advantage of using a eukaryotic organism to express a recombinant eukaryotic protein, the *P. pastoris* expression system (as marketed by Invitrogen) uses a strong inducible promoter with the option of secretion of the protein into the growth medium, giving high yields of active enzyme (Cregg *et al.*, 1993).

P. pastoris can utilise methanol as its sole carbon source by producing the enzyme alcohol oxidase which oxidises ethanol to formaldehyde. There are two alcohol oxidase genes, AOX1, which is responsible for most of the alcohol oxidase activity in the cell, and AOX2, which has far lower activity (Cregg *et al.*, 1993).

The alcohol oxidases have poor affinity for the molecular oxygen used in this oxidation and this poor affinity is compensated by the transcriptionally-regulated overproduction of AOX1 (Invitrogen manual, 1995). The AOX1 promoter has been shown to induce expression of AOX1 to over 30% of the total soluble protein in the cell (Cregg *et al.*, 1993). The strong promoter that regulates the alcohol oxidase production can be harnessed to a foreign gene, in this case AlDH, and expression of the foreign AlDH gene can be induced by growing the recombinant cell culture on methanol as the sole carbon source.

A DNA vector is required to enable the integration of foreign DNA into the *Pichia* genome. pPIC3K is a vector which allows the intracellular expression of the heterologous protein and pPIC9K can be used to bring about secretion of heterologous protein into the growth medium, as shown in Figure 4-1.

Once the foreign gene is cloned into pPIC9K, the vector is linearized and then introduced into the *Pichia* yeast cells through electroporation. The linearized vector can integrate into the *Pichia* genome through homologous recombination at either the AOX1 or HIS4 genes.

Once the integration is confirmed, cultures of the cells can be grown up on glycerol in order to increase the cell-culture mass. Expression of the protein can then be induced by growing the yeast in a fully aerated, buffered, methanol medium.

Many different proteins have been expressed successfully using this *P. pastoris* expression system. Some recent examples, with the yields expressed in mg/L of culture, are: HIV-1 envelope protein, 1250 mg/L culture (Scorer *et al.*, 1993); the penicillin-binding-protein, dacA, 100 mg/L culture (Despreaux & Manning, 1993); tetanus toxin fragment C, 12,000 mg/L culture (Clare *et al.*, 1991). Yields are generally much higher than can be achieved with *E. coli*-based systems.

Overall, many benefits have been found using the *Pichia* expression system. The yeast cells can be grown to a very high cell density on a simple defined medium and the protein is usually folded correctly in the eukaryotic system (Scorer *et al.*, 1993). Electroporating the yeast cells for transformation is a simple and fast process (White *et al.*, 1994), and detection of multiple insertions can be carried out with an antibiotic screening technique. The expressed enzyme can be secreted into the growth medium in large amounts, effectively a first step in the purification process as very few native proteins are secreted (Cregg *et al.*, 1993).

With the potential of this expression system it was hoped to have a high yield of expressed aldehyde dehydrogenase and for this to be purified using the DEAE-Sephadex and *p*-hydroxyacetophenone-Sepharose affinity columns. This strategy would not avoid the problem of purifying mutagenic proteins with an altered active site, as mentioned in Section 1.4.1.

In this work the sheep class 1 AlDH cDNA was cloned into both the secreted and intracellular plasmids as described in Chapter 4. The cDNA inserts in the pPIC9K-shAlDH and pPIC3K-shAlDH clones were fully sequenced, electroporated into *Pichia pastoris*, and the integration confirmed by PCR and G418 screening. These *Pichia* cultures were grown up and induced, and the preliminary expression results are also presented in Chapter 4.

1.5 Requirements for an Effective Expression System.

In deciding which expression system best suits a particular protein many factors need to be considered including:

- level of expression,
- ability to purify to homogeneity,
- minimisation of degradation and modifications to the protein such as glycosylation,
- correct folding,
- ease of scale-up,
- ability to purify mutated enzymes,
- time scale for entire process,
- cost effectiveness.

The three expression systems discussed were tested for expression and purification against these criteria, in order to ascertain the best system for the expression and purification of class 1 aldehyde dehydrogenase.

1.6 AIMS FOR THIS THESIS.

The main aim was to consolidate and extend the work already done in this area of research, by testing three types of expression and purification strategies, detailed above; the T7 expression system with purification using the ion-exchange and affinity columns, the GST fusion protein expression and purification strategy, and the *Pichia pastoris* expression system. As the human class 1 cDNA was already cloned into the pT7-7 and pGEX4T-3 plasmids, my work was to express and purify the enzyme. The *Pichia pastoris* kit was used for the first time in this department, so the yeast methods needed to be established, as well as appropriate cloning strategies.

It was hoped that once the best method of expression and purification was established, site-directed mutagenesis could be undertaken to investigate the role of Lys-272 as the putative general base assisting in the acid-base reaction in the catalytic process. Preliminary work in the production of K272A, K272H and K272L mutants is presented in Chapter 5.

2. Materials And Methods.

2.1 Molecular Biology And Cloning Techniques.

2.1.1 Buffers.

TE (pH 8.0) 10 mMTris-HCl, 1 mM EDTA

TAE (pH 8.0) 40 mM Tris-HCl, 20 mM acetic acid, 2 mM EDTA.

2.1.2 Agarose Gels.

Agarose gels (0.8 % in TAE) were used to separate, size and quantitate DNA fragments. Ethidium bromide was added to the gel mix to a final concentration of 1 μg/mL. DNA samples were diluted five parts to one part bromophenol blue loading dye (Sambrook *et al.*, 1989, Section 6.12) and up to 12 μL of each sample was loaded. At least one well was loaded with a 1 kb DNA ladder (BRL); this ladder contains DNA fragments of ~500 bp, 1.0 kb, 1.6 kb, and then approximately every 1kb from 2kb to 12 kb. The gel was electrophoresed until the bromophenol blue dye had migrated one third of the length of the gel. The DNA was visualised using UV light (310 nm) and a photograph taken either with polaroid film or with the gel-documentation system (AlphaImager™, Alpha Innotech Corporation, Biolab Scientific Ltd.).

2.1.3 Polymerase Chain Reaction (PCR).

The polymerase chain reaction is a method used for amplifying a specific region of DNA, by using Taq DNA polymerase to extend two annealed primers. A fresh PCR mixture was prepared which contained 0.05 units Taq polymerase per reaction (Gibco BRL), 3 mM $MgCl_2$, 0.3 mM of each dNTP, 1x Taq polymerase buffer, and was then made up to the appropriate volume (14 μL multiplied by the number of reactions) with sterile MilliQ water. This mixture was aliquoted out (14 μL per reaction) into 0.5 mL Eppendorf tubes.

The required forward and backward primers were made up to a concentration of 5 pmol/μL and 2 μL of each primer was added to the 14 μL reaction mixture. Template DNA (pBluescript-shAlDH plasmid, also known as D9) was diluted to a concentration of 1 ng/μL and 2 μL of this was added to the mixture to give the 20 μL final volume.

Negative controls included all of the above components, excluding the template DNA. A positive control used primers that annealed within the sheep AlDH cDNA and had been previously shown to yield a PCR product.

Each reaction mix was overlaid with mineral oil and placed in the thermocycler (Omn-E, SciTech) preheated to 95 °C. The temperature cycle was composed of 94 °C for 30 seconds, 60 °C for an annealing time of 45 seconds and an extension time of 2 minutes at 72 °C. This was carried out for 30 cycles. The final cycle had a 72 °C extension of 5 minutes to complete the extension of all strands being polymerised. After the cycling was complete the DNA was either visualised on an agarose gel, directly purified, or stored at 4 °C.

2.1.4 Purification of DNA.

Purification of DNA, either from an agarose gel or directly from a restriction digest or PCR, was performed using the Bio-Rad Prep-A-Gene DNA kit. The protocol provided was followed exactly, and the purified DNA was usually resuspended in 50 μL of TE buffer.

2.1.5 Precipitation of DNA.

DNA was concentrated and desalted using ethanol precipitation (Sambrook *et al.*, 1989, Section E.10-E.11). The DNA solution was mixed with 0.1 volumes of 3 M sodium acetate (pH 5.5) and 2 volumes of 95 % ethanol. The tube was placed at -70 °C for at least 15 minutes and then centrifuged at maximum speed in a microfuge for 10 minutes at 4 °C. The supernatant was carefully removed, 100 μL of 70 % ethanol added and the tube centrifuged

as before for 2 minutes at 4 °C. The DNA pellet then was dried down by carefully removing most of the ethanol, then leaving the open Eppendorf on the benchtop until the residue liquid had evaporated. The pellet was resuspended in 50-100 μL of TE buffer.

2.1.6 Restriction Endonuclease Digests.

Digestion of DNA with restriction endonucleases was carried out according to the manufacturers' instructions, using the following guidelines:

- The volume of the digestion reaction was determined such that the total DNA concentration never exceeded 0.1 μg/μL.
- 5-10 units of enzyme was used for every μg of DNA.
- The time of digestion was usually 2-3 hours at 37 °C, unless otherwise specified by the manufacturer.

To check for completeness of the DNA cleavage, digested DNA was run on an agarose gel with uncut DNA in an adjacent lane for a comparison.

2.1.7 T4 DNA Polymerase.

After PCR and purification, the PCR products were treated with T4 DNA polymerase. During the extension phase of the PCR cycle Taq polymerase often 'overruns' the end of the template strand, adding one or two extra A nucleotides on the 3' end of the newly synthesised strand. The 3' to 5' exonuclease activity of T4 polymerase is used to remove these overhangs, creating a DNA molecule with blunt, ends.

The reaction was carried out in a 40 μL volume containing 1x T4 polymerase buffer (Boehringer Mannheim), 1.5 units of enzyme (1 unit/μL, Boehringer Mannheim), 0.1 mM of each dNTP, half of the PCR product DNA, and sterile water to make up the volume. This was left at 12 °C for 30 minutes.

At the end of the reaction the enzyme and buffer were removed from the DNA by the direct purification method described in Section 1.1.3.

2.1.8 Dephosphorylation of Vector.

Thermosensitive alkaline phosphatase (Gibco BRL) was used to remove the 5' phosphates from the vector fragments prior to ligation, and so reduce the likelihood of the vector fragments reannealing to each other. The phosphatase reaction was carried out according to the manufacturer's instructions.

1 pmole of the ends of DNA molecules is equivalent to 5.7 μg of a 9.3 kb fragment. Approximately 1 μg of DNA was added to a mixture containing 2 μL 10x buffer, 0.1 μL of phosphatase enzyme (1 unit/μL) and made up to 20 μL with sterile water. This was incubated at 65 °C for 15 minutes, then 2 μL of stop buffer (Gibco BRL) added. A further incubation at 15 minutes at 65°C was carried out and 1-2 μL of this prepared vector DNA was used in subsequent ligation reactions.

2.1.9 Ligation.

5 μL of insert DNA (prepared as described in Section 2.1.7) was added to 2 μL of prepared vector DNA so that the molar ratio of insert to vector was approximately three to one. Two microlitres of 5x ligase buffer and 0.5 units of T4 DNA ligase (Gibco BRL) were also added. This mixture was made up to 10 μL with sterile water and incubated overnight at 16 °C. Control reactions were also prepared: vector, with no insert or ligase; vector and ligase, with no insert.

If the DNA to be ligated had blunt ends rather than overhangs, 1 μl more ligase was added and the mixture was incubated for a further 3-4 hours at room temperature.

2.1.10 Plasmid Transformation of *E. coli.*

Transformation was carried out using "ultracompetent" calcium competent XL-1 Blue cells, prepared as described by Inoue *et al.* (1990).

Two methods of transformation were used.

In the original method, a 100 μL aliquot of competent cells (stored at -70 °C) was thawed on ice and 5 μL of the ligation mix, or 15 ng of purified plasmid stock, was added. The cells were incubated on ice for 15-30 minutes and then heat-shocked at 42 °C for 45 seconds. The cells were chilled on ice, and 0.9 mL of LB broth (without antibiotic) was added. The mixture was incubated at 37 °C in a shaking incubator for 45 minutes to 1 hour. LB-Ampicillin plates (100 μg of ampicillin per mL LB broth/agar) were used, with 100-200 μL of the cell cultures being plated onto each plate. Transformation controls included plating cells without DNA onto LB-Ampicillin plates to check for correct selection, and onto LB plates to check the survival of the cells through the procedure.

During the course of this work a quicker method of transformation was described (Pope and Kent, 1996) and this method was adopted. A 100 μL aliquot of competent cells was thawed on ice for each transformation and 5 μL of a ligation mix, or 15 ng of purified plasmid stock, was added. The cells were kept on ice for 15-30 minutes, then plated onto LB-Ampicillin agar plates that had been prewarmed at 37 °C.

2.1.11 Small-Scale Preparation of DNA.

DNA minipreps were conducted by the "rapid boil" technique as described in Sambrook *et al.* (1989, Section 1.29-1.30) and resuspended in 50 - 100 μL TE.

2.1.12 Large-Scale Preparation of DNA.

'Maxiprep' DNA was prepared using the 'Wizard Maxipreps DNA Purification System' (Promega). 1-2 mg of plasmid DNA was usually purified from a 400 mL cell culture and the DNA was stored in aliquots at -70 °C.

2.1.13 DNA Sequencing.

DNA sequencing was required to check that clones produced using PCR did not contain any mutations incorporated by the action of Taq polymerase. Samples were sent to the automated sequencing laboratory at Otago University, Dunedin, New Zealand. Seven primers (each at a concentration of 0.8 pmol/μL) were sent with the DNA (200 ng/μL) allowing the complete 2.1 kb aldehyde dehydrogenase cDNA to be sequenced.

The seven sequences were merged to form a single overlapping sequence using the GCG Fragment Assembly program. This was then compared to the published sequence for sheep cytosolic aldehyde dehydrogenase (Staynor and Tweedie, 1995).

The pGEX-hAlDH plasmids that had undergone site-directed mutagenesis (to alter the Lys-272 residue) were manually sequenced using the 'Sequenase DNA Sequencing Kit' (United States Biochemicals). Single-stranded template DNA was prepared from 60 μL of a 100 μL 'rapid boil' plasmid miniprep solution, using alkaline denaturation, and the primer was annealed using the standard Sequenase protocol. [^{35}S]-dATP (ICN Biomedicals) was used to label the extended DNA strand and manganese-containing buffer was included to aid in sequencing close to the primer, as directed in the Sequenase protocols.

The 6 % acrylamide gel used for DNA sequencing was prepared as specified in the Sequenase manual, and polymerisation occurred for at least two hours. 1x TBE buffer was used as the electrophoresis buffer and the gel was pre-

electrophoresed at 65 W for at least 30 minutes. Three microlitres of each sequencing reaction was loaded and the gel was electrophoresed for approximately 80 minutes.

The gel was then soaked in 10 % ethanol / 10 % acetic acid for 10 minutes to remove urea, and then transferred onto 3MM paper at 80°C under a vacuum. Radioactivity of the dried gel was estimated using a Geiger counter, and the gel laid down against X-ray film for the appropriate time, usually 1-3 days.

The sequence was read manually from the developed autoradiograph and compared to the known DNA sequence.

2.2 EXPRESSION OF AlDH IN *E. COLI*.

2.2.1 Solutions and Buffers.

PBS (pH 7.3)	140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4.
PBSED	PBS with the addition of 1 mM EDTA, 1 mM DTT
PBS-Tween	PBS with the addition of 2 % Tween-20.
TED (pH 8.2 with HCl)	50 mM Tris, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 10 % glycerol.
MI medium (pH 7.2)	15 g/L casamino acids, 5 g/L yeast extract, 0.04 M Na_2PO_4, 10 mL/L solution A, 1 mL/L solution B. Solution A: 1 M NH_4Cl, 50 mM K_2SO_4, 50 mM $MgSO_4$ and 2 mM $CaCl_2$. Solution B: 100 mM HCl, 10 mM $FeSO_4$, 2 mM $MnCl_2$, 2 mM $ZnSO_4$, 0.2 mM $CoSO_4$, 0.1 mM $CuSO_4$ and 0.1 mM $NiCl_2$.
Bradford dye mix	100 mg Coomassie Blue G-250 dissolved in 50 mL ethanol, 100 mL 88 % (w/v) orthophosphoric acid, made up to 1 L with water.
PMSF (100 mM)	17.4 mg dissolved in 1 mL isopropanol.

2.2.2 Culture Growth and IPTG Induction.

After transformation with the expression plasmid, single *E. coli* colonies were picked from the LB-Ampicillin plate. These colonies were cultured overnight in 2 mL volumes, and used to inoculate 0.5 - 3 L volumes. The culture was grown until the A_{600nm} was between 0.6 and 1.0. Induction involved the addition of IPTG from a 100 mM stock to the appropriate final concentration (usually 1.0 mM) and the culture grown for a further 3 - 5 hours. The cells were centrifuged in an SC3 rotor for 20 minutes at 4 °C and 4000 g and the pellets were stored at -20 °C until required.

2.2.3 Cell Lysis.

The cell paste was suspended in 6 volumes of either PBS or 50 mM Tris buffer (pH 8.2) and kept on ice. 1 mM EDTA / DTT was added to these buffers to stabilise the expressed recombinant AlDH.

The cells were lysed either with the French press or by ultra-sonication. The latter was carried out in 30 second bursts, with the cells kept on ice between bursts. After lysis, PMSF was added to a final concentration of 1 mM to inhibit serine proteases.

2.2.4 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE).

The following stock solutions were made up and stored at 4°C in foil-covered bottles:

- 32 % acrylamide / 0.2 % bisacrylamide Resolving Gel Mix
- 10 % acrylamide / 0.05 % bisacrylamide Stacking Gel Mix

The 12 % SDS-PAGE gels were prepared as described in Sambrook *et al.* (1989, Section 18.47-18.59). After polymerisation was complete they were wrapped in cling-film and stored at 4°C for up to a month.

Electrophoresis was carried out by placing the gel in the electrophoresis apparatus (Mighty Small II, Hoefer Scientific Instruments) and submerging in electrophoresis buffer (Sambrook *et al.*, 1989, Section 18.53). The samples were diluted in an equal volume of 2x SDS sample buffer (Sambrook *et al.*, 1989, Section B.25), boiled for three minutes and up to 20 µL of the sample was loaded into a well. The gels were electrophoresed at 12 mA until the sample entered the resolving gel, when the current was increased to 18mA.

A set of proteins (SDS-7, Sigma Co.) was run on each gel. The seven proteins had molecular masses of 66, 45, 36, 29, 24, 20 and 14 kDa.

The gel was stained in a solution of 0.2 % Coomassie Brilliant Blue R-250 in 30 % methanol / 10 % acetic acid for ten minutes and then destained in three 30 minute washes in 30 % methanol / 10 % acetic acid. The gel was stored in 'Welcome' solution (5 % methanol / 5 % acetic acid) until it was dried onto 3MM paper under vacuum at 80°C.

2.2.5 Determination of Protein Concentration.

The Bradford method of protein concentration determination was used (Bradford, 1976). A stock of 0.5 mg/mL BSA protein was diluted appropriately to give a standard curve covering the range of 0 to 25 μg of protein. The Bradford dye solution (0.95 mL) was added to 50 μL of the protein standard or appropriately diluted sample, and the mixture allowed to stand for 20 minutes. The absorbence was read at 595 nm, and the standards plotted on a graph.

2.2.6 Activity Assay For AlDH.

The dehydrogenase activity of AlDH was assayed by combining 2.6 mL sodium pyrophosphate-HCl (0.1 M, pH 9.3), 250 μL NAD^+ (20 mg/mL) and 50-200 μL enzyme sample in a cuvette.

The spectrophotometer (Ultraspec 3000, Pharmacia Biotech) was blanked at 340 nm, then 150 μL of acetaldehyde (200 mM) was added. The absorbance increase at 340 nm was followed for ten minutes and the rate of increase was calculated and converted to the number of μmoles of acetaldehyde oxidised per minute per mL of original sample.

2.2.7 Western Blotting.

Western Blotting was carried out as described in Sambrook *et al.*, (1989, Section 18.60-18.75). The primary antibody used was a 1:5000 dilution of rabbit IgG anti-AlDH (raised by K. Jones, in this laboratory) and the secondary antibody was a 1:10,000 dilution of goat anti-rabbit IgG with a peroxidase conjugate (Sigma, product number A-0545). After an hour

incubation with the appropriate antibody the filter was washed in PBS-Tween for three successive 30 minute washes. After the final wash the nitrocellulose filter was stained in a solution of 3 mg/mL 3, 3'-diaminobenzidine (DAB, Sigma) and 2 μL/mL H_2O_2 for 1 minute, then dried.

2.3 YEAST MANIPULATION AND GROWTH.

2.3.1 Growth Media for the Growth of *P. pastoris.*

The Yeast Nitrogen Base (YNB) required is without amino acids and with ammonium sulphate (Difco Laboratories, product number 0919-15-3)

.

YPD
1 % yeast extract
2 % peptone
2 % glucose

Minimal Dextrose (MD)
1.34 % YNB
1 % glucose
0.4 mg/L biotin

Minimal Glycerol (MGY)
1.34 % YNB
1 % glycerol
0.4 mg/L biotin

Minimal Methanol (MM)
1.34 % YNB
0.5 % methanol
0.4 mg/L biotin

Buffered Glycerol-complex (BMGY)
Buffered Methanol-complex (BMMY)
1 % yeast extract
2 % peptone
100 mM potassium phosphate, pH 6.0
1.34 % YNB
0.4 mg/L biotin
1 % glycerol or 0.5 % methanol

2.3.2 *Pichia pastoris* KM71.

This strain has the following genotype and phenotype:

- Genotype: *arg4 his4 aox1::Arg4*
- Phenotype: $Mut^{S}Arg^{+}$

2.3.3 Characteristics of *Pichia pastoris* KM71.

The yeast *Pichia pastoris* grows at 28 - 30 °C. All yeast cultures were stored at 4 °C and a glycerol stock (Invtrogen manual, p.15) was made of each and stored at -70 °C. The doubling time of log phase cultures is 2 hours on YPD medium and 18 hours on MM medium.

2.3.4 Direct PCR From Yeast Cells.

Genomic DNA was extracted from the yeast cell by taking approximately 10 µL of a plated colony, adding 2 µL zymolyase (3 mg/mL), and incubating at 30 °C for 15 minutes. This mixture was added as template to the complete PCR mix, including primers, buffer and Taq Polymerase, and the PCR protocol was followed as described in Section 1.1.2. The primers used were shAlDH-2 and shAlDH-5:

shAlDH-2 5' GGAAACACAGTGGTTGGAG 3'

shAlDH-5 5' TATAGGATCCTAGTTTATGAGTTCTTCGGAG 3'

2.3.5 G418 Screening.

YPD agar plates were prepared to which G418 antibiotic (Geneticin®, Life Technologies, product number 11811-023) had been added at concentrations ranging from 0 to 2.0 mg/mL.

The method for screening for multiple inserts was taken from 'Method 2', pPIC9K Instruction Manual (Version B, p. 16, Invitrogen). All positive recombinant colonies were grown up in YPD broth and 10 µL of each culture was inoculated 190 µL YPD in a single well of a microtitre plate. The cultures were grown to the same cell density by three successive 20 µL inoculations into 180 µL fresh YPD in new microtitre plates, and incubating at 30 °C for 48 hours. A 10 µL sample of each culture from the final microtitre plate was spotted onto each of the series of varying G418 concentration YPD-G418 agar plates. After incubation at 30 °C, the plates

were compared to see which of the original recombinant colonies showed the greatest resistance to the G418 antibiotic.

2.3.6 Expression of shAlDH.

The *Pichia pastoris* colonies selected for expression trials were grown up in 300mL BMGY in a 2 L flask at 30 °C for 2 days. The cells were harvested by centrifugation in a GS3 rotor at 4000g for 8 minutes and resuspended in 100 mL BMMY in a 1 L baffled flask, and incubation was continued at 30 °C at a fast shake for four days. A sample was drawn off at 24 hour intervals, and pure methanol was added to give 0.5 % (v/v) methanol.

Each sample was pelleted by centrifugation at maximum speed in a microfuge and the supernatant transferred to a fresh tube. Both supernatant and pellet were snap-frozen in liquid nitrogen and stored at -70°C. To check for expression, these samples were analysed by SDS-PAGE as described in Section 1.2.3.

3. EXPRESSION OF hAlDH IN *E. COLI.*

Escherichia coli was used as the host for two types of expression system tested in this work. The first to be discussed uses the bacteriophage T7 RNA polymerase/promoter system to over-express the heterologous protein, which is then purified by conventional ion-exchange and affinity columns. In the second type of expression system a fusion protein is produced where the heterologous protein is linked to a protein, in this case glutathione S-transferase, which is easily purified by affinity chromatography.

3.1 T7 RNA POLYMERASE / PROMOTER SYSTEM.

This expression system had been used in this laboratory previously and it was tested once in this work to act as a comparison with the two other expression systems tested, the GST fusion system and expression using the yeast, *Pichia pastoris*.

In 1985, Tabor and Richardson first over-expressed proteins using the RNA polymerase from the bacteriophage T7, which specifically recognises the 23 base pair $\phi 10$ promoter. If a heterologous cDNA was cloned under the control of this specific promoter, then the T7 polymerase could be used to exclusively express the heterologous protein. The timing of induction was controlled by placing the gene for T7 RNA polymerase under the control of an inducible promoter. Induction switched on expression of T7 RNA polymerase, which recognised the T7 promoter, and in turn switched on the expression of the heterologous protein (illustrated in Figure 3-1).

3.2 EXPRESSION AND PURIFICATION OF hAlDH USING THE T7 SYSTEM.

The cDNA for human class 1 aldehyde dehydrogenase was kindly supplied by Professor H. Weiner (Department of Biochemistry, Purdue University Indiana, USA) and was cloned into pT7-7 by means of PCR, and had not

been resequenced. The pT7-7 plasmid containing human class 1 AlDH was named pTCAD.

In previous work at this laboratory, pTCAD was transformed into *E. coli* SRP84/pGP1-2 cells where the resident pGP1-2 plasmid carries the T7 RNA polymerase gene under the control of the λ promoter, pL, along with the temperature sensitive λ repressor (Jones *et al.*, 1995). Heat shock of the transformed cells inactivates the λ repressor, allowing transcription of T7 polymerase, which further allows the transcription of the foreign cDNA (illustrated in Figure 3-1).

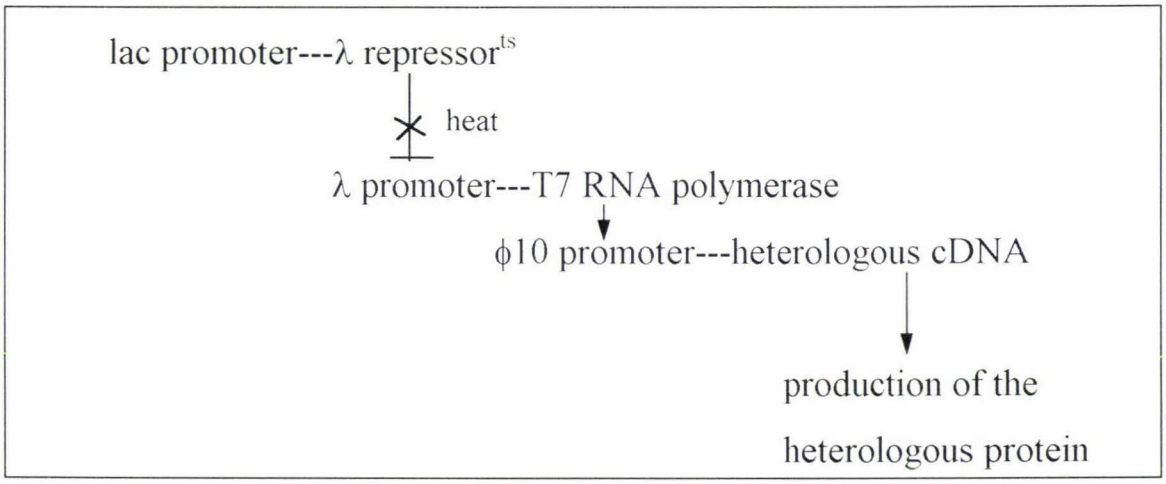

Figure 3-1: Schematic diagram of the induction process of the T7 polymerase expression system.

IPTG induction has also been achieved by transforming pTCAD into *E. coli* BL21 (DE3) cells (unpublished data, this work). In this case, the T7 polymerase gene is integrated into the *E. coli* chromosome under the control of a *lac*^{kv5} promoter, which is induced by addition of IPTG.

Purification of the over-expressed recombinant aldehyde dehydrogenase has previously been achieved by treating the disrupted cells with protamine sulphate, then purifying by using ion-exchange chromatography (CM-Sephadex), and a *p*-hydroxyacetophenone-Sepharose affinity column (Jones *et al.*, 1995). Expression and purification of the human class 1 aldehyde

dehydrogenase by this T7 method has previously yielded 3 mg/L after purification with a specific activity of 0.34 μmol/min/mg (Jones *et al.*, 1995).

3.2.1 Expression of hAlDH.

The pTCAD plasmid had been previously transformed into the *E. coli* SRP84/pGP1-2 cells and a colony grown up and stored as a glycerol stock at -70 °C (by E. M. Loughnane, in this laboratory). The following improved protocol for the expression and purification of AlDH from this system had been tested by Mrs. C. Flyger, also working in this laboratory.

In this work, expression and purification of class 1 hAlDH was carried out by this method in order to establish a measure for comparison with the other expression and purification systems tested.

A sample of the *E. coli SRP84*/pGP1-2/pTCAD glycerol stock was plated onto LB-KanAmp plates (where the kanomycin acts as a selection for cells carrying the resident plasmid pGP1-2) and incubated overnight at 30 °C. A single, isolated colony was selected and transferred to 2 mL MI medium (Section 2.2.1) containing 100 μg/mL ampicillin, 50 μg/mL kanomycin and 10 mM glucose. This 2 mL culture was used to inoculate 600 mL supplemented MI, in a 2 L flask. The optical density (A_{650}) of this 600 mL culture equalled 6.2 after an overnight incubation at 30 °C, on a fast shaker.

Heat induction was carried out by adding 400 mL of supplemented MI, preheated to 70 °C, to the 600 mL *E. coli* culture. Once combined the temperature of the 1 L culture was 42.5 °C. This culture was then shaken at room temperature for five minutes in order to mix thoroughly, then cooled in an ice-bath for a further minute.

After induction the flask was returned to the 30 °C shaking incubator for 3.5 hours, then the cells were harvested by centrifugation with an SC3 rotor for

20 minutes at 4000 revs/min (4 °C). The cell pellet was stored at -20 °C until required.

3.2.2 Purification of HALDH.

All buffers contained 1 mM EDTA and 3 mM DTT. The cell pellet produced as described in Section 3.2.1 was thawed on ice and 15 mL of 20 mM phosphate buffer was added. The cells were lysed in the French press and the insoluble material spun down in an SS34 rotor at 12,000 revs/min for 15 minutes at 4 °C. The supernatant was decanted and the pellet discarded. Protamine sulphate (12.5 mg) was added to the supernatant, allowed to dissolve and the solution centrifuged as before, removing the precipitated nucleic acid material. Any remaining protamine sulphate was removed by overnight dialysis against 20 mM phosphate buffer (pH 6.1).

The dialysate was loaded onto the DEAE-Iontosorb column (ICS Identifikanci Systemy A. S., Prague, Czech Republic) which had been equilibrated at pH 5.8 with 20 mM phosphate buffer. The column (bed volume 7.5 mL) was washed with 20 mM phosphate buffer until the A_{280} of the eluate was less than 0.01. At this stage a salt gradient (100 to 350 mM sodium chloride) was run through the column, and the eluted fractions that displayed aldehyde dehydrogenase activity were pooled.

The pooled eluate was dialysed overnight in 30 mM phosphate buffer (pH 7.4), then loaded onto a *p*-hydroxyacetophenone-Sepharose affinity column (bed volume 9.5 mL) which had been equilibrated at pH 7.4 with 30 mM phosphate buffer containing 50 mM NaCl. The column was washed with the same buffer until the A_{280} was less than 0.01, and a gradient of *p*-hydroxyacetophenone (0-10 mM) was then applied. Eluted fractions showing aldehyde dehydrogenase activity were pooled and dialysed against the 30 mM phosphate buffer to remove the *p*-hydroxyacetophenone.

3.2.3 Results and Comparisons.

The yield of hAlDH from this preparation was 3 mg/L with a specific activity of 1.2 μmol/min/mg (1.2 units/mg).

This specific activity of 1.2 units/mg compared very favourably with the other human-liver aldehyde dehydrogenase purifications. The first purified human class 1 aldehyde dehydrogenase enzyme from human autopsy liver had a specific activity of 0.58 units/mg (Greenfield and Pietruszko, 1977) and from later purifications the specific activity was found to range between 0.22 and 0.54 units/mg (MacGibbon *et al.*, 1979; Hempel *et al.*, 1982 a and b; Yoshida *et al.*, 1993) The purified human enzyme from the T7 expression system as described above has given specific activities of 0.25 (Zheng *et al.*, 1993) and 0.34 units/mg (Jones *et al.*, 1995).

3.3 GST Gene Fusion System.

Expression of a foreign gene cDNA in frame with the DNA encoding a protein 'tag' produces a fusion protein, where the fused protein can enable straight-forward purification by affinity chromatography. Examples of fusion peptides that have been used include a His_6 peptide that binds to a nickel-chelating resin and glutathione-S-transferase (GST) which has a strong affinity for glutathione.

Glutathione-S-transferase has 220 amino acid residues (26 kDa) and catalyses the reaction of potential alkylating agents with the -SH group of glutathione in the first step of mercapturic acid production (Habig *et al.*, 1974).

The GST fusion protein expression and purification system (Pharmacia Biotech) is used for expression in *E. coli*. The expressed fusion protein can be purified by passing the induced-culture lysate through an GST-affinity matrix, glutathione-Sepharose (Figure 3-2). The matrix produced by Pharmacia Biotech has an expected capacity of 5.0 mg glutathione S-

transferase per mL of drained gel, which is roughly equivalent to 15 mg of the GST-hAlDH fusion protein per mL of matrix.

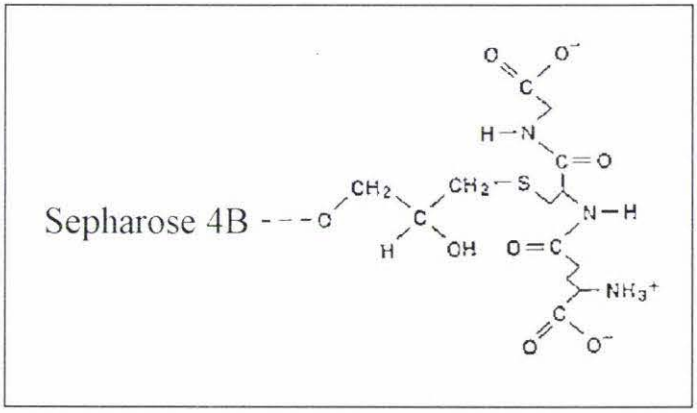

Figure 3-2: The structure of the glutathione-Sepharose 4B matrix (Pharmacia Biotech). From the GST Gene Fusion System manual (revision 2, 1996, p.4)

There are two approaches for the cleavage of the fusion protein and the removal of the GST protein from the heterologous protein. The first involves incubating the fusion protein/matrix with thrombin, resulting in the GST peptide being bound to the matrix while the heterologous protein is released and eluted with the buffer. The second approach involves elution of the GST fusion protein from the matrix by competition with free glutathione, cleaving with the protease thrombin in solution, and then removing the GST fragment from the solution by passing back down a glutathione-Sepharose column.

3.3.1 The Expression Vectors, pGEX.

The pGEX plasmids supplied as part of the GST Gene Fusion System kit (Pharmacia Biotech) allow the inserted cDNA to be cloned in frame with the GST cDNA. The plasmid used in this work, pGEX-4T-3, has the following features:

- protein expression under the control of the tac transcription promoter, and induced by IPTG, a lactose analog,
- a thrombin cleavage site to allow cleavage of GST protein from the heterologous protein after affinity purification,

- the β-lactamase gene to confer ampicillin resistance on recombinant *E. coli* colonies,
- the pBR322 origin of replication,
- the *lacI*^{q} gene that produces the lac repressor protein to reduce expression before induction through a leaky promoter.

The cDNA for human class 1 aldehyde dehydrogenase has previously been cloned into pGEX-4T-3 and the plasmid map of the resulting construct is shown in Figure 3-3. In this work the resulting construct was transformed into *E. coli* and expression was induced. Various experiments were then performed to establish a protocol for the expression and purification of human class 1 aldehyde dehydrogenase using the GST fusion protein system.

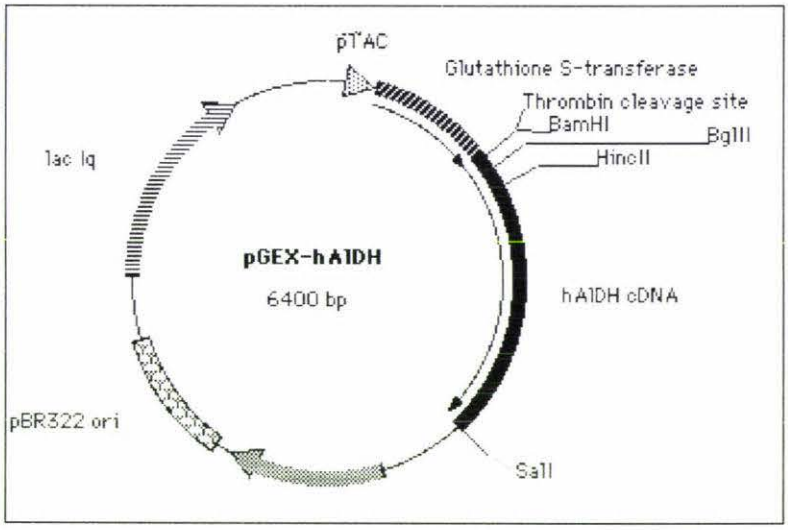

Figure 3-3: Plasmid map of the pGEX-4T-3-hAlDH construct.

The expressed product from this construct should be approximately 78 kDa, which after thrombin cleavage yields a 26 kDa GST protein and a 52 kDa hAlDH enzyme. The N-terminus sequence predicted from the cDNA sequence for hAlDH is Met-Ser-Ser-Ser-Gly (Hsu *et al.*, 1989), with the methionine residue normally removed in vivo and the N-terminal serine normally acetylated (Weiner *et al.*, 1995). The hAlDH produced from the

cleaved fusion protein is predicted to have a glycine residue in place of the acetylated serine, giving an N-terminal sequence of Gly-Ser-Ser-Gly.

3.3.2 Transformation of *E. coli* with pGEX-4T-3.

In this work, the pGEX-4T-3-hAlDH construct was retransformed into *E. coli* XL-1 Blue cells and plasmid minipreps were prepared from four of the transformants. A diagnostic digestion was performed using *Bst*EII to confirm the identity of the construct. *Bst*EII cuts once in the hAlDH cDNA and once within the pGEX plasmid and two fragments will be produced of sizes 3.6 and 2.7 kb if the hAlDH is correctly inserted into the pGEX plasmid.

The identity of the construct was confirmed, and a large-scale plasmid preparation was undertaken (Section 2.1.12). The pGEX-hAlDH construct was stored at -70°C in aliquots with a concentration of 900 ng/μL. The DNA from this large-scale preparation was used to transform *E. coli* XL-1 Blue cells, and the transformants used for the following expression and purification experiments.

3.4 Initial Expression Experiments.

3.4.1 Effect of A_{600} At Induction.

For the first expression experiment with the pGEX-hAlDH construct, the effect of the cell density on the production of fusion protein was investigated. Cultures of transformant *E. coli* XL-1 Blue containing the pGEX-hAlDH construct were grown in 2 mL LB-Amp broth to varying cell densities, then induced with 1 mM IPTG. The cultures were incubated at 30 °C and 500 μL samples taken for analysis 1.75 and 3.5 hours after induction. A culture of XL-1 Blue transformed with the parental plasmid, pGEX-4T-3, was also grown and induced for comparison.

The samples were spun down and resuspended in 150μL Phosphate Buffered Saline (PBS, Section 2.2.1) and sonicated for 15 seconds. The

lysate was centrifuged to separate cell debris and insoluble protein from soluble protein and samples of both pellet and supernatant were analysed by SDS-PAGE (one of the two gels is shown in Figure 3-4).

The results from this initial experiment show that the GST protein was expressed - a strong GST band (26 kDa) was produced from the control, pGEX, and a strong band (78 kDa) was produced from the expression of the pGEX-hAlDH construct.

The samples on this gel are not loaded equally, but after taking this into account, it appears as though around 40 % of the fusion protein produced is soluble. However, sonication may not have been complete as the protein distribution for both pellet and supernatant lanes is similar. As incubation time after induction increases, so does the proportion of insoluble fusion protein and induction of the culture at an A_{600} of 0.6 produced more fusion protein than the cultures with lower cell density at induction.

This experiment showed that high levels of expression could be gained, and we decided to trial different IPTG concentrations in order to maximise the solubility of the fusion protein. We also decided to sonicate the cells longer and ensure equal loadings on the gel.

3.4.2 Effect of IPTG Concentration on Induction.

The pGEX-hAlDH colony used for the last experiment was used to inoculate four more 2 mL solutions of fresh LB-Amp medium. When the A_{600} reached 0.55, IPTG was added to each culture to a final concentration of 0, 0.04, 0.1 and 1.0 mM. The cultures were grown on a shaking rack at 30 °C for a further 3 hours. The lower incubation temperature is known to reduce the amount of insoluble heterologous protein produced, by reducing the amount of protein that is packaged into inclusion bodies.

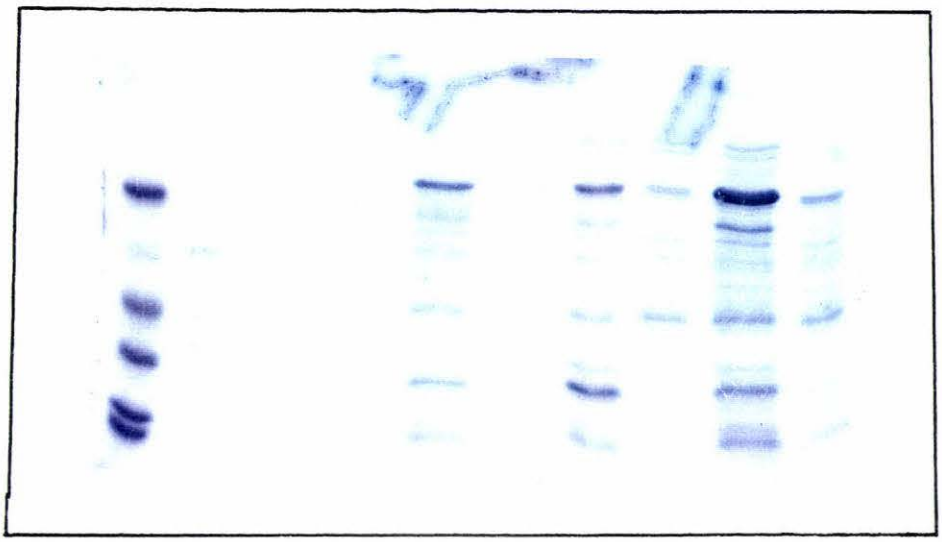

Figure 3-4: SDS-PAGE showing the expression of the the GST-hAlDH protein after IPTG induction at different cell densities. The cell culture has been lysed by sonication and centrifuged. The expected product is 78 kDa.

Lane 1: molecular markers (SDS7, Sigma); Lane 2: pure hAlDH; Lanes 3--6: culture induced at A_{600} equal to 0.135, supernatant and pellet at time of 1.5 hours, and 3.5 hours post-induction respectively; Lanes 7-10; culture induced at A_{600} equal to 0.600, pellet and supernatant at time of 1.5 hours, and 3.5 hours post-induction respectively.

The cells were harvested by centrifugation at maximum speed in a microfuge and resuspended into 300 μL PBS. The cells were sonicated for three 10 second bursts until the cell suspension cleared. The lysate was centrifuged for ten minutes at 4 °C to separate soluble proteins from cell debris and insoluble proteins, and samples of the supernatant and pellet for each of the four IPTG concentrations were analysed by SDS-PAGE (Figure 3-5).

The gel showed that the sonication was more effective this time; the protein distribution was very different for the soluble and insoluble fractions. The band corresponding to the fusion protein of GST-hAlDH was prominent in the 1mM IPTG sample and not noticeable in the samples with the lower IPTG concentrations. It was estimated that 65 % of the fusion protein produced was soluble.

With soluble GST-hAlDH being produced, it was time to tackle the purification of hAlDH using the glutathione-Sepharose matrix.

3.5 Initial Purification Experiments - Batch Mode.

As mentioned in Section 3.3, there are two methods for the purification and cleavage of the fusion protein; one involves thrombin cleavage while the protein is bound to the matrix, the other uses competitive elution with glutathione, then thrombin cleavage in solution. Both of these were tested in parallel in small-scale purifications.

3.5.1 Preparation Of The Cell Lysate.

A one litre culture of XL-1 Blue cells transformed with pGEX-hAlDH was grown up in LB-Amp until the A_{600} was 0.65, then induced with 1mM IPTG. After an incubation of 3.5 hours shaking at 28 °C, the culture was spun down to sediment the cells, and the cell pellet was stored at -70 °C until required.

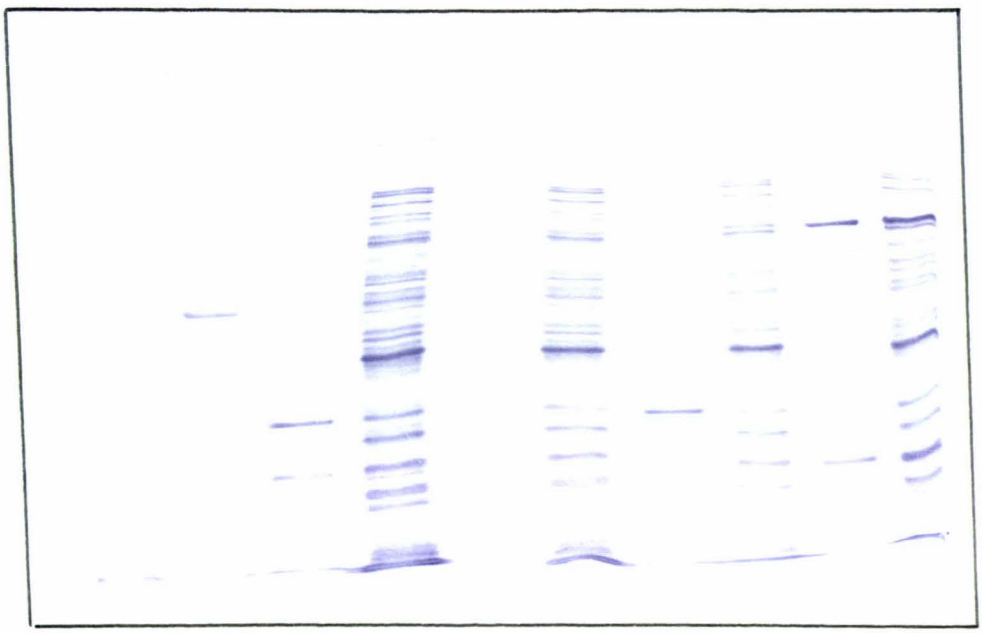

Figure 3-5: SDS-PAGE showing the expression of the GST-hAlDH protein after induction by different concentrations of IPTG. The cell culture has been lysed by sonication and centrifuged. The expected band is 78 kDa.

Lane 1: pure hAlDH; Lanes 2 and 3: pellet and supernatant with no IPTG added; Lanes 4 and 5: 0.04 mM IPTG added, pellet and supernatant; Lanes 6 and 7: 0.1 mM IPTG added, pellet and supernatant; Lanes 8 and 9: 1.0 mM IPTG added, pellet and supernatant.

The pellet (1.6 g) was resuspended into 8 mL PBS with 1 mM of each of EDTA, DTT and PMSF added. The cells were lysed by passing through a French press and centrifuged using an SS34 rotor at 15,000 rev/min for 15 min at 4 °C. This enabled the cell debris and insoluble material to be removed from the soluble protein fraction.

Approximately 7 mL of soluble lysate was collected, and this was split into 0.5 mL aliquots and stored at -70 °C until required. Each 0.5 mL aliquot was therefore equivalent to around 0.1 g of cell mass or 70 mL of the original 1 L culture.

Analysis by SDS-PAGE showed that lysis of the cells using the French press was not complete, the protein distribution visible on the SDS-PAGE was the same for both the soluble and insoluble material, showing that some of the soluble material had been spun down in intact cells. The French press was chosen as the method for cell lysis for this experiment as the volume of the resuspended pellet (8 mL) was larger than used previously. However this volume can still be lysed by sonication, and sonication was chosen as the method of choice for the following experiments.

3.5.2 Binding of the Fusion Protein to the Matrix.

One 0.5 mL aliquot of the lysate, produced as described in Section 3.5.1, was thawed on ice and added to the glutathione-Sepharose 4B matrix (1 mL of 50 % slurry, bed volume of 0.5 mL). The matrix had been washed and prepared as specified in the GST Gene Fusion System manual (Pharmacia Biotech).

The cell lysate/matrix mixture was incubated in a 1.5 mL Eppendorf tube at room temperature on a 'rotating wheel' to provide gentle agitation for 30 min and then centrifuged to sediment the matrix and bound fusion protein. The pellet was resuspended, washed with 2 mL of PBS buffer and re-centrifuged, for a total of six cycles in order to remove unbound, contaminating protein. Finally the pellet was resuspended in 2 mL PBS.

SDS-PAGE analysis showed that most of the fusion protein had bound to the matrix and nearly all other contaminating proteins were washed away, as shown by a single band of 78 kDa present after the matrix was washed (gel not shown).

At this stage the sample was split in half; one half was used to evaluate thrombin cleavage while the fusion protein is bound to the matrix, the other used free glutathione to elute the fusion protein before cleavage by thrombin.

Fractions were collected from various stages of the two purification processes and run on SDS-PAGE for analysis.

3.5.3 Thrombin Cleavage While Matrix-Bound.

The thrombin (Pharmacia Biotech) had a specific activity of greater than 7500 units/mg of protein, and one cleavage unit was expected to digest 90 % of 100 μg of a test fusion protein in 16 hours at 22 °C (GST fusion protein manual, Pharmacia Biotech, p. 29). The incubation time for thrombin cleavage should be kept short to reduce the non-specific digestion of the desired protein; the amount of thrombin added to cleave the amount of fusion protein present needs to be balanced with the requirement for a short incubation time.

In this case 25 units of thrombin was added to one half of the fusion protein/matrix solution (250 μL matrix suspended in 1.5 mL PBS). The solution was placed on the 'rotating wheel' at room temperature and 50 μL samples were removed for analysis after 4 hours and 16 hours.

Thrombin digestion was complete within the four hour timespan as seen by the disappearance of the 78 kDa band after the thrombin digestion and the coincident appearance of two bands, one with a mass of 26 kDa and the other 52 kDa (gel not shown). A time-course of digestion was required in

order to establish the best length of incubation for these small-scale purifications.

3.5.4 Glutathione Elution.

The other half of the fusion protein/matrix solution (250 μL) was treated with 1 mL glutathione elution buffer (10 mM reduced glutathione in 50 mM Tris-HCl, pH 8.0) so that the fusion protein would be competitively eluted from the matrix. After a ten minute incubation at room temperature the matrix was centrifuged and the supernatant, containing fusion protein, decanted. This was repeated three times to ensure that as much of the fusion protein was removed as possible.

Most of the fusion protein was competitively eluted from the matrix as shown by analysis on SDS-PAGE (gel not shown). There was only a small proportion of the fusion protein remaining bound to the matrix.

3.5.5 Thrombin Cleavage In Solution.

The next purification experiment used glutathione elution followed by a thrombin timecourse to follow up on the results seen in Sections 3.5.3 and 3.5.4. The protocol was followed as described in Section 3.5.3, then 25 units of thrombin was added. Samples were taken at timepoints of 0, 10, 20, 30, 60, and 90 minutes. The fusion protein eluted from the matrix was cleaved by the thrombin within ten minutes as shown by SDS-PAGE analysis (gel not shown). The 78 kDa band in the eluted solution, corresponding to the GST-hAlDH fusion protein, was not present in the samples taken after the 10 minute timepoints; two new bands appear, 26 and 52 kDa, which correspond to the GST and hAlDH proteins.

At this stage a Western blot was carried out (Section 2.2.7) in order to confirm the identity of the 78 kDa band as the fusion protein and the 52 kDa bands as the cleaved hAlDH protein. As expected both of these bands were

recognised by the rabbit IgG anti-AlDH antibody and gave a positive result after staining.

These results showed that a small-scale trial of this expression/purification system was successful. The next step was to scale up to a purification from a litre culture of cells, using column purification.

3.6 Large-Scale Expression and Purification.

3.6.1 Preparation Of Cell Lysate.

A 250 mL culture of XL-1/pGEX-hAlDH was grown overnight at 37 °C and used to inoculate 750 mL LB-Amp. This one litre culture was grown for a further 30 minutes at 37 °C so the cells would re-enter log-phase, and then induced with 1 mM IPTG, at which point the A_{600} was 0.695. The induced culture was incubated at 28 °C for four more hours (final A_{600} = 1.02), then the cells harvested by centrifugation using the SC3 rotor for 20 minutes at 4 °C and 4000 g. The cell paste was stored at -20 °C until required.

The cell paste was suspended into 40 mL ice-cold PBSED buffer (PBS buffer with 1 mM EDTA and DTT added) with 10 % glycerol. The cells were lysed by sonication in a series of one minute bursts interspersed with one minute incubations on ice and PMSF (1 mM) was added. The sonicate was centrifuged using an SS34 rotor at 15,000 g for fifteen minutes at 4 °C to separate insoluble protein and cell debris from the soluble protein.

3.6.2 Purification by Cleavage on the Glutathione-Sepharose Column.

The Glutathione-Sepharose 4B matrix was washed and prepared as described in Procedure 7 of the GST Gene Fusion System manual (Pharmacia Biotech). One mL of 50 % matrix slurry was mixed with the 40 mL sonicate solution and incubated on ice for 20 minutes. This matrix/sonicate mixture was poured into a 5 mL empty column (Biorad) and as the matrix settled, with the bound fusion protein, the flow-through was

kept for further analysis. The column was washed three times with 10 mL PBSED to remove non-bound proteins.

The thrombin solution, containing 15 units of thrombin in 1 mL PBSED, was loaded on the top of the column, drained into the matrix and left for 2 hours at room temperature. After this time the thrombin was inactivated by the addition of 1 mM PMSF, washed down the column with 2 mL PBSED. PBSED was used to wash cleaved hAlDH through the column and all of the eluate was collected (15 mL) and then concentrated to a volume of 1 mL.

Samples were collected at all points of the purification procedure and these were analysed by SDS-PAGE (Figure 3-6). The results show that the majority of the fusion protein is soluble (80 %), but that not all of it binds to the matrix (compare lane 4 - supernatant with lane 6 - flow-through). Some of the flow-through fusion protein may be incapable of binding to the matrix because of folding aberrations, but some may not have been able to bind through overloading - the bed volume used may have been too small to allow all of the fusion protein to bind. A small proportion of full length fusion protein was found in the eluted product, this may be due to incomplete washing of the matrix. The sample of eluate after the washes, but before thrombin cleavage (lane 7), shows that unbound protein was still being washed from the matrix.

Activity assays and protein determination assays were carried out and the results were:

[protein]	2.6 mg/mL
total protein	2.6 mg
activity:	1.02 U/mL
Specific Activity:	0.39 U/mg

This yield of 2.6 mg/L was most encouraging and the specific activity was comparable to other preparations of hAlDH (Section 3.2.3). Given that much of the fusion protein produced was not bound to the matrix and was instead found in the flow-through fraction, we decided to take this fraction and pour

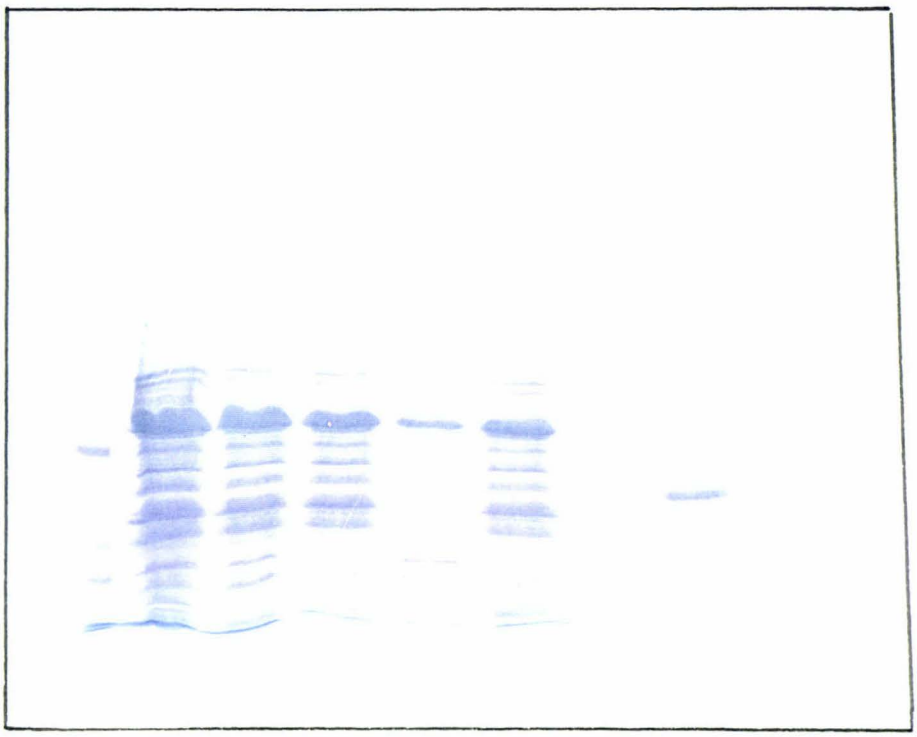

Figure 3-6: SDS-PAGE of fractions collected in the purification using the cleavage on the column method.

Lane 1: molecular markers (SDS7, Sigma); Lane 2: whole cells before sonication; Lane 3: cells after lysis; Lane 4: insoluble fraction after lysis; Lane 5: soluble fraction after lysis; Lane 5: flow-through fraction collected; Lane 6: sample collected after the final wash of the column; Lane 7: eluted hAlDH.

it through a fresh glutathione-Sepharose column, this time with a bed volume of 2 mL.

3.6.3 Secondary Purification - from the flow-through fraction.

2 mL of the glutathione-Sepharose matrix was added to the 40 mL of sonicate that passed straight through the last 0.5 mL column without binding. This was then poured into an empty 5 mL column (Biorad) and the flow-through collected. The unbound proteins were removed from the column with three 20 mL PBSED washes and the matrix was allowed to drain completely between each wash.

In the previous purification 15 units of thrombin cleaved approximately 4 mg of fusion protein (after the cleavage there were 2.6 mg of hAlDH) in 2 hours at room temperature. Fifty units of thrombin was added this time to ensure that all of the fusion protein bound to the matrix would be cleaved. The thrombin was kept in the column for 3.5 hours at room temperature and PBSED buffer, with the addition of 1 mM PMSF, was added to the column to wash through the released hAlDH protein.

Samples were taken from various steps of the purification procedure and analysed by SDS-PAGE (gel not shown). The protein distribution showed that the majority of the fusion protein has bound to the matrix; little is to be seen in the flow-through from this column. However, a large proportion of fusion protein remained on the matrix uncleaved and some of the cleaved hAlDH was not eluted from the column.

The activity and protein determination assays were carried out with the following results:

[protein]	0.23 mg/mL
total protein	1.1 mg
activity:	0.107 U/mL
Specific Activity:	0.46 U/mg

These results show that a total of 3.7 mg of hAlDH was purified overall from this procedure from a 1 L culture, which is a good yield from a bacterial system, equivalent to the yield from the purification using the DEAE and hydroxyacetophenone columns. However, there were two problems with this method of purification; incomplete cleavage of the fusion protein and incomplete elution of the hAlDH protein.

To investigate the nature of the binding of the hAlDH to the glutathione matrix, a high salt wash (PBSED with 0.5 M NaCl) was washed through the column. The SDS-PAGE gel of the fractions collected showed that the high salt elution had not altered the proportion of free hAlDH bound to the column; the distribution of the hAlDH protein was the same before and after the salt wash (gel not shown).

3.6.4 Glutathione elution.

Because of the difficulties in completely cleaving the fusion protein and in completely eluting the cleaved hAlDH, we decided to test elution of the intact fusion protein with glutathione, then cleaving the fusion protein in solution. This may solve the problems of incomplete cleavage and incomplete elution as discussed in Section 3.6.3, as this method should give a higher efficiency of elution given the strong competitive effects of adding free glutathione, and the thrombin should have easier access to the cleavage site in solution.

This procedure adds extra steps to the purification procedure, as along with hAlDH the eluted solution will contain GST, thrombin and free glutathione. The GST contaminant can be removed by another column purification on glutathione-Sepharose, after dialysis to remove the free glutathione from the solution.

A solution of free glutathione is acidic, a 10 mM solution will have a pH of around 4 at room temperature. Thrombin has a pH optimum of around 8.5 and so thrombin cleavage will be very slow in an acidic environment. PBS

buffer was not a strong enough buffer to hold the pH of the glutathione solution high enough for thrombin activity, therefore we used a 50 mM Tris buffer, containing 150 mM NaCl, at a pH of 8.1 at room temperature (pH 8.6 at 4 °C). This should ensure a pH for efficient thrombin cleavage.

A fresh cell culture was prepared for this experiment, two litres of the pGEX-hAlDH-transformed *E. coli* XL-1 Blue cells were grown and induced as described in Section 3.6.1 and stored as four pellets at -70 °C. Two of these pellets (2.52 g in total) were thawed on ice, resuspended in six volumes of TED buffer (pH 8.6, Section 2.2.1). Sonication was also carried out as described in Section 3.6.1.

The sonicate, thus derived from a one litre cell culture, was added to 2 mL of freshly prepared glutathione-Sepharose and poured into a clean, empty 5 mL column (Bio-Rad). The column was washed with TED buffer until the A_{280} reached zero. Free glutathione (final concentration 20 mM) was added to 6 mL TED buffer and Tris added to the solution until the pH reached 8.2. The glutathione was then applied to the column, allowed to drain in and left overnight at 4 °C. Next morning, the eluate was collected in 1 mL fractions, and a further 2 mL glutathione was applied to the column. After 30 minutes at room temperature, the column was drained and all eluate collected.

Thrombin (50 units) was added to the eluted fusion protein solution and cleavage allowed to take place overnight at room temperature. The solution was dialysed for six hours against TED buffer (no glycerol added) to remove the free glutathione and the dialysate was poured down a fresh 2 mL column.

The samples taken from throughout this purification were run on an SDS-PAGE gel as shown in Figure 3-7. Approximately 80 % of the fusion protein produced was soluble (compare lane 2 with lane three) and half of this bound to the matrix as shown by the proportion of GST-hAlDH in the flow-through. All of the contaminating proteins were washed off, as shown by the

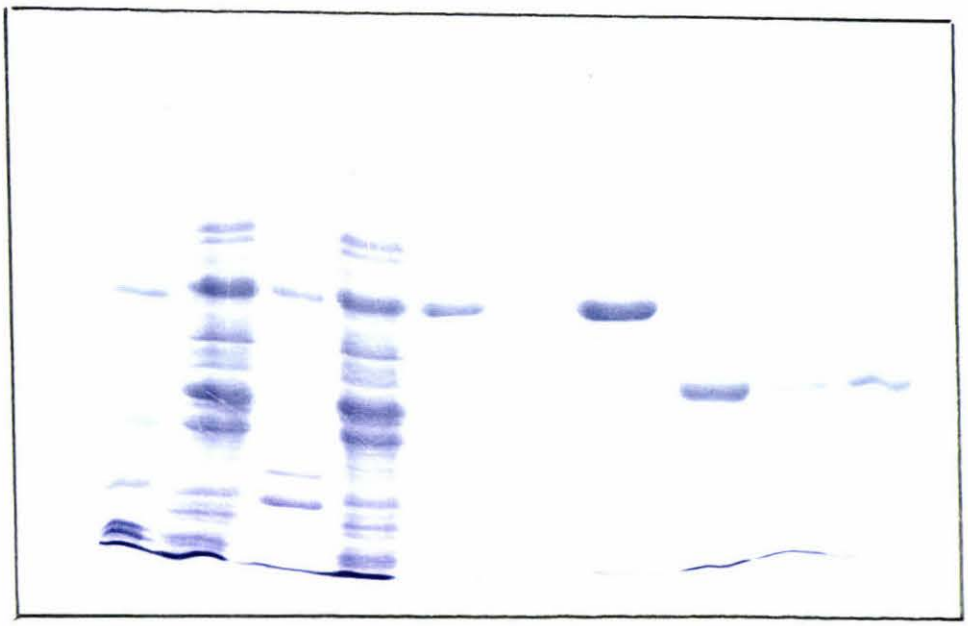

Figure 3.7: SDS-PAGE showing the purification of hAlDH from the fusion protein, using glutathione elution followed by thrombin cleavage in solution.

Lane 1: molecular markers (SDS7, Sigma); Lane 2: soluble material after lysis; Lane 3: insoluble material after lysis; Lane 4: flow-through fraction after loading onto the glutathione-Sepharose matrix; Lane 5: matrix-bound protein; Lane 6: matrix after glutathione elution; Lane 7: eluant after glutathione elution; Lane 8: eluant after thrombin cleavage; Lane 9: matrix after loading cleaved eluant onto second column; Lane 10: eluant from the second column, purified hAlDH.

sample of the matrix after the wash (lane 5). Nearly all off the fusion protein was eluted with the glutathione wash, and nearly all of this was cleaved.

After dialysis the cleaved protein was poured down a second column to remove the GST peptide from solution. Unexpectedly, half of the hAlDH bound to the glutathione-Sepharose column, as shown by the hAlDH band in lane 9.

The cleaved hAlDH eluted from the column was assayed and the results were:

[protein]	0.18 mg/mL
total protein	1.26 mg
activity:	0.0392 U/mL
Specific Activity:	0.196 U/mg

The specific activity for the purified hAlDH was low. This may be due to the extended preparation time needed to carry out this procedure, for example the addition of the dialysis step. Denaturation of the hAlDH protein over time may lead to decreased activity of the enzyme.

Activity assays carried out on samples taken from throughout the purification showed that the GST-hAlDH fusion protein has dehydrogenase activity. This may mean that the GST protein forms a distinct globule attached to each subunit of the tetramer. The GST peptide may not affect the active site, allowing the active site to be folded as in the native molecule. If it is assumed that the fusion protein is able to have the same activity as the cleaved product then the loss of activity through the purification procedure can be tracked.

Three possible reasons for the binding of hAlDH to the glutathione matrix are presented as follows.

1. The tetramerisation of hAlDH may affect its binding to the column. If, for example, three of the four subunits bound in the tetramer were cleaved from their corresponding GST tag, but the fourth one still had its GST

attached, then the whole tetramer is likely to bind to the column. On the denaturing SDS-PAGE gel, the tetramer will dissociate into its monomers, and the ratio of the uncleaved GST-hAlDH monomer to the cleaved hAlDH monomer would be able to be seen. At least a quarter of the protein present would need to be uncleaved GST-hAlDH for this to be a valid explanation, but there appears to be only a trace of uncleaved fusion protein present (lane 9, Figure 3-7).

2. As the addition of an eluting buffer with a high ionic strength did not appear to cause the further elution of hAlDH, ionic interactions are probably not the cause of the hAlDH protein binding to the matrix.
3. The possibility of hydrophobic interactions has not been excluded.

3.6.5 Replication of the 'cleavage on the column' experiment.

As a concluding experiment for this piece of work, we decided to try again the purification of hAlDH using thrombin cleavage on the column, to see if we could reproduce the results described in Section 3.6.2, where the yield of purified protein was 2.6 mg purified from a one litre culture, and the specific activity was 0.39 units/mg.

A 1.2 g cell pellet (equivalent to 0.5 L) was sonicated, as described in Section 3.6.4. The cleared sonicate was loaded onto a fresh 2 mL glutathione-Sepharose column and the column washed with TED until the A_{280} was less than 0.01.

Thrombin (50 units) was added, drained into the column and left for three hours at room temperature. At the end of the incubation, fractions were collected (around 0.5 mL each) and run on an SDS-PAGE gel to determine which fractions contained hAlDH (Figure 3-8).

Fractions 1 to 4 were pooled and assayed for protein concentration and dehydrogenase activity. The results were:

[protein]	0.48 mg/mL
total protein	1.44 mg

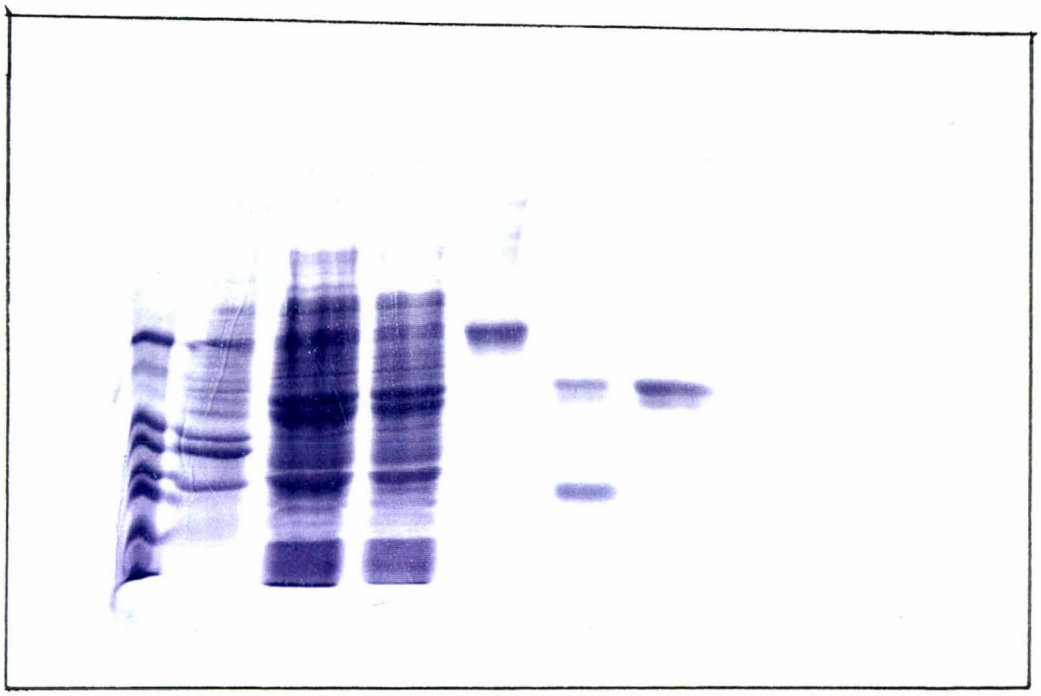

Figure 3-8: SDS-PAGE showing fractions collected from purifcation using the cleavage on the column method.

Lane 1: molecular markers (SDS7, Sigma); Lane 2: insoluble material after lysis; Lane 3: soluble material after lysis; Lane 4: flow-through fraction collect after the loading of the column; Lane 5: matrix-bound protein; Lane 6: matrix-bound protein after cleavage; Lane 7: eluted hAlDH.

activity: 0.059 U/mL
Specific Activity: 0.147 U/mg

Only 500 mL of a cell culture was used for this purification so the lower yield was expected. However, there was a lower specific activity than expected. The activity assay showed that in the sonicate hAlDH specific activity was 0.016 units/mg climbing to a specific activity of 0.147 units/mg by the end of the procedure, a 90-fold purification factor. In contrast the specific activity for the purified enzyme from Section 3.6.2 was 0.39 units/mg, more than twice than in this purification.

The entire purification was carried out in one day and all steps were strictly pH controlled and carried out at 4 °C, except for the three hour room temperature incubation for thrombin cleavage. PMSF was added after sonication and after thrombin cleavage to stop protease activity.

The major changes between this preparation and the one described in Section 3.6.4, with 2.6 mg purified from a one litre culture, are:

- different cell culture used,
- 500 mL culture instead of one litre,
- TED buffer used instead of PBSED,
- 2 mL bed volume instead of 0.5 mL.

These may be factors that altered the activity level of the purified enzyme.

3.7 SUGGESTIONS FOR FUTURE WORK.

The best expression of the GST-hAlDH fusion protein was obtained by growing up a 2 L culture of *E. coli* XL-1/ pGEX-hAlDH, inducing at an A_{600} of 0.6 with 1 mM IPTG and lysing the cells by sonication (Section 3.4).

Table 3.1: Purification of human cytosolic AlDH from the GST fusion protein.

Purification step	volume (mL)	Total Activity (U)	Protein (mg)	Activity (U/mg)	Purification Factor
soluble fraction	12	1.56	96	0.016	1
flow-through	12	0.816	48	0.017	n.a.
matrix after cleavage	2	0.156	0.4	0.39	24
eluted hAlDH	4	0.236	1.6	0.147	91

AlDH activity and protein determination was assayed according to the methods to the methods given in Sections 2.2.5 and 2.2.6.

Approximately 20-35 % of the expressed AlDH was insoluble, perhaps having formed inclusion bodies.

Two different methods for the purification of the recombinant enzyme were tested. The first was the thrombin cleavage of the fusion protein while bound to the glutathione-Sepharose affinity matrix, with the pure aldehyde dehydrogenase being eluted in a single step. The second method involved the competitive elution of fusion protein, then thrombin cleavage in solution.

This first method of on-the-column cleavage (bed volume 0.5 mL) yielded 2.6 mg/L of aldehyde dehydrogenase with a high specific activity of 0.39 units/mg (Section 3.6.2). Analysis of the fractions collected from various steps of the procedure showed that some of the fusion protein washed straight through the column, not binding to the matrix (Figure 3-6). When this flow-through fraction was loaded onto a second column (bed volume 2 mL) and the procedure repeated, a further 1.1 mg of aldehyde dehydrogenase was eluted, with a high specific activity of 0.46 units/mg. The flow-through from this secondary purification still contained fusion protein, which may not have been able to the affinity matrix through an alteration in conformation.

The potentially higher yields that could be gained from this system could be seen by the amount of protein which was remaining uncleaved and bound to the matrix and in the incomplete elution of the cleaved hAlDH (Section 3.6.3).

Given the incomplete cleavage and incomplete elution of the recombinant protein, we decided to test the second method of purification. This involved the elution of the intact fusion protein from the affinity matrix with free glutathione, cleaving the GST-hAlDH in solution, dialysing to remove the glutathione, and reloading the dialysate onto a second column to remove the GST protein from solution. A higher efficiency of elution should be gained

due to the strong competitor, and once in solution the thrombin should have easier access to the cleavage site within the fusion protein.

The yield of protein was 1.26 mg with a specific activity of 0.196 units/mg from a 500 mL cell culture. SDS-PAGE analysis (Figure 3-7) showed that approximately only 50 % of the soluble protein was binding to the glutathione-Sepharose (2.0 mL bed volume), the rest being found in the flow-through. The glutathione elution and thrombin cleavage steps were very efficient. However, during the second column step (where the GST fragment is removed from solution), half of the cleaved hAlDH bound to the GST affinity column which was unexpected as there should be no affinity of the hAlDH protein with the glutathione-Sepharose. Three possible reasons for this action have been explored (Section 3.5.4); effects of tetramerisation on binding, ionic interactions, and hydrophobic interactions.

This last possibility of hydrophobic interactions could be explored by the utilisation of a different type of GST affinity matrix. A glutathione-Sepharose 6B matrix and a sulfobromophthalein-Sepharose 6B matrix have been kindly donated by Dr Alan Clark (School of Biological Sciences, Victoria University of Wellington, NZ). This glutathione-Sepharose column has a different linker region, which should have less hydrophobicity associated with it (A Clark, pers. comm.).

The initial trial of the cleavage-on-the-column purification worked well, with as good a yield as with the T7 expression and DEAE/*p*-hydroxy-acetophenone purification.

The purification using glutathione elution, then cleavage in solution was comparable in yield and activity gained, but is not as efficient with respect to length of time and monetary cost.

On the basis of these results, the purification protocol of the hAlDH using the GST fusion protein system should involve the loading of the sonicate

onto the glutathione-Sepharose column, washing off unbound proteins, loading the thrombin, incubation with thrombin at room temperature for an hour and then elution of the cleaved hAlDH.

The GST fusion protein system (Pharmacia Biotech) has produced pure and active protein in a single step affinity chromatography, and does not rely on the integrity of the AlDH active site for the purification.

4. *PICHIA PASTORIS*

4.1 INTRODUCTION.

The methylotrophic yeast, *Pichia pastoris*, has been used to produce large amounts of an heterologous protein, with an average yield around 200 mg per litre, with yields of up to 8 g/L being reported (Sberna *et al.*, 1996). This potential for high yields was the major reason for testing this expression system, as crystallography and kinetic studies, such as stopped-flow experiments, require large quantities of high purity protein for characterisation of the enzyme.

Under growing conditions where the sole carbon source is methanol, the yeast produces alcohol oxidase (AOX) which catalyses the oxidation of methanol to formaldehyde. There are two *AOX* genes, producing isoenzymes AOX1, which is responsible for most of the activity, and AOX2 which is less efficient (Cregg *et al.*, 1989). AOX1 is transcriptionally regulated under the control of a very strong promoter, induced by the presence of methanol. Under inducing conditions up to 30% of total cell protein will be alcohol oxidase (Sberna *et al.*, 1996). *Pichia pastoris* is also able to be grown to a very high cell density, especially in a fermentor system, adding for the high capacity of production of AOX1 (Cregg and Higgins, 1995).

By placing the cDNA of a foreign gene directly downstream from the *AOX1* promoter, heterologous gene expression can be specifically induced by growing the recombinant yeasts on methanol.

4.2 THE *PICHIA PASTORIS* VECTORS, pPIC3K AND pPIC9K.

In order for the foreign gene to be correctly inserted into the yeast genome, a shuttle vector needs to be used. Two such vectors that have been constructed are pPIC9K and pPIC3K (Invitrogen). The first vector allows the heterologous protein to be exported into the growth medium through the use of a secretion signal, while the second gives intracellular expression. An

example of a vector used to direct the secretion of the heterologous protein, pPIC9K, is shown in Figure 4-1.

Secretion of the expressed protein has been shown to act effectively as a purification step, due to the low proportion of naturally secreted protein (Sberna *et al.*, 19960. However, it has also been found that proteolysis of secreted proteins and inefficient secretion of complex proteins have made the intracellular expression an attractive alternative (Romanos, 1995).

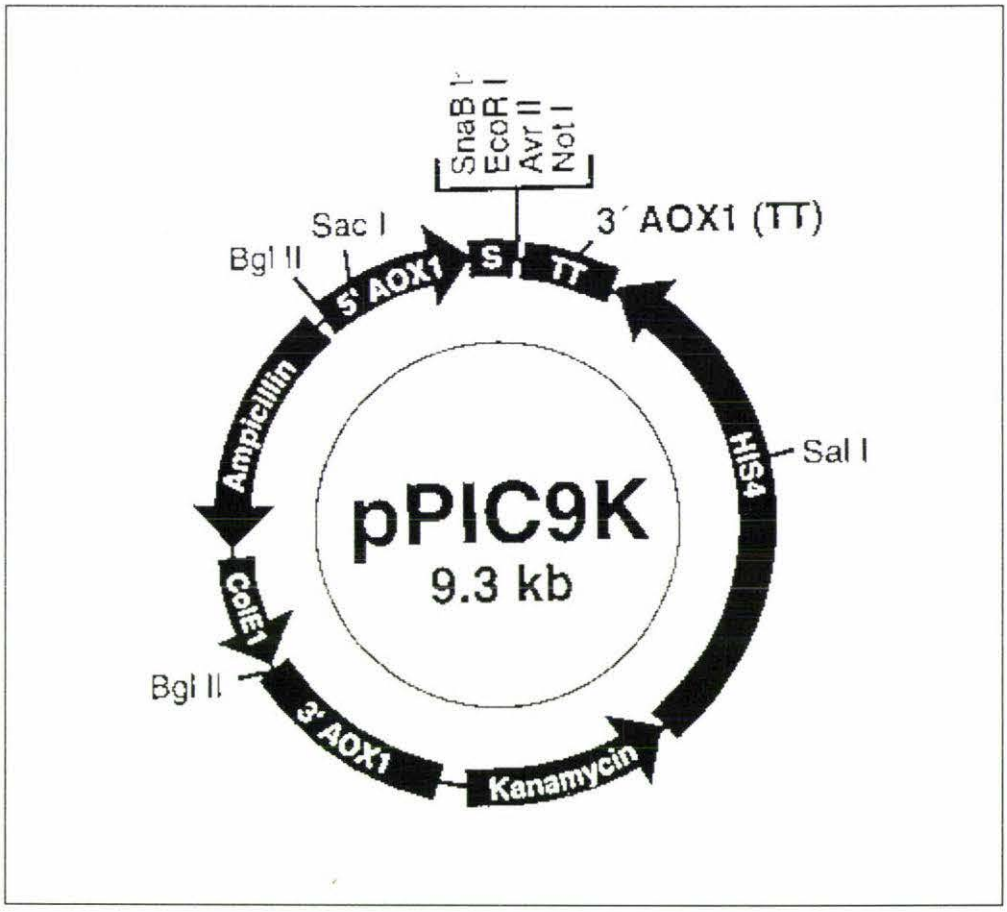

Figure 4-1: Schematic diagram of pPIC9K (Invitrogen). The expressed heterologous protein will be secreted.

Features of these plasmids include:

- The 5' promoter and 3' terminating sequences for the alcohol oxidase (*AOX1*) gene to ensure correctly regulated expression,

- The native *AOX1* transcription-terminating and polyadenylation signals;
- A multiple cloning site;
- Ampicillin resistance gene and *E. coli* origin of replication;
- The *Pichia* wild-type *HIS4* gene, used to complement the *Pichia his4* mutant yeast strain, acting as a *Pichia* selectable marker;
- The bacteriophage f1 origin of replication to allow the production of single-stranded DNA for mutagenesis,
- A kanomycin/G418 resistance gene that allows selection of multicopy integrants.

In addition, the pPIC9K vector has the *Saccharomyces cerevisiae* α-factor secretion signal immediately 5' to the multiple cloning site, which targets the protein for extracellular secretion. This signal is cleaved from the expressed protein in the Golgi body apparatus (Cregg and Higgins, 1995).

Pichia pastoris does not have any natural stable episomal vectors (Romanos, 1995). Therefore the information carried on the pPIC plasmids is integrated into the yeast genome, through homologous recombination, in four ways:

- Gene replacement by a double crossover event between the plasmid *AOX1* sequences and the genomic *AOX1* gene.
- A single crossover event between the *AOX1* sequences on the plasmid and the genomic *AOX1*,
- A single crossover event between the *HIS4* sequences on the plasmid and the genomic *his4* gene,
- Rarely, multiple gene insertion events occur where many gene copies are integrated at the same locus.

If the heterologous gene copy inserts within the genomic *AOX1* the ability for the yeast to grow well on a methanol medium will be reduced, but not completely knocked out due to the activity of the less efficient AOX2 protein. This phenotype has been termed Mut^S or methanol utilisation sensitive. However, if the gene inserts into the *his4* gene, the AOX1 activity is retained, termed Mut^+.

The $Mut^{+/S}$ phenotype needs to be determined as the strategy used for expression differs depending on the ability for the yeast to grow on methanol, the expression-inducing substrate.

4.3 CLONING OF SHEEP AlDH INTO pPIC3K AND pPIC9K.

The sheep homologue of the class 1 aldehyde dehydrogenase was chosen for this expression study. More data is available with the native sheep-liver-purified enzyme which will enable direct comparison with the recombinant *Pichia*-produced enzyme, to ensure that studies with the recombinant enzyme are a valid model for enzymatic action *in vivo*.

The sheep class 1 aldehyde dehydrogenase cDNA had previously been cloned into the *Eco*RI site of the multiple cloning site of pBluescript SK+/- (Staynor and Tweedie, 1995). This included 600 bp of 5' and 3' untranslated region and 1500 bp of coding region. The coding region of this cDNA was placed in between the *Sna*BI and *Eco*RI restriction sites in the multiple cloning site of pPIC9K, and in between the *Bam*HI and *Avr*II (also known as *Bln*I) restriction sites of pPIC3K.

4.3.1 Preparation of the Secreted-Expression Construct, pPIC9K-shAlDH.

The pPIC9K-shAlDH construct was prepared by amplifying the shAlDH cDNA using PCR, and cutting the PCR product with *Eco*RI to produce a molecule with a 5' blunt end and a 3' *Eco*RI end. The pPIC9K vector was prepared by cutting with *Eco*RI and *Sna*BI and the insert cDNA was ligated in. The plasmid map of the resulting construct is shown in Figure 4-4.

The α-factor signal sequence has its own ATG translational start point 250 bp upstream from the multiple cloning site. The native AlDH translational start point was removed and the inserted shAlDH cDNA was cloned in frame with the α-factor signal sequence.

Polymerase Chain Reaction (PCR) was used to selectively amplify the 2.1 kb shAlDH insert from the pBluescript-shAlDH plasmid. A 5' forward primer was designed to the first 23 base pairs immediately after the initiating ATG and annealed to the non-coding strand:

PCRF 5' TCGTCCTCAGCCATGCCAGA 3'

The 3' 'reverse' primer used was complementary to the T7 region of the pBluescript plasmid and annealed to the coding strand:

T7 5' GTAATACGACTCACTATAGGGC 3'

PCR was carried out as described in Section 2.1.3.

The products of the PCR were run on an 0.8 % agarose gel (Section 2.1.2) and a single band of the expected size of 2.1 kb was seen (Figure 4-2) and excised from the gel. The DNA was purified from the agarose/ethidium bromide slice using the Biorad "Prep-a-gene" kit, and resuspended in 50 μL TE. The PCR product was treated with T4 DNA polymerase, then purified using Wizard "PCR Preps" (Promega) and resuspended in 50 μL TE. The DNA was quantified by eye, by comparison with standards on an agarose gel, and estimated to be 10 ng/μL.

The blunt-ended 2.1 kb PCR fragment was then digested with the *Eco*RI restriction endonuclease to give a *Eco*RI 'sticky' end at the 3' end of the gene. The DNA was purified, again using the Biorad "Prep-a-gene" kit, and resuspended in 50 μL TE. The insert DNA therefore had a blunt 5' end and an *Eco*RI compatible sticky 3' end.

The pPIC9K vector was prepared by digesting the plasmid with *Sna*B1 (to prepare a blunt end), purifying the linearised plasmid DNA from the agarose gel, then digesting with *Eco*RI. The vector then had the correct ends for ligation with the insert DNA.

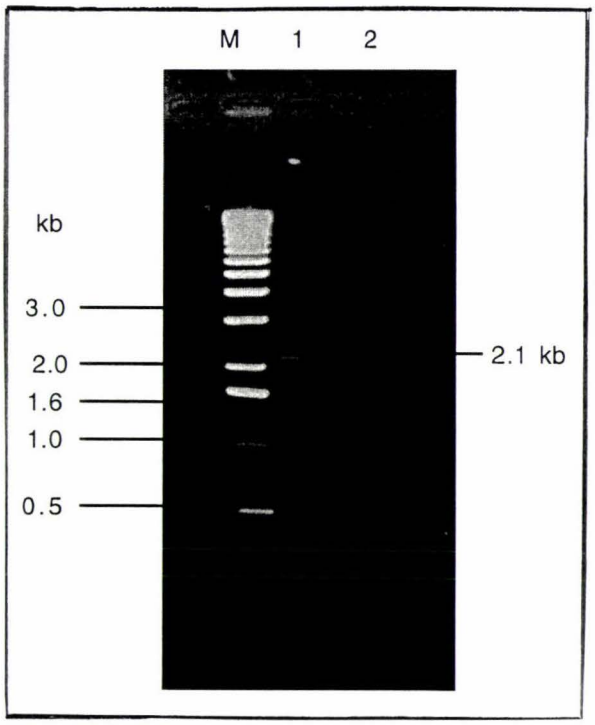

Figure 4-2: PCR amplification of the hAlDH cDNA from the pBluescript-hAlDH template. The expected size of product is 2.1 kb.

Lane M: molecular markers (1 kb DNA ladder, BRL); Lane 1: sample from the PCR reaction containing the pBluescript-hAlDH template; Lane 2: PCR with no template added, negative control.

The 5' phosphates from the linearized vector DNA were removed using thermosensitive alkaline phosphatase (Section 2.1.8) making the vector less likely to self-ligate in the ligation reaction.

Prepared vector DNA (90 ng) and prepared insert DNA (35 ng) were ligated as described in Section 2.1.9 and incubated overnight at 16 °C. A further 1 µL of ligase was added, and the mixture left at room temperature for three more hours.

Three microlitres of the ligation mix was added to 100 µL calcium-competent XL-1 Blue cells. Transformation was performed as in Section 2.1.10 but the control reactions with 'vector alone' or 'vector + insert alone' were omitted. The results of these transformations are shown in Table 1.1.

Table 4.1: The number of transformants after ligation of shAlDH into pPIC9K.

DNA	Ligation Reaction	Number of Colonies
pPIC9K	1. vector + insert + ligase	16
	2. vector + insert + ligase	25
Stock pPIC9K (15 ng)		>500
No DNA		0

Sixteen of these transformant colonies were picked and the DNA was isolated using rapid-boil minipreps (Section 2.1.10). Three of the sixteen ran at a higher molecular weight than the stock pPIC9K DNA and gave the expected pattern in a diagnostic digestion using *Bgl*II. *Bgl*II cuts twice in the pPIC9K vector and once in shAlDH. If there is no insert in the vector, there will be two fragments, 6.8 and 2.5 kb in size. However, if the insert is present and in the correct orientation there will be three fragments, 7.8, 2.5 and 1.2 kb in size. This diagnostic digestion is illustrated in Figure 4-3, where the digestion pattern for the pPIC9K-shAlDH clone is compared with the pattern for pPIC9K.

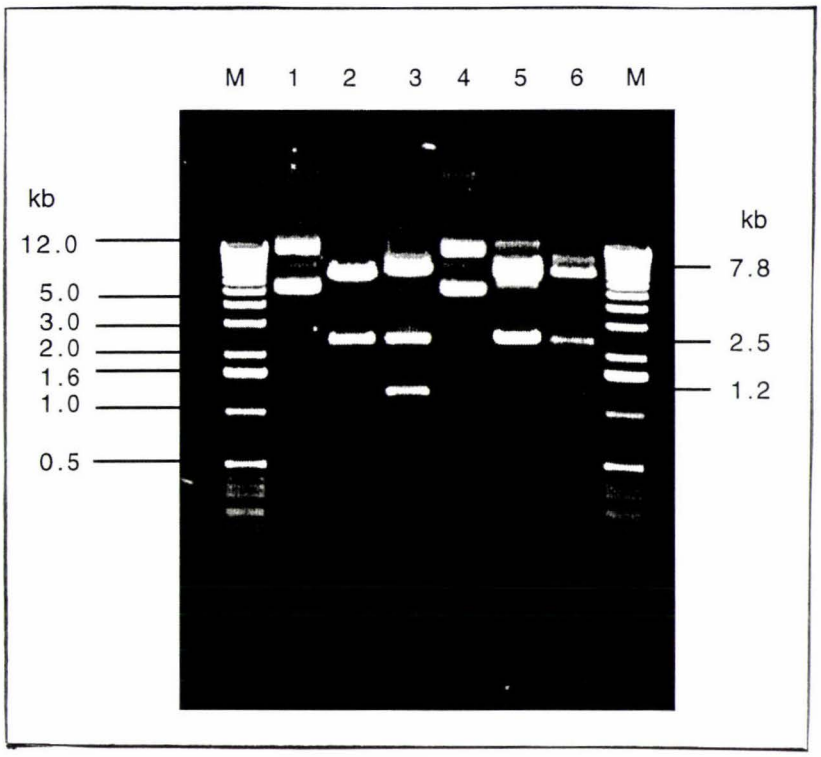

Figure 4-3: *Bgl*II Diagnostic digestion of the pPIC-hAlDH constructs compared to the original pPIC vectors. The correct insertion of the hAlDH DNA will give three fragments of 7.8, 2.5 and 1.2 kb after digestion with *Bgl*II digestion.

Lanes M: molecular markers (1 kb DNA ladder, BRL); Lane 1: pPIC9K vector uncut; Lane 2: pPIC9K cut with *Bgl*II; Lane 3: pPIC9K-hAlDH cut with *Bgl*II; Lane 4: pPIC3K vector uncut; Lane 5: pPIC3K cut with *Bgl*II; Lane 6: pPIC3K-hAlDH cut with *Bgl*II.

A large-scale DNA MaxiPrep was prepared (as described in Section 2.1.11) of two of these clones:

- pPIC9K-shAlDH 1 648 ng/μL
- pPIC9K-shAlDH 2 660 ng/μL

which were stored in aliquots at -70 °C. pPIC9K-shAlDH 1 gave the expected banding pattern after digestion with *Bgl*II (Figure 4.3). The plasmid map of the pPIC9K- shAlDH clone is shown in Figure 4-4.

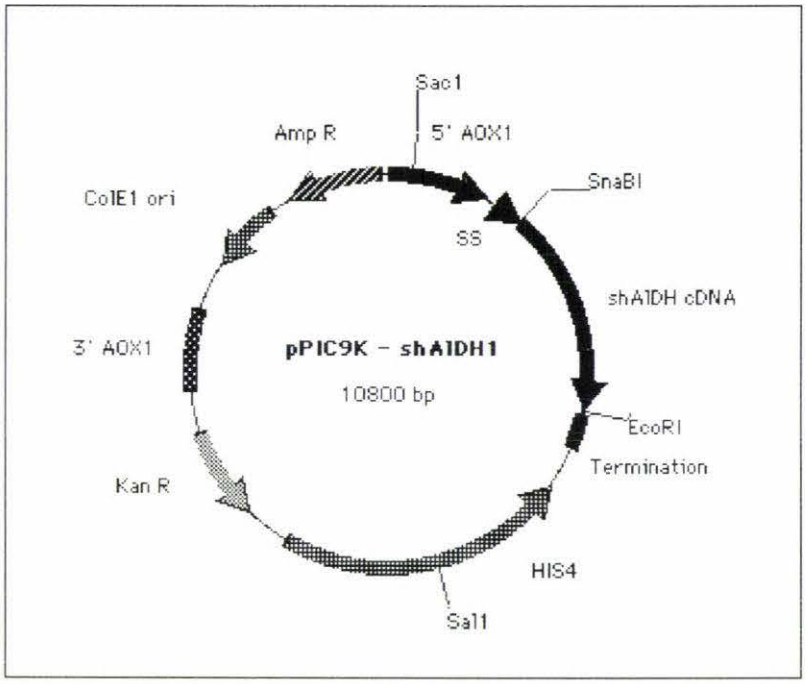

Figure 4-4: Schematic diagram of the pPIC9K-shAlDH construct. SS: Secretion Signal.

As the pPIC9K-shAlDH1 clone was constructed using PCR, which can not be guaranteed to be error-free, the shAlDH insert was sequenced (Sanger Sequencing Laboratory, Biochemistry Department, University of Otago, Dunedin). Seven sequencing primers were used that gave overlapping sequence information in both directions.

Around 500 bp were clearly read from each primer and the fragments were assembled together into a contiguous sequence using Fragment Assemble from the GCG package of software.

The sequencing showed that the PCR-generated shAlDH DNA cloned into the pPIC9K plasmid would give a protein product with an identical amino-acid sequence to the published class 1 sheep AlDH sequence (Staynor and Tweedie, 1995), except for one change. Ile-25 is coded by the DNA codon ATA in the published shAlDH sequence. However, in the PCR-generated sequences this codon is consistently altered to the AAA codon, which codes for a lysine. This change also occurred in the sequence of another pPIC9K-shAlDH clone, constructed separately, which may suggest that the error lies in the original sequencing, and not in a Taq-induced alteration. The primary sequence data used to generate the published shAlDH sequence (Staynor and Tweedie, 1995) has not been analysed to establish the cause of the sequence discrepancy.

4.3.2 Preparation of the Intracellular-Expression Construct, pPIC3K-shAlDH.

The pPIC3K-shAlDH vector, giving intracellular expression of shAlDH, was prepared differently as the native shAlDH translation start point was required. A straight-forward 'cut and paste' cloning procedure was used by cutting the insert and plasmid with *Bam*HI and *Avr*II restriction enzymes, purifying the required fragments from an agarose gel, then ligating the insert within the plasmid. The plasmid map of the resulting construct is shown in Figure 4-5.

*Bam*HI cuts the pBlue-shAlDH plasmid twenty bases upstream from the translational start point and *Avr*II cuts within the 3' untranslated region 285 bp downstream from the translational stop point. This gives a shAlDH insert size of 1850 bp, containing the 1500 bp coding region. Both the *Bam*HI and *Avr*II restriction sites are in the multiple cloning site of the pPIC3K vector.

The shAlDH insert was prepared by cutting 1 μg of the pBluescript-shAlDH plasmid with 15 units *Avr*II and digesting for 2 hours. The digested mixture was run on an agarose gel and the band equivalent to the 5.1 kb linearised plasmid was excised and purified. This fragment was then cut with *Bam*HI giving two fragments, pBluescript (3.5 kb) and shAlDH (~1.8 kb). The 1.8 kb fragment was excised from the agarose gel and purified.

The pPIC3K vector was prepared by *Bam*HI digestion, purification of the linearised vector from uncut vector by electrophoresis followed by band excision, then cut with *Avr*II digestion. The DNA was purified and the vector fragment treated with thermosensitive phosphatase as in Section 2.1.8.

The ligation reaction was carried out as described in Section 2.1.9, with a room temperature incubation of 5.5 hours. Three microlitres of the ligation mix was added to 100 μL calcium-competent *E. coli* XL-1 Blue cells and the transformation was carried out as described in Section 2.1.10.

Table 4.2: The number of transformants after ligation of shAlDH into pPIC3K.

DNA	Ligation Reaction	Number of Colonies
pPIC3K	1. vector alone	13
	2. vector + insert, no ligase	18
	3. vector + insert + ligase	45
Stock pPIC3K (15 ng)		>500
No DNA		0

Sixteen of the transformant colonies from ligation reaction 3 were grown up and the DNA was isolated by 'rapid boil' minipreps (Section 2.1.11). One of the four successful minipreps clearly showed a higher molecular weight and

it gave the correct pattern with the diagnostic *Bgl*II digest, as described in Section 4.3.1.

This miniprep DNA was subsequently retransformed into XL-1 Blue cells in order to grow up a culture for a large-scale plasmid preparation (Section 2.1.12).

This plasmid was renamed pPIC3K-shAlDH1 (1.7μg/μL) and was stored in aliquots at -70 °C. pPIC3K-shAlDH1 gave the expected pattern after digestion with *Bgl*II as shown in Figure 4-3.

As this cloning procedure did not involve PCR, sequencing of the clone was unnecessary.

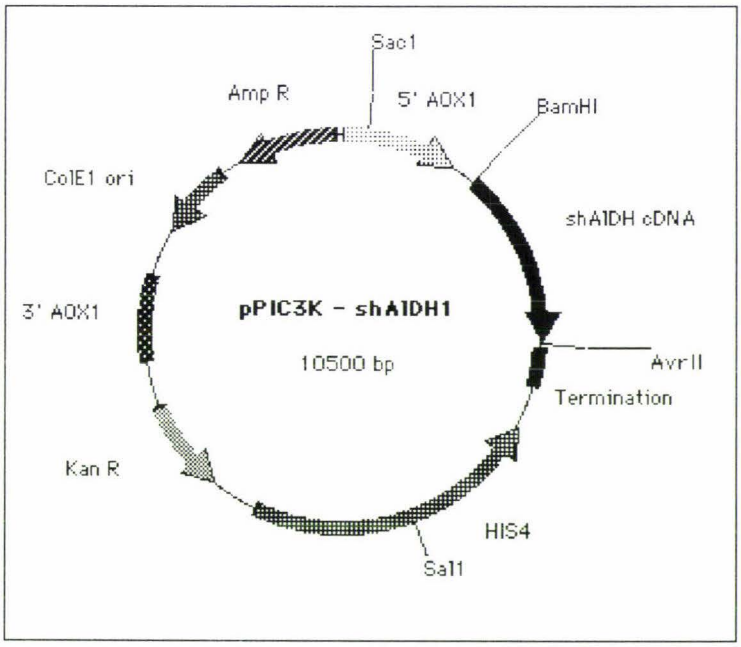

Figure 4-5: Schematic diagram of the shAlDH-pPIC3K construct.

4.4 TRANSFORMATION INTO *PICHIA PASTORIS*.

The pPIC vectors do not contain an origin of replication recognised by the yeast, so the plasmid DNA needs to be incorporated into the yeast genome in order to be maintained through cell division. The ways in which the

plasmid DNA is able to be inserted into the *Pichia* genome are described in Section 4.2.

The plasmid DNA can replace, or insert within, the native genomic *AOX1* gene, or it can insert into the genomic *his4* gene. If the native *AOX1* gene is disrupted then the yeast will not grow well on the methanol medium used to induce expression, giving the phenotype Mut^S. On the other hand, if the vector DNA inserts within the genomic *his4* gene the AOX1 activity is retained and the phenotype will be Mut^+. An additional round of screening is required to determine the $Mut^{S/+}$ status of each clone as the strategy for expression will be different for Mut^S and Mut^+ cultures, given the different rates of growth on the methanol medium. Mut^+ transformants are less likely to become poisoned by the methanol, while Mut^S cultures are not likely to be as affected by limited oxygen.

To bypass this additional round of screening, KM71 was the *Pichia pastoris* strain used for the following expression experiments. KM71 has the *AOX1* gene knocked out through the insertion of the *ARG4* gene, therefore the strain is already known to be Mut^S and the growth pattern will not change due to the site of insertion of the foreign gene. Therefore all recombinant cultures can be treated as Mut^S in the expression trials. KM71 has also been reported to have a higher transformation frequency with electroporation than other *Pichia* strains (Romanos, 1995).

In order for the foreign gene to be integrated into the yeast genome, the pPIC plasmid needs to be linearised by cutting at unique restriction sites within the recombinant vector. If the restriction site used is in the *AOX1* region of the plasmid then the vector is likely to be directed to the *AOX1* gene in the genome, and likewise, if the site used is in the *HIS4* gene the integration site is likely to be within the genomic *his4*. The pPIC9K-shAlDH1 plasmid has a unique *Sac*I site within the 5' *AOX1* sequence (as does pPIC3K-shAlDH1), and this was used to linearise the plasmid.

There are two high efficiency methods for transforming plasmid DNA into *Pichia pastoris* cells, electroporation and spheroplasting. The former involves inducing the uptake of DNA with a high voltage current, while the latter involves partially digesting the *Pichia* cell walls with zymolyase and incubating the cells with the linearised DNA.

Spheroplasting was tried with no success and given the lengthier time needed to carry out the spheroplasting protocol, electroporation was the chosen method of transformation.

4.4.1 Electroporation.

Electrocompetent *P. pastoris* KM71 cells were prepared as directed in the *Pichia pastoris* Instruction Manual (Invitrogen). A mixture of 5-20 μg of linearised, clean plasmid DNA and 80 μL of prepared cells was placed in a ice-cold 0.2 cm electroporation cuvette and pulsed using the Bio-Rad GenePulser, with the following parameters:

- Charging voltage of 1500 V
- Capacitance of 25 μF
- Resistance of 400 Ω.

The time constant for the pulse generally ranged between 5.0 and 8.1 ms. Ice-cold 1M sorbitol (500 μL) was added immediately after the pulse and 50-200 μL aliquots of the cells were plated onto MD agar plates and incubated at 30°C until colonies appeared, usually four days.

Recombinant colonies will have the wild-type *HIS4* gene from the integrated plasmid, and so are capable of growing on a histidine deficient medium. The yeast cells without the incorporated plasmid are not able to survive on the minimal MD medium.

After several of these electroporations there were adequate numbers of recombinant yeast colonies to carry through into screening trials.

4.5 Screening for Integration and Multi-copy Insertion Events.

By screening all of the *Pichia* recombinant colonies for multi-copy insertion events, the clones that are likely to produce higher amounts of heterologous protein can be isolated. This can lead to more efficient expression trials.

The integration of the plasmid DNA into the yeast genome should also be confirmed before expression, as it is possible that recombination has occurred within the plasmid during integration. This can be achieved using PCR to amplify the inserted region. The PCR screening was carried out after the multi-copy screening to reduce the number of PCR reactions needing to be carried out.

4.5.1 G418 Screening for Multi-copy Integration.

The G418 resistance of a particular recombinant colony is proportional to the number of Kan^r genes inserted into the host genome. As there is one Kan^r gene per recombinant vector transformed, the level of G418 resistance is proportional to the number of plasmid copies inserted as a multi-copy integrant (Scorer *et al.*, 1994). An increase in the number of the foreign-gene copies in the genome may increase the level of expression of the heterologous protein.

To screen the recombinant *Pichia* colonies with G418, an equivalent cell density needs to be plated onto YPD plates with various concentrations of G418. To do this there are two methods set out in the Invitrogen manual.

The first involves the pooling of all positive recombinant colonies and plating a titre of 10^5 cells onto YPD plates containing different concentrations of G418. While this is quick it involves an amount of uncertainty as, given that the colonies are pooled, it is impossible to tell if colonies growing on a certain concentration of G418 are from different integration events, or have arisen from the same original positive colony.

The second method uses microtitre plates to grow all positive recombinant colonies up to the same cell density by successive inoculations of fresh medium. A 10 μL sample from each colony is spotted onto each LB agar plate containing a different concentration of G418, in the range of 0 to 2 mg/mL. After incubation at 30 °C, the plates are compared to see which original recombinant colony can grow on the highest G418 concentrations. This method takes more time but is more useful as the clonal origin of each colony on the G418 plates is known.

Both methods were tested in this research and the results from the screening with the second method were used for subsequent expression experiments. Results are presented in Table 1.3, and a representative agar plate from this procedure is shown in Figure 4-6.

Table 4.3: Results from the G418 screening.

DNA	[G418] mg/mL	Transformants that grew
pPIC9K-shAlDH1	0	All
	0.25	All
	0.38	Most of the transformants
	0.75	18 of the transformants
	1.5	9A 9B 9C 9D 9E 9G 9H 9L 9M
	2.0	9B 9C 9G 9H 9L 9M

Of all the colonies screened, nine pPIC9K-shAlDH1 clones appeared to have multicopy insertions. These were selected for further study.

4.5.2 PCR Amplification of shAlDH From the Yeast Genome.

Polymerase Chain Reaction (PCR) can be used to find out if the foreign gene insert has integrated into the genome without deleterious recombination. PCR was carried out with genomic DNA isolated from the yeast cultures and the selection of primers that amplify the integrated

plasmid sequences. The presence of a band of the expected size suggests correct integration.

Genomic DNA can be extracted from the yeast cell as described in Section 2.3.4, added to the PCR mix, and the PCR protocol followed as described in Section 2.1.3.

The primers used, shAlDH2 and shAlDH5, annealed within the shAlDH insert and should give an expected product of 950 bp in size. As shown in Figure 4-7, four transformants gave a clear 950 bp product, and show that the shAlDH gene is integrated into the yeast genome. These four transformants were 9B, 9G, 9H and 9L derived from the pPIC9K-shAlDH1 plasmid.

4.6 Expression of the Recombinant *Pichia* Clones.

4.6.1 Small-scale Cultures.

The first expression trial was carried out with 3 transformants: pPIC9K-shAlDH1 clones 9A, 9C and 9D. Two more cultures were grown up, pPIC9K-shAlDH1 (no insert control) and pPIC9K-shAlDH1 transformant 9B, but the cultures were inadvertently destroyed while centrifuging on the day of methanol induction.

The cultures were picked from single colonies on the 0.75 mg/mL G418 plate, and grown up in 10 mL of BMGY medium in a 50 mL 'nunc' tube.

Three days later the cultures were induced to produce the heterologous protein. The cultures were spun down and resuspended in 10 mL BMMY medium (0.5 % methanol). A sample of 150 µL was taken (Day 0). The cultures were placed in the 28 °C room onto a fast-shaking benchtop shaker.

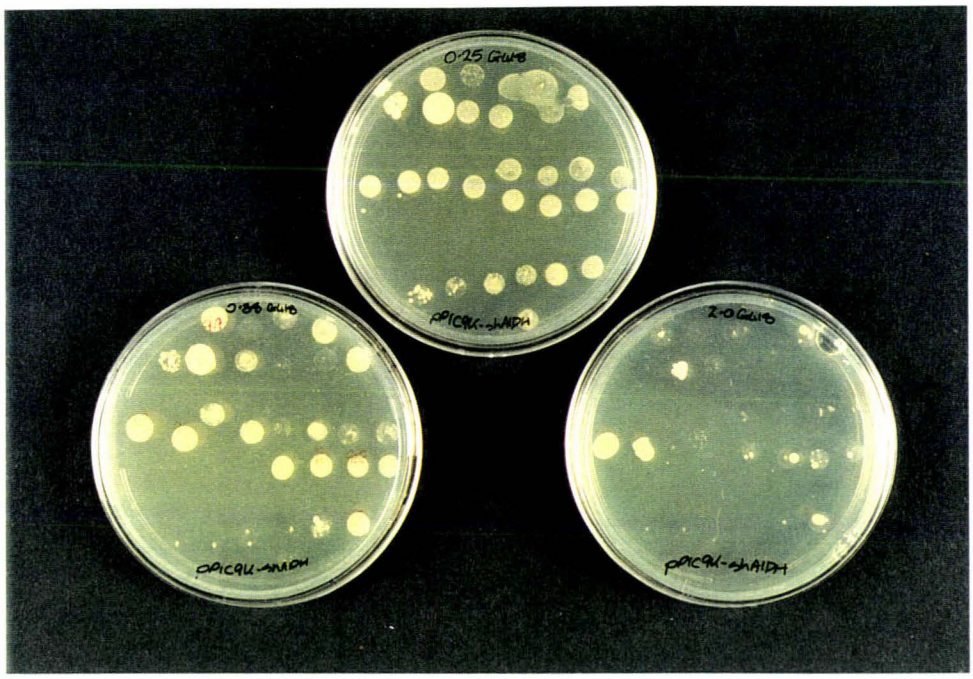

Figure 4-6: A representive example of multi-copy screening, showing the *Pichia* transformants capable of growing on G418 plates after 5 days at 30 °C. Anticlockwise from top: G418 concentration 0, 0.38 and 2.9 mg/mL.

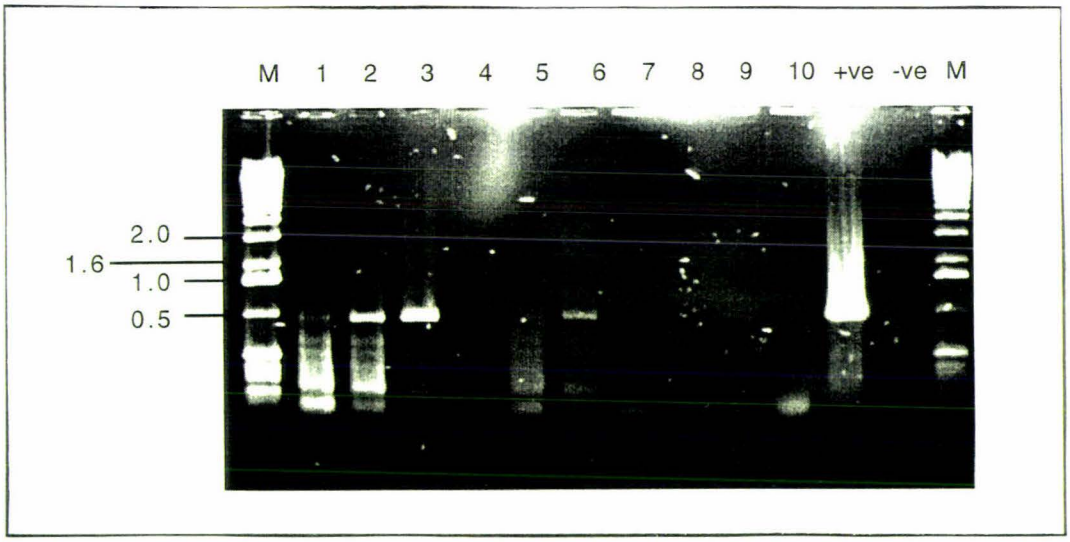

Figure 4-7: PCR amplification of the hAlDH cDNA within the yeast genome. The expected size of product is 0.5 kb.

Lane M: molecular markers (1 kb DNA ladder, BRL); Lanes 1-10: *Pichia*-hAlDH recombinants G, H, L, M, A, B, C, D, E, and F respectively; +ve control: pBluescript-hAlDH template; -ve control: no DNA template added.

Samples were taken from each culture every 24 hours for three days (Days 1-4). After the 150 μL sample was drawn off, 100 % methanol was added to give a final concentration of 0.5 % in the medium, replacing the methanol that was metabolised or had evaporated in the previous 24 hours.

The *Pichia* samples were spun down to separate the cells from the growth medium. If the secretion of the shAlDH product is successful, a 52 kDa band should be present in the lane corresponding to the supernatant (culture medium) but if the signal processing is not occurring the shAlDH protein should be present in the lane corresponding to the pellet (*Pichia* cells).

A Western blot was undertaken to see if a low level of expression had occurred. Western blots are capable of detecting a lower amount of shAlDH than the Coomassie staining of the SDS-PAGE gel. While the positive control (stock hAlDH) was clearly visible on the western blot, no band was visible in either the supernatant or pellet lanes.

4.6.2 Large-scale Cultures.

We decided to repeat the previous expression experiment using a 100 mL culture in a 1 L baffled flask to improve aeration. This time the 9B culture was tested; the 9B clone had a strong positive PCR product and grew on the highest concentration of G418 antibiotic, showing that it had the potential for the shAlDH to be expressed.

A 500 mL BMGY culture of the 9B clone was grown up for 24 hours at 28°C, by which time the A_{600} was 1.7. The cells were harvested by centrifugation and the cell pellet was resuspended into 100 mL BMMY and placed in a 1 L baffled flask. A 500 μL sample was taken (Day 0) and the culture was placed at 28 °C on a fast shaker. Samples (500 μL) were taken at intervals of 24 hours for three days, at which time 100 % methanol was added to give a final methanol concentration of 0.5 %.

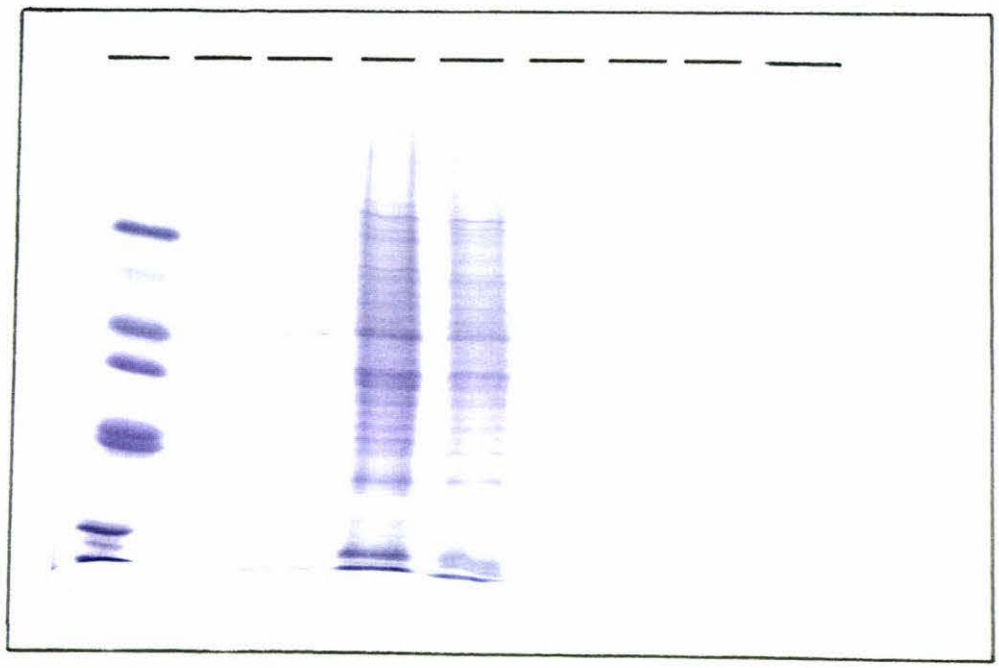

Figure 4-8: SDS-PAGE of samples taken from throughout the expression of the *Pichia* clone 9B.

Lane M: molecular markers (SDS7, Sigma); Lane 1: sample of pure AlDH; Lanes 2-4, cell sample from days 0, 2 and 3 respectively; Lanes 5-8, sample from the medium from days 0, 2 and 3 respectively.

The pellet and supernatant samples were analysed by SDS-PAGE (Figure 4-8) which clearly showed the absence of a band corresponding a protein 52 kDa in size appearing after induction of the *Pichia* culture.

4.7 SUGGESTIONS FOR FUTURE WORK.

The preliminary expression trials described in Section 4.6 did not produce any detectable AlDH.

Laroche *et al.*, (1994) employed a screening technique where large numbers of transformants were grown up and induced in a 96 well tissue culture plate. The transformants secreting the recombinant protein, tick anticoagulant peptide (TAP), were selected after an activity assay was performed on the supernatant from each culture. By using a combination of SDS-PAGE, Western blots and AlDH activity assays, this may be a useful way of identifying the transformants with a high level of secreted expression.

Intracellular expression needs to be explored as an alternative to secreted expression. Intracellular expression may prevent inactivation by oxidation of the active site thiol, which may occur with the secreted enzyme due to the highly aerated medium. Also, the native class 1 AlDH protein is a cytosolic enzyme, and so secreted enzyme may not be processed correctly, resulting in incorrectly folded and/or inactive enzyme. This may compromise the ease of purification found with the secreted product, which could be overcome by the expression of a fusion protein, incorporating the GST tag into the expressed *Pichia*-produced product.

Growth medium composition has a marked effect on expression and the stability of the expressed heterologous enzyme. Proteolysis can be minimised by decreasing the medium pH to 3.0 (Scorer *et al.*, 1993), adding amino acid/peptide supplements such as 1-2 % casamino acids (Clare *et al.*, 1991; Brankamp *et al.*, 1995) or by using a protease-deficient strain of *Pichia* (Clare *et al.*, 1991). Other growth parameters, for example the

culture density at time of induction and the length of induction, also need to be optimised.

It has also been shown that sometimes the mRNA is truncated, due to the action of yeast transcriptional terminators which prematurely terminate the mRNA (Romanos *et al.*, 1992) This has previously been solved by gene synthesis to increase the GC content of the foreign gene (Romanos, 1995). A northern blot could be carried out to identify possible transcription/ translation problems in the expression of the heterologous protein.

The production of very high yields of class 1 AlDH is still feasible using the *Pichia pastoris* expression system, in particular by using intracellular expression. The capacity of this system has not yet been explored to its full potential.

5. SITE DIRECTED MUTAGENESIS.

5.1 INTRODUCTION.

Identification of the residues at the active site and analysis of the mechanism of catalysis of an enzyme have traditionally been achieved through methods such as chemical modification studies and kinetic experiments. In combination with the new technologies available with molecular biology and x-ray crystallography, more information can be gained about the structure and function of an enzyme.

Site directed mutagenesis is a method for exploring structure and function relationships, where the specific substitution of an amino acid residue can impair the activity of the enzyme. This method is more sensitive than chemical modification as any changes are solely due to the substituted side-chain, and not to the addition of chemical groups to particular residues. However, the results of mutagenesis studies must be interpreted with caution, as a change in catalytic efficiency or specificity could be due to many varied events and not completely due to changes at the active site (Leatherbarrow and Fersht, 1986). For example, any one residue may be involved in subunit association/dissociation, stability, cofactor binding or conformational changes (Weiner *et al.*, 1995).

5.1.1 The Significance of Lys-272.

At least three active site amino acid residues are involved in the catalytic mechanism of aldehyde dehydrogenase (Figure 1-2). A nucleophile attacks the carbonyl group of the aldehyde, forming a thiohemiacetal intermediate which is then oxidised. A general base (histidine, lysine or arginine) accepts a proton from the covalently bound, oxidised intermediate in a deacylation reaction. Glutamate-268 has been shown to be an important residue through conservation analyses and site-directed mutagenesis and is speculated to be part of a charge relay system for the removal of a proton from an active site

residue (Weiner *et al.*, 1995). This charge relay system may be part of either the deacylation process or the activation of the active site nucleophile.

A number of site-directed mutagenesis experiments on rat class 2 aldehyde dehydrogenase have been carried out in the laboratory of H. Weiner (Purdue University, Indiana, USA). For example, the cysteine-302 residue is probably the active site nucleophile, as shown through the mutagenesis experiments carried out by Farrés *et al.* (1995), where the cysteine residue was altered to serine, a weak nucleophile, and alanine. The C302A enzyme lacked catalytic activity. The C302S enzyme retained some activity, although the k_{cat}/K_m was lower by a factor of 7×10^5 (Farrés *et al.*, 1995). Cys-302 is also strictly conserved through all of the available aldehyde dehydrogenase sequences, strongly supporting its role in the active site.

Confirmation of the presence of the putative general base has proved more elusive. The candidates include histidines, lysines and arginines. There are no strictly conserved histidines and none of the highly conserved histidines are essential parts of the catalytic process, as shown by mutagenesis experiments (Zheng and Weiner, 1993). There is one strictly conserved arginine (residue 84) and one strictly conserved lysine at position 193 (Hempel *et al.*, 1993). A lysine at position 272, next to two glycines, is part of the conserved sequence (conserved amino acid sequence, with the strictly conserved glycine shown in bold, VTLEL**G**GKSPICV; Hempel *et al.*, 1993). This sequence is only thirty residues upstream from the Cys-302 in the primary sequence and the Lys-272 residue may play the role of proton acceptor at the catalytic site.

We decided to alter Lys-272 to each of the following amino acid residues: alanine, arginine, histidine and leucine. Alanine is a common substitution; it shows the relative importance of the wild-type residue within the catalytic process. Substitution by an alanine removes ionic charge and reduces the size of the sidechain. Arginine and histidine substitutions of a lysine residue will retain the ionic charge yet alter the steric conformation, with the

replacement by a guanidinium group or a imidazole ring respectively. Substitution of a lysine by a leucine removes the ionic charge yet retains most of the bulk of the sidechain.

5.1.2 Methods for Site-Directed Mutagenesis.

The method of mutagenesis previously used in this laboratory was based on the Kunkel method (Kunkel, 1987), which uses single-stranded DNA with a high incorporation of uracil to act as a template (produced using the M13 vector or the pGEM-3 phagemid in *dut⁻ ung⁻* strain of *E. coli*). A mutagenic primer is annealed, *in vitro*, and a complementary strand is synthesised, incorporating the mutagenic region of the primer. Transformation of this plasmid into a *dut⁺ ung⁺* strain of *E. coli* causes the breakdown of the uracil containing strand (by the action of uracil N-glycosidase) resulting in a strong selection against the wild-type, non-mutated strand. This method was found to be largely unsuccessful in this laboratory, due to an inability to produce enough clean single-stranded template.

Therefore, a site-directed mutagenesis kit named 'Unique Site Elimination', marketed by Pharmacia Biotech, was used in attempts to generate the mutations within the DNA sequence of hAlDH. This system has several advantages over the Kunkel method, for example; single-stranded template is not required, the DNA to be altered does not need to be subcloned into a special vector and straight-forward laboratory procedures are employed.

This mutagenesis strategy utilises a two-primer system, where the first primer eliminates a unique restriction site within the recombinant plasmid and the second primer introduces the mutation into the heterologous DNA (Deng and Nickeloff, 1992). The selection and mutagenic primers must anneal to the same strand of the denatured plasmid. These primers are annealed to the single-stranded DNA derived from the denatured plasmid, extended using T4 DNA polymerase and the newly synthesised strand is ligated. The altered plasmid with two regions of mismatch is then transformed into a repair-deficient *E. coli* strain (NM522 Mut^S), and

approximately half of the resulting DNA preparation should contain altered plasmids. Plasmids with the primers incorporated, and the unique restriction site deleted, will be resistant to digestion with that restriction enzyme and so will transform the *E. coli* strain of choice more efficiently. Any parental plasmids without the incorporated primers will be linearised by the restriction endonuclease and will be transformed with an efficiency between 10 and 1000 fold lower than with circular closed plasmids (Conley and Saunders, 1984).

According to the manufacturers, the mutagenesis frequency should be greater than 85 % (USE Mutagenesis Kit Manual revision 5, Pharmacia Biotech, 1996).

5.2 The 'Unique Site Elimination' Method of Mutagenesis

5.2.1 Design of the Selection and Mutagenic Primers.

The primers used to introduce the mutations need to be designed according to the instructions stated in the manual for the USE kit (Pharmacia Biotech). The primers should be 25-35 bases in length with the mismatch in the centre and a minimum GC content of 40 %, 5' phosphorylated, and available at a high purity.

The selection primer designed to specifically anneal over the *Nar*I restriction site within the *lac*I^q gene in the pGEX expression plasmid was supplied by Pharmacia Biotech. The pGEX-hAlDH recombinant plasmid was therefore able to be used for mutagenesis, enabling straight-forward expression and purification of the altered protein without the need for subcloning. This selection primer substituted the *Nar*I restriction site with a *Nhe*I restriction site.

The sequence of this selection primer, with the altered bases in bold and the new *Nhe*I restriction site underlined, is:

5' pGCGTATTGGGCGCTAGCGTGGTTTTTCTTTTCACC 3'

Two primers were designed to anneal over the Lys-272 region of the cDNA, each incorporating a mismatch which would give the desired mutations within the DNA. The codon that translates to a lysine residue is AAG. By altering this to GCG an alanine is introduced, and by altering to CNC one of His, Arg, Leu or Pro will be produced. The first primer (mutK-A) would give a K272A altered protein, and the other primer (mutK-HRL) would give the desired K272H, K272R and K272L substituted proteins, as well as the K272P mutant protein.

The mutagenic primers also annealed the coding strand. The sequences for these primers compared to the wildtype sequence are given, with the altered bases in bold, and the Lys-272 codon underlined:

Wildtype	5'	GGAGCTTGGAGGAAAGAGCCCTTGCATTGTTG 3'
	3'	CCTCGAACCTCCTTTCTCGGGAACGTAACAAC 5'
MutK-A	3'	CCTCGAACCTCCTCGCTCGGGAACGTAACAACp 5'
MutK-HRL	3'	CCTCGAACCTCCTGNGTCGGGAACGTAACAACp 5'

Note that these primers are written 3' to 5', against convention.

5.2.2 Positive Control.

A positive control was provided as part of the USE kit, which consisted of the pUC18 plasmid and a selection primer eliminating the unique *Sca*I site. The second mutagenic primer would alter the DNA encoding the lacZ protein such that the lacZ peptide will be inactivated. After plating onto LB-Amp agar plates coated with 40 μL X-gal (20 mg/mL in dimethylformamide) and 4 μL IPTG (200 mg/mL), transformant colonies carrying the mutation will be colourless while non-mutant colonies will be blue. The positive

control was performed in parallel with the Lys-272 mutagenesis as described in Section 5.2.3, using *Sca*I as the selection endonuclease, and provided an estimate of the frequency of mutagenesis by comparing the numbers of wildtype and mutated colonies.

5.2.3 Methods and Results.

At first the site-directed mutagenesis was carried out essentially as described in the USE manual, Pharmacia Biotech (1996). This protocol is described in this section, and subsequent changes to the protocol are discussed in Section 5.4.

The annealing of the primers to the single-stranded plasmid DNA is achieved by making a 20 µL solution of 1x One-Phor-All buffer PLUS (OPA^+, Pharmacia Biotech), 0.025 pmol plasmid DNA (105 ng of pGEX-hAlDH) and 1.25 pmol of each primer, giving a molar ratio of 50 primer molecules to one plasmid molecule. The plasmid was denatured by boiling in a water bath for five minutes, then immersing in ice for five minutes. Seven µL of nucleotide mix (2.86 mM of each dNTP, 4.34 mM ATP and 1.43x OPA^+, Pharmacia Biotech) and 3 µL of reaction mix (T4 DNA ligase, T4 DNA polymerase, and T4 Gene 32 Protein, Pharmacia Biotech) were added. The extension and ligation reactions were carried out for one hour at 37 °C and the reactions stopped by heating to 85 °C for 15 minutes.

The *Nar*I site will be present in parental, non-altered pGEX-hAlDH, but will be replaced by a *Nhe*I site in mutated plasmids. Any double-stranded plasmids that formed through the reannealing of the parental strands will be linearised by the *Nar*I restriction enzyme, and will be less likely to be transformed into the *E. coli* host. Ten units of *Nar*I (Boehringer Mannheim) was added to the DNA mixture and the endonuclease digestion was carried out for two hours at 37 °C.

Competent *E. coli* NM522 MutS cells were prepared according to the protocol in the USE Mutagenesis Kit manual (Pharmacia Biotech, 1996) and

the transformation of the DNA mixture was performed by combining 10 μL of the DNA mixture with 100 μL of NM522 competent cells as described in Section 2.1.10.

Samples from this transformation were spread onto LB-Amp plates and the rest of the transformation mix was added to a 2 mL LB-Amp liquid medium. The number of colonies counted on the LB-Amp plates, with the proportion of blue to white transformants from the positive control reaction, are presented in Table 5.1.

Table 5.1: The number of colonies on each plate, resulting from the transformation into NM522 after primary digestion.

DNA	Amount of Transformation Mix Plated	% White Colonies	Number of Colonies
Positive Control	100μL	90	112
pGEX-hAlDH	10μL	n.a.	25
	100 μL	n.a.	60

After an overnight incubation, plasmid DNA was prepared from the 2 mL cell culture (Section 2.1.11), giving a mixture of wild-type and mutated plasmids, and the DNA mixture was digested with *Nar*I to linearise the wild-type plasmids. This DNA, with circular mutated plasmids and linear wild-type plasmids, was transformed into competent *E. coli* XL-1 Blue cells (Section 2.1.10) and the cells used to inoculate a 2 mL LB-Amp broth and plated onto LB-Amp plates. The results from this transformation are shown in Table 5.2.

Table 5.2 The number of colonies on each plate, resulting from the transformation into XL-1 Blue after secondary digestion.

DNA	Amount of Transformation Mix Plated	% White Colonies	Number of Colonies
Positive Control	100 µL	100	6
pGEX-hAlDH	10 µL	n.a.	2
	10 µL	n.a.	3
	100 µL	n.a.	7
	100 µL	n.a.	1

Ten of the thirteen colonies that resulted from the transformation of the potentially mutated pGEX-hAlDH plasmid were grown up into 2 mL cultures. The DNA was purified by plasmid miniprep (Section 2.1.11) and a sample from each of the minipreps was digested with *Nar*I and *Nhe*I in separate reactions. DNA with the same sequence as the parental plasmid will be linearised with *Nar*I and uncut with *Nar*I. The reverse is true if the selection primer was incorporated into the plasmid, altering the DNA sequence.

Out of the ten minipreps, seven clearly showed the restriction pattern expected if the mutagenic primer has been incorporated. The DNA from these seven putative mutants was manually sequenced according to the protocol stated in Section 2.1.13. The primer used for the sequencing was:
hAlDH Top 1441 5' GACAAAGTAGCTCACA 3'

This sequencing showed that while the DNA digestion by *Nar*I and *Nhe*I gave the correct pattern for a plasmid mutated at the *Nar*I site, all of the plasmids isolated had the wild-type sequence, rather than a mutation over the Lys-272 region.

5.2.4 Changes Made to the Standard Protocol.

This system has been used extensively in other laboratories around the world with varying success with the protocol suggested by the manufacturers. A posting to the Internet newsgroup 'bionet.molbio.methds-reagnts' brought four responses, with contrasting ideas on what to change and test:

1. Low efficiency reactions occurred with a primer annealing to a coding region containing repeated codons and a high GC content; primers must be phosphorylated and deletions often occur within the mutated sequence, perhaps through homologous recombination. The mutagenesis frequency ranged from 10 % to 100% (M. Brandt, posting 26 August, 1996).
2. Do not freeze a stock of non-competent NM522 cells, but prepare enough competent cells from the fresh cells to last the number of reactions required; try the mutagenesis step at 42 °C rather than 37 °C, and before sequencing the putative mutant colonies, carry out a diagnostic digestion with the appropriate endonuclease(s) (S. Secker 1996, e-mail 20 August, 1996).
3. Use a 400-fold molar excess of each primer; denature the double-stranded template by boiling for ten minutes; carry out the ligase/polymerisation step overnight at room temperature; and concentrate the DNA before each transformation (H. Moss, posting, 18 October 1996)
4. Recommendations from the USE kit manual worked, at a mutagenesis efficiency of 60 % (D. Thomas, posting, 22 October 1996).

As a result of this information, experiments incorporating changes to the standard protocol were carried out. Each of the following was tried in separate experiments:

- increasing the molar ratio of primer to template from a 50-fold excess of primer to a 400-fold excess of primer.
- carrying out the ligase/polymerase step overnight at room temperature.

- following the annealing protocol as used in the Sequenase dideoxy DNA sequencing kit.
- performing a 'tertiary' digestion by repeating the miniprep/digestion step with the 2 mL culture broth of the second transformation (XL-1 Blue), and transforming the DNA into XL-1 Blue again.
- only using the mutagenic primer, not annealing the selection primer, and removing reannealed wildtype plasmid by a digestion before polymerisation. The endonuclease (*Nco*I) will only cut double-stranded parental molecules, and not the single-stranded plasmid-primer molecules, linearising plasmids without the primer annealed.

None of these strategies were successful. Sequencing of the plasmids showed that all had the wildtype sequence at the Lys-272 region of the cDNA.

5.3 Suggestions For Future Work.

While the strategies tested above tried to allow for most of the variables, no conclusion was reached as to why the mutagenesis was not succeeding.

The selection primer had clearly been incorporated into the plasmid, as the screening restriction digestions showed a pattern indicative of the expected mutation. This selection primer was supplied by Pharmacia Biotech and so was known to perform in the expected manner.

The mutagenic primer had been designed as part of this work, manufactured and 5' phosphorylated by Gibco BRL. It is possible (although unlikely) that this primer does not anneal at all, or is annealing to other places on the plasmid. This possibility could be checked by using PCR to amplify a region of the AlDH cDNA, using this primer in combination with other AlDH primers. The products produced could be analysed to estimate the site(s) of annealing, and to eliminate secondary structure as a possible reason for non-annealing.

Only one batch of the MutS NM522 cells had been used for all of these experiments. It is possible that the cells were not functioning correctly as repair-detective cells, and so another batch of competent cells should be made up from a fresh dried cell culture, as provided by Pharmacia Biotech.

Another method of site-directed mutagenesis may prove to be required to generate the mutants required for this research. There has been considerable discussion on the merits of various methods within bionet.molbio.methds-reagnts, with some people finding that different methods are successful for various types of mutagenesis. Exploration into other techniques could prove worthwhile, for example the Altered Sites kit from Promega, Quik-Change (Stratagene), and various PCR-based methods. The Kunkel method of generating mutants may be worth revisiting, as revised protocols may now be available for the efficient production of clean single-stranded template.

6. DISCUSSION

6.1 THE AIM FOR THIS THESIS.

The aim of this research was to investigate different types of expression system in order to develop a standard protocol for the expression and purification of a high yield of active, pure recombinant class 1 aldehyde dehydrogenase. This would enable further work in crystallographic, kinetic and mutagenic studies to be achieved.

6.2 WHY WAS THIS RESEARCH ATTEMPTED

6.2.1 Importance of AlDH as a Research Area

Aldehydes are generated through metabolic processes, for example in the metabolism of proteins, carbohydrates and vitamins, ingested in food, and inhaled in smoke and smog (Lindahl, 1992). Aldehydes are highly reactive molecules due to their carbonyl group which can react with nucleic acids and protein nucleophiles, and are therefore toxic molecules (Schauenstein *et al.*, 1977).

Aldehyde dehydrogenase reduces the likelihood of cytotoxicity, mutagenicity, genotoxicity and carcinogenicity from aldehydes by oxidising the reactive aldehyde to its respective carboxylic acid (Lindahl, 1992). Aldehyde dehydrogenase has a very broad specificity, enabling it to remove most aldehyde species, including aromatic and aliphatic aldehydes. The oxidation is an irreversible reaction and uses NAD^+ as a co-factor.

This enzyme is involved in, for example, the metabolism of ethanol, lipid peroxidation in the cornea and in the oxidation of retinal to the developmentally important retinoic acid. Understanding the functions of aldehyde dehydrogenase could help to explain the clinical manifestations of these and other related biochemical processes in humans.

6.2.2 Progress Made So Far.

As discussed in Section 1.3, research into aldehyde dehydrogenase is at present centred on determining the active site structure of the enzyme. The first X-ray structure of an aldehyde dehydrogenase has only recently been determined (the rat class 3 isozyme, Wang *et al.*, University of Pittsburgh, USA) and as yet the coordinates are unpublished. However, information on the structure and function of aldehyde dehydrogenase has been gathered from kinetic and mutagenic experiments, and chemical modification of the enzyme.

A model for the mechanism of catalysis was put forward in Section 1.3 (Figure 1-2). In this model NAD^+ binds first to the enzyme first, then a nucleophile attacks the carbonyl group of the aldehyde, forming an thiohemiacetal intermediate. Oxidation occurs with the formation of the acyl-enzyme intermediate which is then deacylated, probably with the assistance of a general base. The release of NADH, the rate determining step, happens last.

Cys-302 has been found to be the active site nucleophile through many experiments, but the general base has not yet been identified (Section 1.3). Glu-268 has been implicated in the active site mechanism, perhaps being involved in a charge relay system in the activation of the Cys-302 residue or in the deacylation reaction (Wang and Weiner, 1995).

6.2.3 What Is Needed for Progression In the Area.

A three-dimensional structure of the class 1 aldehyde dehydrogenase protein would provide many answers into how this enzyme acts and the effects of naturally-found mutations on the structure and function. The Class 1 aldehyde dehydrogenase has been crystallised and work is continuing on the determination of a three dimensional structure (Prof. E.N. Baker, pers. comm.). A consistent source of pure aldehyde dehydrogenase is required for the crystallisation studies to be effective.

Stopped-flow kinetic experiments use large quantities of an enzyme and so also require a consistent source of pure and active enzyme.

A recombinant enzyme, produced in a heterologous host such as *E. coli*, is able to be studied at the nucleic acid level, with site-directed mutagenesis enabling the structure-function relationship of the enzyme to be investigated. In addition, when human proteins are involved, autopsy tissue may not be as available as readily as the recombinant enzyme.

For these reasons we decided to investigate methods of expressing and purifying aldehyde dehydrogenase in heterologous expression systems, in the hope that we could produce higher yields of highly active enzyme in a way which was reproducible for wild-type and altered aldehyde dehydrogenase protein.

6.2.4 Approaches Used Previously to Investigate This Area.

As more of the cDNAs for various forms of aldehyde dehydrogenases have been isolated, recombinant technology has been utilised to express the proteins in *E. coli*.

The cDNA coding for the mature rat liver class 2 aldehyde dehydrogenase was cloned into the pT7 vector and the protein expressed in the *E. coli* strain, BL21 (DE3)/plysS, and purified using DEAE-cellulose then Cibachron blue-Sepharose chromatography (Jeng and Weiner, 1991). Subsequently the same method of expression was used, with the purification was carried out by Cibachron blue chromatography followed by affinity chromatography using *p*-hydroxyacetophenone-Sepharose (Ghenbot and Weiner, 1992).

This method, with expression using the T7 system and purification using either an ion exchange step or a Cibachron blue step followed by *p*-hydroxyacetophenone-Sepharose affinity chromatography, has been used for

human class 1 and 2 aldehyde dehydrogenases (Zheng *et al.*, 1993a; Jones *et al.*, 1995), for the conserved-histidine-mutated (Zheng and Weiner, 1993b), serine-74-mutated (Rout and Weiner, 1994) and conserved-cysteine-mutated (Farrés *et al.*, 1995) rat liver class 2 enzyme, the Glu-487-mutated rat and human class 2 enzyme (Farrés *et al.*, 1994) and the Glu-268-mutated human class 2 enzyme (Wang and Weiner, 1995).

6.3 The Approaches I Used To Achieve This Aim.

6.3.1 Purification Using Ion-Exchange/Affinity Chromatography.

The traditional method of expression and purification of aldehyde dehydrogenase, as described above, was tested in order to have a valid comparison with the other expression and purification methods investigated in this work.

The pT7-hAlDH plasmid was transformed into *E. coli* SRP84/pGP1-2 cells and a 1 L culture of these cells was grown up, heat-induced, and lysed by a passage through a French press. The cell lysate was passed through a DEAE-Sepharose column and bound proteins were eluted with a salt gradient. The active fractions were pooled, dialysed, loaded on the *p*-hydroxyacetophenone-Sepharose affinity column and aldehyde dehydrogenase was competitively eluted with free *p*-hydroxyacetophenone. The active fractions were pooled and assays determined a specific activity of 1.2 units/mg of protein, with a total of 3 mg being purified.

This specific activity was more than double than that previously published (Section 3.2.3) and the protein was re-assayed giving the same results. This enzyme has since been crystallised and three-dimensional structure is currently being worked on (Prof. E.N. Baker, pers. comm.).

This purification was therefore very successful giving highly active, pure and crystallisable recombinant human class 1 aldehyde dehydrogenase.

However, a bigger yield would be necessary to carry out the stopped-flow kinetic experiments and for crystallography work.

The mechanism behind the specific binding of the ligand to aldehyde dehydrogenase, presumably occurring at the active site, is unknown (Section 1.4.1). Therefore, it is possible that an aldehyde dehydrogenase with an altered active site, through site-directed mutagenesis, would not bind to the *p*-hydroxyacetophenone-Sepharose matrix. Consequently, purification of the altered enzyme may not be obtained with this method.

For these reasons, we decided to investigate another method of expression and purification, for which the purification did not depend on the affinity of the aldehyde dehydrogenase itself to the matrix, but instead depended on the affinity of a protein 'tag' (GST) for a specific matrix (glutathione-Sepharose).

6.3.2 GST Fusion.

The GST fusion protein expression system was chosen for reasons discussed in Section 1.4.2 and Chapter 3 and this expression and purification strategy had also been successfully used for other applications within our laboratory.

Within this work many expression and purification trials were held to establish the best method for the production of a large quantity of active and pure class 1 aldehyde dehydrogenase.

The best expression of the GST-hAlDH fusion protein was obtained by growing up a 2 L culture of *E. coli* XL-1/ pGEX-hAlDH, inducing at an A_{600} of 0.6 with 1 mM IPTG and lysing the cells by sonication (Section 3.4). Approximately 20-35 % of the expressed AlDH was insoluble, perhaps having formed inclusion bodies.

Two different methods for the purification of the recombinant enzyme were tested. The first was the thrombin cleavage of the fusion protein while bound

to the glutathione-Sepharose affinity matrix, with the pure aldehyde dehydrogenase being eluted in a single step. The second method involved the competitive elution of fusion protein, then thrombin cleavage in solution.

This first method of on-the-column cleavage (bed volume 0.5 mL) yielded 2.6 mg/L of aldehyde dehydrogenase with a high specific activity of 0.39 units/mg (Section 3.6.2).

Cleavage-on-the-column purification worked well, with as good a yield as with the T7 expression and DEAE/*p*-hydroxyacetophenone purification. The specific activity was comparable to other preparations using the traditional method, but not as high as the specific activity gained using the traditional method in this work, as discussed in Section 3.2.3. The purification using glutathione elution, then cleavage in solution was comparable in terms of yield and activity gained, but is not as efficient with respect to length of time and monetary cost.

The cleaved hAlDH was not eluting efficiently from the glutathione matrix, but a large proportion was remaining bound, lowering the yield of hAlDH. Three possible reasons for the binding of the hAlDH to the glutathione matrix are discussed in Section 3.6.4. Effects of the tetramerisation of the hAlDH and ionic interactions have been discounted, but hydrophobic interactions between hAlDH and matrix may be hindering the elution of hAlDH. This could be explored by the utilisation of a different type of GST affinity matrix, as kindly donated by Dr Alan Clark (School of Biological Sciences, Victoria University of Wellington, NZ).

On the basis of these results, I recommend that the purification protocol of the hAlDH using the GST fusion protein system should involve the loading of the sonicate onto the glutathione-Sepharose column, washing off unbound proteins, loading the thrombin, incubation with thrombin at room temperature for an hour and then elution of the cleaved hAlDH.

In order to scale up to purify larger amounts of hAlDH, I would suggest using a separate 2 mL glutathione-Sepharose column for each 1 L of bacterial culture, with a secondary purification of AlDH from the flow-through from each of these columns.

The GST fusion protein system (Pharmacia Biotech) has produced pure and active protein in a single step affinity chromatography, and does not rely on the integrity of the AlDH active site for the purification. Problems are not anticipated with the purification of enzymes with an altered active site using the GST fusion protein system.

6.3.3 *Pichia pastoris.*

We also decided to test the *Pichia pastoris* expression system, as discussed in Section 1.4.3 and Chapter 4. The methylotrophic yeast, *Pichia pastoris*, has been shown to give very high yields of heterologous proteins, in the range of grams per litre, which far exceeds other expression systems. This high yield would be extremely useful in crystallographic studies and in stopped-flow kinetic experiments. The secretion of the heterologous protein into the culture medium, through the use of a secretion signal, can act as a very efficient purification step, as the yeast secretes few of its own proteins.

In the course of this work I have successfully cloned the sheep class 1 aldehyde dehydrogenase into two different vectors (Invitrogen), enabling expression of AlDH once integrated into the *Pichia pastoris* genome. The sheep form was used as it is more extensively characterised than the human form, allowing for direct comparison between the native liver enzyme and the recombinant *Pichia*-produced enzyme, to validate the use of the *Pichia*-produced enzyme as a model for the action of the enzyme *in vivo*.

The two vectors used were pPIC9K and pPIC3K. The first allows the secretion of the enzyme through the use of a secretion signal, the *S. cerevisiae* α-MF prepro leader sequence, which is removed in the Golgi apparatus (Cregg and Higgins, 1995). Secretion of the expressed protein has

been shown to act effectively as a purification step, due to the low proportion of naturally secreted protein (Sberna *et al.*, 1996). The vector that gives intracellular expression, pPIC3K, was also used. Intracellular expression is an attractive alternative because of the potential for proteolysis of secreted proteins and inefficient secretion of complex proteins (Romanos, 1995). Also, AlDH is likely to be susceptible to inactivation by oxidation of the active site thiol, making the intracellular expression more likely to produce active AlDH.

The protocol for transformation by electroporation of the vector into the *Pichia* cells was established (Section 4.4.1) and the continued isolation of more transformants after electroporation will be desirable in order to identify a shAlDH-expressing clone.

Two preliminary expression trials were performed with four *Pichia*/pPIC9K-hAlDH transformants that were identified as having correct integration (Section 4.5.2) and high multi-copy integration (Section 4.5.1). The first expression trial involved small 10 mL cultures of three transformants, grown in BMMY medium over three days. Analysis by SDS-PAGE of daily samples of each culture did not show any shAlDH being produced. A Western blot of the same samples, a more sensitive detection technique, also showed no positive result. As aeration of the induced culture has been found very important in the expression of the heterologous protein, it was decided to grow the fourth culture in a larger 100 mL volume of BMMY in a 1 L baffled flask. Again, analysis showed that expression of shAlDH was undetectable.

In Section 4.7 some suggestions are made in order to achieve expression of AlDH using *Pichia pastoris*. These include:

- A screening technique in which large numbers of the *Pichia*/pPIC9K-hAlDH transformants are grown in a 96 well microtitre plate has been used by Laroche *et al.* (1994). By using a combination of SDS-PAGE, Western blots and AlDH activity assays, this may be a useful way of

identifying the transformants with a high level of expression and secretion of AlDH.

- Intracellular expression may prove to be more successful than secreted expression, especially because of potential inactivation of secreted AlDH in the aerated environment. Also, the native class 1 AlDH protein is a cytosolic enzyme, and so secreted enzyme may not be processed correctly, resulting in incorrectly folded and/or inactive enzyme. Incorporating the GST tag into the expressed *Pichia*-produced product will enable easier purification of AlDH.
- Culture conditions such as the composition and pH of growth medium, culture density, length of induction and aeration can greatly affect the results of the expression trials.
- A northern blot of the RNA produced by *Pichia* cells after induction could be carried out to identify possible transcription/translation problems in the expression of the heterologous protein.

In this work, the shAlDH was successfully cloned into the expression vectors, the plasmid transformed into *Pichia pastoris*, and the screening protocols established. As explained above, the expression trials carried out as part of this work are preliminary only; there are many parameters yet to be explored in order to produce efficient expression of AlDH.

The expression of heterologous proteins in *Pichia pastoris* has been shown to successful in many cases, giving yields in the grams per litre range. However, *Pichia* is not guaranteed to be successful in all cases. The Invitrogen manual (1995, p. 50) states that "based on available data, there is approximately a 50 to 75 % chance of expressing protein of interest".

6.4 Site-Directed Mutagenesis of Lys-272.

Site-directed mutagenesis has been a tool used for the determination of active site residues and the investigation into the structure-function relationships of specific residues within the protein (Section 1.3.4). Various

residues of class 1 and 2 aldehyde dehydrogenases have been altered and any change in the activities of the modified enzyme was determined (Section 1.3.4).

Lysine-272 was selected as the subject for a site-directed mutagenesis as it is a candidate for the role of general base at the active site, which accepts a proton in the deacylation of the acyl-enzyme intermediate (Section 1.3).

As the Kunkel method ((Kunkel, 1987) had been tried in this laboratory with varied success, a new method of site-directed mutagenesis was employed. The Unique Site Elimination kit (Pharmacia Biotech) uses a two primer system; one primer eliminates a unique restriction site within the plasmid (selection primer) and the mutagenic primer introduces the desired mutation into the coding sequence of the gene, in this case the class 1 hAlDH cDNA. The selection for the mutated plasmid occurs by the digestion of the plasmid mix with the restriction endonuclease whose site was eliminated. Wild-type plasmids will be linearised and so will transform the *E. coli* host less efficiently.

The advantages of this USE method of mutagenesis are claimed to be: the expression plasmid can be used without the need for subcloning; a claimed mutagenesis frequency of up to 85 %; and the use of standard protocols.

Site-directed mutagenesis was carried out using the pGEX-hAlDH as a template. This would enable the straight-forward purification of the altered protein by way of the GST fusion protein purification, as described in Chapter 3.

After several attempts I was unable to isolate any plasmids with the desired mutation over the Lys-272 region, although several gave positive results after the diagnostic restriction digestion. Various alterations to the protocol were tested after helpful suggestions were received from people subscribing to the 'bionet.molbio.methds-reagnts' newsgroup over the Internet (Section 5.4). None of the revised methods met with success; all plasmids isolated had the wild-type sequence.

This method for generating mutations was unsuccessful and I do not recommend it. In future it may be worth revisiting the Kunkel method, which is now supplied in kit form (Muta-gene, Bio-Rad), perhaps in combination with colony hybridisation screening for the desired mutations (Farrés *et al.*, 1995). In addition, there are new/improved methods for the generation of mutant proteins, for example PCR-based techniques, the Altered Sites system from Promega and Quikchange (Stratagene).

6.5 SUMMARY.

The results presented in this thesis show that the expression and purification of class 1 aldehyde dehydrogenase has been achieved in two different *E. coli*-based strategies, and the groundwork has been established for the expression and purification using the methylotrophic yeast *Pichia pastoris*.

The T7 system of expression in *E. coli*, followed by ion-exchange and *p*-hydroxyacetophenone-affinity chromatography was successful, producing a yield of 3 mg/L with a specific activity of 1.2 units/mg.

The production of the GST-hAlDH fusion protein in *E. coli* was achieved and the conditions of expression and methods required to purify active hAlDH were investigated. The cleavage of the GST-hAlDH while bound to the glutathione-Sepharose was found to be the best method of purification, with 2.6 mg/L of homogenous, active hAlDH (specific activity 0.39 units/mg) produced by single-column chromatography in one day. Suggestions for improvements are presented (Section 3.7).

The potential for high yields of class 1 AlDH from *Pichia pastoris* has not yet been realised, but the cloning of the AlDH cDNA into two *Pichia* vectors has been achieved and transformation and screening techniques have been established. Suggestions for the expression of the AlDH cDNA from the *Pichia* transformants have been made in Section 4.7.

A method for site-directed mutagenesis (USE) was also tested as part of this work, but no mutants were generated.

REFERENCES

Abriola, D.P., MacKerell, A.D. and Pietruszko, R. (1990) Correlation of loss of activity of human aldehyde dehydrogenase with reaction of bromoacetophenone with glutamic acid-268 and cysteine-302 residues. *Journal of Biochemistry* **266**, 179-187.

Blackwell, L.F., Buckley, P.D. & MacGibbbon, A.K.H. (1989) Aldehyde dehydrogenases - kinetic characterisation, in "*Human Metabolism of Alcohol*" (Crow, K.E. & Batt, R.D., Eds), Vol II, pp. 89-104, CRC Press, Boca Raton, Florida.

Blatter, E.E., Abriola, D.P. & Pietruszko, R (1992) Aldehyde dehydrogenase: covalent intermediate in aldehyde dehydrogenation and ester hydrolysis. *Journal of Biochemistry* **282**, 353-360.

Bradford, M.M. (1976) A rapid and sensitive method for quantitation of microgram quantities of protein utilising the principle of protein-dye binding. *Analytical Biochemistry* **72,** 248-254.

Brankamp, R.G., Sreekrishna, K., Smith, P.L., Blankenship, D.T. and Cardin, A.D. (1995) Expression of a synthetic gene encoding the anticoagulant-antimetastatic protein ghilanten by the methylotrophic yeast *Pichia pastoris. Protein Expression and Purification* **6**, 813-820.

Clare, J.J., Rayment, F.B., Ballantine, S.P., Sreekrishna, K. and Romanos, M.A. (1991) High-level expression of tetanus toxin C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. *Bio/Technology* **9**, 455-460.

Conley, E.C. and Saunders, J.R. (1984) Recombination dependant recircularization of linearized pBR322 plasmid DNA following transformation of *Escherichia coli*. *Molecular and General Genetics* **194**, 211-218.

Cregg, J.M., Madden, K.R., Barringer, K.J., Thill, G.P. and Stillman, C.A. (1989) Functional characterisation of the two alcohol oxidase genes from the yeast *Pichia pastoris*. *Molecular and Cellular Biology* **9**, 1316-1323.

Cregg, J.M. and Higgins, D.R. (1995) Production of foreign proteins in the yeast *Pichia pastoris*. *Canadian Journal of Botany* **73**, S891-897.

Cregg, J.M., Vedvick, T.S. and Raschke, W.C. (1993) Recent advances in the expression of foreign genes in Pichia pastoris. *Bio Technology* **11**, 905-910.

Deng, W.P. and Nickeloff, J.A. (1992) Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. *Analytical Biochemistry* **200**, 81-88.

Despreaux, C.A. and Manning, R.F. (1993) The *dacA* gene of *Bacillus stearothermophilus* coding for D-alanine carboxypeptidase: cloning, structure and expression in *Escherichia coli* and *Pichia pastoris*. *Gene* **131**, 35-41.

Dickenson, F.M. (1985) Studies on the mechanism of sheep cytosolic aldehyde dehydrogenase. *Journal of Biochemistry* **225**, 159-165.

Duester, G. (1996) Involvement of alcohol dehydrogenase, short-chain dehydrogenase/reductase, aldehyde dehydrogenase, and cytochrome P450 in the control of retinoid signaling by activation of retinoic acid synthesis. *Biochemistry* **35**, 12221-12227.

Duester, G., Shean, M.L., McBride, M.S. and Stewart, M.J. (1991) Retinoic acid response element in the human alcohol dehydrogenase gene ADH3: implications for regulation of retinoic acid synthesis. *Molecular Cell Biology* **11**, 1638-1646.

Farrés, J., Wang, T.T.Y., Cunningham, S.J. and Weiner, H. (1995) Investigation of the active site cysteine residue of rat liver mitochondrial aldehyde dehydrogenase by site-directed mutagenesis. *Biochemistry* **34**, 2592-2598.

Ghenbot, G. and Weiner, H. (1992) Purification of liver aldehyde dehydrogenase by *p*-hydroxyacetophenone-Sepherose affinity matrix and the coelution of chloramphenicol acetyl transferase from the same matrix with recombinantly expressed aldehyde dehydrogenase. *Protein Expression and Purification* **3**, 470-478.

Greenfield, N.J. and Pietruszko, R. (1977) Two aldehyde dehydrogenases from human liver: Isolation via affinity chromatography and charaterization of the isoenzymes. *Biochimica et Biophysica Acta* **483,** 35-45.

GST Gene Fusion System (1996) Version 2, 2nd ed., Pharmacia Biotech, New Jersey, USA.

Habig, W.H., Pabst, M.J. and Jakoby, W.B. (1974) Glutathione S-transferases: The first enzymatic step in mercaptic acid formation. *Journal of Biological Chemistry* **249**, 7130-7139.

Hempel, J., Pietruszko, R., Fietzek, P. and Jörnvall, H. (1982a) Identification of a segment containing a reactive cysteine residue in human liver cytoplasmic aldehyde dehydrogenase (isoenzyme E_1). *Biochemistry* **21**, 6834-6838.

Hempel, J., Reed, D.M. and Pietruszko, R. (1982b) Human aldehyde dehydrogenases: Improved purification procedure and comparison of homogenous isoenzymes E_1 and E_2. *Alcoholism: Clinical and Experimental Research* **6**, 417-425.

Hempel, J., Harper, K. and Lindahl, R. (1989) Inducible (class 3) aldehyde dehydrogenase from rat heptocellular carcinoma and 2,3,7,8-tetrachloridibenzo-*p*-dioxin-treated liver: distant relationship to the class 1 and 2 enzyme from mammalian liver cytosol/mitochondria. *Biochemistry* **28**, 1160-1167.

Hempel, J., Nicholas, H. and Lindahl, R. (1993) Aldehyde dehydrogenases: widespread structural and functional diversity within a shared framework. *Protein Science* **2**, 1890-1900.

Hill, J.P., Blackwell, L.F., Buckley, P.D. and Motion, R.L (1991) Steady-state and pre-steady-state kinetics of propionaldehyde oxidation by sheep liver cytosolic aldehyde dehydrogenase at pH 5.2. Evidence that the release of NADH remains rate-limiting in the enzyme mechanism at acid pH values. *Biochemistry* **30**, 1390-1394.

Hill, J.P., Motion, R.L., Buckley, P.D. and Blackwell, L.F. (1994) The effect of *p*-(chloromercuri)benzoate modification of cytosolic

aldehyde dehydrogenase from sheep liver. Evidence for a second aldehyde binding site. *Archives of Biochemistry and Biophysics* **310**, 256-263.

Hsu, L.C., Chang, W.-C. and Yoshida, A. (1989) Genomic structure of the human cytosolic aldehyde dehydrogenase gene. *Genomics* **5**, 857-865.

Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of *Esherichia coli* with plasmids. *Gene* **96**, 23-28.

Jeng, J. and Weiner, H. (1991) Purification and characterisation of catalytically active precursor of rat liver mitochondrial aldehyde dehydrogenase expressed in *Escherichia coli*. *Archives of Biochemistry and Biophysics* **259**, 214-222.

Johnson, K.S., Harrison, G.B.L., Lightowlers, M.W., O'Hoy, K.L., Vinton, J.G., Heath, D.D. and Rickard, M.D. (1989) Vaccination against ovine cysticercosis using a defined antigen. *Nature* **338**, 585-587.

Jones, K.M., Kitson, T.M., Kitson, K.E., Hardman, M.J. and Tweedie, J.W. (1995) Human Class 1 aldehyde dehydrogenase: Expression and mutagenesis, in "*Enzymology and Molecular Biology of Carbomyl Metabolism 5* " (Weiner, H. Ed.), pp 17-23, Plenum Press, New York.

Kaelin, W.G., Pallas, D.C., DeCaprio, J.A., Kaye, F.J. and Livingston, D.M. (1991) Identification of cellular proteins that can interact specifically with the T/E1A-binding region of the retinoblastoma gene product. *Cell* **64**, 521-532.

Kaelin, W.G., Krek, W., Sellers, W.R., DeCaprio, J.A., Ajchenbaum, F., Fuchs, C.S., Chittenden, T., Li,Y., Farnham, P.J., Blanar, M.A., Livingston, D.M. and Flemington, E.K. (1992) Expression cloning of a cDNA encoding a retinoblastoma-binding protein with E2F-like properties. *Cell* **70**, 351-364.

Kitson, T.M. (1989) Reactions of aldehyde dehydrogenase with disulfiram and related compounds, in "*Human Metabolism of Alcohol*" (Crow, K.E. & Batt, R.D, Eds), Vol II, pp. 117-145, CRC Press, Boca Raton, Florida.

Kitson, T.M. and Kitson, K.E. (1995) Probing the active site of cytoplasmic aldehyde dehydrogenase with a chromophoric reporter group. *Journal of Biochemistry* **300**, 25-30.

Knowles, J.R. (1987) Tinkering with enzymes: What are we learning? *Science* **236**, 1252-1258.

Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) Rapid and efficient mutagenesis without phenotypic selection. *Methods of Enzymology* **154**, 367-382.

Laroche, Y., Storme, V., De Muetter, J., Messens, J. and Lauwereys, M. (1994) High-level secretion and very efficient isotopic labelling of tick anticoagulant peptide (TAP) expressed in the methylotrophic yeast *Pichia pastoris*. *Bio/Technology* **12**, 1119-1124.

Lindahl, R. (1992) Aldehyde dehydrogenases and their role in carcinogenesis. *Critical Reviews in Biochemistry and Molecular Biology* **27**, 283-335.

Lindahl, R. and Hempel, J. (1991) Aldehyde dehydrogenases: what can be learned from a baker's dozen sequences? in "*Enzymology & Molecular Biology of Carbonyl Metabolism 3*" (Weiner, H., Wermuth, B. & Crabb, D.W., Eds.), pp. 1-8, Plenum Press, New York.

Loomes, K.M., Midwinter, G.G., Blackwell, L.F. and Buckley, P.D. (1990) Evidence for reactivity with *trans*-4-(*N*,*N*-dimethylamino) cinna-maldehyde during oxidation by the cytoplasmic aldehyde dehydrogenase from sheep liver. *Biochemistry* **29**, 2070-2075.

MacGibbon, A.K.H., Buckley, P.D. and Blackwell, L.F. (1977) Evidence for two-step binding of reduced nicotinamide-adenine dinucleotide to aldehyde dehydrogenase. *Journal of Biochemistry* **165**, 455-462.

MacGibbon, A.K.H., Motion, R.L., Crow, K.E., Buckley, P.D. and Blackwell, L.F. (1979) Purification and properties of sheep-liver aldehyde dehydrogenases. *European Journal of Biochemistry* **96**, 585-595.

Mangelsdorf, D.J. and Evans, R.M. (1995) The RXR heterodimers and orphan receptors. *Cell* **83**, 841-850.

Pichia Expression Kit: A Manual of Methods for Expression of Recombinant Proteins in *Pichia pastoris*. (1995) Version 3.0. Invitrogen Corporation, San Diego, USA.

Pope, B. and Kent, H.M. (1996) Multiplex co-amplification of 24 retinoblastoma gene exons after pre-amplification by long distance PCR. *Nucleic Acids Research* **86**, 536-537.

Romanos, M. (1995) Advances in the use of *Pichia pastoris* for high-level gene expression. *Current Opinion in BioTechnology* **6**, 527-533.

Romanos, M., Scorer, C.A., and Clare, J.J. (1992) Foreign gene expression in yeast: a review. *Yeast* **8**, 423-488.

Rout, U.K. and Weiner, H. (1994) Involvement of serine 74 in the enzyme-coenzyme interaction of rat liver mitochondrial aldehyde dehydrogenase. *Biochemistry* **33**, 8955-8961.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989*) Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, Cold Spring Harbor, NY.

Sberna, G., Cappai, R., Henry, A. and Small, D.H. (1996) Advantages of the methylotrophic yeast *Pichia pastoris* for high-level expression and purification of heterologous proteins. *Bio Technology* **6**, 82-87.

Schauenstein, E., Esterbauer, H. and Zollner, H. (1977) *Aldehyde in Biological systems: their Natural Occurrence and Biological Activities*, Pion, London.

Scorer, C.A., Buckholz, R.G., Clare, J.J. and Romanos, M.A. (1993) The intracellular production and secretion of HIV envelope protein in the methylotropic yeast *Pichia pastoris*. *Gene* **136**, 111-119.

Scorer, C.A., Clare, J.J., McCombie, W.R., Romanos, M.A. and Sreekrishna, K. (1994) Rapid transformation using G418 of high copy number transformants of *Pichia pastoris* for high-level foreign gene expression. *Bio/Technology* **12**, 181-184.

Smith, D.B. and Johnson, K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67**, 31-40.

Staynor, C.K. and Tweedie, J.W. (1995) Cloning and characterisation of the cDNA for sheep liver cytosolic aldehyde dehydrogenase, "*Enzymology and Molecular Biology of Carbomyl Metabolism 5*" (Weiner, H. Ed.), pp 61-66, Plenum Press, New York.

Tabor, S. and Richardson, C.C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Science USA* **82**, 1074-1078.

Toye, B., Zhong, G., Peeling, R. and Brunham, R.C. (1990) Immunological charaterization of a cloned fragment containing the species-specific epitope from the major outer membrane of *Chlamydia trachomatis*. *Infection and Immunity* **58**, 3903-3913.

Vallari R.C and Pietrusko, R. (1981) Kinetic mechanism of the human cytoplasmic aldehyde dehydrogenase E_1. *Archives of Biochemistry and Biophysics.* **212**, 9-19.

von Bahr-Lindstrom, H., Jeck, R., Woenckhaus, C., Sohn, S., Hempel, J. and Jornvall, H. (1985) Characterization of the coenzyme binding site of liver aldehyde dehydrogenase: differential reactivity of coenzyme analogues. *Biochemistry* **24**, 5847-5851.

Wang, X. and Weiner, H. (1995) Involvement of glutamate 268 in the active site of human liver mitochondrial (class 2) aldehyde dehydrogenase as probed by site-directed mutagenesis. *Biochemistry* **34**, 237-243.

Weiner, H., Farrés, J., Rout, U.J., Wang, X. and Zheng, C.-F. (1995) Site directed mutagenesis to probe for active site components of liver mitochondrial aldehyde dehydrogenase, in "*Enzymology and Molecular Biology of Carbomyl Metabolism 5*" (Weiner, H. Ed.), pp 1-7, Plenum Press, New York.

White, C.E., Kempi, N.M. and Komives, E.A. (1994) Expression of highly disulfide-bonded proteins in *Pichia pastoris*. *Current Biology: Structure* **2**, 1003-1005.

Yin, S.-J., Wang, M.-F., Han, C.-L. and Wang, S.-L. (1995) Substrate binding pocket structure of human aldehyde dehydrogenase, in "*Enzymology and Molecular Biology of Carbomyl Metabolism 5*" (Weiner, H. Ed.), pp 9-16, Plenum Press, New York.

Zheng, C.-F., Wang, T.T.Y. and Weiner, H. (1993a) Cloning and expression of the full-length cDNAs encoding human liver class 1 and class 2 aldehyde dehydrogenase. *Alcoholism: Clinical and Experimental Research* **17**, 828-831.

Zheng, C.-F. and Weiner, H. (1993b) Role of the highly conserved histidine residues in rat liver mitochondrial aldehyde dehydrogenase as studied by site-directed mutagenesis. *Archives of Biochemistry and Biophysics* **305**, 460-466.