Cloning, sequencing and expression of glucose dehydrogenase from Thermoplasma acidophilum

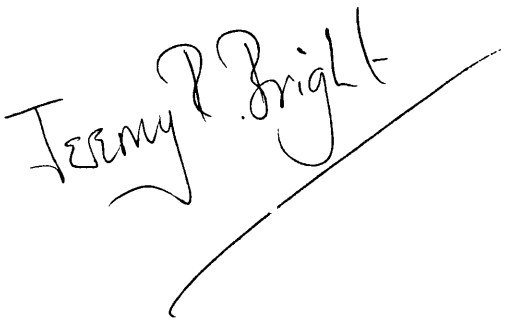
Bright, Jeremy R.

Award date:
1991

Awarding institution:
University of Bath

CLONING, SEQUENCING AND EXPRESSION OF GLUCOSE DEHYDROGENASE FROM *Thermoplasma acidophilum*

Submitted by Jeremy R. Bright
for the degree of Ph.D. of the
University of Bath, 1991



Jeremy R. Bright

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

CONTENTS

ABBREVIATIONS

A_{260}	absorbance at 260 nm
ADP	adenosine diphosphate
ATP	adenosine triphosphate
BSA	bovine serum albumin
CIAP	calf intestinal alkaline phosphatase
DMF	dimethyl formamide
dNTP	deoxynucleotide triphosphate
dsDNA	double-stranded deoxyribonucleic acid
DYT	double yeast tryptone
EDTA	ethylenediamine tetraacetic acid
FPLC	fast protein liquid chromatography
HCl	hydrochloric acid
IPTG	isopropylthiogalactoside
kb	kilobase pairs
KCl	potassium chloride
LB	Luria-Bertani (growth medium)
Mb	megabase pairs
M_r	relative molecular mass
NAD^+	nicotinamide adenine dinucleotide (oxidised form)
NADH	nicotinamide adenine dinucleotide (reduced form)
$NADP^+$	nicotinamide adenine dinucleotide phosphate (oxidised form)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced form)
NaOH	sodium hydroxide
NBRF	National Biomedical Research Foundation
O.R.F.	open reading frame
PQQ	pyrollo-quinoline quinone
r-RNA	ribosomal ribonucleic acid
SDS	sodium dodecyl sulphate
SDS-PAGE	polyacrylamide gel electrophoresis in the presence of SDS
SSC	salt / sodium citrate
ssDNA	single stranded deoxyribonucleic acid
TBE	Tris-borate-EDTA
TE	Tris-EDTA
$t\text{-}RNA^{Met}$	methioninyl transfer ribonucleic acid
X-gal	5-bromo-4-chloro-3-indolyl-β,D-galactopyranoside

SUMMARY

The N-terminal amino acid sequence of the dual-nicotinamide-cofactor specific enzyme, glucose dehydrogenase (EC 1.1.1.47), from the thermoacidophilic archaebacterium *Thermoplasma acidophilum* (Smith *et al.*, 1989) was used to design two redundant 17-mer oligonucleotide probes. Genomic DNA from *Tp.acidophilum* was digested with four different hexanucleotide-recognising restriction enzymes, both singly and in pairs, and the restriction fragments were separated on agarose gels and transferred to hybridisation membranes. The two probes were radiolabelled, and were used to probe the *Tp.acidophilum* DNA for the glucose dehydrogenase-encoding gene (*gld*); the resulting pattern of hybridisation was used to create a restriction map. From this map, two restriction fragments were identified as being likely to contain the *gld* gene.

One of these, a ≈1.7 kb EcoRI/BamHI fragment, was cloned in pUC19 (construct pTaGDH1) and *Escherichia coli* TG1 cells, its nucleotide sequence determined and the derived N-terminal amino acid sequence shown to be identical to that previously determined (Smith *et al.*, 1989). Computer assisted analysis of the nucleotide sequence failed to reveal the presence of a promoter sequence, any transcription termination sequence or a ribosome-binding site associated with the gene; assay of cell-free extracts failed to reveal any glucose dehydrogenase activity above that of controls. The cloned fragment was subcloned in pUC18 (construct pTaGDH2)/*E.coli* TG1, so that the IPTG-inducible P_{lac} promoter could be used to transcribe the gene and assay of cell-free extracts revealed a 1000-fold enhancement of glucose dehydrogenase activity (equivalent to ≈1% soluble cell protein), indicating that a ribosome-binding site was associated with the gene.

The derived amino acid sequence was used to probe the NBRF database; this failed to pinpoint any sequences with high overall similarity, but did highlight a small segment of high similarity with a PQQ-linked glucose dehydrogenase and a larger region of lower similarity to a number of other dehydrogenases. Closer examination of this region led to the identification of what is almost certainly part of the nicotinamide cofactor-binding domain.

PCR mutagenesis was used to change one nucleotide in a portion of the original cloned fragment which had been tentatively identified as being the native ribosome binding site; this construct (pTaGDH4) generated a level of expression of glucose dehydrogenase equivalent to ≈10 % soluble cell protein. A large-scale purification procedure was devised entailing heat and solvent denaturation of the *E.coli* proteins followed by Mono Q anion exchange chromatography. However, SDS-PAGE analysis of the purified enzyme revealed the presence of two translation products. In an attempt to overcome this, it was recommended that the cloned

fragment from pTaGDH4 be subcloned in pMEX8, a vector lacking the *lacZ* gene, which was thought to be contributing to the formation of the second translation product.

Subsequent to this project, the cloned fragment from pTaGDH4 was subcloned in pMEX8/*E.coli* TG1. This did express a single translation product, which was purified in sufficient quantity to be used in crystallisation trials. Crystals were hexagonal prisms and X-ray diffraction analysis yielded sufficient data to calculate the unit cell dimensions and to determine the packing arrangement of the asymetric unit.

CHAPTER 1 INTRODUCTION

1.1 The Archaebacteria

1.1.1 Classical and Modern Taxonomy

Historically, extant life was classified as either Animal or Plant. However, the discovery of the Bacteria and the subsequent examination and comparison of the three cell types led to the realisation that animal and plant cells were considerably closer to each other than either was to bacterial cells. A re-classification was therefore made, dividing extant life into the *Eukaryotae* and *Prokaryotae*, depending on whether or not a cell had a true nucleus. This quite reasonable system of classification persisted, but it became apparent that some "prokaryotic" cells had features that were either atypical, *eg.* unusual membrane lipids, or typically eukaryotic, *eg.* DNA-associated histone-like proteins.

Taxonomic divisions had always been made on the basis of comparative morphology, physiology or biochemistry of cells. However, a number of research groups felt that gene or polypeptide sequence comparisons would provide more fundamental criteria, allowing the resolution of taxonomic anomalies and generating a true picture of the phylogenetic inter-relationships of living organisms. The advent of improved methods of polynucleotide and polypeptide sequence determination has led to the publication, by many groups, of the details of phylogenetic relationships between organisms, based on such sequence comparisons, but one of the first such publications was based on a comparative study of 16S ribosomal RNA sequences and proposed the system of classification currently in use (Woese & Fox, 1977).

The 16S ribosomal RNA molecule (18S in eukaryotes) is ubiquitous amongst living organisms and is functionally highly conserved. When choosing a molecule for such comparative studies, these two criteria must be satisfied in order that valid phylogenetic comparisons can be made (Woese, 1987). Another criterion, that makes 16S/18S ribosomal RNA a particularly good phylogenetic marker, is that it contains regions important for function that gather mutations slowly and structural regions that gather mutations more quickly. The former provides good points of comparison for distantly related species whereas the latter enables the assessment of the relatedness of phylogenetically closer organisms.

Woese and Fox demonstrated a degree of relatedness amongst a number of methanogenic bacteria, which distinguished them from the other bacteria (the *Eubacteria* as they are now termed). They went on to propose that this group of methanogenic bacteria were sufficiently different from the *Eubacteria* and

Eukaryotae to merit their classification as members of a third primary evolutionary kingdom. This third kingdom was termed the Archaebacteria because of "The apparent antiquity of the methanogenic phenotype plus the fact that it seems well suited to the type of environment presumed to exist 3 - 4 billion years ago..." (Figure 1.1).

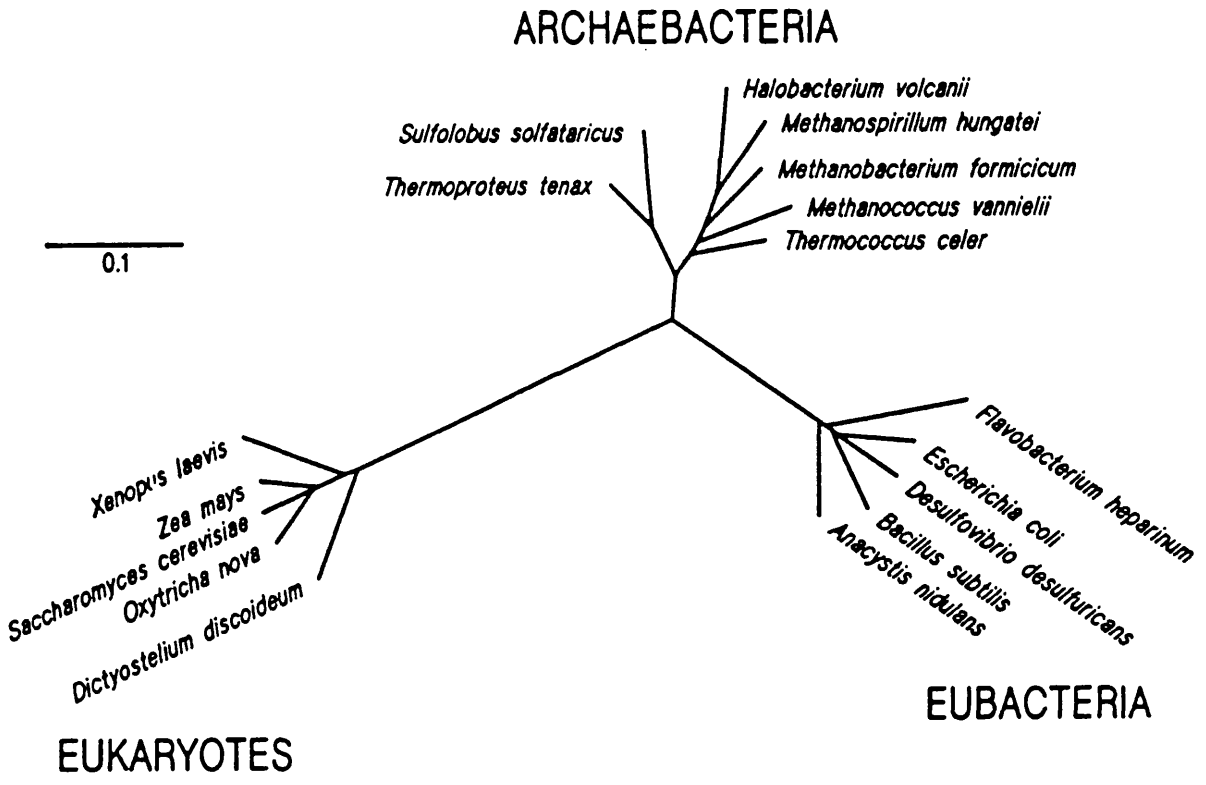


Figure 1.1: Unrooted phylogenetic tree, based on 16S/18S rRNA sequence comparisons. Reproduced from Woese & Olsen (1986).

1.1.2 Three Archaebacterial Phyla

Since 1977, Woese and his associates have investigated archaebacterial phylogeny more extensively and demonstrated that the kingdom is composed of at least three major phenotypic groups: the methanogens, the extreme halophiles and the sulphur-dependent extreme thermophiles (Lane *et al.*, 1985; Woese, 1987).
The Methanogens: as was mentioned earlier, examination of these organisms led to the inception of the *Archaebacteria* as a primary evolutionary kingdom; they are characterised by their ability to reduce carbon dioxide, or simple C1 compounds such as methanol or formate, to methane, and by the unusual coenzymes they are found to possess (De Rosa *et al.*, 1977; Noll *et al.*, 1986; Whitman, 1985).

The Extreme Halophiles: distinguished by their adaptation to extremely saline environments (Larsen, 1973; Schleifer *et al.*, 1982), they have intracellular salt concentrations equating to 5M potassium ions (Kushner, 1985) and therefore contain proteins equipped to cope with such conditions; they are also the only photosynthetic archaebacteria.

The Sulphur-Dependent Extreme Thermophiles: these grow in the temperature range 55°C - 110°C and utilise sulphur as an electron acceptor (Stetter & Zillig, 1985), hence their sulphur-dependence. Members of the genus *Pyrodictium* (the most thermophilic organisms so far isolated - Stetter *et al.*, 1983) grow **optimally** at 110°C and consequently possess some of the most highly thermostable cellular components of all living organisms.

Genera representative of these three phyla are summarised in Table 1.1, along with the environments in which they are found; the unusual nature of these organisms and their cellular components provide tremendous scope for biotechnological application, as well as study, and a number of reviews discuss such applications (Coolbear *et al.*, 1991; Da Costa *et al.*, 1989; Hough & Danson, 1989; Ng & Kenneally, 1986).

Phenotype	Genera	Typical environments		
Halophilic	*Halobacterium* *Halococcus*	30°–40°C,	pH 6–7,	3–5 mol/l NaCl, aerobic
	Natronobacterium *Natronococcus*	30°–40°C,	pH 9–11,	3–5 mol/l NaCl, aerobic
Thermophilic	*Thermoplasma*	50°–60°C,	pH 1–3,	aerobic
	Sulfolobus	60°–90°C,	pH 1–5,	aerobic and anaerobic species
	Acidothermus	85°–93°C,	pH 1–5,	facultatively aerobic
	Thermoproteus	70°–110°C,	pH 2–7,	anaerobic
	Thermofilum	55°–100°C,	pH 3–7,	anaerobic
	Desulfurococcus	80°–100°C,	pH 5–7,	anaerobic
	Thermococcus	75°–90°C,	pH 5–6,	anaerobic
	Thermodiscus	75°–98°C,	pH 5–7,	anaerobic
	Archaeoglobus	70°–92°C,	pH 7,	anaerobic
	Pyrobaculum	74°–102°C,	pH 5–7,	anaerobic
	Pyrococcus	70°–103°C,	pH 5–9,	anaerobic
	Pyrodictium	85°–110°C,	pH 5–7,	anaerobic
Methanogenic	*Methanococcus* *Methanothrix* *Methanosarcina* *Methanolobus* *Methanomicrobium* *Methanogenium* *Methanoplanus* *Methanospirillum* *Methanobacterium* *Methanobrevibacter* *Methanothermus*	All species are anaerobic with pH optima pH 6–8 and most are mesophilic (30°–40°C) except: Extremely thermophilic (>80°C) e.g. *Methanothermus fervidus* *Methanococcus jannaschii* Moderately thermophilic (65°–80°C) e.g. *Methanococcus thermolithotrophilus* *Methanobacterium thermoautotrophicum*		
	Halomethanococcus	30°–40°C,	pH 6–8,	anaerobic, 2–3 mol/l NaCl
	Methanohalophilus	30°–45°C,	pH 9,	anaerobic, 0·2–2 mol/l NaCl

Table 1.1: Typical characteristics of some archaebacterial genera. Reproduced from Hough & Danson (1989).

In addition to these three phyla, three archaebacterial organisms have been isolated, which do not lie phylogenetically where their phenotypes suggest they might. These are *Thermococcus celer*, which branches between the methanogenic cluster and the extreme thermophile cluster, and two *Thermoplasma* species, *Tp.volcanium* and *Tp.acidophilum*, which branch from amongst the methanogenic cluster but remain quite distantly related to them (Figure 1.2); *Tp.acidophilum* is the organism of study in this thesis.

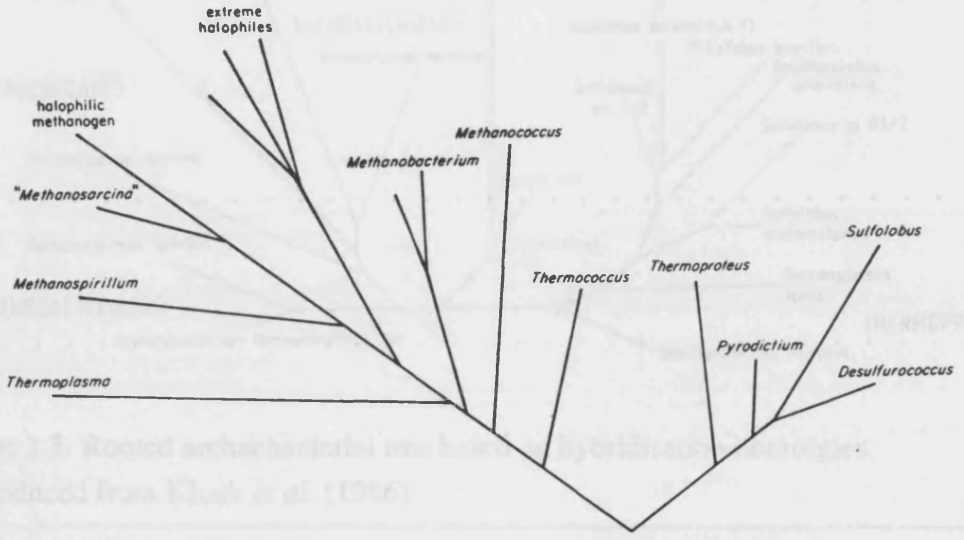


Figure 1.2: Rooted dendrogram showing the phylogenetic relationship of *Thermoplasma* and *Thermococcus* to other archaebacterial genera. Reproduced from Woese (1987).

1.1.3 Current Views and Controversy

As has already been presented, these relationships have been established using 16S ribosomal RNA sequence comparisons, but other comparative studies have demonstrated the same or similar relationships. One study took 17 archaebacterial species and determined their phylogenetic relationships on the basis of the hybridisation homologies displayed by their r-RNA (Klenk *et al.*, 1986). This confirmed the findings of Woese and his associates, but also suggested that the *Thermococcales* represented a third distinct branch of the archaebacteria (Figure 1.3). Other examples of such studies arose during the research performed on *Tp.acidophilum* and are therefore dealt with later (1.2).

The dendrogram shown in Figure 1.2 is the currently accepted view of archaebacterial phylogeny, though Woese recently published a proposal to rename the methanogenic/halophilic cluster and the extreme thermophilic cluster as the

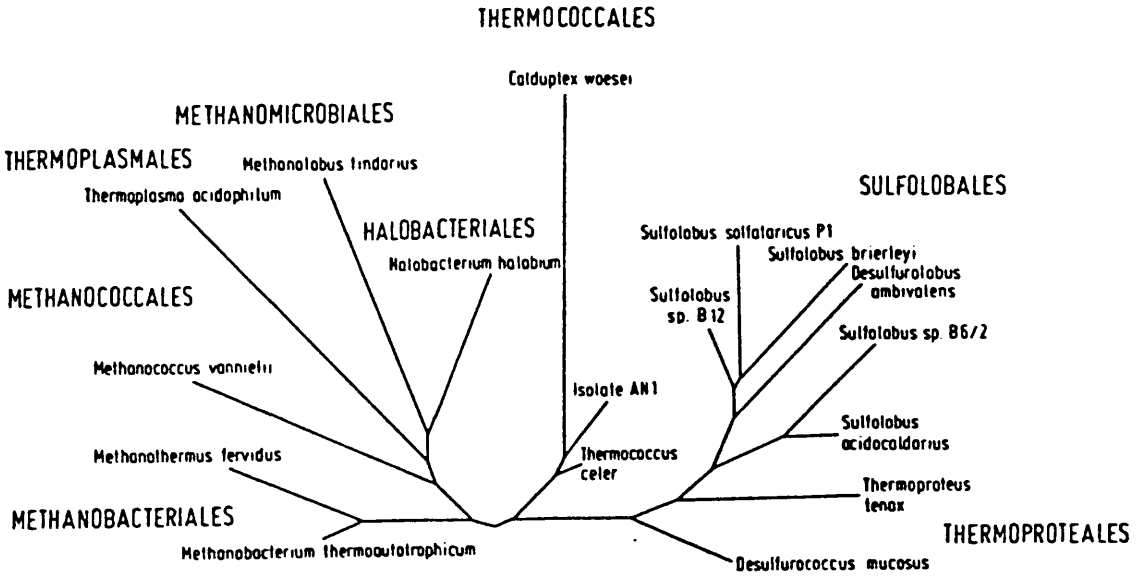


Figure 1.3: Rooted archaebacterial tree based on hybridisation homolgies. Reproduced from Klenk *et al.* (1986)

Euryarchaeota and the *Crenarchaeota*, respectively, and to rename the Archaebacteria the Archaea, in order to emphasise their difference from the Eubacteria (Woese *et al.*,1990).

The most significant opposition to Woese's views on taxonomy has come from Lake and his associates. Lake presented ribosome structural data which suggested that the sulphur-dependent extreme thermophiles should be regarded as a fourth primary evolutionary kingdom, the *Eocyta*, rather than part of the Archaebacteria (Lake *et al.*, 1984), and that the eocytes were more closely related to eukaryotes than archaebacteria. He later substantiated these claims using 16S/18S ribosomal RNA sequence comparisons and went on to say that a fundamental evolutionary divide existed between the eubacterial/archaebacterial cluster and the eukaryotic/eocytic cluster (Lake,1989; Lake,1991). The phylogenetic tree he presented as "an alternative to archaebacterial dogma" is shown in Figure 1.4.

The reason why Lake obtained this different view of phlyogeny from Woese, whilst apparently using the same sequence data, was that he used slightly different regions of the rRNA molecule for his comparisons and he employed different

algorithms to Woese in construction of his phylogenetic tree. Lake claimed that the algorithms used by Woese did not account for the "unequal rate effect". If the

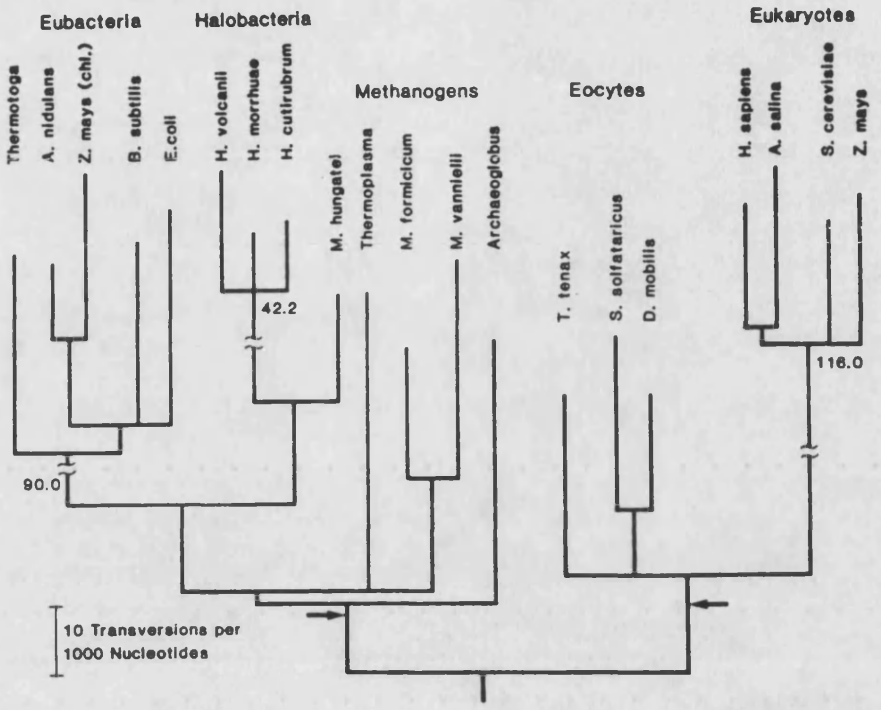


Figure 1.4: Lake's view of phylogeny. Reproduced from Lake (1991).

mutation rates of any given nucleotide are measured within different phylogenetic branches, these rates often differ; Lake claimed that Woese's failure to take account of this effect led to the clustering of unrelated organisms.

Other data, apart from the evidence Lake has presented, would tend to support his view of phylogeny, *eg*. the presence of intronic DNA in the t-RNA genes of *Sulfolobus solfataricus* (Kaine, 1987), an erstwhile uniquely eukaryotic feature. Lake and Woese have openly criticised each other's views (Lake, 1989; Olsen & Woese, 1989), have both highlighted anomalies in the other's arguments and undoubtedly this rather acrimonious debate will continue, even though, to outsiders, it seems to be an exercise in pusillanimity. More positively, the phylogenetic proposals and subsequent debate have encouraged a tremendous amount of research on these organisms and clearly the debate will not be resolved without more evidence.

Having discussed the archaebacteria in general terms, it would seem appropriate to review some of the biochemistry and molecular biology of *Tp.acidophilum*, because this not only provides a relevant introduction to the

organism of study, but also provides specific examples of the sort of taxonomic contradictions which led to the postulation of the existence of the archaebacteria as a kingdom phylogenetically distinct from the eubacteria and eukaryotes.

1.2 *Thermoplasma acidophilum*

1.2.1 "I'm sure it's a Prokaryote"

This organism was first isolated by Gary Darland and Thomas Brock from a coal refuse pile at the Friar Tuck mine in southwestern Indiana (Darland *et al.*, 1970). The temperature of the pile was 56°C, and a sample of it, mixed with an equal mass of water, had a pH of ≈2. Darland managed to culture the organism by incubating a modified version of Allan's basal salts medium (Allan, 1959) - [adjusted to pH 2 and supplemented with yeast extract and glucose], inoculated with samples of the pile, at 56°C. Experiments to define the organism's optimal growth conditions demonstrated growth in the temperature range 45°C - 62°C, with the optimum at 57°C, and in the pH range 1.0 - 3.5, with the optimum lying between pH 1 and 2. For a number of years, these man-made refuse piles remained the only site from which the organism could be isolated, which began to raise the suspicion that *Tp.acidophilum* was something of an evolutionary curiosity. However, this suspicion was dispelled when Segerer *et al.* (1988) reported isolating *Tp.acidophilum* and its close relative *Tp.volcanium* from solfataric fields in the Azores, Iceland, Indonesia, Italy and the United States of America, *ie.* similar, but geographically very diverse, natural environments.

Darland *et al.* (1970) attempted to classify the organism, describing it as a thermoacidophilic prokaryote, which lacked a cell wall, reproduced by budding and displayed other traits consistent with it being one of the *Mycoplasmas*. The first of these traits was the organism's growth response to culture in the presence of various antibiotics; inhibitors of cell wall synthesis, such as vancomycin and ristocetin, had no effect on growth, neither did cycloheximide, the specific inhibitor of protein synthesis in eukaryotic cells, but novobiocin, an antibiotic known to inhibit the growth of mycoplasmas, did prevent its growth. The other feature of the organism which they reported was the very low G+C content of its DNA, another common trait of mycoplasmas, although this was later shown to be an underestimate.

Christiansen *et al.* (1975) and Searcy and Doyle (1975) estimated the G+C composition of *Tp.acidophilum* as 46 %, almost double the original estimate, but still not a feature inconsistent with the organism being a mycoplasma. Also they both calculated the genome size of the organism to be 8 - 10 x 10^8 daltons (≈1.3 - 1.5 Mb),

an extremely small genome for a free living organism; the only other organisms known to have comparable genome sizes are members of the *Acholeplasmataceae*, one of the two *Mycoplasma* families. At this time, therefore, there seemed no reason to question the classification of *Tp.acidophilum* as a thermoacidophilic mycoplasma.

1.2.2 "No it isn't, but what is it?"

One of the observations which called this classification into question stemmed from investigations into the nature of the organism's cell membrane. Belly and Brock (1972) found that, at pH 2, the cells would not lyse even if they were boiled, but they would lyse rapidly at low temperatures if the pH was increased to ≥7.5. This finding suggested a novel membrane structure, which was born out by subsequent biophysical characterisation of the membrane lipid.

Firstly, the polar head groups are joined to the membrane lipids by ether rather than ester-linkages (Langworthy *et al.*, 1972), a logical discovery, when the susceptibility of ester bonds to acid hydrolysis is considered; this is a feature found in all archaebacteria. Secondly, the non-polar membrane lipids have a highly unusual structure (Langworthy, 1977). The basic structure is a 2,3-linked diglycerol tetraether, with saturated C_{40} components made from two phytanyl chains linked head to head; this C_{40} component is present in an acyclic and monocyclic form (Figure 1.5) in *Tp.acidophilum* in an approximate ratio of 2:1.

Langworthy concluded that such a structure was consistent with a novel type of membrane assembly, namely a monolayer, which was not something previously identified in other cell types. Similarly, another observation, which called into question the classification of *Tp.acidophilum* as a eubacterium, was the discovery of a sort of amoeboid cytoskeleton, regulated by calcium ions, and composed of proteins resembling eukaryotic actin and myosin (Searcy *et al.*, 1978; Searcy,1986).

At this time, although several taxonomic anomalies had been described, no attempt had been made to reclassify *Tp.acidophilum*. As was discussed earlier, the archaebacteria were proposed as a third distinct kingdom by Woese and Fox (1977) and subsequent to this, there were a number of comparative studies performed to try to pinpoint the phylogenetic relationship of *Tp.acidophilum* to other cells. One of the first entailed an investigation of unusual base modifications within transfer RNA molecules from a variety of archaebacteria. This showed that as a group they contained similar base modifications but some of these were quite distinct from those found in eubacterial and eukaryotic t-RNA molecules (Gupta & Woese, 1980). This was probably the first evidence confirming the suspicion that *Tp.acidophilum* was an archaebacterium not a eubacterium.

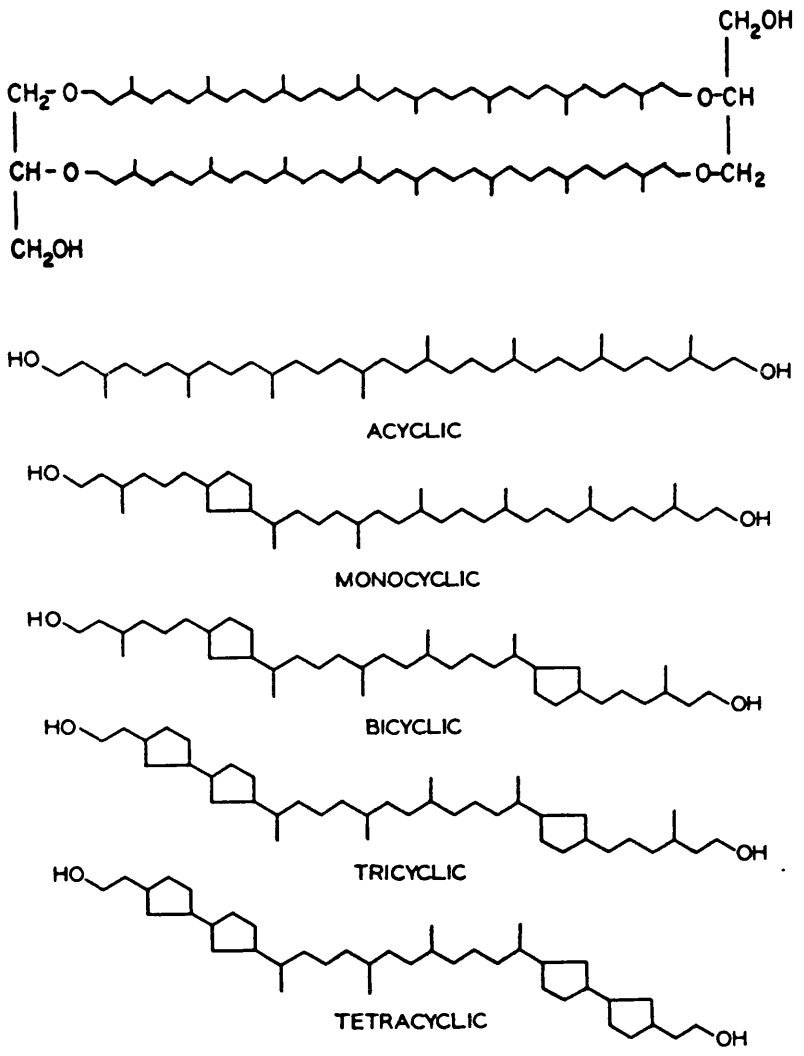


Figure 1.5: The novel non-polar lipids found in archaebacterial membranes. Reproduced from Brock (1974) and Langworthy (1977).

The more familiar approach of using nucleotide sequence comparisons to determine taxonomic relationships was first applied to *Tp.acidophilum* when sequence information became available for the t-RNA^{Met} molecule. Its initial isolation and sequencing (Kilpatrick & Walker, 1981a & b) revealed it to be less than 70% homologous with any known t-RNA^{Met} and that it contained a sequence of three modified nucleotides thought to be unique to archaebacteria (Kilpatrick & Walker, 1982). Work independent of this generated nucleotide sequences for the t-RNA^{Met} molecules from *Sulfolobus acidocaldarius*, *Halococcus marrhuae* and *Tp.acidophilum*, and demonstrated that they all had common unique structural features differing from both eubacterial and eukaryotic t-RNA^{Met} molecules (Kuchino *et al.*, 1982). Finally a comparison of r-RNA and t-RNA sequences from eight

Mycoplasma spp. and from *Tp.acidophilum* confirmed that the latter was not a true mycoplasma (Walker, 1983).

Another body of work which questioned the classification of *Tp.acidophlum* as a eubacterium stemmed from an early observation that *Tp.acidophilum* contained histone-like proteins associated with its DNA (Searcy, 1975). At the time this was regarded as a typically eukaryotic feature, and it was postulated that these proteins were present to prevent the complete dissociation of the DNA under extremes of temperature, and to promote its reassociation when more normal conditions returned (Stein & Searcy, 1978). Determination of the amino acid sequence of the histone-like protein from *Tp.acidophilum* (HTa) and comparison of this sequence with eukaryotic histone amino-acid sequences and with the sequences of two DNA-associated proteins from *E.coli* (HU-1 and HU-2), demonstrated a small but statistically significant similarity of HTa to both the eubacterial and eukaryotic proteins (DeLange *et al.*, 1981) confirming the protein's identity as a form of histone.

Subsequent sequence comparisons with such proteins from *Bacillus stearothermophilus* and *Clostridium pasteurianum* (Kimura & Wilson, 1983; Kimura *et al.*, 1984a) demonstrated that all the eubacterial sequences were more similar to each other than to the *Tp.acidophilum* sequence (Table 1.2) (Kimura *et al.*, 1984a) and this low level of homology between HTa and the eubacterial sequences was consistent with *Tp.acidophilum* not being a eubacterium. However, the study of HTa and DNA-binding proteins from other archaebacteria did not provide corroborative evidence that *Tp.acidophilum* should be grouped with the archaebacteria, because although amino acid sequences have been determined for DNA binding proteins from the archaebacteria *Sulfolobus solfataricus* and *Methanosarcina barkeri*, they are apparently unrelated to amino acid sequences determined for any of the other DNA-binding proteins (Kimura *et al.*, 1984b; Laine *et al.*, 1986).

	Degree of homology				
	C. pasteurianum	*B. stearothermophilus*	*E. coli* NS1	*E. coli* NS2	*T. acidophilum*
C. pasteurianum	–				
B. stearothermophilus	60	–			
E. coli NS1	47	58	–		
E. coli NS2	45	59	69	–	
T. acidophilum	27	32	27	27	–

The percentage of identical amino acids at identical positions is given for each pair of the various DNA-binding proteins

Table 1.2: Sequence homologies displayed between the histone-like protein from *Tp.acidophilum* and DNA-binding proteins from other organisms. Reproduced from Kimura *et al.* (1984a).

1.2.3 "It's definitely an Archaebacterium"

The first study to establish a phylogenetic relationship between *Tp.acidophilum* and other archaebacteria was based on r-RNA sequence comparisons which divided the archaebacteria into two major divisions: the sulphur-dependent extreme thermophile branch and the branch consisting of the methanogens and the halophiles; *Tp.acidophilum* was shown to be peripherally associated with the latter branch (Yang *et al.*, 1985). An alternative method of comparing 16S r-RNA sequence data also drew the conclusion that the archaebacteria could be divided into two major branches and that *Tp.acidophilum* was more closely related to the halophile/methanogen branch (McGill *et al.*, 1986; J.A.Lake, personal communication). More detailed analyses of complete 16S r-RNA sequence data positioned *Tp.acidophilum* and the halophiles in the centre of the cluster of methanogens (Woese & Olsen, 1986).

Apart from the use of sequence comparisons, corroborative evidence for the phylogenetic arrangement of the archaebacteria and the placement of *Tp.acidophilum* within the kingdom has come from several sources: the biophysical and biochemical analyses of ribosomes (Teichner *et al.*, 1986; Cammaranno *et al.*, 1986), an examination of the biochemical requirements for polypeptide synthesis amongst the thermoacidophilic archaebacteria, (Londei *et al.*, 1986) and two studies pursuing in greater depth the original work on antibiotic sensitivity (Sanz *et al.*, 1987; Oliver *et al.*, 1987).

Studies on *Tp.acidophilum* which might provide more fruitful information about its relationship to other organisms, in other words, information about the general molecular genetics of the organism, are only just beginning to appear. Four genes have been sequenced, namely those encoding the 16S r-RNA (Ree *et al.*, 1989) the elongation factor EF-Tu (Tesch & Klink, 1990), citrate synthase (Sutherland *et al.*, 1991) and the large α-subunit of the proteosome (Zwickl *et al.*, 1991) and, in Chapter 7, information from these publications about DNA control sequences is presented along with consensus information from other archaebacteria. Studies on the ribosomal RNA genes of *Tp.acidophilum*, revealed it to be highly unusual in that its three ribosomal genes (encoding 5S, 16S and 23S r-RNA) are present in only single copies (Tu & Zillig, 1982), and are transcribed independently of one another (Ree & Zimmerman, 1990), rather than forming an operon.

The only other organism reported to contain an unlinked r-RNA gene is *Thermococcus celer* (Culham & Nazar, 1988), which, as was stated earlier, is the only other member of an archaebacterial genus which does not group within one of the three established archaebacterial phyla. The currently accepted picture of the phylogenetic relationship of *Tp.acidophilum* to other archaebacteria was shown

earlier (Figure 1.2 - based on comparisons of 16S r-RNA sequences - Woese, 1987), but confirmation of this should become possible by comparison of nucleotide or amino acid sequences from cellular components other than ribosomes.

1.3 Glucose Dehydrogenase from *Thermoplasma acidophilum*

Having discussed at length the archaebacteria in general and *Tp.acidophilum* in particular, it is now important to consider the protein that has been the focus of this study, both from the viewpoint of its rôle in metabolism and from the viewpoint of its structural and functional features which led to the inception of this research project. The following section serves to introduce the topic of glucose metabolism in *Tp.acidophilum* and the rôle of glucose dehydrogenase within that, but for more extensive reviews please refer to Payton and Haddock (1985), Cooper (1986) and, pertaining to archaebacterial cells in particular, Danson (1989).

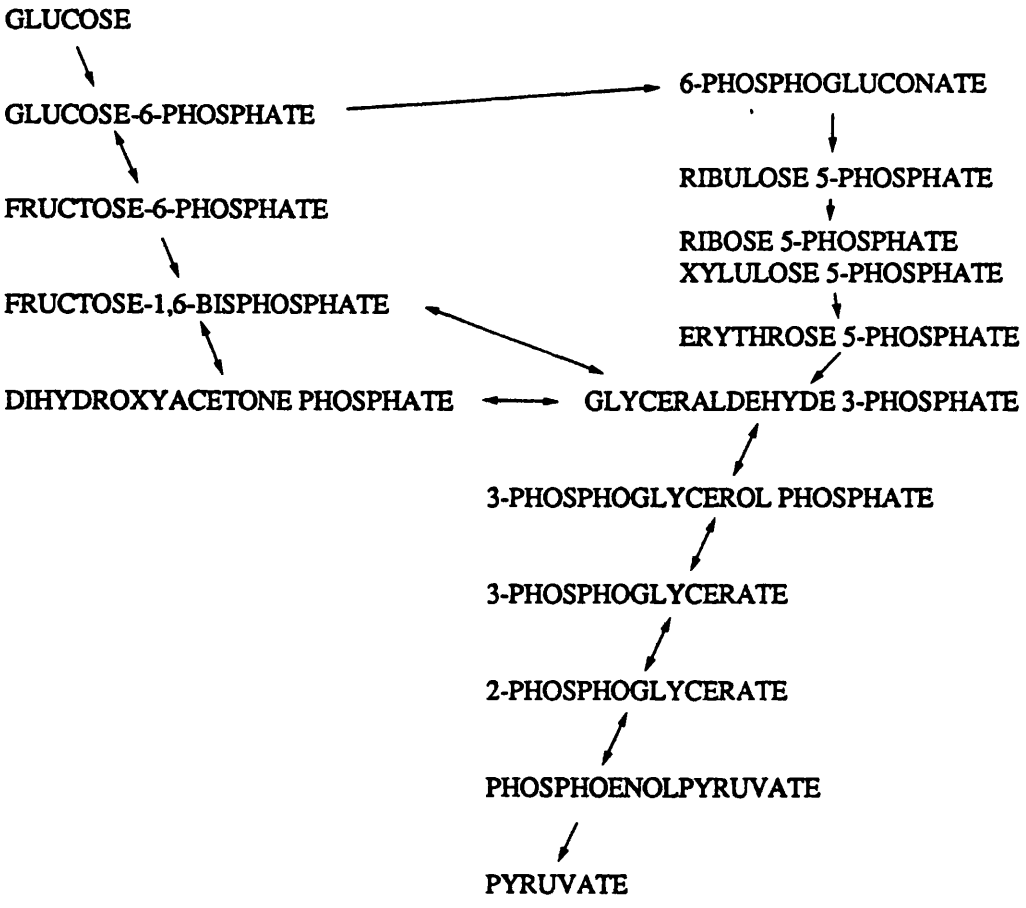


Figure 1.6: The Embden-Meyerhof and pentose phosphate pathways

1.3.1 Metabolism of Glucose to Pyruvate in the Cell

The Embden-Meyerhof and pentose phosphate pathways (Figure 1.6) are characteristic of eukaryotes and a number of anaerobic and facultatively anaerobic eubacteria. However, many strictly aerobic eubacteria lack the enzyme 6-phosphofructokinase and therefore catabolise glucose via the Entner-Doudoroff pathway (Figure 1.7 - Entner & Doudoroff, 1952).

As illustrated in Figure 1.7, the conventional Entner-Doudoroff pathway consists of elements of both the Embden-Meyerhoff and pentose phosphate pathways, but employs the enzymes 6-phosphogluconate dehydratase and 2-keto 3-deoxy 6-phosphogluconate aldolase to perform the reaction that cleaves the hexose intermediate into two triose catabolites. Studies on the archaebacterium *Halobacterium saccharovorum* (Tomlinson *et al.*, 1974) demonstrated that it employed a modified version of the Entner-Doudoroff pathway (Andreesen & Gottschalk, 1969),

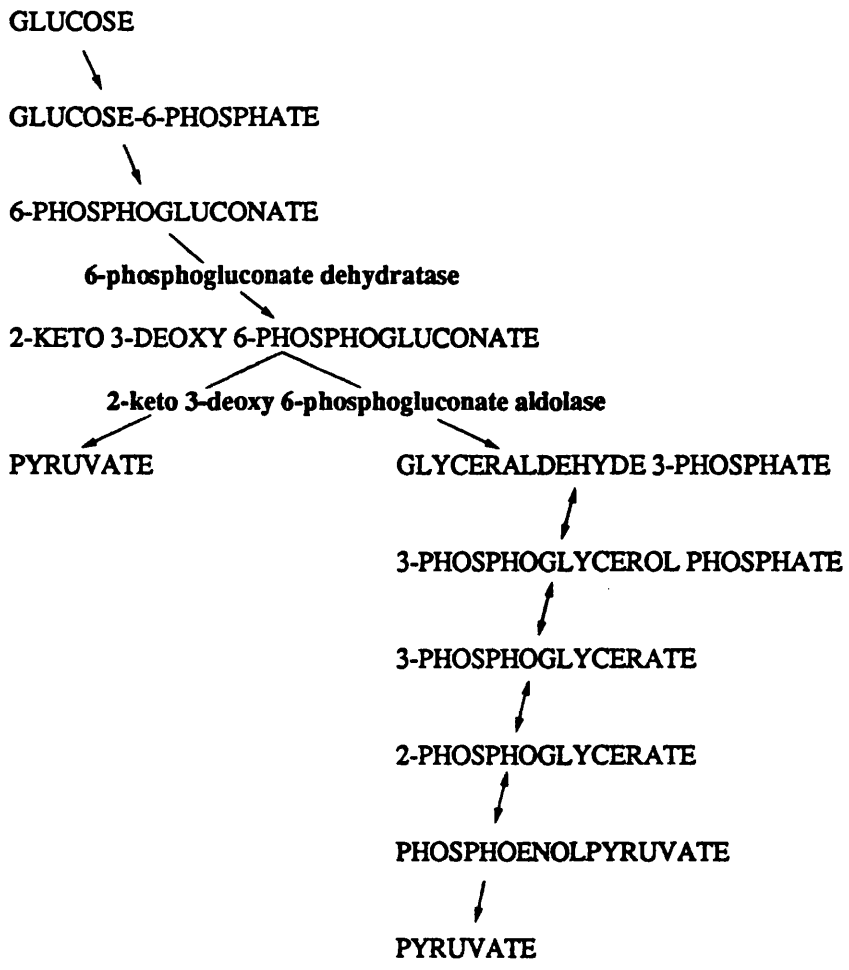


Figure 1.7: The conventional Entner-Doudoroff pathway

which differs from the normal Entner-Doudoroff pathway in that it utilises some non-phosphorylated intermediates (Figure 1.8). It commences with the direct oxidation of glucose to gluconate by glucose dehydrogenase, followed by the elimination of water by gluconate dehydratase, before the catabolite is phosphorylated by the enzyme 2-keto 3-deoxygluconate kinase to form 2-keto 3-deoxy 6-phosphogluconate; thereafter, the pathway proceeds in an identical fashion to the conventional Entner-Doudoroff pathway.

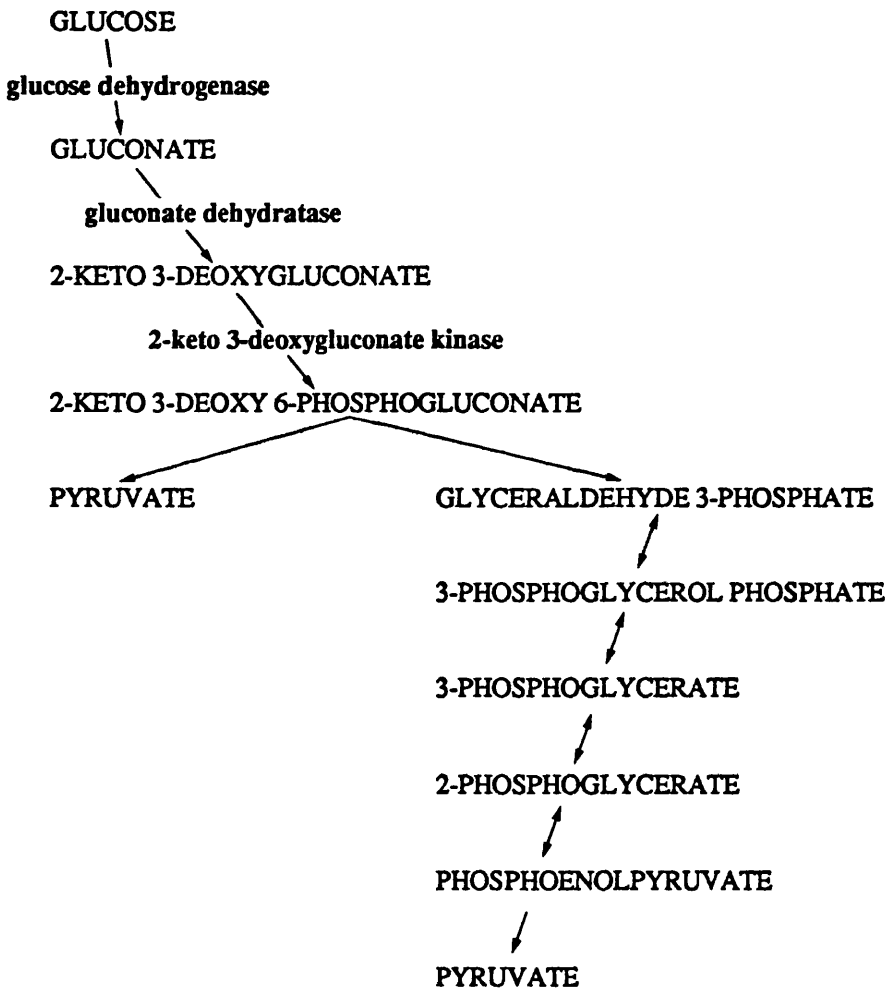


Figure 1.8: The "Modified" Entner-Doudoroff pathway

Examination of glucose metabolism in the thermoacidophile *Sulfolobus solfataricus* (De Rosa *et al.*, 1984) demonstrated another modification of the Entner-Doudoroff pathway, in which the aldol cleavage occurs using a non-phosphorylated intermediate to form pyruvate and glyceraldehyde (Figure 1.9). Studies of glucose metabolism in *Tp.acidophilum* (Budgen & Danson, 1986) revealed the same "Non-

Phosphorylated" pathway as far as the formation of glyceraldehyde. Beyond this point, De Rosa *et al.* (1984) examined the possibility that glyceraldehyde was phosphorylated to form glyceraldehyde-3-phosphate, after which the usual route to pyruvate would be followed. They demonstrated a glyceraldehyde kinase activity but failed to demonstrate glyceraldehyde-3-phosphate production. Budgen and Danson (1986) found evidence of a novel route in *Tp.acidophilum* where an $NADP^+$-dependent glyceraldehyde dehydrogenase oxidised glyceraldehyde directly to glycerate which was then phosphorylated to form 2-phosphoglycerate (Figure 1.9).

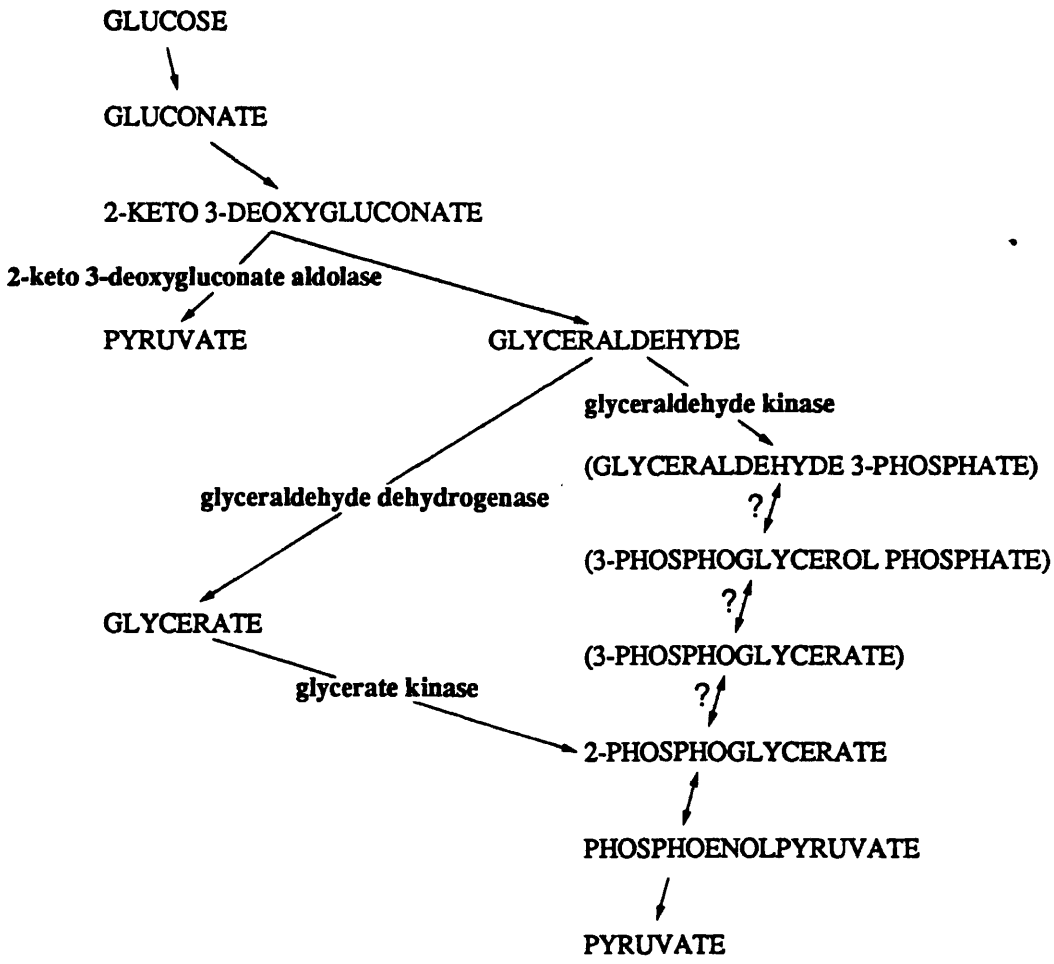


Figure 1.9: The "Non-Phosphorylated" Entner-Doudoroff pathway and the different routes employed by *Tp.acidophilum* and *S.solfataricus*

The remaining archaebacterial phylum, the Methanogens, cannot use glucose as a carbon source but they have been shown to take up and metabolise radiolabelled glucose (Evans *et al.*, 1985, 1986). It was concluded that the normal glycolytic rote was in operation, except that phosphofructokinase has never been found.

Budgen (1988) reported fully the characterisation of the "Non-Phosphorylated" Entner-Doudoroff pathway then went on to report a preliminary characterisation of the first enzyme involved in the pathway, namely glucose dehydrogenase. This analysis, together with a subsequent one by Smith (1989), revealed that the enzyme displayed thermotolerance, significant resistance to denaturation by organic solvents and dual nicotinamide cofactor specificity. The latter feature was also noted about the glucose dehydrogenase from *S.acidocaldarius* (De Rosa *et al.*, 1984) and is the topic under consideration in the next section.

1.3.2 Dual Nicotinamide Cofactor Specificity

Many dehydrogenases which utilise nicotinamide cofactors vary in their specificities for the utilisation of both NAD and NADP. It has been suggested (Levy,1979; Tsai *et al.*, 1989) that nicotinamide-dependent dehydrogenases could be conveniently classsified as belonging to one of five groups. These groups are:
i) obligatory specificity for NAD
ii) preferential specificity for NAD
iii) ambidextrous specificity for NAD(P)
iv) preferential specificity for NADP
v) obligatory specificity for NADP
and the classifications are made on the basis of the activity ratio of NAD utilisation vs. NADP utilisation. If the ratio is greater than 10^3 the enzyme is regarded to have obligatory specificity, if between 10 and 10^3, preferential specificity, and less than 10, ambidextrous specificity. On this basis, glucose dehydrogenase from *Tp.acidophilum* would be said to show preferential specificity for NADP.

The ability to use either NAD or NADP as a cofactor is quite common amongst the dehydrogenases from the archaebacteria and although not unique to the archaebacteria, it is relatively uncommon amongst eubacterial and eukaryotic dehydrogenases. Other archaebacterial dehydrogenases which display this trait are: isocitrate dehydrogenase from *Sulfolobus acidocaldarius* (Danson & Wood, 1984), malate dehydrogenase from *Methanobacterium hungatei* (Sprott *et al.*, 1979), *Tp.acidophilum* and *S.acidocaldarius* (Grossebeuter *et al.*, 1986), glutathione reductase from *S.acidocaldarius* (Smith *et al.*, 1988) and glucose dehydrogenase from *S.solfataricus* (Giardina *et al.*, 1986).

Whichever category a dehydrogenase falls into, it is impressive that the presence or absence of a 2'-phosphate group can determine the ease with which the cofactor binds. The structures of NAD and NADP are essentially identical and their employment as electron carriers is evolutionarily highly conserved (Rossmann *et al.*, 1974) so it is perhaps logical that structural analyses (such as X-ray crystallographic

studies) of nicotinamide cofactor binding sites have revealed several conserved structural features of the nicotinamide cofactor binding site (Rossmann *et al.*, 1974; Ohlsson *et al.*, 1974). The basic structure of the nicotinamide cofactor-binding domain is shown in Figure 1.10; it consists of two halves, each composed of: β-strand - α-helix - β-strand - α-helix - β-strand, the first half of which forms the ADP binding domain and the second half of which forms the nicotinamide binding domain. As has been mentioned, a high level of structural conservation exists between the dinucleotide binding domains of dehydrogenases but the level of sequence conservation within these is low.

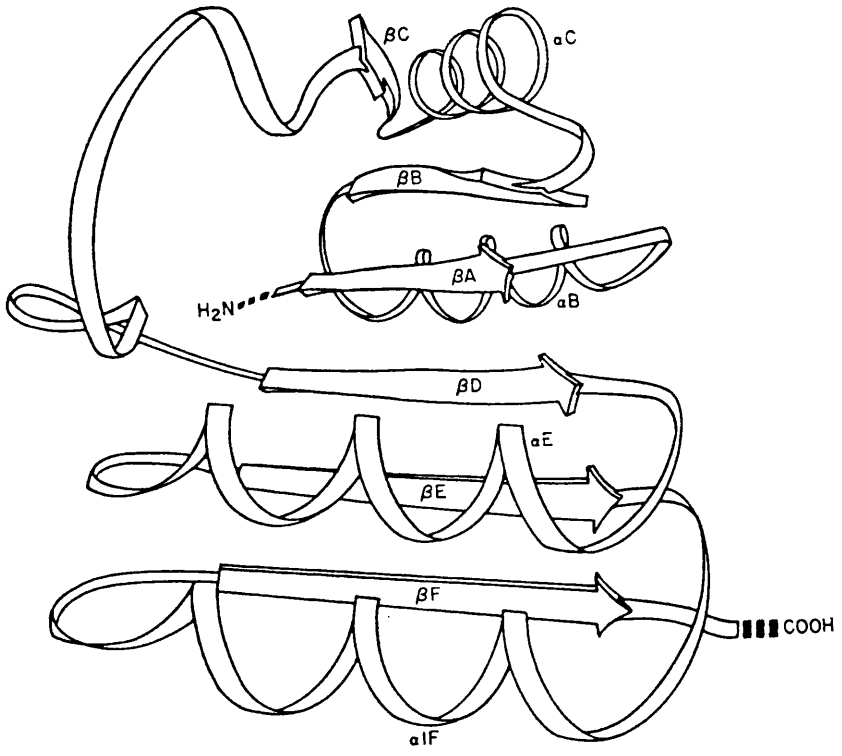


Figure 1.10: Schematic representation of the nicotinamide cofactor binding domains in dehydrogenase enzymes. Reproduced from Ohlsson *et al.*, 1974.

The adenosine moiety of the dinucleotide interacts with a compact fold that sits in a hydrophobic crevice such that the pyrophosphate bridge is located near the N-terminus of helix αB; the nicotinamide ring sits in a cavity which is hydrophobic on one side and hydrophilic on the side of the substrate binding site (Wierenga *et al.*, 1985). In NAD-dependent dehydrogenases, the free 2'-hydroxyl group on the adenosinyl ribose of NAD interacts with an aspartate residue (Adams *et al.*, 1973; Eklund *et al.*, 1974), whereas, in NADP-dependent dehydrogenases, positively charged arginine residues interact with the phosphate group at the 2' position of the

adenosinyl ribose of NADP (Matthews *et al.*, 1979; Rice *et al.*, 1984). These factors serve to explain how dehydrogenases exhibit either NAD or NADP specificity, but they do not explain the observations of cofactor preference or ambidexterity.

Reports exist of enzymes whose nicotinamide cofactor specificity has been altered by mutagenesis. Feeney *et al.*, (1990) altered an aspartate to a serine residue in NADH-linked lactate dehydrogenase from *B.stearothermophilus*. The aspartate residue is thought to interact with the free 2' hydroxyl group mentioned earlier and its replacement with a serine residue was designed to reduce steric hindrance of the 2' phosphate upon binding of NADPH, and to remove the negatively charged side chain from the vicinity. The mutated protein exhibited slightly enhanced catalytic efficiency with NADPH (with respect to the wild type enzyme) but also demonstrated an attenuated catalytic efficiency with NADH, so overall, this mutation achieved a significant shift in the cofactor specificity of the enzyme.

A similar, but more extensive exercise was carried out by Scrutton *et al.* (1990) who generated mutants of the NADP-linked enzyme glutathione reductase, which then exhibited a preference for NADH. Two arginine residues, identified as being those involved in the binding of the 2' phosphate of NADPH were altered to a methionine and a leucine residue. This accomplished the reverse of what Feeney *et al.* (1990) achieved, *ie.* the K_M for NADPH rose substantially and the k_{cat} dropped to about 5% of the wild type value; some enhancement of the catalytic efficiency of the enzyme with NADH was noted at this stage but five further mutations had to be made to swap the enzyme's cofactor preference completely. All of these five mutations resided in a 'fingerprint' region commencing at the end of the first β-strand; this region also contains the few highly conserved amino acid residues in the ADP-binding domain. The essence of these mutations was to remove any positively charged residues in the vicinity of the ADP-binding fold and to substitute one valine residue, previously identified as corresponding to that which should interact with the free 2'-hydroxyl of NAD, with a glutamate.

Both of these reports illustrate a knowledge-based approach to changing the cofactor specificity of dehydrogenases. One further report deals with how an NAD-linked glyceraldehyde-3-phosphate dehydrogenase was mutated to show dual NAD(P) cofactor specificity, using sequence information from an existing dual NAD(P) cofactor specific glyceraldehyde-3-phosphate dehydrogenase (Corbier *et al.*, 1990). The enzyme that was the subject of the mutagenesis and the enzyme structure that provided the information for choosing the mutations, were more than 60 % identical and the majority of residue changes in the cofactor binding domains were conservative. Consequently, the number of possible mutations to be made in that region was quite small at the outset. This provided an ideal scenario to attempt to

pinpoint the basis of dual-cofactor specificity. The obvious differences between the two enzymes were a leucine and a proline residue in the cofactor binding domain of the NAD-specific enzyme and corresponding alanine and serine residues in the dual cofactor-specific enzyme. These mutations were duly made and NADP was found to bind to the mutated enzyme. Interestingly, the wild type dual cofactor specific enzyme contained no positively charged amino acid side chains in the vicinity of the 2'-phosphate-binding region, which Scrutton *et al.* (1990) had to remove to affect NADP-binding. Therefore, the NADP-binding exhibited by the mutant enzyme could only be promoted by the introduction of the smaller side chains of alanine and serine, reducing steric hindrance, and by the fact that the resulting environment was more hydrophilic and therefore more attractive to NADP.

These three excellent reports demonstrate how well characterised some of these systems are becoming; it will be of interest to see if the molecular basis of the dual cofactor specificity exhibited by *Tp.acidophilum* glucose dehydrogenase conforms to these systems or whether, being an archaebacterial enzyme, a very different system is in operation.

1.3.3 Thermal and Solvent Stability of Proteins

As was mentioned at the end of section 1.3.1, the characterisation of glucose dehydrogenase carried out by Budgen (1988), and Smith (1989) revealed its thermotolerant nature and its unusually high resistance to denaturation in organic solvents. The coexistence of such properties in proteins is quite common and seems to be an outcome of similar structural features.

Thermotolerance in proteins has been the focus of study of many research groups for many years and some general rules about what confers this property on a protein have emerged; this is the subject of an excellent review by Klibanov & Ahern (1987). Under "normal" active conditions of temperature, *ie.* 0°C - 60°C, proteins are in constant internal motion; whether this motion is simply vibration or bond rotation, or something more extensive, like the displacement of a whole segment of a protein, is largely temperature-dependent, usually reversible and sufficiently limited so as not to disrupt the overall "native" nature of the protein. As long as the motion of the protein remains within these constraints, the concerted movements implicit in catalysis and regulation remain possible because overall secondary and tertiary structure of the ligand-binding sites remain intact. Paraphrasing, thermodenaturation of a protein requires a certain amount of energy because it has to overcome the non-covalent and covalent interactions that confer ordered secondary and tertiary structure.

As temperature is increased, a point is reached where the weaker non-covalent interactions (Van der Waals forces, hydrophobic interactions and hydrogen bonds) are counteracted by the energy being introduced, resulting in unfolding of the protein; at this point, structures such as ligand-binding sites are inevitably disrupted, interfering with catalytic function, but for many globular proteins this is not an irreversible effect. However, continued temperature increase will cause the protein to be irreversibly inactivated. This "thermoinactivation" of the protein is contributed to by several factors:

(i) as concentrated protein solutions cool, unfolded molecules can aggregate to bury hydrophobic regions revealed by the heating process,

(ii) as dilute protein solutions cool, the proteins may refold without aggregating, but adopt catalytically inactive tertiary conformations that they would be thermodynamically incapable of forming at the normal growth temperature of their native host,

(iii) in mildly acidic solutions "asp - X" and some "X - asp" peptide bonds (where X is any amino acid) become very labile as temperature increases,

(iv) in mildly basic solutions, disulphide bond cleavage can occur, and

(v) at a variety of pH values, increased temperature can cause the deamidation of asparagine and glutamine residues.

Since (iii) - (v) are solvent-mediated effects, substitution of water with a less polar solvent such as glycerol or sorbitol has been effective in overcoming thermal denaturation by these means, ($T_{0.5}$ (inactivation) of enzymes in molten sorbitol, *ie.* ≈130°C, can be upto 1 hour; R.M. Daniel, personal communication), but this does not address the issue of what confers thermostability on a protein.

Naturally-thermotolerant enzymes often exhibit a rigid structure which overcomes the problem of conformational thermodenaturation. This has been revealed by taking highly homologous genes encoding the same enzyme from mesophilic and thermophilic organisms, creating chimaeric genes from different regions of the two and investigating the effect on the thermostability of the gene product. Biro *et al.* (1990) carried out such an exercise with glyceraldehyde-3-phosphate dehydrogenase from the mesophilic archaebacterium *Methanobacteriun bryantii* and from the thermophilic archaebacteria *Methanothermus fervidus* and *Pyrococcus woesii.* They found that a region near the C-terminus of the polypeptide, that lay at the interface of the two domains of the enzyme subunit, was largely responsible for conferring thermostability on the enzyme. Further mutagenesis experiments concluded that it was hydrophobic interactions between this region and neighbouring parts of the protein that conferred a rigidity upon the structure which in turn caused the enzyme's thermostability.

Ionic interactions can also achieve the rigidification of protein structures. For example, comparison of the nucleotide sequences encoding kanamycin nucleotidyl transferase from *B.stearothermophilus* and *Staphylococcus aureus* reveals a single base change resulting in the substitution of a threonine with a lysine residue. Without any significant change in the three-dimensional structure, this effects the introduction of a salt bridge, which dramatically enhances thermostability (Matsumura *et al.*, 1984). Similarly, the engineering of disulphide bonds into proteins has been successful in their stabilisation (Perry & Wetzel, 1984). The chemical changes outlined above that are a result of thermodenaturation can also be overcome by mutagenesis. Rather than changing the solvent conditions, asparagine residues can be substituted with glutamine residues, which are functionally similar but less readily deamidated; alternatively, if it doesn't impair catalytic function, they can be substituted with residues that are merely a similar size, such as isoleucine or threonine (Ahern *et al.*, 1987).

The proposed explanation for the observation that thermotolerant proteins often display resistance to denaturation by organic solvents, bears some similarity to the explanation of protein thermostability in that it is the rigidity of the structure of the protein that prevents it from unfolding and in turn conveys thermal/solvent stability upon it. Denaturants such as guanidinium hydrochloride or urea probably cause protein unfolding because they are much better at solvating a variety of amino acid side-chains than water (Arakawa & Timasheff, 1984). Similarly, ionic interactions within a protein will be destabilised by interaction with non-polar solvents and the hydrophobic portions of the protein will be stabilised by interaction with them, causing the protein to unfold. Within a rigid structure this process is not prevented, but occurs less readily.

1.4 Aims of the Project

1.4.1 Background

As has been discussed in section 1.3, work on glucose dehydrogenase from *Tp.acidophilum* began because it was identified by Budgen & Danson (1986) as being the first enzyme of a "Non-Phosphorylated" Entner-Doudoroff pathway, and was unusual in that it could utilise either NAD or NADP as a cofactor. Preliminary characterisation studies on the purified enzyme (Budgen, 1988) revealed that the enzyme exhibited significant thermal and solvent stability. Smith (1989) investigated these features more extensively and demonstrated that its $T_{0.5}$ (inactivation) was >40 hours at 55°C (the organism's normal growth temperature), that it was 85 % active in

4M urea and that it was 50 % active in 40 % (v/v) methanol (Smith, 1989). This work was performed with a view to assessing the enzyme's suitability for use in an immobilised form, in a biological reactor as part of cofactor-regeneration system.

1.4.2 Cofactor Regeneration

Industrial processes that utilise an oxidoreductase require a supply of that enzyme's cofactor, *eg.* $NAD(P)^+$ or NAD(P)H. As these are reduced or oxidised by the enzyme they need replacing, so a better approach is to include a second enzyme that will oxidise/reduce the cofactor for use by the first enzyme. Additionally, if the second enzyme displays dual cofactor-specificity, it is more versatile in this rôle than enzymes which can only utilise a single cofactor. For the process to be effective, the enzyme must satisfy a number of criteria:

(i) it must be immobilisable on an insoluble matrix so it can be readily separated from the reaction mixture

(ii) it must retain most of its activity when immobilised,

(iii) it must be stable in this immobilised state

(iv) it must utilise a cheap substrate that is converted to a harmless product, neither of which interfere with the process, and

(v) it must be readily obtainable, pure and in large quantities.

These are the salient points for this application, but this is an oversimplification of a problem which is more extensively discussed by Smith (1989) and reviewed by Wang & King (1979). Smith studied the glucose dehydrogenases from both *Tp.acidophilum* and *Sulfolobus acidocaldarius*, and concluded that the enzyme from *Tp.acidophilum* would be the more appropriate for such an application.

1.4.3 Aims of this Project

A major reason for the inception of this project was the desire to produce large quantities of this enzyme, cheaply. Smith investigated growing *Tp.acidophilum* cells in large fermenters, but with little success. Therefore, a logical solution to the problem was to overexpress the enzyme in a recombinant form. The practical aims of this project were consequently quite specific, namely to clone, sequence and express the gene encoding *Tp.acidophilum* glucose dehydrogenase in *E.coli*. Beyond this, the aim was to manipulate the expression system to provide sufficient recombinant protein for possible biotechnological application.

Implicit in this work, was the desire to derive information about the molecular genetics of the organism and simply to discover whether a thermophilic archaebacterial gene and protein could be cloned and expressed in a mesophilic eubacterial host. If the latter was achieved, the logical progression would be to purify

sufficient recombinant protein for structural analysis and then to embark on knowledge-based site-directed mutagenesis studies in order to probe the molecular basis of the enzyme's thermotolerance, solvent stability and dual-cofactor specificty.

CHAPTER 2 MATERIALS

All reagents used were of analytical grade or of the finest grade commercially available.

2.1 Organisms

E.coli TG1 [δ(lac-pro)thi supE[Res$^-$Mod$^-$(k)]F' (traD36proA$^+$B$^+$ lacI^{q}ZδM15)] was originally from A.Bankier, MRC Cambridge and was used throughout as the host for cloning. *Thermoplasma acidophilum* (DSM 1728) was obtained from the Deutsche Sammlung von Mikroorganismen culture collection, Gottingen, West Germany.

2.2 Vectors

Plasmids pUC18 and pUC19 (Yanisch-Perron *et al.*, 1985), pBR322 (Bolivar *et al.*, 1977), M13mp18 and M13mp19 (Norrander *et al.*,1983; Yanisch-Perron *et al.*,1985) were obtained from Northumbria Biologicals Ltd., U.K.

2.3 Enzymes

Restriction endonucleases and DNA modifying enzymes were obtained from Northumbria Biologicals Ltd., U.K. RNase A, proteinase K, lysozyme and calf intestinal alkaline phosphatase were from Boehringer, Mannheim, Germany. Sequenase™ nucleotide sequencing kits were obtained from United States Biochemicals, Cleveland, Ohio, U.S.A.

2.4 Reagents

General reagents and salts for growth media, unless otherwise specified, were from BDH, U.K. or Fisons, U.K.; yeast extract, tryptone and agar were from Difco, Michigan, U.S.A. $NADP^+$ (grade 1) and ATP (grade 1) were from Boehringer, Mannheim, Germany. Phenol was from Rathburn, Walkerburn, Scotland. DNA synthesis reagents were from Glenn Research, Sterling, Virginia, U.S.A., Cruachem, Livingston, Scotland and Rathburn. Agarose (grade II-A) was from Sigma, U.K. and Sequagel™ was from National Diagnostics. Hybond N, [γ-^{32}P]ATP (185 TBq mmol^{-1}; >5000 Ci mmol^{-1}) and deoxyadenosine 5'-[α-^{35}S] thiotriphosphate (37 TBq mmol^{-1};

>1000 Ci mmol^{-1}) were from Amersham International, U.K. Genescreen+ hybridisation membrane and Nensorb 20™ reverse phase columns were from NEN Research Products, Boston, MA, U.S.A. Chromatographic media were from Pharmacia, U.K., or Whatman Ltd., U.K.

2.5 Equipment

Milli-Q water was from the Milli-Q™ reagent-grade water system from Millipore Corp., Bedford, MA, U.S.A. Oligodeoxyribonucleotides were synthesised on an Applied Biosystems 381A DNA Synthesiser; DNA sequencing apparatus was from BRL, Gaithersburg, Maryland, U.S.A., the DNA Thermal Cycler™ was from Perkin-Elmer-Cetus, Instruments Division, Norwalk, CT 06856, U.S.A., Fast Protein Liquid Chromatography (FPLC) equipment was from Pharmacia, Uppsala, Sweden and Micrograd™ protein minigel apparatus was from Gradipore Ltd., Pyrmont, Australia.

2.6 Computer Hardware and Software

Wordprocessing was performed under licence on ICL DRS-M40 personal computers using the Microsoft™ Word 5.0 package. DNA and amino acid sequence information was compiled, stored and analysed on a VigII Viglen personal computer using the Amersham Staden-Plus™ package. DNA and amino acid sequence database searching was performed on a Microvax 3300, using the VMS operating system and the Wisconsin GCG package (Devereux *et al.*, 1984).

CHAPTER 3 METHODS

This chapter has been subdivided as follows: the first section (3.1) details techniques involving the manipulation of microorganisms, the second (3.2) with DNA probe production and techniques for screening DNA, and the third (3.3) with methods of DNA manipulation. Section 3.4 covers methods pertaining to the recombinant protein and the final section (3.5) details any reagents whose composition is not given in the text.

3.1 Manipulation of Microorganisms

3.1.1 Storage and Culture of Escherichia coli TG1

E.coli TG1 cultures were stored on minimal agar at 4°C and subcultured every 4 - 8 weeks; cells were grown by incubation at 37°C in DYT or LB broth, shaking, or on DYT agar or nutrient agar plates.

3.1.2 Competent E.coli TG1 Cells

Competent *E.coli* TG1 cells were prepared using the calcium chloride method described by Maniatis *et al.* (1982).

3.1.3 Plasmid Transformation of Competent E.coli TG1 Cells

Competent cells were transformed with plasmid or recombinant plasmid using the method described by Maniatis *et al.* (1982), but performing the heat-shock treatment at 50°C for 1 min.

3.1.4 Colour Selection of Recombinant Clones

Prior to spreading transformed cells on the nutrient agar/ ampicillin plates, 10 µl of 0.1 M aqueous IPTG and 50 µl of 2% (w/v) X-gal in DMF were spread on each plate.

3.1.5 Storage of Recombinant E.coli TG1 Cells

Recombinant colonies were picked into wells of 96-well titre plates, containing sterile DYT broth + ampicillin (100 mg/l) and incubated at 37°C for 24 h. Sterile glycerol was added, aseptically, to a final concentration of 15% (v/v) and the plates were then sealed with adhesive-backed plastic and stored at -20°C.

3.1.6 M13 R^E Transformation of Competent E.coli TG1 Cells

300 µl of competent cells containing samples of ≈10 ng of M13 replicative form were incubated in wet ice for 30 min then at 50°C for 1 min. IPTG and X-gal (3.1.4) were added to the cells and the mixture added to 3 ml of molten top agarose, standing in a heating block at 45°C. The agarose was mixed by inversion, poured over DYT agar plates, and incubated overnight at 37°C.

3.1.7 Preparation and Storage of Recombinant M13 Phage

5 ml aliquots of DYT broth were inoculated with recombinant M13 phage plaques and incubated for 18 h , shaking, at 37°C. The cells and cell debris were removed from 1 ml samples of each culture by centrifugation at 10,000 x *g* for 5 min. The supernatant was transferred to a sterile 1.5 ml eppendorf tube, 50 µl of chloroform added, and stored in the dark, at room temperature.

3.1.8 Production of Cell-Free Extracts from E.coli Cultures

E.coli cells were harvested from overnight cultures by centrifugation at 4,000 x *g* for 10 min and resuspended in 0.001 volumes of ice-cold 50 mM sodium phosphate buffer pH 7.0. Cells were lysed by sonication on wet ice in 4 x 15 s bursts, leaving 45 s between bursts to permit cooling, the cell debris removed from the sonicated extract by centrifugation at 5,000 x *g* for 10 min and the supernatant removed and stored at 4°C.

The cell pellet was resuspended in half the original volume of buffer, sonicated again, cell debris removed and discarded and the supernatant pooled with the original supernant. Finally, the supernatant pool was clarified by centrifugation at 20,000 x *g* for 20 min and the pellet discarded.

3.1.9 Culture of Tp.acidophilum

Starter culture: Cells were obtained as a freeze-dried pellet and stored at 4°C. 0.5 ml of sterile glucose solution was added to 25 ml of growth medium (3.5.1) and 1 ml of this used to resuspend the dry cells. The remaining medium, in a flat sided bottle, was inoculated with the cell suspension and incubated at 55°C for 5 days with gentle shaking (50 rpm).

Inoculation: 2 litre conical flasks containing 1 litre of sterile growth medium were warmed to 55°C; 20 ml of sterile glucose solution and 3 ml of the starter culture were added, aseptically, per litre of medium and the culture grown, shaking (200 rpm), at 55°C; every 2-3 days, 3 ml of culture were used to inoculate 1 litre of fresh medium.

3.1.10 Production of Cell-Free Extracts from Thermoplasma acidophilum

Tp.acidophilum cells from a 3.5 day culture were harvested by centrifugation at 6,000 x *g* at 4°C for 10 min, resuspended in 0.001 volumes of sterile growth medium and lysed by adding 0.004 volumes of ice-cold 20 x TE and incubating at 0°C for 30 min. Cell debris was pelleted by centrifugation at 10,000 x *g* at 4°C for 15 min and discarded.

3.2 DNA Probes and Screening Methods

3.2.1 Synthesis of Oligodeoxyribonucleotides

Both redundant and single-sequence oligodeoxyribonucleotide probes were made using ß-cyanoethyl phosphoramidites on an Applied Biosystems 381A DNA Synthesiser. Redundant probes were designed such that the total number of different component sequences did not exceed 100.

Probes were synthesised on solid phase supports in 0.2 µmole amounts and eluted by incubation with ammonia solution (sp.gr. 0.880) for 1 h at room temperature; the eluate was subsequently incubated at 55°C for 16 h to remove any remaining protecting groups, then vacuum-dried and stored desiccated at 4°C.

3.2.2 Cleaning of Oligodeoxyribonucleotides

Dry probes were dissolved in 400 µl of sterile 0.4 M sodium chloride and 1 ml of absolute ethanol was added; the sample was then mixed and incubated at -20°C for 90 min. Precipitated probe was harvested by centrifugation at 10,000 x *g* for 5 min at 0°C, washed in ice-cold absolute ethanol, vacuum-dried and dissolved in 400 µl of sterile Milli-Q water; insoluble contaminants were removed by centrifugation at 10,000 x *g* for 2 min at 4°C and the supernatant stored at -20°C.

3.2.3 ^{32}P-Labelling of Oligodeoxyribonucleotides

Probes were 5'-^{32}P-phosphorylated using γ-^{32}P ATP [initial activity ≥5000 Ci/mmol] - (3.3.15). Unincorporated γ-^{32}P ATP and other reactants were removed from the radiolabelled probe by purification on a Nensorb 20™ reverse phase column (3.3.6).

3.2.4 Southern Transfer

DNA fragments in agarose gels were transfered to and bound to Genescreen+ hybridisation membrane according to the manufacturer's instructions (capillary blot

procedure - Southern, 1975); the dry membrane was sealed in a polythene bag and stored at -20°C.

3.2.5 Hybridisation of the Labelled Probe to the Membrane

10 ml of freshly-prepared hybridisation solution were added to the dry membrane in the polythene bag and any air bubbles were removed; the bag was resealed and incubated for 6 hours at 37°C. 5 mg of denatured salmon sperm DNA, the ^{32}P-labelled oligodeoxyribonucleotide and 15 ml of hybridisation solution were added to the membrane, the bag was resealed and incubated shaking at 37°C for 16 h.

3.2.6 Washing and Autoradiography of the Labelled Membrane

Initially, the membrane was washed at room temperature for 5 min in 6 x SSC and then again in 6 x SSC, 1% (w/v) SDS; a β-particle monitor was used to assess the amount of bound radiolabel. The membrane was subsequently washed in 6 x SSC, 1% SDS for 15 min at increasing temperatures until the activity was observed to drop sharply.

The membrane was sealed in a plastic bag, sandwiched between two pieces of pre-flashed X-ray film, placed in an autoradiography cassette lined with intensifying screens and incubated at -70°C for 24 h. One piece of X-ray film was removed and developed, and the cassette replaced at -70°C; the exposure time necessary for the second piece of film was estimated based on the intensity of the image on the first piece of X-ray film.

3.2.7 Colony Hybridisation

Hybond N hybridisation membrane was placed on nutrient agar/ ampicillin plates and a 48-toothed replica plating device, designed to accommodate half a microtitre plate, was used to transfer inocula of the recombinant stock cultures (3.1.5), in a regular 6 x 8 matrix pattern, to the hybridisation membranes, which were then incubated for 24 h at 37°C.

The binding of colony DNA to the membranes and hybridisation with 5'-^{32}P labelled oligodeoxyribonucleotide probes (3.2.3) were performed according to the manufacturer's instructions, except the hybridisation temperature was reduced to 37°C. Washing and autoradiography of the membranes were performed as described (3.2.6).

3.3 General DNA Methods

3.3.1 Determination of DNA Concentration and Purity

DNA concentration in aqueous solutions was determined spectrophotometrically by measurement of absorbance at 260 nm, using the following approximations:

1 absorbance unit ≡ 50 µg ds DNA/ml
1 absorbance unit ≡ 30 µg ss DNA/ml

DNA purity was assessed spectrophotometrically by measurement of absorbance at 260 nm and 280 nm and calculation of the A_{260}/A_{280} ratio; a value ≥1.8 for ds DNA and ≥1.6 for ss DNA were the criteria for purity.

3.3.2 Agarose Gel Electrophoresis

Agarose was dissolved in boiling TBE buffer and allowed to cool to <50°C; ethidium bromide was added to a final concentration of 400 µg/l and the solution used to pour slab gels. The agarose concentration was chosen according to the size of the DNA under examination:

[Agarose] (w/v)	DNA size range (kb)
0.5%	1 - 50
1.0%	0.2 - 20

DNA in sample buffer, at a concentration ≤200 ng DNA/ml, was loaded onto slab gels and separated by electrophoresis in TBE, containing 400 µg ethidium bromide/l, at a constant current of 200 mA for analytical purposes or 50 mA for preparative purposes. HindIII-cut, PstI-cut or uncut lambda DNA were used as molecular weight markers.

3.3.3 Isolation of DNA fragments from Agarose Gels

Freeze-Squeeze procedure was performed as described by Towner (1991a). Electroelution onto DE-81 paper was performed as described by Perball (1988).

3.3.4 Proteinase K Treatment of Nucleic Acids

Nucleic acid preparations were incubated at a concentration of 0.05 mg of proteinase K/mg nucleic acid, for 15 min at 45°C.

3.3.5 RNase A Treatment of Nucleic Acids

Nucleic acid solutions were incubated with DNase-free RNase at a final concentration of 0.05 mg RNase A/mg nucleic acid, at 65°C for 15 min.

3.3.6 Purification of DNA Fragments

Phenol/chloroform extractions were performed as described by Maniatis *et al.* (1982). Nensorb 20™ reverse phase columns were used according to the manufacturer's instructions. Sepharose CL-6B spun columns were used as described by Towner (1991b).

3.3.7 Equilibrium Density Gradient Centrifugation

Caesium chloride equilibrium density gradient centrifugation was performed according to Maniatis *et al.* (1982), omitting the final dialysis step. In place of this step, an equal volume of sterile Milli-Q water was added, followed by absolute ethanol (at room temperature) until the DNA was observed to precipitate. DNA was then harvested by centrifugation at 10,000 x *g* for 10 min at 0°C, washed in ice-cold, 70% (v/v) ethanol, vacuum-dried and dissolved in TE.

3.3.8 Isolation of Nucleic Acids by Spooling

Nucleic acids were precipitated from crude extracts by the gentle addition of 2 volumes of absolute ethanol to the extract, such that an organic/aqueous interface was generated. Precipitated nucleic acids (predominantly large DNA molecules) were harvested from this interface by spooling with a clean glass rod, vacuum-dried and redissolved in the minimum volume of TE.

3.3.9 Precipitation of Nucleic Acids

Nucleic acids were precipitated using ethanol or isopropanol as described by Maniatis *et al.* (1982).

3.3.10 Velocity Gradient Centrifugation

Size fractionation of DNA by centrifugation through sodium chloride gradients was performed according to Kaiser and Murray (1985); no phosphatase treatment was included.

3.3.11 Partial Restriction of DNA

DNA was incubated at 37°C with 0.4U enzyme/μg DNA in the appropriate restriction buffer; samples were taken at 5 min intervals and the reaction terminated by the addition of disodium EDTA (pH 8.0) to a final concentration of 15 mM.

Samples were analysed by electrophoresis in 1% agarose gels (3.3.2) to determine the optimum period of digestion needed to generate DNA fragments of the required size.

3.3.12 Restriction of DNA

DNA was incubated with restriction enzyme(s), at a final concentration of 2U enzyme(s)/μg DNA, at 37°C for 90 min, in an appropriate restriction buffer. If required, the restricted DNA was purified by phenol/chloroform extraction (3.3.6) and ethanol precipitation (3.3.9).

3.3.13 CIAP-Treatment of DNA

DNA was incubated at a concentration of 0.4 μM 5' DNA termini with 0.05U CIAP/pmol of 5' DNA termini, in CIAP buffer, at 37°C for 60 min. The reaction mixture was extracted with phenol/chloroform (3.3.6), the DNA precipitated with ethanol (3.3.9), dried and dissolved in TE.

3.3.14 Ligation of DNA Fragments

For sticky-ended ligations, the two DNA fragments, in equimolar amounts, were incubated together at a final concentration of 1 μM total DNA, with 0.1 U T4 DNA ligase/pmol of sticky ends, in the presence of ligase buffer, for 16 h at 15°C.

3.3.15 Kination of DNA Fragments

5'-dephosphorylated DNA was incubated with ATP, in a molar ratio of 1:3, with 1 U of T4 polynucleotide kinase/pmol of 5'-DNA termini, in kinase buffer for 30 min at 37°C.

3.3.16 Plasmid DNA 'Minipreps'

Plasmid DNA 'Minipreps' were performed according to the boiling method described by Maniatis *et al.* (1982).

3.3.17 Plasmid DNA 'Maxipreps'

Plasmid DNA 'Maxipreps' were performed according to the alkaline lysis method descibed by Maniatis *et al.* (1982).

3.3.18 DNA Sequencing

Single-stranded template (M13mp18/19) and double-stranded template (pUC18/19) DNA sequencing were performed by the dideoxynucleotide chain termination method, using Sequenase™ kits and α-^{35}S-dATP, according to the manufacturer's instructions (Sanger *et al.*, 1977; Messing, 1983).

Sequences were separated by electrophoresis in TBE through denaturing 6% acrylamide Sequagel™ gels. Linear 0.35 mm gels were run at 30 mA constant current and 0.2 - 1.0 mm wedge gels were run at 50 mA constant current. Gels were fixed and the urea eluted by incubation for 30 min (linear gels) or 45 min (wedge gels) in 5% (v/v) acetic acid / 5% (v/v) methanol in distilled water.

Excess fixing solution was drained from them, the gels transferred to Whatman 3MM paper covered with Saran-wrap and vacuum-dried at 80°C for 30 min (linear gels) or 60 min (wedge gels). X-ray film was sandwiched between the dry gel and a glass plate inside an autoradiography casette and incubated at room temperature for 16 h prior to development.

3.3.19 Klenow Fragment Treatment of DNA

DNA was incubated with 100 pmol of each appropriate dNTP, and 0.1 U of Klenow fragment/pmol of sticky ends, in Klenow 'fill-in' buffer at a final concentration of 0.5 mM sticky ends, for 15 min at 25°C.

3.3.20 Nuclease S1 Treatment of DNA

DNA was incubated with 0.1 U of nuclease S1/pmol of sticky ends in nuclease S1 buffer at a final concentration of 0.5 mM sticky ends, for 5 min at 25°C.

3.3.21 T4 DNA Polymerase Treatment of DNA

DNA was incubated with 200 pmol of each dNTP and 0.1 U of T4 DNA polymerase/pmol of sticky ends in T4 DNA polymerase buffer at a final concentration of 0.5 µM sticky ends, for 5 min at 25°C.

3.3.22 Polymerase Chain Amplification of DNA

100 µl reaction mixtures were prepared in AmpliTaq™ buffer containing target DNA at a final concentration of 10^5 - 10^6 molecules/100 µl (≈1 pg DNA/kb of DNA/reaction), amplification primers at a final concentration of 0.5 µM, dNTPs at a final concentration of 0.2 µM each dNTP, AmpliTaq™ Taq DNA polymerase at a final concentration 0.05 U/µl of reaction, and magnesium chloride at a final concentration of 1, 2, 3, 4 and 5 mM (in order to establish optimum Mg^{2+} concentration for the reaction); 100 µl of liquid paraffin were pipetted onto the surface of each reaction mix prior to placing the samples in the Perkin-Elmer-Cetus DNA Thermal Cycler™. Amplification was performed using 30 cycles of the following incubation protocol:

Melting step:	1 min at 94°C
Annealing step:	1.25 min at 37°C
Reaction step:	1 min at 74°C.

3.4 Protein methods

3.4.1 Glucose Dehydrogenase Assay

Glucose dehydrogenase activity was determined spectrophotometrically by following the change in absorbance at 340 nm, corresponding to the reduction of $NADP^+$. The standard 1 ml assay mixture contained 0.4 mM $NADP^+$ and 0.5 M D-glucose in 50 mM sodium phosphate buffer pH 7.0, and was performed at 55°C.

3.4.2 NADPH Oxidase Assay

NADPH oxidase activity was determined spectrophotometrically by following the change in absorbance at 340 nm, corresponding to the oxidation of NADPH. The standard 1 ml assay mixture contained 0.1 mM NADPH in 50mM sodium phosphate buffer pH 7.0, and was performed at 55°C.

3.4.3 Protein Estimation

Protein concentration was determined spectrophotometrically using the method of Bradford (1976), by comparison with a standard curve of absorbance at 595 nm vs. protein concentration over a protein concentration range of 0 - 20 µg bovine serum albumin/ml.

3.4.4 Thermal Denaturation

Samples were incubated at a range of temperatures from 50°C to 75°C for 10 min; after cooling by incubation at -20°C for 10 min any precipitated proteins were removed by centrifugation at 20,000 x g for 5 min at 4°C.

3.4.5 Solvent Denaturation

Samples, in wet ice, were mixed with ice-cold alcohol, incubated at -20°C for 10 min, and any precipitated material removed by centrifugation at 20,000 x *g* for 10 min at 4°C.

3.4.6 FPLC™ Mono Q Anion Exchange Chromatography

Samples containing up to 30 mg of total protein in 50mM sodium phosphate buffer, pH 7.0, were loaded onto the FPLC™ Mono Q column at a flow rate of 1

ml/min; this flow rate was used throughout. The loaded column was washed with 10 ml of 50 mM sodium phosphate buffer pH 7.0 and the bound protein was then eluted using a 30 ml 0 - 200 mM sodium chloride gradient in the same buffer. Any remaining bound protein was eluted with 5 ml of 2 M sodium chloride in buffer and the column washed with 5 ml of buffer prior to loading another sample. All buffers and samples were ultra-filtered prior to use.

3.4.7 SDS-Polyacrylamide Gel Electrophoresis

Discontinuous SDS-polyacrylamide gel electrophoresis was performed according to the method of Laemmli (1970), or using the Flowgen™ Minigel System with ready-made discontinuous Tris-Tricine gels, according to the manufacturers instructions.

Gels run according to the method of Laemmli were fixed and stained overnight at 37°C in Fix solution (aqueous 40 % (v/v) methanol/5 % (v/v) acetic acid), containing 0.1 % (w/v) Coomassie Blue, and destained in Fix Solution. When adequately destained, they were stored in aqueous 12.5 % (v/v) isopropanol/ 5 % (v/v) methanol.

Gels run using the Flowgen™ Minigel sytem were fixed for 1 hour in 10 % (w/v) aqueous trichloroacetic acid, rinsed with distilled water, stained for 1 h in the manufacturer's catalytic stain solution and destained overnight in distilled water. Gels were stored in distilled water.

3.5 Solutions and Media

Solutions were sterilised either by autoclaving at 120°C, 20 psi for 15 min, or by filter sterilisation using millipore 0.22μm syringe filters. All solutions were stored at room temperature in colourless glass or plastic bottles unless otherwise stated.

3.5.1 Growth Media

All media were made using distilled water and were sterilised by autoclaving. 16% (w/v) agar was included for solid phase culture and 0.8% agarose was included to make top agarose for M13 culture.

Thermoplasma growth medium:
Trace element solution (in distilled water):

$FeCl_3.6H_2O$	1.930 g/l	$CuCl_2.2H_2O$	0.005 g/l
$Na_2B_4O_7.10H_2O$	0.450 g/l	$VOSO_4.5H_2O$	0.004 g/l
$MnCl_2.6H_2O$	0.180 g/l	$NaMoO_4\ 2H_2O$	0.003 g/l
$ZnSO_4.7H_2O$	0.022 g/l	$CoSO_4.7H_2O$	0.002 g/l

Growth medium (in distilled water):

$(NH_4)_2SO_4$	1.320 g/l	$MgSO_4.7H_2O$	0.247 g/l
KH_2PO_4	0.372 g/l	$CaCl_2.2H_2O$	0.074 g/l
Trace element solution	10.0 ml/l		

Growth medium was adjusted to pH 2.0 with 5 M sulphuric acid and stored at room temperature. 2.0 g of yeast extract were added per litre of medium immediately prior to autoclaving; 20 ml vials of 50% (w/v) D-glucose in distilled water were also autoclaved.

Minimal medium (per litre): 10.5 g dipotassium hydrogen phosphate, 4.5 g potassium dihydrogen phosphate, 1.0 g ammonium sulphate, 1.0 g sodium citrate, autoclaved. Sterile stock solutions were then added aseptically to include the following: 0.2 g magnesium sulphate heptahydrate, 2.0 g D-glucose and (for *E.coli* TG1 cells) 5 μg thiamine.

DYT medium (per litre):	20 g tryptone, 10 g yeast extract, 10 g sodium chloride.
LB medium (per litre):	10 g tryptone, 5 g yeast extract, 10 g sodium chloride.

3.5.2 General Reagents

All were dissolved in Milli-Q™ water unless stated otherwise.

Chloroform:	24:1 (v/v) chloroform:isoamyl alcohol.
Ethidium bromide:	10 mg/ml and 1 mg/ml stocks, stored dark at 4°C.
Hybridisation solution:	50 mM Tris/HCl pH 7.5, 0.9 M sodium chloride, 1% (w/v) SDS, used immediately.
Phenol:	obtained distilled and stored at 4°C, equilibrated with 2 vols of 50 mM Tris pH 8.0 before use, then stored at -20°C.
Phenol/chloroform:	equal volumes of equilibrated phenol and chloroform/isoamyl alcohol, used immediately.
Potassium acetate:	3 moles of potassium acetate and 2 moles of glacial acetic acid per litre.
RNase A, DNase-free:	10 mg/ml stock in x10 TE, incubated at 95°C for 15 min and stored at -20°C.
Salmon sperm DNA, denatured:	5 mg/ml solution, denatured by the addition of 50ml of 5M NaOH, neutralised by addition of 2M HCl, then incubated at 100°C for 10 min; used immediately.

SSC (x20):	3 M sodium chloride, 0.3 M sodium citrate in d.w.
TBE (x10):	0.89 M Tris base, 0.89 M boric acid, 0.02 M EDTA in d.w.
TE (x10):	100 mM Tris/HCl and 1 mM EDTA pH 8.0, autoclaved.

3.5.3 DNA Restriction and Modifying Enzyme Buffers

AmpliTaq™ buffer (x10):	0.1 M Tris/HCl pH 8.3, 0.5 M KCl, filter sterilised, stored at -20°C.
CIAP buffer (x10):	0.5 M Tris/HCl pH 9.0, 10 mM magnesium chloride, 1 mM zinc chloride, 10 mM spermidine, filter sterilised, stored at -20°C.
Kinase buffer (x10):	0.5 M Tris/HCl pH 8.0, 0.1 M magnesium chloride, filter sterilised and stored at -20°C.
Klenow 'fill-in' buffer (x10):	0.5 M Tris/HCl pH 7.5, 0.1 M magnesium sulphate, 1 mM dithiothreitol, 500 µg BSA/ml, filter sterilised, stored at -20°C.
Ligase buffer (x10):	0.5 M Tris/HCl pH 8.0, 0.1 M magnesium chloride, 0.2 M dithiothreitol, filter sterilised and stored at -20°C.
Nuclease S1 buffer (x10):	0.2 M Tris/HCl pH 7.5, 0.5 M sodium chloride, 1 mM zinc chloride filter sterilised and stored at -20°C.

Restriction buffer, High salt (x10):	for use with BamHI, EcoRI, PstI and Sau3A; 0.1 M Tris/HCl pH 7.5, 1 M sodium chloride, 0.1 M magnesium chloride, 10 mM dithiothreitol, filter sterilised and stored at -20°C.
Restriction buffer, Medium salt (x10):	suitable for use with BsmI and HindIII; 0.1 M Tris/HCl pH 7.5, 0.5 M sodium chloride, 0.1 M magnesium chloride, 10 mM dithiothreitol, filter sterilised and stored at -20°C.
T4 DNA polymerase buffer (x10):	50 mM Tris/HCl pH 8.8, 10 mM magnesium chloride, 10 mM EDTA, 20 mM ammonium sulphate, 10 mM 2-mercaptoethanol, 150 µg/ml BSA filter sterilised and stored at -20°C.

CHAPTER 4 IDENTIFICATION OF THE *gld* GENE

4.1 Attempted plasmid gene library production

4.1.1 Introduction

The initial strategy for isolating the *gld* gene was to create and screen a plasmid library of clones. *Tp.acidophilum* has an extremely small genome for a free living organism and therefore a relatively small number of clones are needed to constitute a library that is representative of the entire genome. The following equation (Figure 4.1) was used to estimate the number of clones required, containing DNA fragments of 2 Kb, to constitute a representative library:

$$n = ln(1-p) / ln(1-x/y)$$

n = no. of clones in library
p = probability of any sequence being present
x = insert size (bases)
y = genome size (bases)

Figure 4.1: The equation used to determine the number of clones in a representative library, for a given level of confidence and insert size (Clarke and Carbon, 1976).

The genome size of *Tp.acidophilum* is reported as 10^9 daltons (≈1.54 Mb) (Christiansen, *et al.*, 1975; Searcy & Doyle, 1975); therefore a library of recombinant clones containing inserts of this size and having a 99% probability of including a particular sequence, should consist of ≈3500 recombinant clones (Figure 4.2).

4.1.2 Isolation of Thermoplasma acidophilum Genomic DNA

Smith (1989) demonstrated that the maximum cell density of growing *Tp.acidophilum* cells was optimised without affecting the length of the growth cycle if the yeast extract concentration in their growth medium was ≈2 g/l. With the yeast extract concentration at 2 g/l, maximum cell density was reached after 3 days (beyond which time cell viability diminished rapidly) and 1 litre of culture would routinely yield 500 mg of cells, wet weight (3.1.9).

A cell-free extract (3.1.10), prepared from 1 g of *Tp.acidophilum* cells, was incubated with proteinase K (3.3.4) and subjected to phenol/chloroform extraction (3.3.6); the nucleic acids were isolated by spooling (3.3.8), RNase A treated (3.3.5), the

DNA recovered by precipitation with ethanol (3.3.9) and redissolved in the minimum volume of TE. This procedure routinely yielded between 5 and 10 mg of genomic DNA.

DNA concentration and purity were determined spectrophotometrically (3.3.1) and, if necessary, the DNA further purified by equilibrium density gradient centrifugation (3.3.7). Size distribution of the purified DNA was assessed by electrophoresis in agarose gels (3.3.2); DNA preparations whose median size was judged to be ≤10 kb were discarded and the DNA preparation repeated. DNA preparations of adequate size and purity were dissolved in TE at a concentration of 1 mg/ml and stored at 4°C.

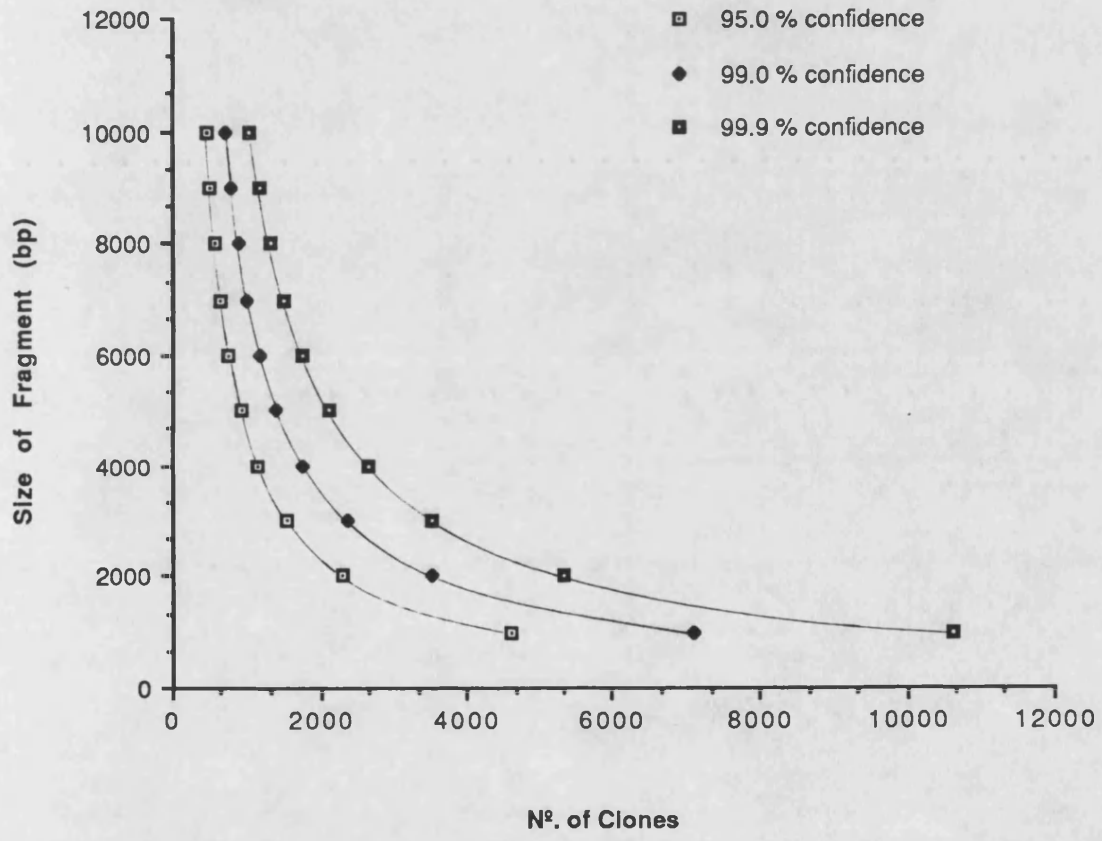


Figure 4.2: A graph demonstrating how the predicted size of a representative *Tp.acidophilum* library fluctuates according to the size of the cloned fragments and the statistical level of confidence required.

4.1.3 Production of a Thermoplasma acidophilum Gene Library

100 µg of genomic *Tp.acidophilum* DNA were subjected to partial digestion (3.3.11) with Sau3A and the partially-restricted DNA was then separated by velocity gradient centrifugation (3.3.10). After fractionation and analysis of the fractions by agarose gel electrophoresis (3.3.2) (Figure 4.3), those fractions containing DNA fragments of 2 - 10 kb were pooled and 10 µg of DNA were recovered by ethanol precipitation (3.3.9).

500 ng of plasmid pBR322 were digested (3.3.12) with BamHI and were then CIAP-treated (3.3.13). Sau3A-restricted genomic DNA and BamHI-restricted, dephosphorylated plasmid were ligated (3.3.14), and the ligation reaction products used to transform competent *E.coli* TG1 cells (3.1.1-3). At best, approximately 100 recombinant colonies were obtained from a single experiment, at a transformation frequency of $\approx 10^4$ colony forming units/µg of DNA.

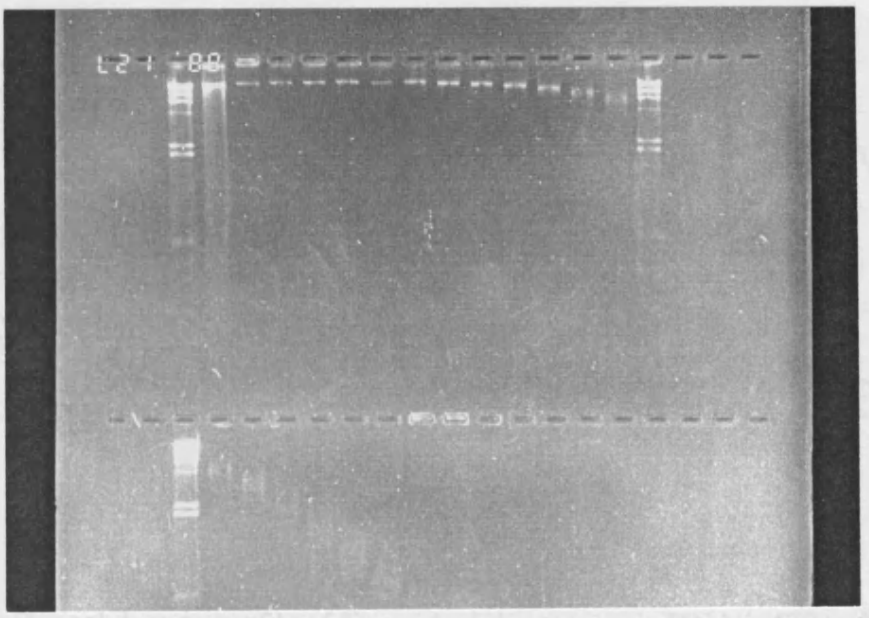


Figure 4.3: TBE/agarose gel electrophoresis of fractions of partially digested *Tp.acidophilum* genomic DNA, after separation by velocity gradient centrifugation.

4.1.4 Discussion

To generate a representative library, it is important that all the clones stem from a single ligation reaction; some DNA fragments ligate into plasmid very readily and can predominate in a library if it is assembled by cloning from several ligation reactions. In order to remove the disparity between the number of recombinant colonies obtained and the number required, at least a 10-fold increase in the observed transformation efficiency would have been necessary and, despite repeated attempts, this was not achieved.

It has been shown that supercoiled plasmid can transform competent cells with transformation efficiencies 10-fold greater than with nicked plasmid that has adopted an open-circular conformation, and that the transformation efficiency of recombinant plasmid drops as the size of the insert DNA increases (Hanahan, 1983). In addition to these factors, the purity of the recombinant plasmid almost certainly has a significant effect on transformation efficiency.

The scanning optical density trace (Figure 4.4) shows no sign of a "shoulder" at 280 nm, indicating that the genomic DNA prepared from *Tp.acidophilum* cells was apparently free from contaminating protein; since the transformation frequency was consistently poor it would have been useful to include a further purification step such as reverse-phase or ion-exchange chromatography, prior to the transformation experiments, in order to establish whether the failure of this approach was merely due to impure DNA or whether there was something more fundamentally wrong.

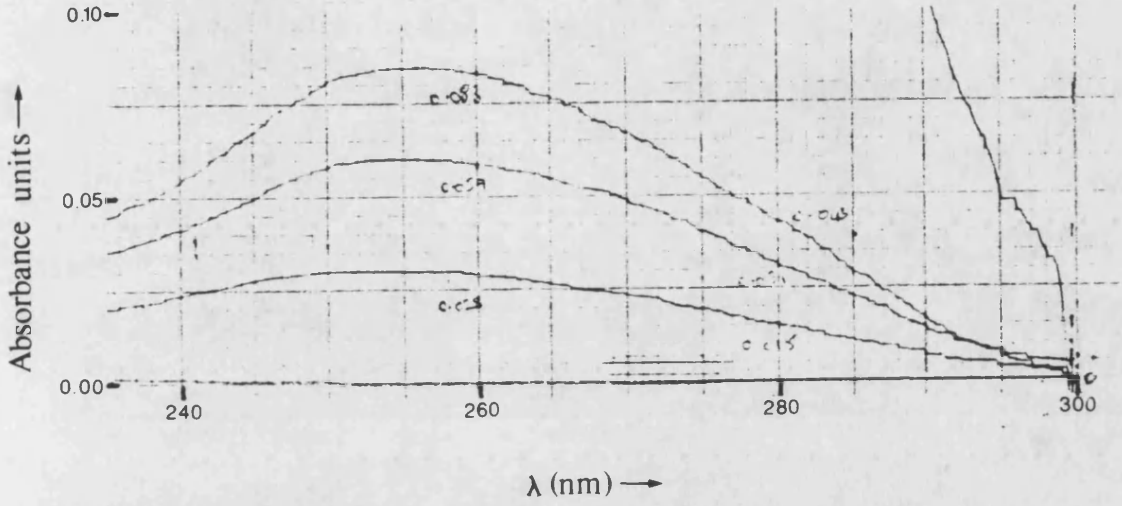


Figure 4.4: Optical density profile of *Tp.acidophilum* genomic DNA in TE buffer.

4.2 Identification of the *gld* Gene by Southern Analysis

4.2.1 Introduction

The shortcomings of the plasmid gene library approach led to the adoption of a more direct approach to identify the *gld* gene; it was felt that using Southern analysis to map the restriction sites in the vicinity of the gene would be the quickest way to pinpoint DNA fragments likely to contain it, allowing the cloning of the gene from a smaller population of DNA fragments and obviating the need to obtain such large numbers of recombinant clones.

4.2.2 Design of Probes

Tp. acidophilum glucose dehydrogenase, of sufficient purity to be used for N-terminal amino acid sequence analysis, was first isolated by Smith *et al.* (1989). Analyses of two purified samples identified amino acid residues 2 → 24 without

ambiguity, so this information was used to design two short, redundant oligodeoxyribonucleotides.

Examination of the reverse-translated nucleotide sequences, which could give rise to the determined N-terminal amino acid sequence, revealed two mixed 17-base oligodeoxyribonucleotides, designated probes 1 and 2 (Figure 4.5), of sufficiently low redundancy to be suitable for use in identifying the *gld* gene. These were synthesised and purified as described (3.2.1,2).

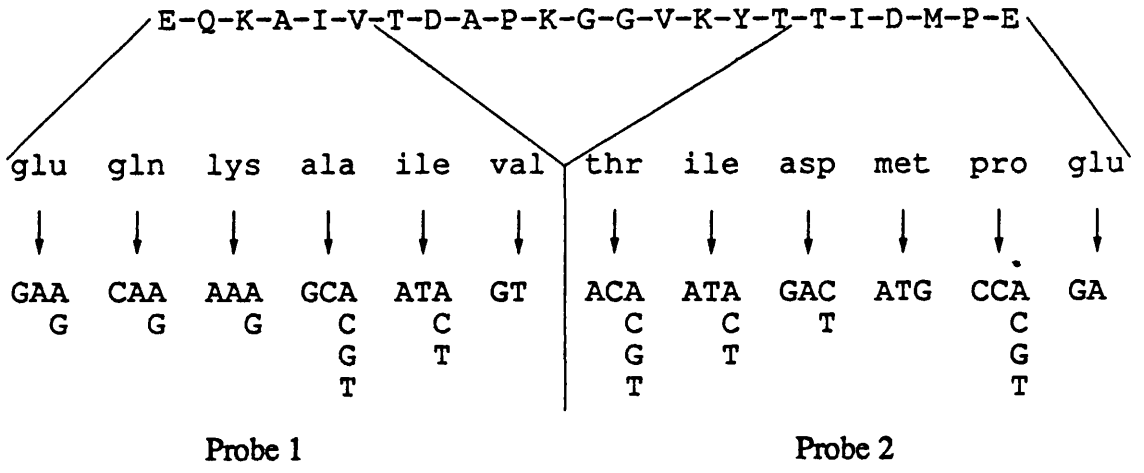


Figure 4.5: Redundant oligodeoxyribonucleotide probes designed from reverse-translated N-terminal amino acid sequence information.

4.2.3 Southern Analysis

Thermoplasma genomic DNA was digested with restriction enzymes BamHI, PstI, HindIII and EcoRI (3.3.12), both individually and in pairs, and the restriction products separated by electrophoresis in agarose-TBE gels (3.3.2); routinely, 5 µg of DNA were included in a restriction digest. The gels were photographed (Figures 4.7a & b), the DNA transferred to hybridisation membranes (3.2.4) and these were then screened (3.2.5,6) with 5'-^{32}P-labelled probes 1 and 2 (3.2.3).

Autoradiographs of these membranes (3.2.6), obtained using the two different probes, had substantially different hybridisation patterns, but a single band was present in each track that was common to both probes as demonstrated in Figure 4.6. Superimposition of these autoradiographs on life-size photographs of the original gels pinpointed restriction fragments to which both probes bound. The sizes of these fragments were determined by comparing their mobilities with the mobilities of HindIII-restricted lambda DNA and the information was used to construct a restriction map (Figure 4.8).

4.2.4 Discussion

The success of Southern analysis hinges on the specificity of the DNA probes used. The major pitfall of using redundant short oligodeoxyribonucleotides as DNA probes is that there is no way of predicting whether the sought DNA sequence will be the one to which the probe mixture binds with highest affinity; hence the use of two probes to provide points of comparison is recommended.

No information was available with regard to codon usage in *Tp.acidophilum*, so it was felt to be essential that all possible third-base positions should be accommodated. It has since been shown that using a longer oligodeoxyribonucleotide, in which third-base positions are assigned in keeping with the total G+C content of the genomic DNA, obviates this problem, even if some of the third-base positions are incorrect (Sutherland, 1991). The smallest restriction fragment to which both probes bound was a ≈0·4 kb EcoRI/HindIII fragment (Figure 4.8), which must therefore contain the extreme 5'-terminal region of the *gld*; however, the orientation of the gene with respect to this fragment could not be defined from Southern analysis alone.

4.2.5 Conclusion

The published subunit molecular weight for glucose dehydrogenase is 37,000 ± 3,000 d (Smith *et al.*, 1989); assuming that *Tp.acidophilum* genes contain no intronic DNA, the size of the *gld* would be ≈1 Kb. The restriction map (Figure 4.8) was examined, knowing the position of the 5'-end of the gene, and it was decided that the gene would be encompassed either by the ≈3.3 Kb HindIII/HindIII restriction fragment or by the ≈1.7 Kb EcoRI/BamHI fragment.

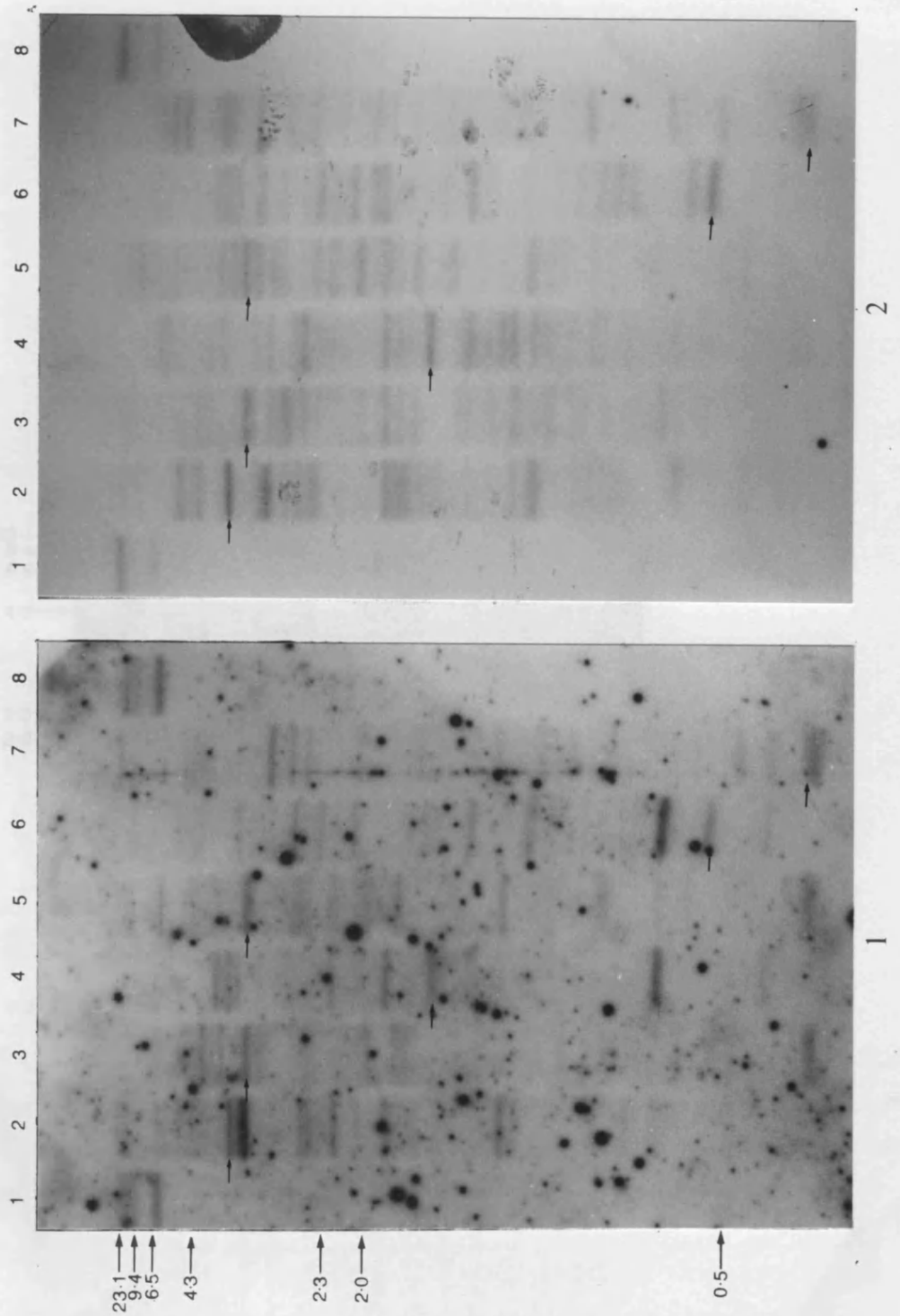


Figure 4.6: Southern analysis of the double restriction digests shown in Figure 4.7b, using redundant 5'-^{32}P-labelled probes 1 and 2. The Southern blots were made from duplicate gels of that shown in Figure 4.7b, so the lane orders are the same; both were washed as described in 3.2.6. at a final temperature of 37°C. (1 & 8) Lambda DNA restricted with HindIII, (2 - 7) *Tp.acidophilum* genomic DNA restricted with (2) BamHI/PstI, (3) BamHI/HindIII (4) BamHI/EcoRI, (5) PstI/HindIII, (6) PstI/EcoRI, (7) HindIII/EcoRI; the approximate sizes of the Lambda markers are given in kilobases. Note the 1.7 kb band in track 4 and the 0.4 kb band in track 7.

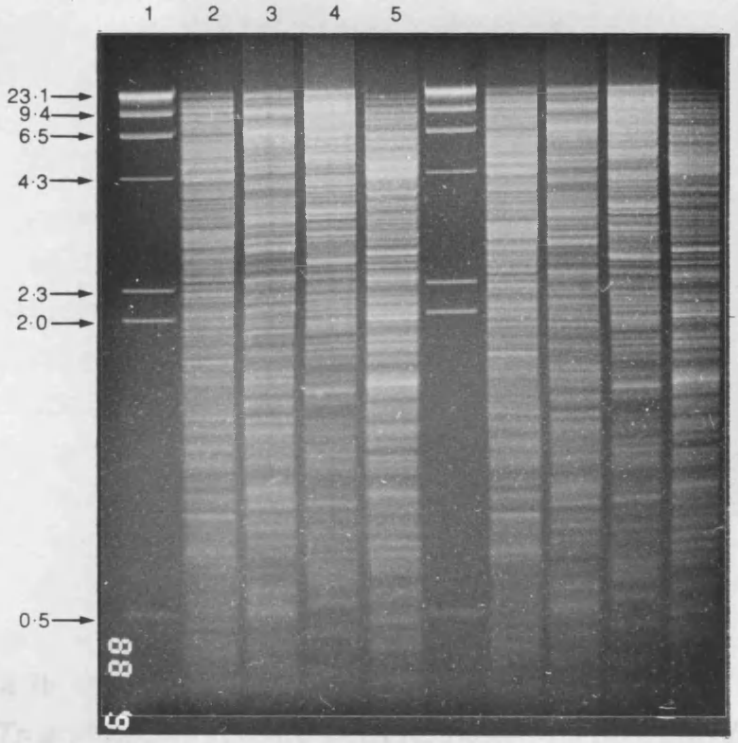


Figure 4.7a: Agarose gel electrophoresis: (1) Lambda DNA restricted with HindIII, (2 - 5) *Tp.acidophilum* genomic DNA restricted with (2) BamHI, (3) PstI, (4) HindIII, (5) EcoRI. The other tracks are duplicates; the approximate sizes of the Lambda markers are given in kilobases.

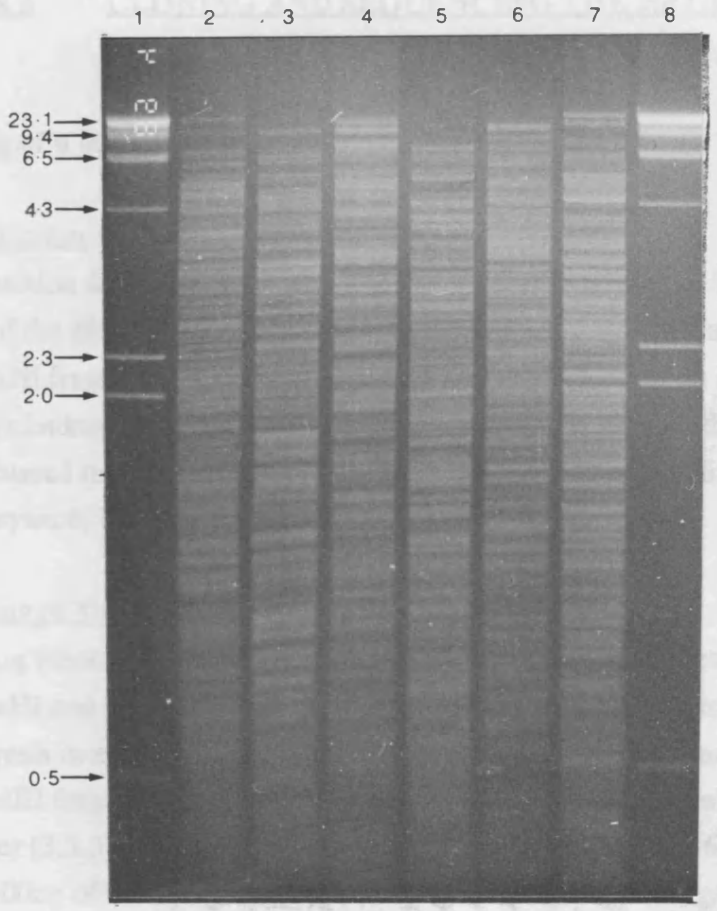


Figure 4.7b: Agarose gel electrophoresis: (1 & 8) Lambda DNA restricted with HindIII, (2 - 7) *Tp.acidophilum* genomic DNA restricted with (2) BamHI/PstI, (3) BamHI/HindIII (4) BamHI/EcoRI, (5) PstI/HindIII, (6) PstI/EcoRI, (7) HindIII/EcoRI; the approximate sizes of the Lambda markers are given in kilobases.

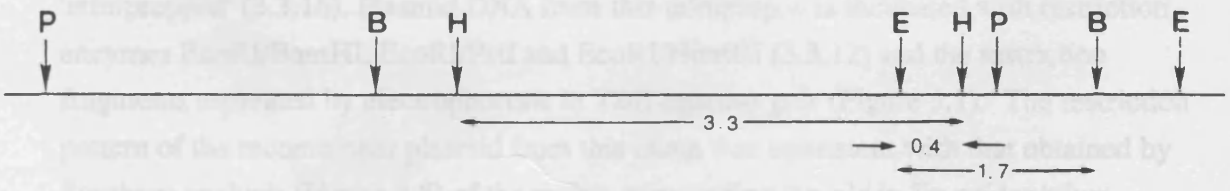


Figure 4.8: Restriction map of the *gld*-bearing fragments: the HindIII/HindIII, EcoRI/HindIII and EcoRI/BamHI fragments are marked.

CHAPTER 5 CLONING AND SEQUENCING THE *gld* GENE

5.1 Cloning of a Restriction Fragment Containing the Gene

5.1.1 Introduction

In section 4.2, three restriction fragments were identified as being likely to contain all or part of the *gld* gene; two of these, the ≈0.4 kb EcoRI/HindIII and the ≈1.7 kb EcoRI/BamHI fragments, had different sticky ends and so could be cloned directionally. Directional cloning is often easier than cloning DNA fragments with the same sticky ends (this is discussed in section 5.1.5) so cloning the ≈3.3 kb HindIII/HindIII fragment was put into abeyance, in favour of the other two DNA fragments.

5.1.2 Cloning of Target Restriction Fragments

10 µg amounts of *Tp.acidophilum* genomic DNA were digested with EcoRI/BamHI and EcoRI/HindIII (3.3.12) and the restriction fragments separated by electrophoresis in agarose-TBE gels (3.3.2). EcoRI/BamHI fragments of ≈1.7 kb and EcoRI/HindIII fragments of ≈0.4 kb were isolated from the gel by electroelution onto DE-81 paper (3.3.3) and were then purified using a Sepharose CL-6B spun column (3.3.6). ≈100ng of each fragment-type were recovered, samples ligated into pUC19 (3.3.14), the ligation reactions purified using columns of Sepharose CL-6B (3.3.6) and 10 ng of total DNA used to transform competent *E.coli* cells (3.1.3); colour selection was used to identify positive recombinants (3.1.4). The EcoRI/BamHI construct generated 100 recombinants and the EcoRI/HindIII construct generated 10, all of which were grown and stored in microtitre plates (3.1.5).

5.1.3 Identification of Positive Recombinants

The stocks of recombinant clones were screened by colony hybridisation (3.2.7) with probe 1 (5'-^{32}P-labelled); two of the 100 EcoRI/BamHI recombinants gave strong hybridisation signals and the recombinant clone giving the marginally stronger signal was 'miniprepped' (3.3.16). Plasmid DNA from this miniprep was incubated with restriction enzymes EcoRI/BamHI, EcoRI/PstI and EcoRI/HindIII (3.3.12) and the restriction fragments separated by electrophoresis in TBE-agarose gels (Figure 5.1). The restriction pattern of the recombinant plasmid from this clone was consistent with that obtained by Southern analysis (Figure 4.8) of the region surrounding the *gld* in *Tp.acidophilum* genomic DNA.

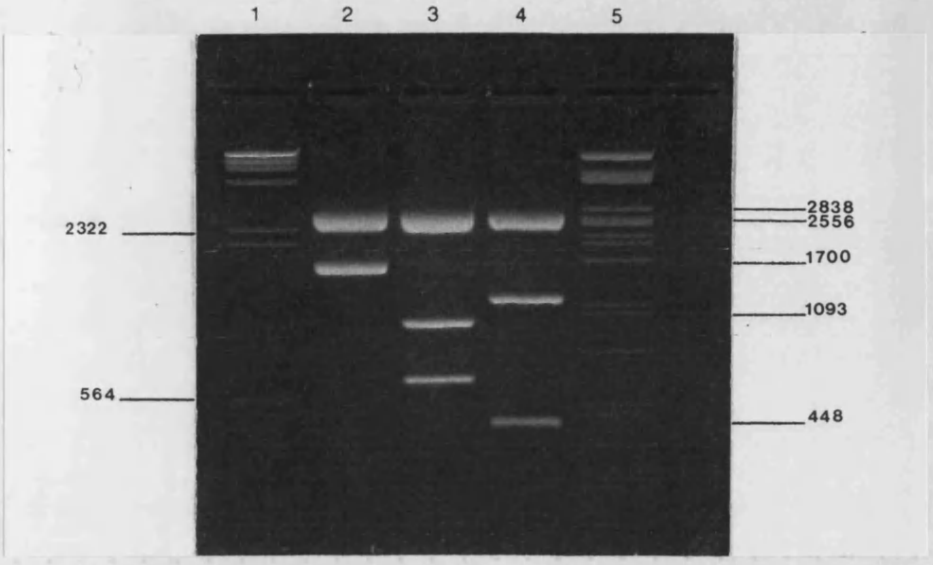


Figure 5.1: Restriction analysis of the cloned construct; (1 & 5) Lambda DNA cut with HindIII and PstI, respectively, (2 - 4) all show a ≈2.6 Kb band of linearised plasmid. (2) EcoRI/BamHI restriction showing a ≈1.7 Kb fragment, (3) EcoRI/PstI showing a ≈1.0 Kb and a ≈0.7 Kb fragment, and (4) EcoRI/HindIII with a ≈1.3 Kb and a ≈0.4 Kb fragment.

5.1.4 Preliminary Sequencing of the Putative Clone

The miniprepped plasmid DNA from this putative clone was purified on a Nensorb20™ reverse phase column (3.3.6), and part of its nucleotide sequence determined (3.3.18) using probe 1 as a sequencing primer. Approximately 150 bases of nucleotide sequence were legible using this primer and these were in an open reading frame. A HindIII restriction site was found in the 3'-region of this portion of determined nucleotide sequence and translation of the 5'-region revealed the remainder of the determined N-terminal amino acid sequence (see Figure 5.3 for the binding sites of probes 1 and 2 and the HindIII site).

5.1.5 Discussion

Cloning of DNA fragments with the same sticky ends is complicated by the tendency of both the vector and the insert to ligate to themselves as well as each other. CIAP-treatment of the vector reduces this problem but any residual phosphorylated plasmid will religate to itself and transform cells much more readily than a plasmid-insert construct; furthermore CIAP-treated plasmid contaminated with CIAP activity will inhibit ligation by continuing to remove terminal phosphate groups.

Directional cloning obviates the need for CIAP-treatment as the different sticky ends of a plasmid will not religate; additionally, if the DNA fragment to be cloned ligates to a second fragment, the resulting dimer will have the same sticky ends and therefore will not ligate into a doubly-restricted plasmid; for these reasons, directional cloning is often felt to be more expedient when attempting to clone particular DNA fragments.

The ambiguity of the results obtained during Southern analysis, caused by the lack of specificity of the redundant DNA probes, was discussed earlier (4.2.4); similarly, this redundancy caused ambiguity in the colony hybridisation experiments. The two colonies that gave strong hybridisation signals with probe 1, were not labelled to the same degree, and many of the other colonies displayed weak hybridisation. This ambiguity was to be resolved by screening the recombinant clones with the second probe, (as was done during Southern analysis); however, the restriction analysis and preliminary sequencing experiments demonstrated that the clone under examination did contain the insert of interest, so further colony hybridisation experiments seemed superfluous.

The results of the preliminary nucleotide sequencing experiments (5.1.4) demonstrated that at least part of a gene, containing a nucleotide sequence compatible with that obtained from N-terminal amino acid sequence analysis, was carried within this ≈1.7 Kb EcoRI/BamHI restriction fragment; furthermore, given the position of this portion of nucleotide sequence with respect to the position of a known HindIII restriction site it was felt, if this was the sought clone, then it probably did contain the entire *gld* gene.

5.2 Determination of the Nucleotide Sequence of the *gld* Gene

5.2.1 Introduction

In order to discover if this cloned 1.7 Kb EcoRI/BamHI fragment did indeed contain the entire *gld* gene, the nucleotide sequences of both strands of the fragment were determined. As was discussed in Chapter 1, little information with regard to the molecular genetics of *Tp.acidophilum* was available, so in addition to obtaining information about the *gld* gene it was hoped that nucleotide sequence data from the DNA surrounding the gene might provide some knowledge of gene structure within the organism.

5.2.2 Design of Sequencing Primers

Primers for DNA sequencing were designed without redundancy, avoiding predicted hairpin loop structures, from the extreme 3'-terminal nucleotide sequence information obtained from employing a preceding primer; these were synthesised and

purified as described (3.2.1,2). The various sequences of these primers are highlighted in Figure 5.3, and those portions of the entire sequence determined using these various sequencing primers are given in Figure 5.2.

5.2.3 Nucleotide Sequencing from the pUC19 Construct

The EcoRI/BamHI fragment in pUC19 was 'maxiprepped' (3.3.17) to provide a stock of pure double stranded DNA template for nucleotide sequencing (3.3.18); the identities of approximately 300 consecutive nucleotides could be assigned using a single sequencing primer. After sequencing both strands of the cloned *Tp.acidophilum* DNA fragment, using the pUC19 construct as a sequencing template, some portions of the nucleotide sequence still contained ambiguities.

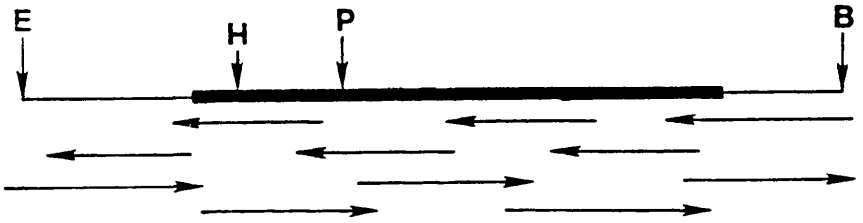


Figure 5.2: The portions of the cloned DNA fragment determined by the various sequencing primers.

5.2.4 Subcloning of the Cloned DNA Fragment in M13mp18 and M13mp19

5µg of the construct in pUC19 was incubated with restriction enzymes BamHI and EcoRI (3.3.12) and the restriction products separated by electrophoresis in TBE-agarose gels (3.3.2). The cloned 1.7 Kb EcoRI/BamHI fragment was isolated from the gel by the freeze-squeeze procedure (3.3.3) and purified using a Nensorb20™reverse phase column (3.3.6).

500 ng of pM13mp18 and pM13mp19 were restricted with EcoRI and BamHI (3.3.12) and purified using a Sepharose CL-6B spun column (3.3.6). The EcoRI/BamHI-restricted 1.7Kb insert was ligated into each of the two restricted vectors and the resulting constructs were used to transform competent *E.coli* TG1 cells (3.1.6); positive recombinants were identified by colour selection (3.1.4), and these were grown and used to make a stock of infective phage (3.1.7).

5.2.5 Nucleotide Sequencing from the M13 Constructs

Single-stranded M13 DNA was prepared from the infective phage stocks of both the M13mp18 and the M13mp19 constructs and was used as the template for nucleotide sequencing (3.3.18) so that the nucleotide sequence of both DNA strands could be determined; the use of ss-DNA templates for nucleotide sequencing resolved the remaining ambiguities in the nucleotide sequence determined using ds-DNA template.

5.2.6 Discussion

Nucleotide sequence ambiguities arising from ds-DNA sequencing are commonly caused by regions of high G+C content and secondary structural features of DNA, such as hairpin loops within DNA strands. Despite denaturation of the template prior to ds-DNA sequencing, these loop regions appear to reform and interfere with the polymerase reaction causing the two following effects:

i) the sequence of the G+C-rich/loop region may appear compressed on sequencing gels, obscuring the identities of the component bases, or,

ii) there maybe a band across all four tracks of the sequencing gel, where the polymerase has been unable to read reproducibly one or two of the template nucleotides within such a region of secondary structure/high G+C content, and therefore has incorporated any dideoxynucleotide.

When performing ss-DNA sequencing, these factors, although not absent, do not seem to impinge as significantly on the polymerase reaction; possibly the presence of the complementary DNA strand assists the re-establishment of secondary DNA structure. Other areas of ambiguity were common in the primer binding regions. As was stated in section 5.2.2, the 3'-terminal sequence information gleaned ≈300 nucleotides from a primer binding site, was used to design the next primer; the 10 - 15 nucleotides immediately after a primer binding site cannot always be determined, so when designing a new primer from a 3'-terminal region of known sequence, it is important to leave a short region of determined sequence beyond its binding site.

With sequence information from ds-DNA sequencing, it is very difficult to resolve more than ≈300 nucleotides from the primer binding site. However, ss-DNA templates readily generate unambiguous sequence information over this distance so the nucleotide sequences around the primer binding regions can be confirmed. The complete nucleotide sequence and primers are shown in Figure 5.3.

5.2.7 Conclusion

The longest open reading frame within the determined nucleotide sequence is 1059 nucleotides, encoding a 352 amino acid polypeptide with M_r 39,264. This derived amino acid sequence has an N-terminal amino acid sequence identical to that determined

from the protein purified from *Tp.acidophilum* cells, and the theoretical molecular weight of the derived amino acid sequence compared very favourably with the purified enzyme subunit molecular weight of 37,000 ± 3,000 d, as determined by SDS-PAGE (Smith *et al.*, 1989). This suggested that the clone under investigation did contain the *gld* gene in its entirety; the recombinant construct was therefore designated pTaGDH1.

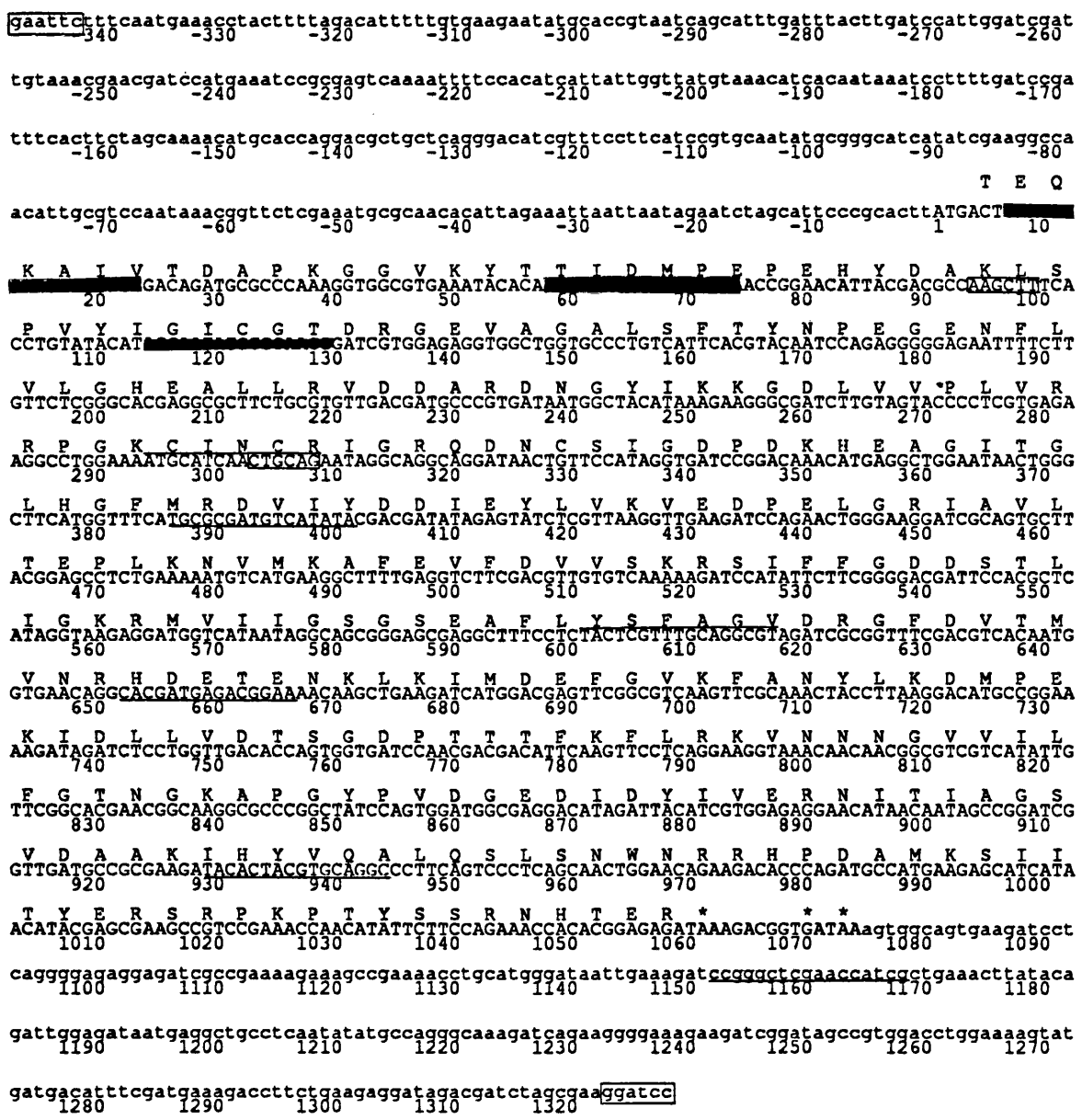


Figure 5.3: The complete nucleotide sequence of the 1.7 kb EcoRI/BamHI DNA fragment and the derived amino acid sequence of the O.R.F. The boxed nucleotides are the unique restriction sites for EcoRI, HindIII, PstI and BamHI respectively; the binding sites for probes 1 and 2 are highlighted in blue and the site for one of the complementary sequencing primers (primer 5) highlighted in green. Other primer binding sites are overlined if sense primers or underlined if antisense primers.

CHAPTER 6 EXPRESSION OF *Thermoplasma acidophilum* GLUCOSE DEHYDROGENASE IN *Escherichia coli*

6.1 Investigation of Expression from the pUC19 Construct, pTaGDH1

6.1.1 Introduction

There is good evidence that pTaGDH1 contains the entire *gld* gene from *Tp.acidophilum* (5.2); however, expression of active *Tp.acidophilum* glucose dehydrogenase, and, if possible, demonstration that its physical and kinetic properties are similar to those of the enzyme purified from *Tp.acidophilum* cells, would provide conclusive proof that the correct gene has been cloned.

6.1.2 Investigation of Glucose Dehydrogenase Activity

Cell free extracts (3.1.8) were prepared from 10 ml cultures of *E.coli* TG1 cells harbouring (i) pUC19 and (ii) pTaGDH1, grown overnight in DYT containing 100 μg ampicillin/ml; these extracts were assayed for $NADP^+$-linked glucose dehydrogenase activity (3.4.1), NADPH oxidase activity (3.4.2), and for total protein (3.4.3). The value obtained for NADPH oxidase activity was used to correct the value obtained for $NADP^+$-linked glucose dehydrogenase activity, prior to calculating the specific activity values for the samples:

pUC19 control	0.0086 ± 0.0004 U/mg
pTaGDH1 construct	0.0086 ± 0.0006 U/mg

Table 6.1: Glucose dehydrogenase specific activities in the cell-free extracts of *E.coli* TG1 cells containing a control plasmid and the recombinant construct, pTaGDH1. Activity values have been corrected for contaminating NADPH oxidase activity.

6.1.3 Discussion

There was no significant difference betweeen the specific activity values determined for the extracts from the two transformed cell types. Several factors may have been responsible for this lack of expression of the active cloned protein:
i) there was no message production from the sense strand of the cloned fragment, either because there was no associated promoter sequence recognised by *E.coli* RNA polymerase, or because the *gld* gene was in the wrong orientation to be transcribed from the P_{lac} promoter in pUC19,

ii) mRNA was being produced from the sense strand of the cloned fragment, but was in some way rendered non-functional, *e.g.* it did not contain a ribosome-binding site recognised by *E.coli* ribosomes, or it was being degraded before protein biosynthesis could take place,
iii) protein was being produced but was being rendered non-functional in some way, *e.g.* it was folding incorrectly, or being degraded.

The first of these possible explanations was investigated by placing a known recognisable promoter upstream of the cloned DNA fragment (6.2). Investigation of the second and third possibilities was postponed, pending the result of this experiment.

6.2 Investigation of Expression from a pUC18 Construct

6.2.1 Introduction

M13-derived plasmids, such as pUC18 and pUC19, contain part of the lactose operon, namely:
(i) the O/P region, which consists of the P_{lac} promoter site followed by the O_{lac} operator site,
(ii) the *lacZ* gene, which has been manipulated so that only the coding region for the first 146 amino acid residues (the α-peptide) remains, and
(iii) the multiple cloning site or polylinker sequence, which interrupts the *lacZ* gene near its 5'-end, but is so designed that it does not affect the O.R.F. of the *lacZ*.

Transcription from the P_{lac} promoter is inducible with IPTG (see 6.2.7) enabling the production of considerable amounts of either the *lacZ* transcript, or transcripts of DNA fragments ligated into the polylinker sequence (sometimes with the *lacZ* transcript appended). pUC18 is identical to pUC19 in every respect, except it is constructed such that the polylinker sequence lies in the opposite orientation; the *gld* gene in pTaGDH1 was lying on the wrong DNA strand to be transcribed from the P_{lac} promoter so the cloned DNA fragment was subcloned into pUC18.

6.2.2 Subcloning of the gld Gene in pUC18

1μg of pTaGDH1 was digested with EcoRI and BamHI (3.3.12) and the restriction products separated by electrophoresis in agarose-TBE gels. A gel slice containing the 1.7 Kb cloned insert was excised from the gel, the DNA was isolated by the freeze-squeeze procedure (3.3.3) and then was purified using a Nensorb20™ column (3.3.6). 100 ng of pUC18 DNA were incubated with EcoRI and BamHI

(3.3.12) and the restricted plasmid purified and separated from the excised polylinker fragment using a Sepharose CL-6B spun column (3.3.6).

The restricted fragment and plasmid were ligated together (3.3.14), the ligation products were cleaned using a CL-6B column (3.3.6), 10 ng of total DNA was used to transform competent *E.coli* TG1 cells and positive recombinants were identified by colour selection (3.1.2-4). No white recombinant colonies were obtained, 3 bright blue colonies were obtained and >100 pale blue colonies were obtained; the latter were assumed to be the positive recombinant colonies (see discussion 7.1.6) so 2 colonies were picked and miniprepped (3.3.16) with a view to restriction analysis.

6.2.3 Restriction Analysis of the pUC18 Recombinants

Restriction of these recombinants with EcoRI and BamHI and analysis of the products by electrophoresis in agarose-TBE gels gave the characteristic 2.8 Kb and 1.7 Kb restriction pattern. One of the recombinants was cultured in 500 ml DYT/ampicillin, a small stock taken for storage (3.1.5) and the remainder maxiprepped (3.3.17). This pUC18 construct was designated pTaGDH2.

6.2.4 Glucose Dehydrogenase Activity in the pTaGDH2-Bearing Strain

Cell-free extracts were prepared from 10 ml cultures of *E.coli* TG1 cells harbouring plasmid pTaGDH2, grown in DYT broth containing 100 µg ampicillin/ml ± 1 mM IPTG. $NADP^+$-linked glucose dehydrogenase specific activity determinations were then performed on the cell-free extracts as described (6.1.2); the data are presented in Table 6.2.

pUC19 control	0.0080 ± 0.0004 U/mg
pTaGDH1 construct	0.0064 ± 0.0012 U/mg
pTaGDH2 construct	4.02 ± 0.30 U/mg
pTaGDH2 + IPTG	4.71 ± 0.23 U/mg

Table 6.2: Glucose dehydrogenase specific activity values in *E.coli* cells bearing pUC19, pTaGDH1 and pTaGDH2, the latter grown with and without IPTG-induction of the P_{lac} promoter.

6.2.5 SDS-PAGE Analysis of the Recombinant Cell Proteins

Cell-free extracts were prepared from the cell-lines harbouring pUC19, pTaGDH1 and pTaGDH2, grown in the presence and absence of IPTG, and were

analysed by SDS-PAGE (Figure 6.1). SDS-PAGE analysis of partially purified samples of each extract is also shown; samples derived from cells grown in the presence of IPTG were heat treated at 55°C for 10 min and the precipitated proteins removed (see 9.1 for an appraisal of heat-treatment as a purification step). The band marked in the figure is only present in the sample from the cell line harbouring pTaGDH2 and runs next to the BSA molecular weight marker. Comparison of its mobility with that of the protein molecular weight markers gave a $M_r \approx 41{,}500$.

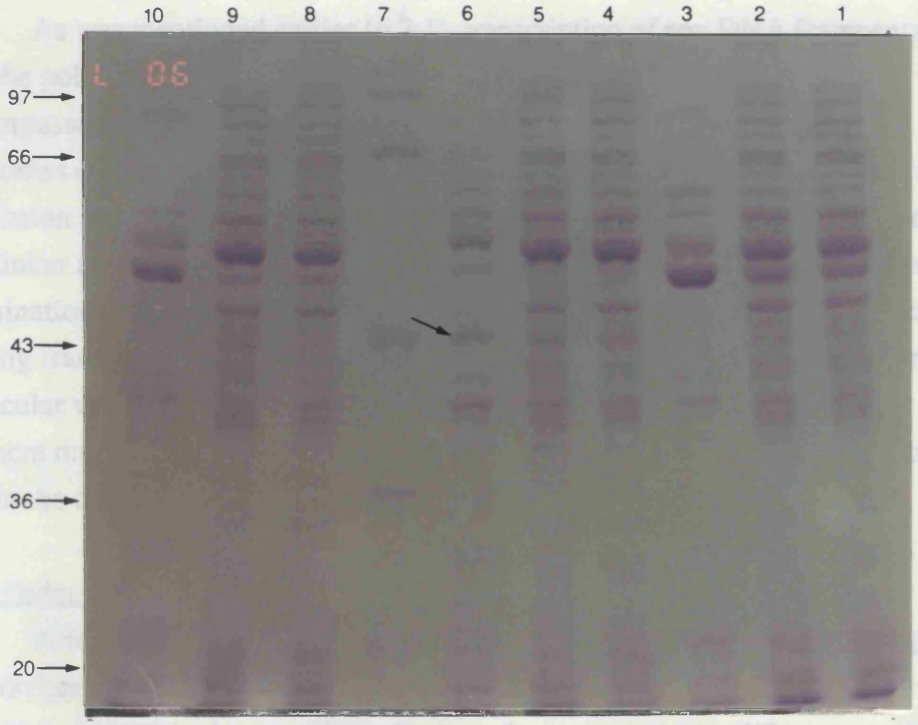


Figure 6.1: SDS-PAGE of cell-free extracts from *E.coli* cells containing pUC19, pTaGDH1 and pTaGDH2. (1), (2) & (3) pUC19 -IPTG, +IPTG and heated, respectively; (4), (5) & (6) pTaGDH2 -IPTG, +IPTG and heated, respectively; (7) molecular weight markers; (8), (9) & (10) pTaGDH1 -IPTG,+IPTG and heated, respectively.

6.2.6 Discovery of Glucose Dehydrogenase Activity

The discovery of high levels of glucose dehydrogenase activity in the cell line harbouring pTaGDH2, provided further evidence that the cloned DNA fragment contained the *Tp.acidophilum gld* gene. The fact that activity was found confirmed: i) that the sense strand of the cloned fragment must be being transcribed and that the transcribed message must remain intact during protein biosynthesis, and

ii) since plasmids pTaGDH1 and pTaGDH2 have identical inserts and activity is only found in the pTaGDH2-bearing strain, there is no promoter region within the cloned fragment which is recognisable to *E.coli* RNA polymerase.

Analysis of the relative mobilities of the molecular weight markers and the candidate recombinant protein band on the polyacrylamide gel (Figure 6.1), indicated that the cloned gene product had a $M_r \approx 41,500$, ≈2,000 more than the calculated M_r 39,264. The published subunit molecular weight for *Tp.acidophilum* glucose dehydrogenase (Smith *et al.*, 1989) is 37,000 ± 3,000, so the apparent subunit molecular weight of the recombinant protein is also marginally outside this range.

As was mentioned earlier (6.2.1), transcription of any DNA fragments inserted into the polylinker of M13-derived plasmids, begins at the P_{lac} promoter and encompasses the start of the *lacZ*. Therefore, if such an O.R.F. exists, *E.coli* ribosomes can bind to the ribosome-binding site upstream of the *lacZ*, protein translation can commence from the start of the *lacZ* and can continue both through the polylinker and the cloned DNA fragment, forming a fusion protein. However, examination of the nucleotide sequence (Figure 5.3) demonstrates that no such open reading frame exists indicating that this is not the reason for the slightly anomalous molecular weight of the recombinant protein and also confirming that:
iii) there must be a ribosome-binding site, recognisable to *E.coli* ribosomes, contained within the cloned DNA fragment, near the start of the *gld*.

6.2.7 Induction of Glucose Dehydrogenase with IPTG

A feature of the pTaGDH2-bearing strain, demonstrated both by the glucose dehydrogenase activity data and the SDS-PAGE data (Figure 6.1), was that the presence of IPTG in the growth medium made no appreciable difference to the level of expression of the cloned enzyme. However, consideration of the genotype of the host strain and of the mechanisms involved in the induction with IPTG of promoter activity at the P_{lac} promoter site, does provide a possible explanation for this.

All the genes for proline biosynthesis and of the lactose operon have been deleted from the chromosomal DNA of *E.coli* TG1 cells (as indicated in 2.1). Instead the cells carry a single episome or F'-factor, which in turn carries a number of genes:
i) the wild-type genes for proline biosynthesis *proA*$^+$*B*$^+$,
ii) a mutated allele of the *traD* gene *traD36*, which is necessary for pilus formation and therefore for the conjugational transfer of DNA or for the uptake of phage,
iii) *lacI*^{q}, a mutant allele of the *lacI* gene that permits overexpression of the gene product, the lac repressor,

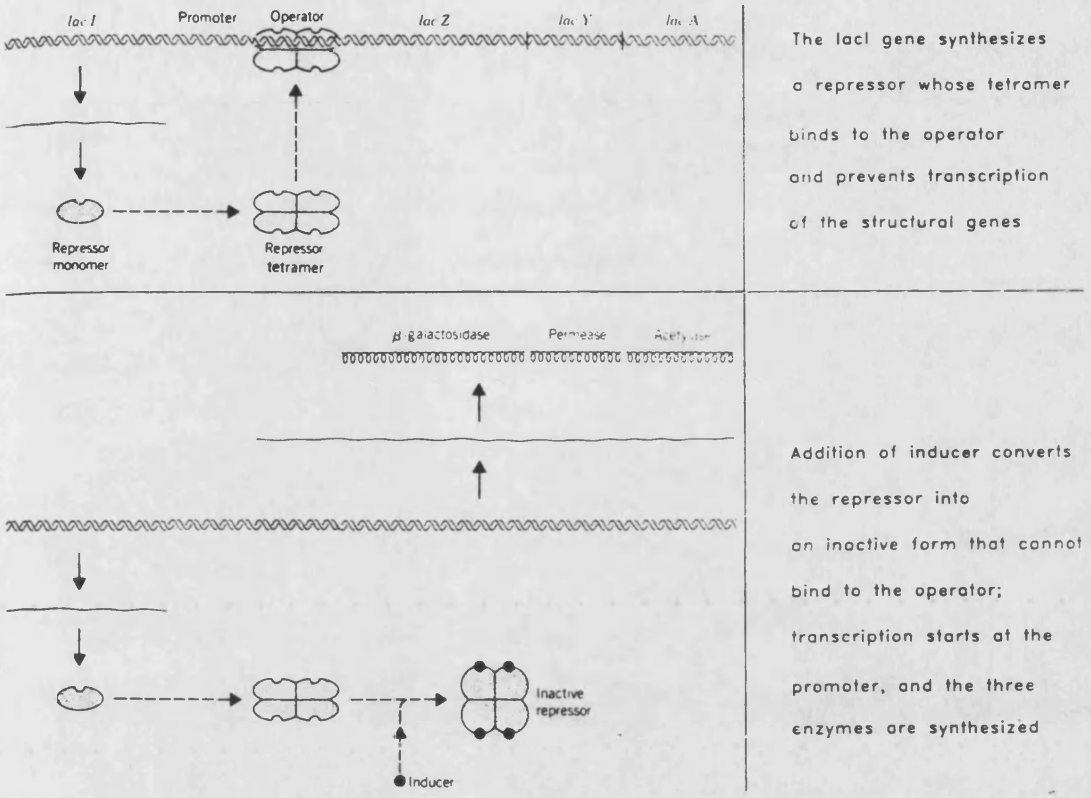


Figure 6.2: Mechanism of induction of the **lac** operon. Reproduced from Lewin (1990).

iv) the *lacZδM15* gene, that contains the O/P region of the lactose operon and a deletion mutant of the *lacZ* gene that lacks the codons for amino acid residues 11 → 41 of the gene product, β-galactosidase.

As was mentioned in 6.2.1, M13-derived plasmids also contain the O/P region of the lactose operon and a mutant allele of the *lacZ* gene. Individually, the gene products of the two mutant *lacZ* genes will not form active tetrameric β-galactosidase, but TG1 cells transformed with M13-derived plasmids have the potential to make both mutant gene products, and these complement to form active enzyme.

Normally, the product of the *lacI*^{q} gene, the lac repressor, assembles in a tetrameric structure and binds to the O_{lac} operator site. *E.coli* RNA polymerase, bound to the P_{lac} promoter region, cannot then move past the repressor to produce any transcript from the *lacZ*. The *lacI*^{q} overexpression allele is present on the episome, because the high copy number of M13-derived plasmids presents the host cell with a large number of O_{lac} operator sites to block. IPTG acts as an inducer molecule by binding to the lac repressor protein and preventing it from binding to the O_{lac} operator sequence. In the absence of repressor protein, *E.coli* RNA polymerase can transcribe

the mutant *lacZ* genes, and any foreign DNA fragments incorporated within the polylinker sequence; this mechanism is illustrated in Figure 6.2.

For this entire mechanism to work, some selection pressure is needed to retain the F'-factor from generation to generation, hence the presence of genes $proA^+B^+$. As long as the cell is deprived of proline, then pilus formation, the IPTG-inducible nature of the P_{lac} promoter and the potential for β-galactosidase-based colour selection of recombinants are retained. When a cell is growing in rich media, the necessity to make proline is lost, and therefore, after a few generations the episome will probably be lost from the majority of cells; so growing the cells in the presence of IPTG causes no increase in the level of expression of the recombinant protein.

Although this was not tested experimentally, it was felt that a higher cell density, obtainable by growing the recombinant cells in rich media, was more advantageous, in terms of the yield of recombinant protein, than growth in supplemented minimal media and the utilisation of IPTG-induction.

6.3 The Level of Expression of Glucose Dehydrogenase

6.3.1 Introduction

The observed $NADP^+$-linked glucose dehydrogenase specific activity of the crude extract from the pTaGDH2-harbouring cell line (6.2.4), expressed as a percentage of the published $NADP^+$-linked specific activity of the purified enzyme (320 U/mg, Smith *et al.*, 1989), is ≈1%; this figure is an indication of the level of expression of the *gld* gene *i.e.* glucose dehydrogenase represented ≈1% of total soluble cell protein. This level of expression, though perfectly adequate in simply diagnostic terms, was not felt to be sufficient if, at some stage, large quantities of the enzyme were to be prepared.

It was already known that IPTG-induction had no appreciable effect on the level of expression; however, no examination had been made of whether the plasmid copy number in the cell may affect expression. When performing maxipreps on some plasmid/host cell combinations the level of plasmid within the cell can be very low at the point when the cells are harvested. To counteract this, extra antibiotic can be added to the medium shortly before harvesting and this seems to boost the plasmid copy number (P.Towner, personal communication); it was felt appropriate to investigate what effect this would have on the level of expression of the cloned gene.

6.3.2 The Effect of Ampicillin on Expression

Using the *E.coli* TG1 cell line harbouring pUC19 as a control, the cells containing pTaGDH2 were grown under conditions designed to test the effects of adding extra ampicillin, on the level of expression of glucose dehydrogenase. Cells were either grown in DYT containing 100 μg ampicillin/ml at zero time and harvested 18 h later or grown under the same conditions with an extra 100 μg ampicllin/ml added 1 h before harvesting. Cell-free extracts were prepared and used for $NADP^+$-linked glucose dehydrogenase specific activity determinations (Table 6.3) as described above (6.1.2).

6.3.3 Discussion

The addition of extra ampicillin about one hour before harvesting the cells from the culture, did enhance the level of expression of the enzyme. It would be useful to gather information to establish whether this effect is peculiar to this host/vector system or whether it could be applied to other sytems in a similar fashion.

	pUC19 (U/mg)	pTaGDH2 (U/mg)	level of expression
Amp at 0 h:	0.0079 ± 0.0013	4.11 ± 0.09	1.3%
Amp at 0 & 17 h:	0.0080 ± 0.0008	8.62 ± 0.35	2.7%

Table 6.3: Demonstration of the effect of ampicillin on the level of expression of the recombinant glucose dehydrogenase.

As was mentioned in 6.3.1, the antibiotic-linked enhancement of expression does seem to be associated with an increase in the plasmid copy number in the cell. The somewhat speculative rationale behind this is that the cells are in the stationary phase of their growth cycle when the second challenge with ampicillin is made, so a cell's net synthesis of protein and nucleic acid will be approximately zero. Even so, the presence of more ampicillin in the growth medium requires the production of more β-lactamase to allow the cell's overall metabolism to be maintained. To accomplish this, nucleic acid and protein biosynthesis are directed towards the plasmids and as well as generating extra β-lactamase, more of the cloned gene product is produced.

6.3.4 Conclusions

Certain growth conditions designed to optimise expression of glucose dehydrogenase in this host/vector system have been examined but the best level of expression is still only 2.7% of soluble cell protein. One of the objectives of this project was to be able to produce the enzyme in large quantities, so that the use of immobilised enzyme on an industrial scale might be possible, and additionally, it was hoped that it would be possible, at some point, to produce enough on the laboratory scale to perform structural studies. This level of expression would have certainly permitted the latter, but it was felt that it could be improved upon.

The P_{lac} promoter is very commonly used for high level expression of cloned proteins and is known to be very efficient, so this is probably not at the root of the problem. As will be discussed in Chapter 7, the nucleotide sequences in *Tp.acidophilum*, which must constitute a ribosome-binding site, show little similarity to Shine-Dalgarno sequences. How this problem was addressed, as a route to enhancement of expression, is presented in Chapter 8.

CHAPTER 7 NUCLEOTIDE AND AMINO ACID SEQUENCE ANALYSIS

7.1 Transcription and Translation Signals in *Thermoplasma acidophilum*

7.1.1 Introduction

In eubacteria, DNA transcription is performed by a single RNA polymerase enzyme and the promoter site to which it binds is characterised by two conserved sequences located 10 bp and 35 bp upstream of the site of transcription initiation (Pribnow, 1975). In contrast, eukaryotes contain three different RNA polymerases for which there are no shared consensus sequences, but of these three enzymes, the consensus sequence for RNA polymerase II (the 'TATA' box) is 5' - TATA$^{T}/_{A}$A$^{T}/_{A}$ - 3' and is found approximately 25 bp upstream of the site of initiation of transcription (Corden *et al.*, 1980). Structural similarities have been highlighted between archaebacterial and eukaryotic RNA polymerases (Huet *et al.*, 1983) and more recently a consensus sequence for archaebacterial promoter regions, resembling the 'TATA' box, has been identified lying approximately 25 bp upstream of the site of initiation of transcription (Thomm & Wich, 1988).

This consensus sequence was identified by sequencing fragments of DNA left after nuclease protection experiments with an archaebacterial RNA polymerase/DNA complex from *Methanococcus vannielii* and comparison of this sequence with upstream sequences from 26 other archaebacterial tRNA and rRNA genes. The primary structure of archaebacterial promoters is described in terms of this consensus sequence - 'Box A', and a second less conserved sequence - 'Box B', which is the site of initiation of transcription; the latter has been identified by primer extension analysis of archaebacterial mRNA. The consensus sequences appear to be, Box A: 5' - TTTA$^{T}/_{A}$ATA - 3' and Box B: 5' - $^{A}/_{T}$TG$^{A}/_{C}$, transcription generally being at the G or a neighbouring purine (Thomm & Wich, 1988; Zillig *et al.*, 1988). Boxes A and B have been determined for three rRNA genes from *Tp.acidophilum*, the consensus for Box A being 5' - CTTATATA - 3' and the Box B sequences being 5' - TTGC -3', 5' - TTCG -3' and 5' - TAC - 3' (Ree & Zimmerman, 1990). Of the other cloned genes from *Tp.acidophilum*, no positive identification of these sequences has been made, but Sutherland *et al.* (1991) tentatively identified Box A as 5' - TTTAATCT - 3', 25 bp upstream of a putative Box B, 5' - ATGT - 3'.

Transcription of archaebacterial genes appears to terminate beyond inverted repeat sequences whose transcripts have the potential to form stem loop structures and which are followed by tracts of thymidine residues (Brown *et al.*, 1989) These thymidine-rich termination sequences resemble the sites at which eukaryotic RNA

polymerase III terminates in the transcription of tRNA and 5S RNA. Identification of these termination regions has been accomplished by sequence analysis of the 3'-termini of mRNA transcripts and examination of these sequences for conserved regions. However, there are instances of transcription termination occurring at discrete sites having no obvious potential for the formation of hairpin-loop structures.

7.1.2 Potential Transcription Initiation and Termination Sites Associated with the *gld* Gene

Examination of the ≈340 bp sequence upstream of the *gld* gene (Figure 7.1) reveals a number of candidate Box B sequences, but these are of little significance without the presence of a corresponding candidate Box A sequence. Only one A/T-rich region exists before the start of the *gld* gene, running from base -34 to base -22 (highlighted in Figure 7.1), but this is very close to the start of the *gld* gene and there is no obvious Box B lying 25 bp downstream of it.

A search for inverted repeat sequences reveals the presence of a number throughout the cloned DNA fragment (Figure 7.2). However, only one of these (between bases 1192 and 1207) lies downstream of the *gld* gene and this is not followed by any thymidine-rich sequences. There are four thymidine-rich sequences in the portion of the cloned fragment upstream of the *gld* (bases -343 - -274) and two of these also coincide with inverted repeat sequences.

```
gaattctttcaatgaaacctacttttagacatttttgtgaagaatatgcaccgtaatcagcatttgatttacttgatccattggatcgat
     -340       -330       -320       -310       -300       -290       -280       -270       -260

tgtaaacgaacgatccatgaaatccgcgagtcaaaattttccacatcattattggttatgtaaacatcacaataaatccttttgatccga
     -250       -240       -230       -220       -210       -200       -190       -180       -170

tttcacttctagcaaaacatgcaccaggacgctgctcagggacatcgtttccttcatccgtgcaatatgcgggcatcatatcgaaggcca
     -160       -150       -140       -130       -120       -110       -100        -90        -80
                                                                                  T  E  Q

acattgcgtccaataaacggttctcgaaatgcgcaacacattagaaattaattaatagaatctagcattcccgcacttATGACTGAACAG
     -70        -60        -50        -40        -30        -20        -10         1         10
```

Figure 7.1: The DNA sequence upstream of the *gld* gene; the A/T-rich region is boxed, 'Box B'-like sequences are underlined; there are no obvious Box A-like sequences associated with any of the latter. The thymidine rich tracts are overlined.

7.1.3 Discussion

Firstly, it is not possible to identify promoter sequences associated with the *gld* without performing the sort of primer extension and nuclease protection experiments described earlier. A single AT-rich region was identified and if this is significant then the apparent lack of a Box B sequence is probably not important,

Upstream

```
                T
               G-G
               T-A
               T-A
               T-G
               T-A
               T-A
               A-T
-324 - TTAGAC  ATGCAC - -298

                A
               C-T
               G-T
               A-T
               C-G
               T-A
               A-T
               A-T
-300 - CACCGT  TACTTG - -274

                T
               A-T
               C-G
               T-A
               A-T
               G-C
               T-G
               T-A
-282 - ATTTAC  TTGTAA - -254
```

Downstream

```
                T
               C-G
               G-C
               G-C
               A-T
               G-C
               T-A
1189 - AGATAA  ATATAT - 1213
```

Figure 7.2: Inverted repeat sequences identified upstream and downstream of the *gld* gene.

side specificity of the recombinant glucose dehydrogenase or, low amounts of or less active forms of the β-galactosidase product of the *lacZ* gene were being produced. Examination of the 3'end of the cloned DNA fragment revealed the presence of an O.R.F. which by chance was in frame with the *lacZ* gene (Figure 7.3). As was mentioned in 7.1.5, two possible start codons were found to be present in this O.R.F., the ubiquitous ATG and the less common GTG, which has been reported being used by some archaebacterial genes.

The presence of this O.R.F. and its putative translation initiation sites provides the simplest explanation for the observation that the recombinant cells were blue, in that a catalytically active fusion protein is being formed. The apparent low level of expression (it was noted in 6.3.3 that the cells transformed with pUC18 alone were considerably darker blue) is supported by the fact that the putative initiation sites do not obviously resemble ribosome-binding site sequences, as was the case with the start of the *gld* gene.

```
                (m) d  l  e  k  y  d  d  i  s (m) k  d  l  l  k  r  i  d  d  l  a  k  d
gaagatcggatagccgtggacctggaaaagtatgatgacatttcgatgaaagaccttctgaagaggatagacgatctagcgaaggat
      1250      1260      1270      1280      1290      1300      1310      1320

 P  L  E  S  T
ccTCTAGAGTCGACC
```

Figure 7.3: The second O.R.F. in the cloned *Tp.acidophilum* DNA fragment; the two possible initiation codons are marked by (m) in the amino acid sequence. Lower case letters refer to the O.R.F. and its derived amino acid sequence, upper case to codons and amino acids from the *lacZ*; putative ribosome-binding sites are underlined.

7.2 Amino Acid Sequence Analysis of *Tp.acidophilum* Glucose Dehydrogenase

7.2.1 Introduction

The following sections are devoted to the computer-assisted analysis of the derived amino acid sequence of *Tp.acidophilum* glucose dehydrogenase. Glucose dehydrogenases have been characterised or isolated from a number of sources. These include *Pseudomonas*, *Serratia* and *Klebsiella* (Matsushita *et al.*, 1980), an alkalophilic *Corynebacterium* (Kobayashi & Horikoshi, 1980), species of *Cyanobacterium* (Pulich *et al.*, 1976) and mammalian liver (Campbell *et al.*, 1982). All these enzymes use nicotinamide cofactors. Additionally glucose dehydrogenases which utilise the cofactor pyrolloquinoline quinone (PQQ) have been isolated from *E.coli* and from *Acinetobacter calcoaceticus* . Unfortunately, there is not a great deal

because it has become clear that there is only a very low level of sequence conservation amongst Box B sequences. What is of greater concern, if this AT-rich region is significant, is its proximity to the start of the *gld* gene, as it barely leaves room for transcription to be initiated upstream of the *gld* gene coding region. However, in *Halobacterium halobium*, transcription initiation of the bacterio-opsin-protein encoding gene (*bop*) does occur only three nucleotides upstream of the coding region of the gene, and occurs adjacent to the start codons of the bacterio-opsin-related gene (*brp*) and the halo-opsin gene (*hop*) (Betlach *et al.*, 1984; Blanck & Oesterhelt, 1987; Dunn *et al.*, 1987).

There are no obvious signs of any transcription termination signals downstream of the *gld* gene, in that the only inverted repeat sequence has no associated tracts of thymidine residues. However, there may be a site of transcription termination at the very start of the cloned DNA fragment, which is quite thymidine-rich, especially if an inverted repeat sequence exists just outside the cloned DNA fragment. Of particular note is the sequence 5' - TTTTTGT - 3' (nucleotides -317 - - 311) which closely resembles a site of transcription termination identified in *Sulfolobus* spp., namely 5' - TTTTT$^{C}/_{T}$T (Reiter *et al.*, 1988).

7.1.4 Archaebacterial ribosome-binding sites

In eubacteria, translation initiation requires the hybridisation of a nucleotide sequence at the 3'-terminus of the 16S rRNA molecule, to a complementary sequence located a few nucleotides upstream of the translation initiation codon; this complementary sequence is the ribosome binding site (Shine & Dalgarno, 1974). The majority of instances of translation initiation in archaebacteria follow this pattern; however, in a few cases the sequence with most complementarity to the 3'-terminus of the 16S rRNA lies just downstream of this initiation codon.

The most frequent site of initiation of translation in archaebacteria is the ATG triplet, as in eubacteria and eukaryotes, but GTG is also used and there is evidence for TGT being employed very occasionally (Bokranz & Klein, 1987; Cubellis *et al.*, 1988). The entire nucleotide sequence of the 16S rRNA from *Tp.acidophilum* is known (Ree *et al.*, 1988) and from that the recognition sequence near the 3'-terminus has been identified as 5' - AUCACCUCC - 3', the complementary binding sequence therefore being 5' - GGAGGTGAT - 3'. The genetic code used by the archaebacteria has not been shown to be different from that used by other organisms and, consistent with this finding, the sites of termination of translation are encoded by the stop codons UGA, UAA and UAG.

7.1.5 Open Reading Frames within the Cloned DNA Fragment and Associated Possible Ribosome Binding Sites

The entire nucleotide sequence of the cloned DNA fragment was searched for O.R.F.s and one was substantially larger than the rest. Contained within this is a nucleotide sequence corresponding to the N-terminal amino acid sequence of the encoded protein (previously determined) and this nucleotide sequence was found adjacent to an ATG codon. At the end of this O.R.F is a single TAA stop codon, but this is followed, in frame, by two consecutive stop codons TGA-TAA, six nucleotides downstream.

Attempts to identify a ribosome binding site based on sequence complementarity to the 3'-terminus of the *Tp.acidophilum* 16S rRNA sequence failed to highlight any such sequence at the start of the *gld* gene. However, for reasons that are discussed in 7.1.6, a ribosome-binding site was known to be present, so a short sequence, 5' - AGCA - 3' (nucleotides -11 - -15), which resembles the Shine-Dalgarno sequence 5' - AGGA - 3', was tentatively identified as being the binding site.

Additionally, for reasons discussed in 7.1.6, a candidate O.R.F. was identified which contained a GTG codon at 1265, an ATG codon at position 1285 and which ran to the end of the cloned fragment (Figure 7.3). Assuming that the conclusion drawn in 7.1.6 is correct, this 'O.R.F.' must also have an associated ribosome-binding site. Again, no sequence complementary to the 3'-terminus of the *Tp.acidophilum* 16S rRNA was evident so sequences upstream of these two possible start codons were tentatively identified as ribosome binding sites: 5' - GATCGGAT - 3' (1243 - 1250) and 5' - ATGATGA - 3' (1271 - 1277).

7.1.6 Discussion

When the *Tp.acidophilum* DNA fragment in construct pTaGDH1 was originally cloned, no glucose dehydrogenase activity, above that of the control, was evident (6.1.2). When its nucleotide sequence was determined, the sequence was analysed using the Amersham Staden-Plus™ software package and this revealed the O.R.F which has been identified as being the *gld* gene (5.2.7). Computerised analysis of the nucleotide sequence revealed several features, but it failed to identify any eukaryotic or eubacterial promoter or ribosome-binding sites. The subsequent expression of glucose dehydrogenase activity using construct pTaGDH2 (6.2) confirmed the absence of any recognisable promoter, but did demonstrate the presence of a ribosome-binding site associated with the *gld* gene.

The positive recombinants chosen when trying to colour-select the pTaGDH2 construct were pale blue (6.3.3); this indicated the presence of low levels of β-galactosidase activity in the recombinant cells. Either this activity was some unusual

of amino acid sequence information about these various enzymes but what is available has been used for sequence comparison.

In Section 1.3.2 the molecular basis of nicotinamide cofactor specificity was discussed and, on the basis of these structural motifs and utilising secondary structural predictions, an attempt is made to identify these motifs in the *Tp.acidophilum* glucose dehydrogenase.

7.2.2 Database Searching

Two algorithms, Wordsearch and Fasta, are provided in the Wisconsin GCG Software (Devereux *et al.*, 1984) for database searching. Wordsearch presents whole sequence alignments, highlighting overall sequence similarity, whereas the algorithm Fasta presents partial sequence alignments where windows of high similarity occur. Of the forty best alignments from the NBRF database using Wordsearch the only dehydrogenase sequence which was pinpointed was leucine dehydrogenase from *Bacillus stearothermophilus* with a level of identity of only 20%. Of the forty sequences giving the best windows of similarity using Fasta, five were dehydrogenase sequences; the partial alignments for these sequences are given in Figure 7.4.

7.2.3 Discussion

With the exception of the small segment of PQQ-linked glucose dehydrogenase sequence, none of these alignments shows greater than 30% identity, so it would be wrong to attach too much significance to any conclusions drawn from this database search. However, the fact that Wordsearch failed to find any other dehydrogenase sequence with a high degree of overall similarity to *Tp.acidophilum* glucose dehydrogenase serves to emphasise the unusual nature of this enzyme and possibly to emphasise the distance between archaebacterial, eubacterial and eukaryotic enzymes.

Two features of the database search using Fasta are striking:

i) the first four enzymes all align to a region of *Tp.acidophilum* glucose dehydrogenase between (approximately) residues 150 and 250

ii) a small segment of the PQQ-linked glucose dehydrogenase from *A.calcoaceticus* aligns with a very high level of identity to *Tp.acidophilum* glucose dehydrogenase.

The first observation will be discussed further in section 7.3.3 along with the results of secondary structural predictions. With regard to the second observation it seems more than just coincidence that two proteins, whose only functional similarity is the ability to bind glucose (their cofactors are quite different), display such a high level of similarity in such a localised area.

Dihydrolipoamide dehydrogenase from *Saccharomyces cerevisiae*:

22.4% identity in a 58 amino acid overlap

```
            170        180          190        200        210        220
Tpac.P AFEVFDVVSKRSIFFGDDSTL-IGK--RMVIIGSGSEAFLYSFAGVDRGFDVTMVNRHDETEN
       : : ::::::: | :: :|| |:|   :||||:|::::: :: ::: ||::: |::::
A30151 MLRIRSLLNNKRAFSSTVRTLTINKSHDVVIIGGGPAGYVAAIKAAQLGFNTACVEKRGKLGG
                10         20         30         40         50         60
```

NADH-ubiquinone oxidoreductase (mitochondrial) from *Neurospora crassa*:

26.3% identity in a 38 amino acid overlap

```
          150        160        170        180
Tpac.P ELGRIAVLTEPLKNVMKAFEVFDVVSKRSIFFGDDSTLIGKR
       || :::::::|| ::|:: ::: :::  :|:|   |||
A23431 YLGLLQAFADALKLLLKEYVALTQANMTLFFLGPVITLIFSL
       50         60         70         80         90
```

NADH-ubiquinone oxidoreductase (mitochondrial) from *Paramecium tetraurelia*:

18.9% identity in a 74 amino acid overlap

```
            170        180        190        200        210        220        230
Tpac.P AFEVFDVVSKRSIFFGDDSTLIGKRMVIIGSGSEAFLYSFAGVDRGFDV-TMVNRHDETENKLKIM-DEFGVKFANYL
       |: ::  | :::::  : | :|: :|::::    ::||::: :: | : |:| ::::: :  ::: :||:||:
S07754 FFGSYSKKSIHAVIYFVAWTQLGSLFVLLAC---LYIYSLTNSTNFFVIKTFVFSKTQAMTIYSLLFVGFGIKFPIWP
       160        170        180        190        200        210        220        230
```

D-lactate dehydrogenase from *Escherichia coli*:

28.3% identity in a 53 amino acid overlap

```
             200        210        220        230        240        250
Tpac.P IGSGSEAFLYSFAGVDRGFDVTMVNRHDETENKLKIMDEFGVKFANYLKDMPEKIDLLVDT
       ||:|||::||::: ::    |:: ||:|: | :  ::  : ::: :::| :||
Deecdl PEEGSKAFLHRFAAAGAAIRYQAVHS-DEVEDILALDIALRRNDTEWYEHLPPEIDSQLVH
         430        440        450        460        470        480
```

Glucose dehydrogenase (PQQ-linked) from *Acinetobacter calcoaceticus*:

61.5% identity in a 13 amino acid overlap

```
       20         30         40
Tpac.P TIDMPEPEHYDAKLSPVYIGICS
          |  |||||: ||:
S00943 SPNAWAPLAYDAKLDIVYVPTGV
         410        420
```

Figure 7.4: Partial sequence alignments of *Tp.acidophilum* glucose dehydrogenase with other dehydrogenase sequences using the Fasta algorithm. Bars mark sequence identities and colons mark conservative mutations (based on mutations most commonly observed in cytochromes).

7.2.4 Alignment of glucose dehydrogenase sequences

Using the Wisconsin GCG algorithm Bestfit, an attempt was made to align *Tp.acidophilum* glucose dehydrogenase with the other glucose dehydrogenase sequences. Little sequence information exists for glucose dehydrogenase enzymes and no archaebacterial glucose dehydrogenases have been sequenced; however, it is available for the two PQQ-linked enzymes from *E.coli* (Cleton-Jansen *et al.*, 1990) and *A.calcoaceticus* (Cleton-Jansen *et al.*, 1988), for a flavoenzyme from *Drosophila melanogaster* (Cavener & Krasney, 1991) and for two dual nicotinamide cofactor

specific enzymes from *Bacillus subtilis* (Lampel *et al.*, 1986) and *Bacillus megaterium* (Jany *et al.*, 1984). The levels of identity arising from aligning these sequences with the *Tp.acidophilum* sequence are given in Figure 7.5.

Tp.acidophilum glucose dehydrogenase versus:	
i) PQQ-linked enzyme from *E.coli*:	16.7 % identity
ii) PQQ-linked enzyme from *A.calcoaceticus*:	19.6 % identity
iii) flavoenzyme from *D.melanogaster*:	19.5 % identity
iv) NAD(P)$^+$-linked enzyme from *B.subtilis*:	17.5 % identity
v) NAD(P)$^+$-linked enzyme from *B.megaterium*:	21.1 % identity

Figure 7.5: Data from pairwise Bestfit alignment of glucose dehydrogenase sequences.

7.2.5 Discussion

It seems reasonable to state, on the basis of these results, that *Tp.acidophilum* glucose dehydrogenase bears no overall sequence resemblance to other known glucose dehydrogenases. This might be expected with the three enzymes that are functionally dissimilar, but it is striking how little resemblance is borne to the functionally similar enzymes from the two *Bacillus spp.* especially as they share the unusual dual nicotinamide cofactor specificity and are tetrameric in their native state. If this attempt at alignment has achieved anything, it has served to emphasise the lack of similarity that exists between archaebacterial and eubacterial genes; a recent extensive comparison of glyceraldehyde-3-phosphate dehydrogenases from all three primary kingdoms (Hensel *et al.* 1989) concluded that there was a much more striking level of homology between eukaryotic and eubacterial enzymes, than between either and archaebacterial dehydrogenases. However, unequivocal homology was observed in functionally important regions.

A tetrameric NAD(P)$^+$-linked glucose dehydrogenase from the thermoacidophilic archaebacterium *Sulfolobus solfataricus* has been isolated and characterised (Giardina *et al.*, 1986) and, when sequence information is forthcoming, it will be of interest to see if this bears any sequence similarity to the *Tp.acidophilum*

enzyme (*S.solfataricus* glucose dehydrogenase subunit M_r ≈30,000, which implies it is more like the *Bacillus* enzymes).

The PQQ-linked enzymes from *A.calcoaceticus* and *E.coli* are very highly homologous (Cleton-Jansen *et al.*, 1990) so it is not surprising that the *E.coli* enzyme also shows similarity to *Tp.acidophilum* glucose dehydrogenase at the same region as the *A.calcoaceticus* enzyme. This small region of homology amongst the three sequences is shown in Figure 7.6; it would be interesting to discover if this does play a rôle in glucose binding, though it was noticeably absent from the *Bacillus* sequences.

```
I D M P E P E H Y D A K L S P V Y I G I C    Tp.a.
P N A W A P L A Y D A K L D I V Y V P T G    A.calc.
P N S W A P A A Y D A K L D L V Y L P M G    E.coli
```

Figure 7.6: A small segment of sequence homology between *Tp.acidophilum* glucose dehydrogenase and the PQQ-linked glucose dehydrogenases from *A.calcoaceticus* and *E.coli*.

7.3 A Possible Cofactor-Binding Domains in *Thermoplasma acidophilum* Glucose Dehydrogenase

7.3.1 Introduction

The Bestfit alignment is of little use in identifying small domains of high similarity, as picked out by the Fasta database search, so in an attempt to define portions of the sequence that strongly resembled portions of the two *Bacillus* sequences, Diagon plots were generated using the Compare algorithm and the Dotplot programme. These data are not presented because they were inconclusive; at low levels of stringency, windows of similarity were scattered all over the plots and, as the level of stringency was increased, it was hoped that windows of similarity remaining would be common to both the comparative plots. This was not the case, and in view of the level of sequence identity between the two *Bacillus* sequences, if anything significant had been present it should have been highlighted by both sequences.

7.3.2 Secondary Structural Prediction

The PeptideStructure algorithm (Jameson & Wolf, 1988) was used in an attempt to gain some information about the structure of *Tp.acidophilum* glucose dehydrogenase. The Chou-Fasman predictions, (designed for globular proteins) for the enzyme are given in Figure 7.7. Additionally, the glucose dehydrogenase sequence was searched manually for the consensus sequence in the fingerprint region of the ADP-binding βαβ fold, namely 'Gly-X-Gly-X-X-Gly'(Wierenga *et al.*, 1985), but no such motif exists. However, a Gly-X-Gly-X-X-Ala motif lies at residues 192 - 197; this region of *Tp.acidophilum* glucose dehydrogenase is shown aligned to some of the aligned sequences from Wierenga *et al.*, (1985) in Figure 7.8. In section 7.3.3 the likelihood of this region of the enzyme being part of the cofactor binding domain is discussed.

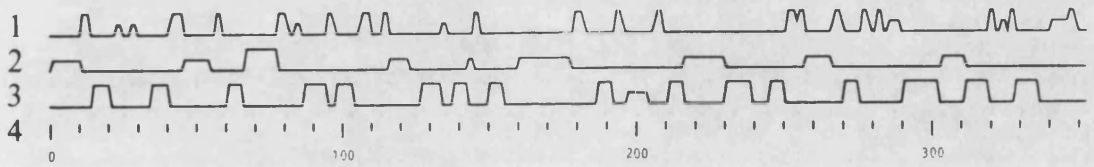


Figure 7.7: Chou-Fasman secondary structural predictions: (1) β-turn, (2) α-helix, (3) β-sheet, (4) scalar

7.3.3 Discussion

The Gly-X-Gly-X-X-Gly motif (marked with a filled circle in Figure 7.8), described by Wierenga *et al.*, (1985) is characteristic of the βαβ fold of the ADP-binding domain of the cofactor binding site. The first glycine is so far entirely conserved as it ensures a tight turn between the first β-strand and the first α-helix. The second glycine is highly conserved because a larger side-chain would cause steric interference with the bound nucleotide. The third is highly conserved because it allows space for a close interaction between the α-helix and the β-sheet; however, in glutathione reductase, an NADP-linked enzyme, an alanine lies at this position causing a deviation of the main chain from that normally found in the fold; such a Gly-X-Gly-X-X-Ala sequence is present in *Tp.acidophilum* glucose dehydrogenase (residues 192 - 197) and is the only one present which resembles this motif. The fact that the partial alignments of the dehydrogenase sequences, arising from the Fasta database search (7.2.2), all aligned to a region in which this sequence lies is encouraging evidence that this might be important in cofactor binding. However,

further features of this 'fingerprint' region discussed by Wierenga *et al.*, (1985), make the evidence for this being the cofactor-binding site very compelling.

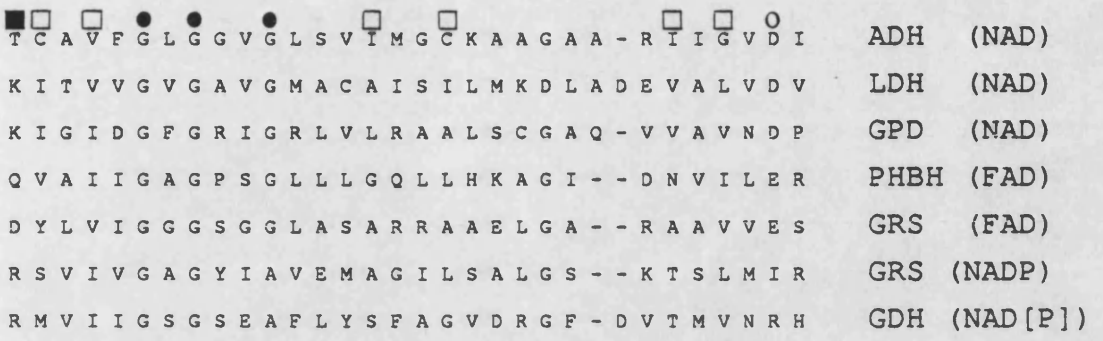


Figure 7.8: A segment of *Tp.acidophilum* glucose dehydrogenase aligned with a multiple alignment of other dehydrogenase sequences taken from Wierenga *et al.* (1985). ADH: alcohol dehydrogenase - horse liver (Jornvall, 1970); LDH: lactate dehydrogenase - dogfish muscle (Taylor, 1977); GPD: glyceraldehyde-3-phosphate dehydrogenase - lobster muscle (Davidson *et al.* 1967); PHBH: *p*-hydroxybenzoate hydroxylase - *Pseudomonas fluorescens* (Weijer *et al.*, 1982); GRS: glutathione reductase - human erythrocytes (Krauth-Siegel *et al.*, 1982); GDH - glucose dehydrogenase - *Thermoplasma acidophilum* (Bright *et al.*, submitted).

The residues marked with squares are predominantly hydrophobic and form the hydrophobic core of the βαβ unit; five out of six of these residues in glucose dehydrogenase are hydrophobic. The exception, the serine residue at position 201, is also found at the equivalent position in NAD-linked alcohol dehydrogenase from *Drosophila melanogaster* (Thatcher *et al.*,1980).

As was discussed in section 1.3.2, an Asp or a Glu residue (marked with an open circle in Fig 7.8) always lies at the end of the second β-strand , the side chain of which interacts with the free 2'-hydroxyl group of NAD. This is replaced by either a neutral or a positively charged residue in NADP-binding enzymes, to permit interaction with the 2'-phosphate of NADP; *Tp.acidophilum* glucose dehydrogenase has an Arg residue at this position (216), which is also found at the equivalent position in dihydrofolate reductase (Bitar *et al.*, 1977; Bolin *et al.*, 1982). Finally a hydrophilic residue is conserved at the amino terminal of the first α-helix (marked with a filled square in Fig.7.8) which interacts with the pyrophosphate bridge. This residue in glucose dehydrogenase is an Arg, which is also found in the NADP-linked glutathione reductase (Scrutton *et al.*, 1990).

The only piece of evidence which detracts from these observations arises from the Chou-Fasman predictions of secondary structure (Figure 7.7). If the above conclusions are correct, an α-helix should be present between residues 194 and 208 and the predictions define this area as β-sheet. However, such predictions are not always accurate; approximately 65% of predicted structural features concur with those shown to be present in determined 3-dimensional structures (G.L. Taylor, personal communication)

Assuming that these conclusions are correct then the underlying reason for the preferential NADP-linked activity, and the undoubted coexistence of NAD-linked activity might be as follows: Pai *et al.* (1988) reported that human NADP-linked glutathione reductase bound NADH in a similar way to NADPH, except that an inorganic phosphate ion substituted for the missing 2'-phosphate group. It was noted above that *Tp.acidophilum* glucose dehydrogenase has an Ala residue which is normally a Gly in other dehydrogenases, and which in human glutathione reductase causes the main chain of the first β-strand and α-helix to adopt a more open conformation (Wierenga *et al.*, 1985). This may provide space for the binding of NAD and an inorganic phosphate moiety. The added presence of an Arg residue (216) at the end of the second β-strand may bind the free inorganic phosphate, permitting apparent NAD-linked activity. The NAD-linked activity of the native enzyme was demonstrated by Budgen (1988) and Smith (1989) using 50 mM sodium phosphate buffer in the standard assay, therefore ample inorganic phosphate would have been present.

7.4 *Thermoplasma acidophilum gld* - Part of an Operon?

7.4.1 Introduction

This short speculative section arises from the apparent existence of an O.R.F. at the very end of the cloned DNA fragment, which was suggested to have been responsible for a low level of β-galactosidase activity, arising from the formation of a fusion protein with the product of the *lacZ* (7.1.3).

7.4.2 Database Searching

It was speculated that the O.R.F. fragment in question could have two possible sites of initiation of translation. The shorter translation product, commencing with an ATG triplet at nucleotides 1285 - 1287, has the derived sequence H_2N-(M)-K-D-L-L-K-R-I-D-D-L-A-K-D-, prior to its fusion with the translation product of the *lacZ*

gene. WordSearch and Fasta algorithms failed to find a high level of similarity between this sequence and any sequences in the NBRF database.

The longer translation product, commencing from the GTG triplet (nucleotides 1255 -1257), has the sequence: H_2N-(V)-D-L-E-K-Y-D-D-I-S-M-K-D-L-L-K-R-I-D-D-L-A-K-D-, prior to β-galactosidase. A Fasta search picked out a galactose-binding protein from *E.coli* (Mahoney *et al.*, 1981) with 41 % identity (Figure 7.8).

```
8 - T I Y K Y D D N F M S V V R K A I E Q D A K - 29  E.coli
        |   | | | |     |     | |   |   |     | |
1 - D L E K Y D D I S M K D L L K R I D D L A K - 22  Tp.a.
```

Figure 7.8: *Tp.acidophilum* O.R.F. fragment, aligned with a galactose-binding protein from *E.coli*.

7.4.3 Discussion

The possibility of *gld* being part of an operon structure arose because of the identification of this O.R.F. fragment only ≈180 nucleotides downstream of the *gld* and because of the apparent lack of any transcription termination sequences between the two. No promoter sequences were highlighted in this region either and it would be unusual for two independently transcribed genes to be so close together. The galactose-binding protein from *E.coli*, picked out by the database search, lies in the periplasmic space of the cell and is responsible for binding dissolved galactose prior to passing it to membrane bound sugar-transport proteins. *Tp.acidophilum* has no cell wall and therefore no periplasm, but conceivably this O.R.F. fragment could encode part of a protein involved in sugar uptake. Smith (1989) demonstrated that *Tp.acidophilum* glucose dehydrogenase can oxidise galactose at about 50 % of the rate for glucose, so possibly this putative 'transporter' could transport either sugar.

7.5 Concluding Remarks

7.5.1 Gene Structure

Without experiments to pinpoint specific gene-structural features such as sites of initiation and termination of transcription, and in the absence of any obvious homology with established archaebacterial consensus sequences, no firm conclusions can be drawn about the nature of these signals in *Tp.acidophilum*. However, the trend amongst the archaebacteria appears to mimic eukaryotic transcription signals. The

expression of *Tp.acidophilum* glucose dehydrogenase in *E.coli* confirms the presence of other features *ie.* sites of initiation and termination of translation. The latter are identical to those in any other cell-type and the former, though less well defined, are broadly similar to eubacterial sites. The possibility of the *Tp.acidophilum gld* gene being involved in an operon associated with glucose uptake and metabolism is highly speculative, but some circumstantial evidence has been presented.

7.5.2 Protein Structure

The sequence comparisons and analyses performed on the derived amino acid sequence of *Tp.acidophilum* glucose dehydrogenase have generated compelling evidence for the presence of the ADP-binding βαβ fold, described by Wierenga *et al.* (1985), between residues 186 and 216. They have also provided a possible explanation for the enzyme's dual nicotinamide cofactor specificity; it will be interesting if these conclusions are born out by structural studies. The evidence presented for a possible glucose binding domain remains no more than a starting point, so it will be interesting to see if structural studies highlight this region as being important in this rôle.

CHAPTER 8 ENHANCEMENT OF EXPRESSION OF *Thermoplasma acidophilum* GLUCOSE DEHYDROGENASE IN *Escherichia coli*

8.1 Translation of Glucose Dehydrogenase from the *lacZ* Ribosome-Binding Site.

8.1.1 Introduction

If one of the reasons for the relatively low level of expression of the *Tp.acidophilum* glucose dehydrogenase is that *E.coli* ribosomes have a low affinity for its associated ribosome-binding site, then redesigning this site may help to enhance expression. As was discussed in Chapter 7, only a tentative identification of this site was made because of its lack of similarity to consensus sequences. It was felt likely that extensive mutagenesis of the cloned fragment would be required to redesign this site, so a simpler approach was adopted; namely, to manipulate the clone so that the ribosome-binding site associated with the start of the *lacZ* gene could be utilised by the ribosomes attempting to translate the *gld* gene.

A strategy was developed to attempt to delete a segment of DNA upstream of the *gld* gene, then to manipulate the resulting sticky ends, so that when the construct was religated, the start codon of the *gld* gene would be in frame with the start of the *lacZ* gene. This chimera would consist largely of the *gld* gene but would also carry the nucleotides important in providing a good ribosome binding site.

8.1.2 Creation of the Deletion Mutants

Figure 8.1 illustrates this stategy designed to place the start codon of the *gld* in frame with the start of the *lacZ*.

i) 2 μg of the pTaGDH2 construct were incubated with EcoRI (3.3.12) to cut the polylinker at the unique EcoRI restriction site

ii) the overhanging T-T-A-A was partially 'filled-in' by incubation with Klenow fragment and dATP (3.3.19), then the manipulated construct was cleaned using a Sepharose CL-6B spun column (3.3.6)

iii) the construct was incubated with BsmI to cut it at a unique restriction site just upstream of the start of the *gld*, then

iv) the overhanging A-A was removed and the remainder of the construct 'proof-read' for nicks, by incubation with T4 DNA polymerase I and dNTPs (3.3.21)

v) the mixture was purified using a Sepharose CL-6B column then the overhanging T-C removed by incubation with nuclease S1 (3.3.20).

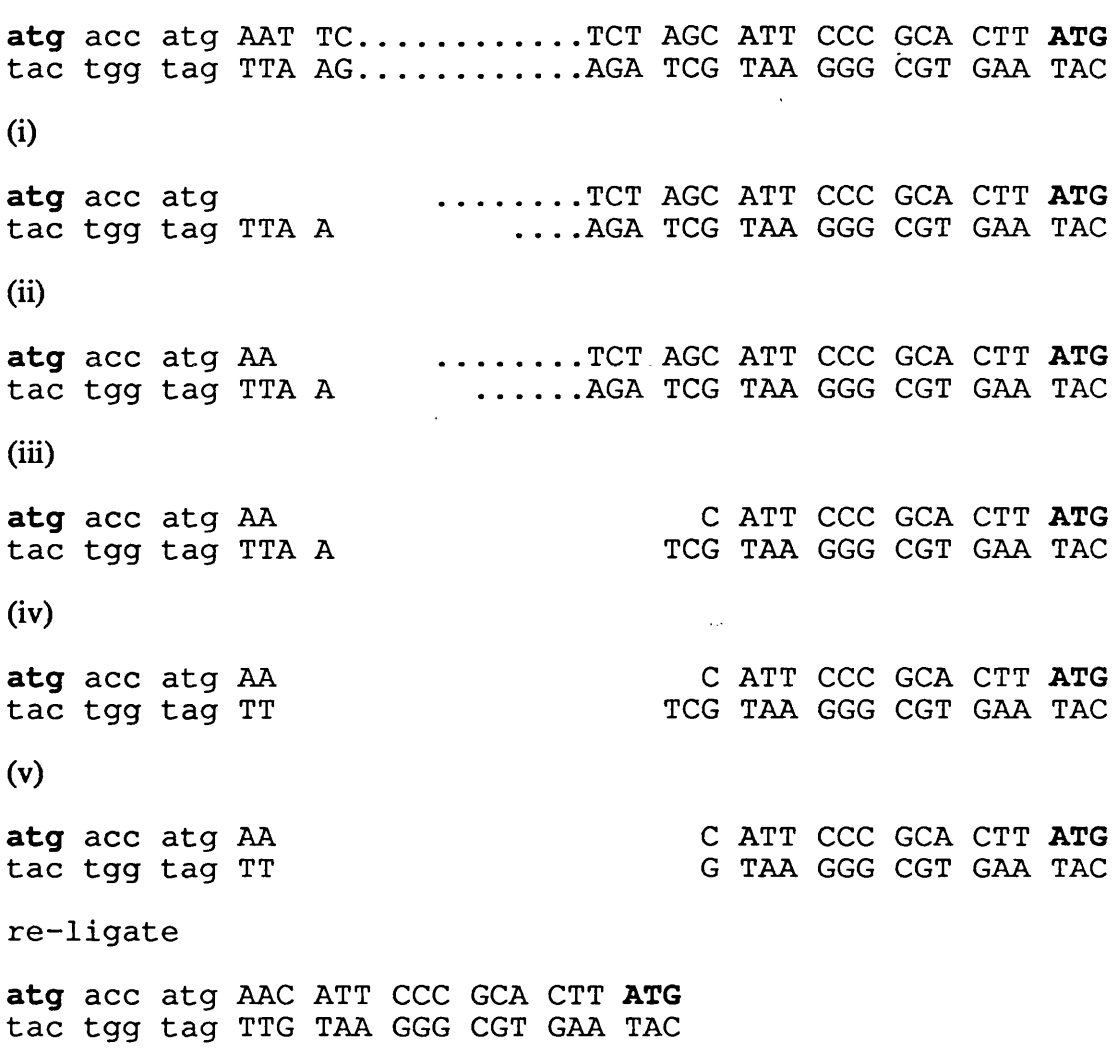


Figure 8.1 Deletion strategy: please refer to 8.1.2 for the various steps involved in the production of this construct. The bold lower case **atg** is the initiation codon of the *lacZ* gene and the bold upper case **ATG** is the initiation codon of the *gld* gene. The dotted lines represent the intervening ≈300 nucleotides which are removed by this procedure.

The manipulated DNA fragment containing the *gld* was separated from the deleted DNA fragment by electrophoresis in a TBE-agarose gel (3.3.2) and isolated from the gel by electroelution onto DE-81 paper (3.3.3).

The manipulated DNA fragment was eluted from the DE-81 paper (3.3.4), purified using a Nensorb20™ column (3.3.6), incubated overnight with T4 DNA ligase and ATP (3.3.14) and the ligation mixture cleaned using a Sepharose CL-6B spun column (3.3.6). One tenth of the cleaned ligation mixture was used to transform

competent *E.coli* TG1 cells (3.1.3) and any positive recombinants identified by colour selection (3.1.4); 14 recombinant clones were picked, these were grown in DYT containing 100µg ampicillin/ml and stored frozen (3.1.5). These putative deletion mutants were designated pTaGDH3A → N.

8.1.3 Glucose Dehydrogenase Expression in Deletion Mutants

E.coli TG1 cells carrying pUC18 only, pTaGDH2 and pTaGDH3A-N were grown for 18 h in 50 ml cultures of DYT containing 100µg ampicillin/ml at the start of incubation, with a further 100 µg ampicillin/ml added after 17 h. As before (6.1.2), cell-free extracts were prepared and $NADP^+$-linked glucose dehydrogenase specific activities were determined (Table 8.1).

	Specific Activity U/mg		Specific activity U/mg
pUC18	0.0080	pTaGDH3G	0.0083
pTaGDH2	8.62	pTaGDH3H	0.0058
pTaGDH3A	0.11	pTaGDH3I	15.4
pTaGDH3B	41.7	pTaGDH3J	0.0079
pTaGDH3C	0.56	pTaGDH3K	0.0087
pTaGDH3D	0.0092	pTaGDH3L	27.0
pTaGDH3E	0.31	pTaGDH3M	0.0094
pTaGDH3F	1.74	pTaGDH3N	0.0079

Table 8.1: Specific glucose dehydrogenase activities in cell-free extracts from cell-lines harbouring deletion mutants pTaGDH3A-N. These are values from single assays only.

8.1.4 Nucleotide Sequence Determination of pTaGDH3A-N

The 14 putative deletion mutants pTaGDH3A-N were 'maxiprepped' (3.3.16), and the nucleotide sequences of the regions of the constructs, where the *gld* gene should meet the *lacZ* gene, were determined by the dideoxy-chain termination method (3.3.18), using primer 5 (see Figure 5.5 for primer binding site); the sequences are presented in Figure 8.2.

8.1.5 Discussion

Assuming that all the positive recombinant clones pTaGDH3A → N did stem from pTaGDH2, then these results demonstrate that only seven of them (A,B,C,E,F,I and L) still retained the primer binding site used in the sequencing reactions.

Deletion of the primer binding site in question would require the removal of the first 110 nucleotides from the *gld* gene and would almost certainly have an effect on the activity of the gene product. Since all of the clones that failed to generate sequence information had very low glucose dehydrogenase activities, indistinguishable from

---

```
int   AGGAAACAGCT ATG ACC ATG ATT ACG AAC ATT CCC GCA CTT atg
act   AGGAAACAGCT ATG AC  ATG ATT ACG AAC ATT CCC GCA CTT atg
A     AGGAAACAGCT ATG AC  ATG               ATT CCC GCA CTT atg
B     AGGAAACAGCT ATG AC  A                             CTT atg
C     AGGAAACAGCT ATG AC  ATG               ATT CCC GCA CTT atg
E     AGGAAACAGCT ATG AC  ATG ATT ACG AA      T CCC GCA CTT atg
F     AGGAAACAGCT ATG AC  ATG                   CCC GCA CTT atg
I     AGGAAACAGCT ATG AC  ATG ATT A         ATT CCC GCA CTT atg
L     AGGAAACAGCT ATG AC  ATG ATT ACG                   •atg
```

Figure 8.2: Nucleotide sequences of the deletion mutants; int: the intended sequence of the construct, act: the actual sequence that would have been obtained due to the absence of a C residue from this stock of pUC18 (Lobet *et al.*, 1989), A,B,C,E,F,I,L: the only recombinants that produced sequence information. The AGGA motif is the core of the ribosome-binding site associated with the *lacZ* gene, the atg is the initiation codon of the *gld* gene and the ATG in each line is the actual site of translation initiation as a result of the deletion procedure.

---

that found in the control, it was felt that the amount of DNA deleted was probably too great to leave a viable construct.

None of the deletion mutants that did display glucose dehydrogenase activities above that of the control had the sequence of the planned construct, which emphasises the lack of specificity of some aspect of the deletion strategy. Earlier unsuccessful attempts to generate these deletion mutants were hampered by the powerful nucleolytic activity of nuclease S1 which removes single stranded DNA both from the termini of DNA fragments and from the midst of a fragment if it contains single strand breaks. The T4 DNA polymerase 'proof-reading' step was introduced in an attempt to overcome this problem.

To explain the different levels of expression observed amongst clones pTaGDH3A,B,C,E,F,I and L, the O.R.F.s of the nucleotides preceding the *gld* gene were examined so as to establish which ATG codon was likely to be acting as the start signal in each construct. In accordance with one of the observations of Shine and

Dalgarno (1975), those with higher levels of expression are those whose start codons are closer to the ribosome-binding site; furthermore, since there are only three possible initiation sites in this region it is not surprising that the levels of expression observed in these clones fall into three ranges.

The specific activity data for those clones displaying elevated glucose dehydrogenase activity demonstrated that the optimum level of expression of the enzyme with this system was ≈13% of cell protein; this is roughly equivalent to 3000 units or ≈10 mg of glucose dehydrogenase per litre of culture. This level of expression was felt to be more than adequate for the purposes of large-scale purification of the enzyme.

8.2 Enhancement of Expression by PCR Mutagenesis of the *gld* Ribosome-Binding Site

8.2.1 Introduction

Having established that an acceptable level of expression of this enzyme was possible with this host/vector combination, it seemed worthwhile to try to manipulate the pTaGDH2 construct in such a way that the glucose dehydrogenase would be expressed alone and not as a fusion protein. The nucleotide sequence near the start of the *gld* gene, thought most closely to resemble that of the highly conserved portion of the Shine-Dalgarno sequence, is A-G-C-A (nucleotides -15 → -12), and it was felt that the use of PCR mutagenesis to make a C → G mutation may go someway to providing the *gld* gene with a better ribosome-binding site.

In order to facilitate this mutagenesis, a primer was synthesised that contained mismatches to a short DNA sequence five nucleotides upstream of the 5'-end of the putative ribosome binding site, as well as to the ribosome-binding site itself, so as to create an EcoRI site at the same time as changing the A-G-C-A sequence. The introduction of an EcoRI site near the start of the gene was done so that the PCR product could be readily cloned, and to permit the accurate deletion of the nucleotides upstream of the *gld* gene.

8.2.2 PCR Mutagenesis of pTaGDH2

Part of the pTaGDH2 construct containing the *gld* gene was amplified and simultaneously mutated using the M13 -40 primer and a mismatched primer, which had been synthesised and purified as described (3.2.1,3). The amplification strategy and magnesium concentrations were as described (3.3.22) and the products were analysed by electrophoresis in TBE-agarose gels. The size of the PCR product was

judged, by comparison with molecular weight markers, to be the predicted size of ≈1.4Kb. The sequence of the mismatched primer and the sequence changes in the amplified product are shown in Figure 8.3.

---

1) a t a g a a t c t a g c a t t c c c g c a c t t A T G
2) C t C g a a t T C a g G a t t c c c g c

Figure 8.3: The sequence changes introduced by PCR mutagenesis of pTaGDH2 to form the fragment carried in pTaGDH4. (1) the sequence immediately upstream of the coding region of *Tp.acidophilum gld*, (2) the mismatched PCR primer sequence introducing an EcoRI site (gaatTC) and a modified ribosome binding site (agGa). The CtC change was introduced to ensure that the 5'-end of the amplified product bound tightly to its complementary strand; if the end of the amplified product 'unzips' restriction of the newly-introduced EcoRI site becomes very difficult.

8.2.3 Cloning of the PCR Product

The 1.4 Kb DNA band (identified in 8.2.2) was excised from the gel and isolated by the freeze-squeeze procedure. ≈100 ng of the PCR product were recovered and then purified using a Nensorb20™ column (3.3.3,6). After incubation with restriction enzymes EcoRI and BamHI (3.3.12), the restricted insert carrying the *gld* gene was separated from the other restriction fragments and buffer salts using a Sepharose CL-6B column. Cloning of the restricted PCR product in pUC18 and *E.coli* TG1 cells and the colour selection of positive recombinants were performed as described (8.2.2).

A positive recombinant was maxiprepped and part of its nucleotide sequence determined using primer 5 and the dideoxy-chain termination method (3.3.18) to ensure that the correct nucleotide changes had been made; this construct was designated pTaGDH4.

8.2.4 Glucose Dehydrogenase Expression from pTaGDH4

E.coli TG1 cells carrying pUC18 and cells carrying the pTaGDH4 construct were grown as described (8.1.3). Cell-free extracts were prepared (3.1.8) and $NADP^+$-linked glucose dehydrogenase specific activity determinations performed (6.1.2); the data are presented in Table 8.2.

	Specific activity (U/mg)	Level of Expression (% soluble cell protein)
pUC18 control	0.0064 ± 0.0018	≈0.0002
pTaGDH2 construct	8.62 ± 0.35	≈2.7
pTaGDH4 construct	52.2 ± 0.9	≈16.3

Table 8.2: Comparative assessment of the level of glucose dehydrogenase expression from cell-lines harbouring plasmid pUC18, or constructs pTaGDH2 or pTaGDH4.

8.2.5 Discussion

The fact that the C → G mutation on its own caused this new construct to display a level of expression comparable to that of the best deletion mutants may have been serendipitous. Unfortunately, any speculation about the changes effected on the *Tp.acidophilum* ribosome-binding site by this mutation, other than that it makes it superficially resemble a Shine-Dalgarno consensus sequence,would be fairly inconclusive without further mutagenesis studies and secondary structural information.

Considerably more sequence and structure-related information is available wtih regard to the translation initiation sites found in prokaryotes; the A-G-G-A recognition sequence is the most conserved feature of such sites, because it lies at the heart of the region of hybridisation between the 16S ribosome and mRNA. It therefore seems quite reasonable, when examining recombinant protein expression in prokaryotic systems, to introduce this feature before anything else.

8.2.6 Conclusions

The reasons for attempting to enhance the level of expression of the cloned enzyme were outlined in section 6.3.4. Having examined whether the chosen host/vector system was capable of this, by creating deletion mutations, PCR mutagenesis was then employed to introduce a single base change in what was felt to be the ribosome-binding site associated with the *gld* gene. The resulting construct fortuitously displayed a level of expression comparable to the best observed using other less satisfactory approaches, so further work on the expression of the cloned enzyme was felt to be unneccessary; purification of the cloned enzyme was seen as the new priority.

CHAPTER 9 PURIFICATION OF RECOMBINANT GLUCOSE DEHYDROGENASE FROM pTaGDH4-BEARING *E.coli* TG1 CELLS

9.1 Thermal denaturation of *E.coli* proteins as a purification step

9.1.1 Introduction

A simple means of separating thermostable proteins from mesostable ones is to denature, aggregate and precipitate the latter by thermal denaturation. If necessary, certain actions can be taken to stabilise the thermostable protein, such as addition of substrates, products or a combination of the two. This technique has been shown to be successful in separating recombinant, thermostable proteins from the majority of *E.coli* proteins with minimal loss of the recombinant protein (R.M. Daniel, personal communication). Cell-free extracts were prepared as described in the experiments designed to examine enzyme expression levels (8.1.3, 8.2.4), and $NADP^+$-linked glucose dehydrogenase activities were determined to monitor the effects of the various purification schemes under investigation.

9.1.2 Thermal Denaturation of E.coli Proteins

100µl samples of cell-free extracts were thermally denatured as described (3.4.4). Pellets were resuspended in 50 mM sodium phosphate pH 7.0, and $NADP^+$-linked glucose dehydrogenase activity determinations performed on the supernatant and the pellet fractions so that the total activity in each could be calculated. Figure 9.1 demonstrates the effect of this heating step on the activity of the protein.

9.1.3 Discussion

A good protein purification step should give a high degree of purification, a high yield of product and be sufficiently gentle to leave the product in its native state. The data displayed in Figure 9.1 demonstrated that this procedure inactivated the enzyme quite extensively above 60°C; however, attempts to separate recombinant *Tp.acidophilum* citrate synthase, expressed in *E.coli*, from *E.coli* proteins and the subsequent analysis of the thermally denatured protein samples by SDS-PAGE, demonstrated that treatment at 70°C caused a significant improvement in the amount of *E.coli* proteins removed (Sutherland, 1990). It was decided that if the approach could be modified to protect the enzyme at temperatures higher than 60°C, thermal denaturation would be a good first-step purification technique to employ.

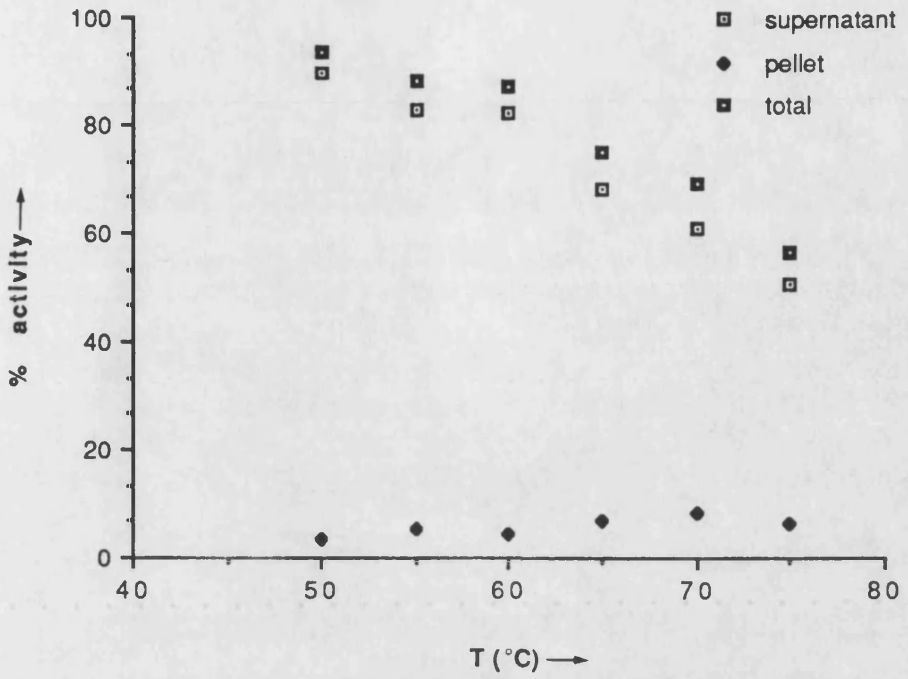


Figure 9.1: The effect of temperature on the activity of recombinant $NADP^+$-linked glucose dehydrogenase; activities are expressed as a percentage of the activity observed in a sample of unheated extract.

9.1.4 Thermal Denaturation in the Presence of Substrate

The addition of substrate to a protein solution, prior to exposing the protein to denaturing conditions, can often afford a considerable stabilising effect (Dignam, 1990). The concentrations of substrate thought likely to be the minimum required to achieve this were 50x - 100x K_M, *ie.* approximately 0.5 - 1 mM $NADP^+$ or 50 - 100 mM D-glucose for *Tp.acidophilum* glucose dehydrogenase (based on the values published by Smith *et al.*,1989). To test this possibility, the heat denaturation experiment was carried out as before (9.1.2) but, prior to the incubation, D-glucose was added to the cell-free extract to a final concentration of 50 mM; the data are presented in Figure 9.2.

9.1.5 Discussion

The graphs in Figures 9.1 and 9.2 show the effect of the presence of glucose on the thermal stability of the enzyme. Thermal denaturation in the presence of substrate clearly continued to affect the enzyme's activity at all of the incubation temperatures examined, but not to the same extent as in the absence of substrate. The use of $NADP^+$ to protect the enzyme against thermal denaturation was not investigated because of the apparently satisfactory result obtained using glucose.

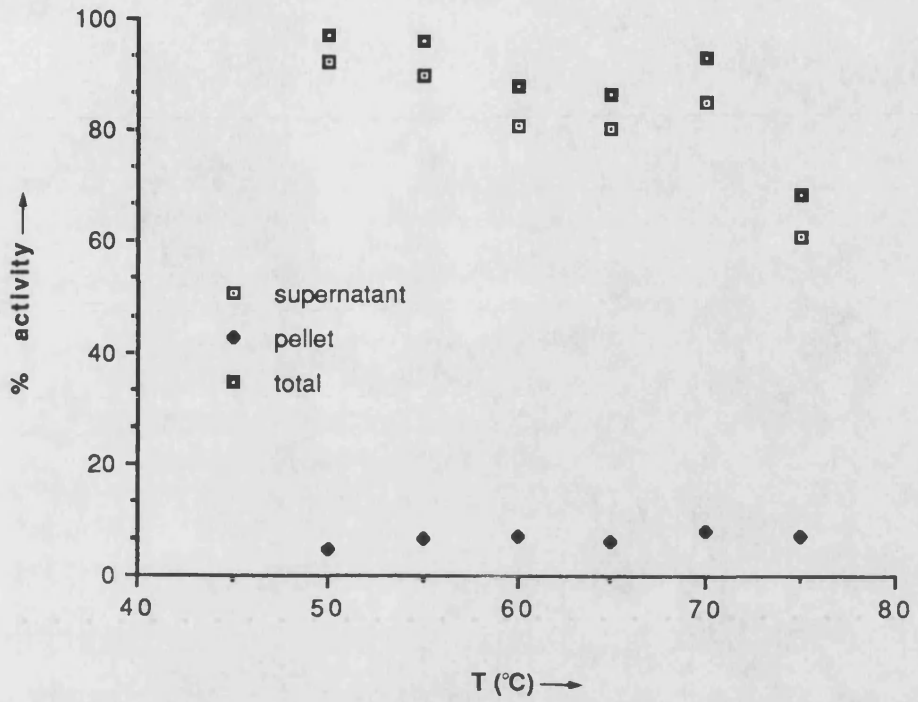


Figure 9.2: The effect of temperature on recombinant $NADP^+$-linked glucose dehydrogenase activity after incubation in the presence of 50mM D-glucose; activities expressed as a percentage of the unheated sample.

From these data it was decided that removal of the majority of the *E.coli* proteins by thermal denaturation, using a 10 minute incubation at 70°C, in the presence of 50mM D-glucose, should be the first step of a large-scale purification scheme.

9.2 Solvent Denaturation of *E.coli* Proteins as a Purification Step

9.2.1 Solvent Tolerance

Smith (1989) investigated the solvent stability of purified *Tp.acidophilum* glucose dehydrogenase and used a methanol precipitation as the first stage of his purification scheme. He found that a substantial proportion of *Tp.acidophilum* proteins was precipitated in 50% (v/v) methanol without loss of glucose dehydrogenase activity. Furthermore, all the glucose dehydrogenase activity was precipitable in 80% (v/v) methanol, but recoverable by resuspension of the precipitate in 20% (v/v) methanol. Therefore, the recombinant glucose dehydrogenase in the supernatant from the heat treatment procedure was examined for its solvent tolerance.

9.2.2 Extraction of the Enzyme with Methanol

Samples of the heat-treated supernatant were solvent denatured as described (3.4.4), using methanol. Pellets were resuspended in 20% (v/v) methanol, both the pellets and supernatants were assayed, and the total activities in each phase were determined. Incubation of the supernatant in the presence of less than 50% (v/v) methanol caused no precipitation of activity, and no overall loss of activity; data from experiments using more than 50% (v/v) methanol are presented in Table 9.1.

methanol (v/v)	%age in sup^t	%age in pellet	%age remaining
50%	123.0	8.0	131.0
60%	89.2	44.2	133.4
70%	0.0	134.2	134.2
80%	0.0	136.3	136.3
90%	0.0	122.5	122.5

Table 9.1: The effect of methanol concentration on recombinant $NADP^+$-linked glucose dehydrogenase activity; activities expressed as a percentage of an untreated sample.

9.2.3 Discussion

It was not felt necessary to perform further solvent denaturation experiments employing substrate protection because of the lack of any detrimental effect on the activity of the enzyme. In the experiments designed to determine the conditions for the removal of *E.coli* proteins and the precipitation of glucose dehydrogenase activity, pelleted proteins were resuspended in 20% (v/v) methanol prior to assay. This was not felt to be satisfactory if the next stage of purification was to employ ion-exchange chromatography, so an attempt was made to recover the pelleted glucose dehydrogenase activity by resuspending the methanolic pellet in 50mM sodium phosphate buffer pH 7.0.

No data are presented for this, but it was possible to recover a high proportion of the precipitated activity by this means. It was concluded that solvent denaturation of the thermally denatured extract should be employed as the second step of a large-scale purification scheme, and should be accomplished by the precipitation of *E.coli* proteins in 50% (v/v) methanol, followed by the precipitation of recombinant glucose dehydrogenase in 80% (v/v) methanol and its recovery by resuspension of the precipitate in aqueous buffer.

9.3 Ion-Exchange Chromatography Using the FPLC System

9.3.1 Introduction

This method of purification was adopted by Smith *et al.* (1989) when purifying *Tp.acidophilum* glucose dehydrogenase from native cell extracts. Smith *et al.* found that the enzyme, loaded on the Mono Q column in 50 mM sodium phosphate buffer pH 7.0, could be recovered with a salt gradient of 0 - 200 mM sodium chloride, and that the glucose dehydrogenase activity was reproducibly eluted at a sodium chloride concentration of ≈120 mM. Therefore, these were the conditions chosen to attempt to purify the recombinant enzyme from the partially-purified extract.

9.3.2 FPLC™ Mono Q Anion-Exchange Chromatography

A typical elution profile is given in Figure 9.3, which demonstrates that the recombinant enzyme behaved in a similar fashion to the native enzyme, in that it did elute at ≈120 mM sodium chloride. Specific activities of fractions displaying glucose dehydrogenase activity were determined and the best of these displayed values in excess of 400 U/mg. However, by pooling fractions with activities >300 U/mg, 65 % of the activity loaded could routinely be recovered.

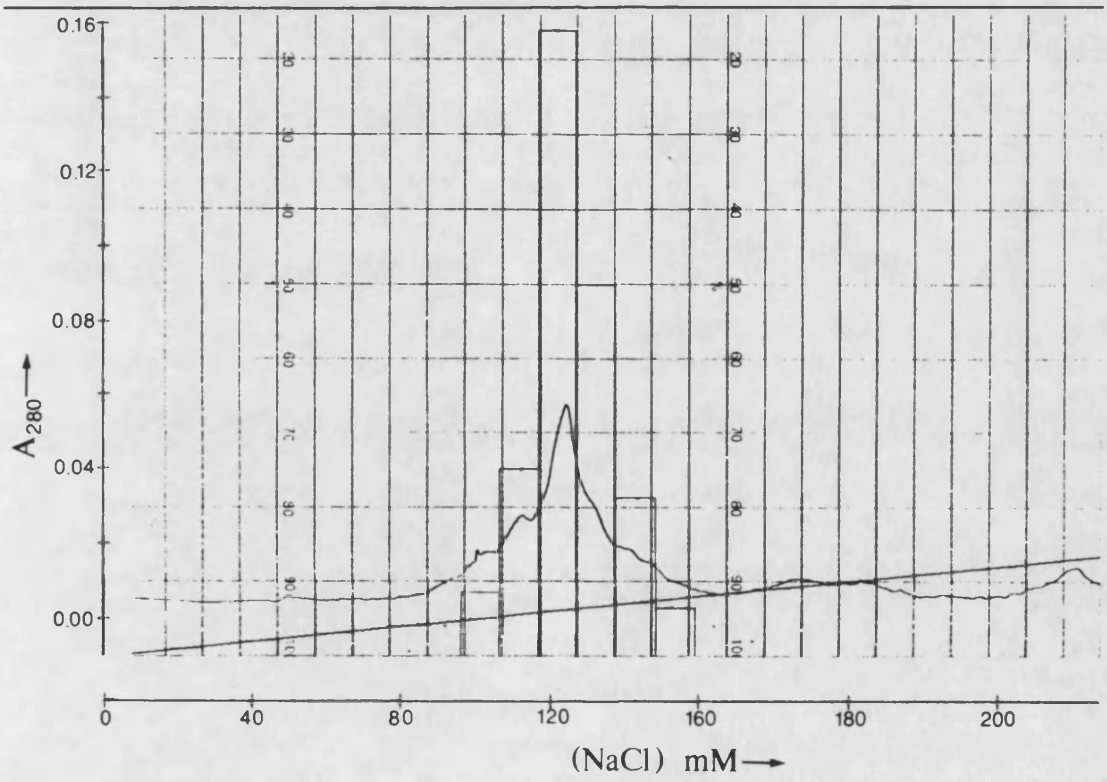


Figure 9.3: Elution profile of glucose dehydrogenase activity from the FPLC™ Mono Q column. The peak of activity elutes at ≈120 mM sodium chloride.

9.3.3 Discussion

Pooling fractions with a specific activity >300 U/mg gave an acceptable yield of the enzyme, which was essentially as pure as that prepared from *Tp.acidophilum* cells by Smith *et al.* (1989) - (320 U/mg), so anion-exchange chromatography was felt to constitute a satisfactory final purification step in a large-scale purification procedure. However, examination of the peak of activity on the elution profile revealed a shoulder on the trailing edge of the peak. It was felt that this might be the result of mild proteolytic activity, occurring whilst the purification procedure was being developed. Efforts to examine this more closely are covered in the subsequent sections of this chapter.

9.4 Large-Scale Purification of Recombinant Glucose Dehydrogenase

9.4.1 Introduction

The various means of purification examined in sections 9.1, 9.2 & 9.3 were used to devise a large-scale purification scheme as follows:

i) to prepare a cell-free extract as described (3.1.8),

ii) to precipitate the majority of host-cell proteins by heat denaturation in the presence of 50 mM D-glucose,

iii) to extract the heat-treated supernatant in 50% (v/v) methanol,

iv) to precipitate the enzyme in 80% (v/v) methanol, harvest the precipitate and recover the activity by resuspending the pellet in 50 mM sodium phosphate buffer, and

vi) to complete the purification by pooling the best fractions from Mono Q anion exchange chromatography

9.4.2 Large-Scale Purification

A cell-free extract was made from 2 litres of culture grown as described (8.1.3, 8.2.4). The heat-denaturation steps, methanol-extraction steps and Mono Q chromatography were performed as described (9.1.2, 9.2.2, 9.3.2), and specific activity determinations were performed on samples from each stage of the procedure. Unfortunately, very little of the enzyme loaded onto the Mono Q column actually bound to it.

The amount of activity retained on the column was determined and the proportion that was eluted and pooled used to determine the yield from the column. This value and the values for specific activity and fold purification are those quoted in Table 9.2, along with the data from the other purification steps. However, the values

for the total activity and total protein eluted from the Mono Q have been extrapolated from the values obtained from the pool of purified enzyme; they therefore appear in parentheses in Table 9.2.

9.4.3 Analysis of the Pure Protein by SDS-PAGE

Figure 9.4 shows SDS-PAGE analysis (using the Flowgen Minigel system) of the purified protein from fractions across the peak of activity eluted from the Mono Q column. Examination of the gel reveals the presence of a doublet where there should be a single band of pure protein; also, this doublet consists of an intense lower band and a less intense upper band.

	Activity (U)	Yield (%)	Protein (mg)	Sp.Ac.(U/mg)	Purification
Extract	11440	100	215	53	-
Heated	6627	58	60	111	2.1.
Methanol	7891	69	34	232	4.4
Mono Q*	(4760)	(42)	(13)	355	6.7

Table 9.2: Purification table for recombinant *Tp.acidophilum* glucose dehydrogenase; values given are for the initial cell-free extract, the extract after heat-treatment, the sample after methanol-treatment and after Mono Q ion exchange chromatography. *Values in parentheses represent the theoretical yield of enzyme after Mono Q ion exchange chromatography, please see 9.4.2.

9.4.4 Discussion

The failure of the Mono Q column to bind all the glucose dehydrogenase activity may be attributable to inappropriate 'scaling-up' of the procedure, though the Mono Q column should be capable of binding approximately 30 mg of protein. Possibly the amount of methanol left associated with the precipitated enzyme contributed to the problem. In any case the SDS-PAGE analysis of the samples eluted from the column revealed a more fundamental problem.

The presence of a 'shoulder' on the activity peak on the Mono Q elution profile was remarked upon in section 9.3.3. This was a consistent feature of the Mono Q elution profiles and, unfortunately, the two protein components revealed by SDS-PAGE analysis of the enzyme do not provide an explanation for this observation, because their relative intensities do not vary across the eluted peak and its shoulder *ie.*

the presence of the shoulder cannot be ascribed to one component eluting marginally later than the other.

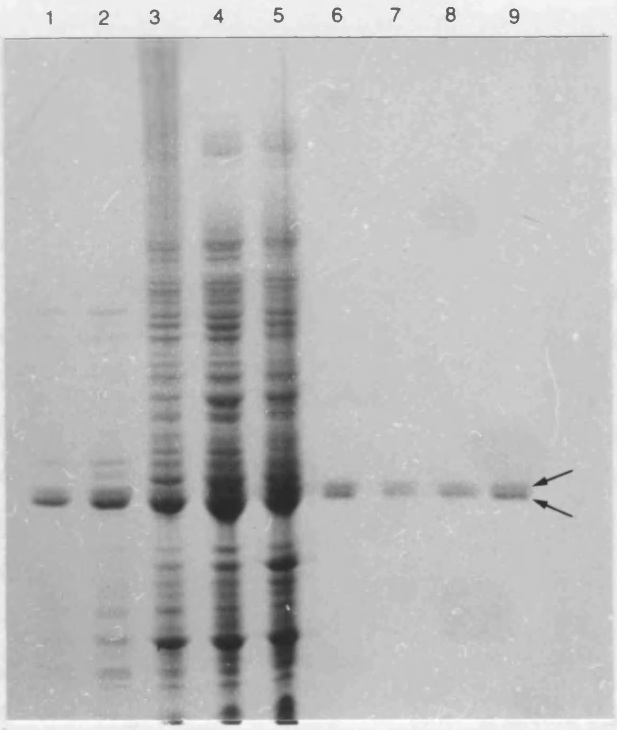


Figure 9.4: SDS-PAGE analysis of the various extracts generated during the purification of recombinant *Tp.acidophilum* glucose dehydrogenase expressed from construct pTaGDH4 in *E.coli*; (1) after methanol extraction, (2) after heat treatment (3) initial extract after centrifugation, (4) & (5) initial extract before centrifugation, (6 - 9) fractions from Mono Q anion exchange chromatography.

For a shoulder to appear at all must mean charge heterogeneity exists within the population of protein molecules. If some proteolysis of the protein had occurred, removing one or two amino acids with charged side-chains, this might explain the presence of the 'shoulder'. Alternatively charge heterogeneity may have been introduced into the protein sample during the heat-denaturation step, as heat can often deamidate asparagine and glutamine residues on a protein surface (1.3.3). Another possible explanation which came forward was the occurrence of glycosylation of surface side-chain amino-groups during heat treatment in the presence of glucose (P. Qasba, personal communication).

The specific activity values (Table 9.2) weigh against the presence of contaminating protein and the SDS-PAGE analysis (Figure 9.4) would tend to confirm this; moreover, whatever the cause of the introduction of charge heterogeneity, it must affect both of the components of the protein doublet equally,

otherwise their relative intensities would vary across the peak and the shoulder. The conclusion that remains is that the two bands present in the doublet are both glucose dehydrogenase subunits.

9.4.5 Gel Filtration Chromatography of the 'Pure' Protein

In order to gather further evidence that the two components of the doublet were both recombinant glucose dehydrogenase, and since the two components of the doublet appeared to co-elute from an anion-exchange column, a small sample of the protein solution was analysed by FPLC™ HiLoad 16/60 Superdex 200 gel filtration chromatography. Although these data are not presented, a single peak of activity was obtained, which when analysed by SDS-PAGE revealed the presence of the two protein components. The likelihood of a second unrelated protein component, which adopted a tetrameric structure in its native state and would therefore co-elute with recombinant glucose dehydrogenase, seemed very remote. Therefore, another explanation was sought as to why two different species of glucose dehydrogenase were being produced from a clonal recombinant cell line.

9.4.6 Alternative Sites of Translation Initiation

The mismatched oligonucleotide that was used to create construct pTaGDH4 by PCR mutagenesis, was designed to accomplish the following:
i) to make the ribosome-binding site associated with the *gld* more recognisable to *E.coli* ribosomes,
ii) to engineer an EcoRI restriction site upstream of the start of the *gld* in order to facilitate the cloning of the mutated product, and
iii) to position this restriction site so that no O.R.F. existed between the start of the *lacZ* and the start of the *gld* when the mutated fragment was ligated into plasmid pUC18.

Examination of the proposed resulting construct (Figure 9.5) reveals that not only is the start codon of the *lacZ* out of frame with the *gld*, but also that the reading frame from the start of the *lacZ* encounters a UGA stop codon at the very start of the *gld*. What is conceivably happening is that protein translation commences both from the ribosome-binding site associated with the *lacZ* and from that associated with the *gld*. When the stop codon at the start of the *gld* gene is encountered, the ribosome must 'wobble', recognise the O.R.F. of the *gld* and continue protein translation in that frame. On the assumption that this is not a very successful procedure, the result would be a heterogeneous protein population, largely composed of the true recombinant protein but containing a significant quantity of this slightly larger putative fusion protein.

```
atg acc atg att acg AAT TCT AGG ATT CCC GAC TTA TGA
```

Figure 9.5: The engineered upstream sequence of construct pTaGDH4. Lower case letters show the start of the *lacZ* gene, upper case letters show the engineered *Tp.acidophilum* sequence joining it at the EcoRI site (gAATTC). Note that the 'atg' of the *lacZ* is not in frame with the 'ATG' of the *gld*.

9.5 An Attempt to Resolve the Problem of Size Heterogeneity

9.5.1 Introduction

One of the aims of purifying the enzyme in large quantities was to use it for crystallisation trials. Discussion about the suitability of this enzyme preparation for such trials led to the decision that any small amounts of charge heterogeneity in the sample were probably unimportant for the purposes of initial trials, but the apparent existence of two protein components of different size would almost certainly disrupt regular crystal formation (G.L. Taylor, personal communication).

Assuming that the upstream gene sequence arrangement of pTaGDH4 was the cause of the formation of the two recombinant products, one of the deletion mutants discussed in Chapter 8 was examined as a possible alternative source of homogeneous recombinant glucose dehydrogenase. Construct pTaGDH3B, cloned in *E.coli* TG1 cells, demonstrated a high level of expression and, from sequence data, would have been predicted to generate a homogeneous population of recombinant *Tp.acidophilum* glucose dehydrogenase, with three extra amino acid residues added on at the N-terminus.

9.5.2 SDS-PAGE Analysis of Purified Glucose Dehydrogenase from pTaGDH3B-bearing *E.coli* TG1 Cells

Recombinant glucose dehydrogenase was prepared from this cell-line as for the pTaGDH4-bearing strain, and fractions eluted from the Mono Q anion exchange column analysed by SDS-PAGE (Figure 9.6). Careful examination of the gel shown in Figure 9.6 demonstrated the presence of two protein bands, although it is extremely difficult to see the lower band of the doublet on the photograph.

9.5.3 Discussion

The two protein components identified from work on this deletion mutant go some way to confirming that the explanation for the presence of the two protein components in the previous experiments was correct. It is notable that the larger

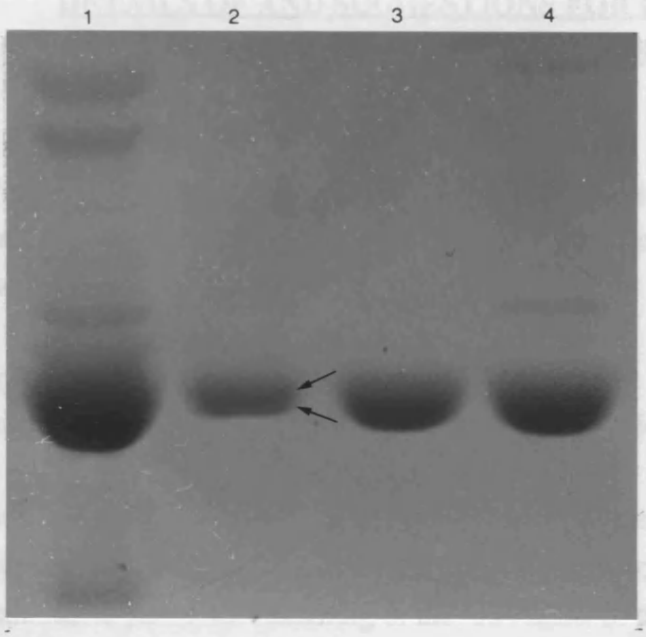


Figure 9.6: Close-up of the SDS PAGE analysis of purified glucose dehydrogenase derived from construct pTaGDH3B. (1) prior to Mono Q chromatography (2 -4) fractions from across the activity peak eluted from the Mono Q column.

protein (presumably glucose dehydrogenase + 3 extra N-terminal amino acids) predominates in the protein preparation from deletion mutant pTaGDH3B. This is consistent with the fact that the alternative translation initiation site (at the very start of the *gld* and presumably responsible for the presence of the smaller recombinant protein) is some way from the ribosome-binding site and therefore would provide a less efficient site of translation initiation (Shine & Dalgarno, 1975).

9.5.4 Concluding Remarks

Unfortunately, because of the presence of size heterogeneity in this second protein preparation, it was felt that the cell-line bearing construct pTaGDH3B was also an unsuitable source of recombinant glucose dehydrogenase for crystallisation studies. This concludes the experimental work performed during this doctoral research project, but a proposal as to how this problem of size heterogeneity could be resolved was made and successfully acted upon. Details of this proposal, the work stemming from it, and suggestions for further work are presented in Chapter 10.

CHAPTER 10 DETAILS OF AND SUGGESTIONS FOR FURTHER WORK; CONCLUDING REMARKS

10.1 Further Work on the Expression of Glucose Dehydrogenase from *Thermoplasma acidophilum* in *Escherichia coli* TG1 cells; kinetic analysis of the recombinant enzyme

10.1.1 Introduction

Experiments involving constructs pTaGDH3B and pTaGDH4 inferred that it was the presence of the start of the *lacZ* gene that was complicating the initiation of translation. To overcome this it was suggested that the DNA fragment used to make construct pTaGDH4 be subcloned into plasmid pMEX8. pMEX8 contains the strong P_{tac} promoter (a chimaeric promoter consisting of the -35 region of the P_{trp} and the -10 region of the P_{lac} promoters), the O_{lac} operator site (required for IPTG-induction experiments), the ribosome-binding site associated with the *lacZ* gene and an adjoining multiple cloning site. However, it does not contain the *lacZ* gene, so initiation of translation can only begin from start codons contained within fragments of DNA ligated into the polylinker sequence. The following results arise from the continuation of this work as a Final Year Project in the Department of Biochemistry, University of Bath; the project was carried out by Ruth Mackness and, for a more extensive appraisal of the work, the reader is referred to Mackness, 1991.

10.1.2 Subcloning of *gld* from *Tp.acidophilum* in pMEX8 and Further Work on Purification of the Recombinant Protein

The recommendation to obviate the expression of chimaeric gene products by subcloning the glucose dehydrogenase gene in pMEX8 was successfully executed. The overall level of expression, though not quite as high as that observed with pUC-derived constructs, was good at ≈8% soluble cell protein. A purification procedure was devised, entailing:

(i) a heat denaturation step in the presence of glucose followed by centrifugation to remove precipitated proteins,

(ii) Mono-Q anion exchange chromatography, and

(iii) HiLoad 16/60 Superdex 200 gel filtration cromatography.

This generated pure protein, as judged by SDS-PAGE, with $M_r \approx 40{,}000$, and free from any apparent size heterogeneity. Analysis of this on an IEF gel showed the protein to have an isoelectric point of pH ≈ 5.5, but also revealed the presence of charge heterogeneity in the sample.

It was felt that this charge heterogeneity may have been introduced by deamidation of asparagine residues during the heat-treatment; this was not felt to be deleterious to the success of crystallisation trials, but the subsequent large-scale protein purification procedure, from which good crystals were made, included a more moderate heat-treatment. A crystal from this experiment was taken for X-ray diffraction analysis (10.2).

Kinetic analysis of the purified recombinant enzyme confirmed that it had similar kinetic characteristics to those determined for the native enzyme (Smith *et al.*, 1989) with respect to its NADP-linked glucose dehydrogenase activity.

10.2 Crystallisation and X-Ray Diffraction Analysis of Glucose Dehydrogenase from *Thermoplasma acidophilum*

10.2.1 Introduction

At various stages in this thesis, the desirability of obtaining structural information about the cloned enzyme has been discussed; for a protein the size of glucose dehydrogenase the only viable method of doing this is by X-ray diffraction analysis. Proteins from *Tp.acidophilum* that have been crystallised include superoxide dismutase (Morris *et al.*, 1985), ferredoxin (Wakabayashi *et al.*, 1985) and malate dehydrogenase (Stezowski *et al.*, 1989) though the three-dimensional structures of these have, at present, not been reported.

The only other glucose dehydrogenase that has been crystallised is that from *Bacillus megaterium* (Pal *et al.*, 1987), the crystals of which were reported to diffract X-rays to 2Å, but, again, the three-dimensional structure of this protein has not been forthcoming. As was mentioned in 10.1.2, crystals of recombinant glucose dehydrogenase were obtained (Mackness, 1991) and taken for X-ray diffraction analysis. Supervision of the crystallisation trials, the manipulation of the crystals and the data handling arising from the X-ray diffraction study were carried out by Dr G.L. Taylor.

10.2.2 Manipulation of Crystals and Unit Cell Data

(refer to **Glossary** for terms in italics)

Crystals were obtained using the hanging-drop method (McPherson, 1990); crystallisation conditions were 4 mg glucose dehydrogenase/ml mother liquor, the mother liquor being 100 mM sodium acetate buffer, pH 5.4, containing 0.4 mM $NADP^+$ and 2% (w/v) polyethylene glycol 4000. Crystals developed after incubation at room temperature for 7 days and were observed to be hexagonal prisms. One

crystal, with dimensions 0.8 x 0.1 x 0.1 mm, was drawn into a fine quartz capillary, along with a little mother-liquor, and the capillary sealed with beeswax. The capillary was mounted on a Siemens Area Detector System, with the crystal ↔ detector distance set at 18 cm, and 190 contiguous frames of diffraction data were collected, oscillating the crystal 0.25° each frame.

Diffraction intensities were measured to a resolution of 4.15 Å, but of the 8,182 unique diffraction intensities obtainable at this level of resolution, only 7,035 were detected from 20,413 diffraction intensity measurements. Calculation of the *Merging R Factors* highlighted the presence of weak data in amongst the intensities that were measured; R = 0.28 and 0.26, for intensity and *F-value* respectively, if all the data were included, whereas omission of the weak data gave R = 0.07 and 0.035 respectively.

Assuming the crystal to have hexagonal symmetry, the *unit cell* dimensions were derived to be 122.2 x 122.2 x 230 Å and the unit cell volume to be $\approx 3 \times 10^6$ Å^3; this unit cell is relatively large. Diffraction intensity is proportional to the volume of the crystal, and inversely proportional to the volume of the unit cell; therefore, the large unit cell and the relatively small crystal explain why some of the diffraction intensities were very weak or absent. The two solutions to this problem are either to obtain bigger crystals or to employ a more intense X-ray source such as a Synchrotron; both of these possibilities are being pursued.

Of the 230 possible *space groups*, only 7 exist that are suitable for protein crystals displaying hexagonal symmetry. Examination of the diffraction intensities with *Miller indices* (0,0,l) revealed a pattern of *systematic absences* consistent with either space group $P6_122$ or $P6_522$; these two groupings are *enantiomorphs*, so whichever is the actual one cannot be determined until some of the protein structure is solved.

Both contain 12 *asymmetric units* within the unit cell, so each asymmetric unit must occupy $\approx 2.5 \times 10^5$ Å^3. The monomeric protein has $M_r \approx 4 \times 10^4$, so if the asymmetric unit was the monomeric protein, the protein would occupy 6.25 Å^3/dalton; similarly, if the assymetric unit consisted of a dimer it would occupy 3.13Å^3/d, or a tetramer, 1.56Å^3/d. The established range for such values is 1.68 - 3.53 Å^3/d and therefore the asymetric unit is most likely to consist of a dimer. This being so, a non-crystallographic 2-fold axis of symmetry should be present where the dimers meet to form the native tetramer; application of the *rotation function* demonstrates the presence of a small peak consistent with this.

Glossary

Assymmetric unit: the smallest repeating unit within the unit cell.

Enantiomorphs: structures which are mirror images of one another.

F-value: a value, proportional to √Intensity values, called the 'Structure factor'.

Merging R-factors: statistical function designed to test the quality of the data:

$R_{merg} = \Sigma_{hkl} | |\overline{I_{hkl}}| - |I_{hkl}| | / \Sigma_{hkl} |I_{hkl}|$

Where I_{hkl} is the mean value of all measurements of the hkl reflection

For perfect data, $R_{merg} = 0$.

Miller indices: indices assigned to all diffraction intensities, of the general form (h,k,l), which signify the plane of electron density giving rise to the reflection.

Space group: symmetry elements in a crystal that relate identical molecules.

Systematic absences: patterns of absent diffraction intensities caused by destructive interference between molecules related by certain kinds of symmetry operations.

Rotation function: a method for searching for non-crystallographic symmetry, using the Patterson function, which can be calculated solely from the observed F-values.

Unit cell: the smallest repeating unit within a crystal.

10.2.3 Concluding Remarks

The data gathered so far are not adequate for protein structure determination, but data collection from crystals with hexagonal space groups is quite rapid because of their high level of symmetry; against this, however, is the problem of solving the structure of an asymmetric unit $M_r = 80{,}000$.

As was discussed in Chapter 7, the attempts to align the amino acid sequences of the functionally similar *Bacillus* and *Tp.acidophilum* glucose dehydrogenases only served to emphasise their dissimilarity. Structurally, they may yet prove to be quite similar; if not, perhaps their differences will yield information with regard to their respective mesophilic and thermophilic natures. Moreover, any differences may serve to highlight their structural similarities and permit the characterisation of such features as the dual-cofactor binding domain and the glucose-binding site.

Acknowledgement

My thanks go to my supervisors, Mike Danson, David Hough and Paul Towner, who have given me tremendous support and encouragement during the course of this project, and to their families for supporting them in this task!

I would like to thank SERC and ICI Biological Products, Billingham for funding me and the project; also David Byrom, my industrial supervisor within ICI, for his continued interest, support and advice throughout the project's execution. Additionally, I am most grateful to Adrian Wolstenholme and Garry Taylor for their respective inputs into the project.

I am certain that everyone in the Biochemistry Department has helped me in some way over the last four years, for which I thank them, but in particular I would like to thank Robert Eisenthal for our many conversations and all the occupants of the Middle and End Labs. for their daily advice, criticism, support and abuse. Jane, Kate and Brian deserve some sort of award for sharing a house with me aswell.

Finally, I wish to thank my family in general, and my parents in particular, without whom none of this would have been possible.

Oh, yes....and Ruth helped too.

REFERENCES:

Adams, M.J., Buehner, M., Chandrasekhar, K., Ford, G.C., Hackert, M., Liljas, A., Rossmann, M.G., Smiley, I.E., Allison, W.S., Everse, J., Kaplan, N.O. and Taylor, S.S. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 1968 - 1972.

Ahern, T.J., Casal, J.I., Petsko, G.A. and Klibanov, A.M. (1987) *Proc. Nat. Acad. Sci. USA* **84**, 675 - 679.

Allan, M.B. (1959) *Arch. Microbiol.* **32**, 270 - 277.

Andreesen, J.R. and Gottschalk, G. (1969) *Arch. Microbiol.* **69**, 160 - 170.

Arakawa, T. and Timasheff, S. (1984) *Biochemistry* **23**, 5924 - 5929.

Belly, R.T. and Brock, T.D. (1972) *J. Gen. Microbiol.* **73**, 465 - 469.

Betlach, M., Friedman, J., Boyer, H.W. and Pfeifer, F. (1984) *Nucleic Acid Res.* **12**, 7949 - 7959.

Biro, J., Fabry, S., Dietmaier, W. Bogedain, C. and Hensel R. (1990) *FEBS Lett.* **275**, 130 - 134.

Bitar, K.G., Blankenship, D.T., Walsh, K.A., Dunlap, R.B., Reddy, A.V. and Freishman, J.H. (1977) *FEBS Lett.* **80**, 119 - 122.

Blanck, A. and Oesterhelt, D. (1987) *EMBO J.* **6**, 265 - 273.

Bokranz, M. and Klein, A. (1987) *Nucleic Acid Res.* **15**, 4350 - 4351.

Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C. and Kraut, J. (1982) *J. Biol. Chem.* **257**, 13650 -13662.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) *Gene* **2**, 95 - 113.

Bradford, M. (1976) *Anal. Biochem.* **72**, 248 - 254.

Bright, J.R., Byrom, D., Danson, M.J., Hough, D.W. and Towner, P. (1991) *Eur. J. Biochem.* submitted.

Brock, T.D. (1974) 'The Genus *Thermoplasma*', pp. 92 - 116, in 'Thermophilic Microorganisms and Life at High Temperatures', Springer Verlag, New York.

Brown, J.W., Daniels, C.J. and Reeve, J. N. (1989) *CRC Crit. Revs. in Microbiol.* **16**, 287 - 338.

Budgen, N. (1988) 'The Catabolism of Glucose by the Thermoacidophilic Archaebacterium *Thermoplasma acidophilum*', Ph.D. Thesis, University of Bath.

Budgen, N. and Danson, M.J. (1986) *FEBS Lett.* **196**, 207 - 210.

Cammaranno, P., Teichner, A. and Londei, P. (1986) *Syst. Appl. Microbiol.* **7**, 137 - 146.

Campbell, D.P., Carper, W.R. and Thompson, R.E. (1982) *Arch. Biochem. Biophys.* **215**, 289 - 301.

Cavener, D. and Krasney, P. (1991) *Mol. Biol. and Evol.* **8**, 144 - 150

Christiansen, C., Freundt, E.A. and Black, F.T. (1975) *Int. J. Syst. Bacteriol.* **25**, 99 - 101.

Clarke, L. and Carbon, J. (1976) *Cell* **9**, 91 - 99.

Cleton-Jansen, A.M., Goosen, N., Vink, K. and van de Putte, P. (1989) *Mol. Gen. Genet.* **217**, 430 - 436.

Cleton-Jansen, A.M., Goosen, N., Fayet, O. and van de Putte, P. (1990) *J. Bacteriol.* **172**, 6308 - 6315.

Coolbear, T., Daniel,R.M., Morgan, H.W. (1991) in press.

Cooper, R.A. (1984) *Ann. Rev. Microbiol.* **38**, 49 - 68.

Corbier, C., Clermont, S., Billard, P., Skarzynski, T., Branlant, C., Wonacott, A. and Branlant, G. (1990) *Biochemistry* **29**, 7101 - 7106.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) *Science* **209**, 1406 - 1414.

Cubellis, M.V., Rozzo, C., Nitti, G., Arnone, M.I., Marino, G. and Sannia, G. (1989) *Eur. J. Biochem.* **186**, 375 - 381.

Culham, D.E. and Nazar, R.N. (1988) *Mol. Gen. Genet.* **212**, 382 - 385.

Da Costa, M.S., Duarte, J.C. and Williams, R.A.D. (Eds.) (1989) 'Microbiology of Extreme Environments and its Potential for Biotechnology', FEMS Symposium N°. 49, Elsevier Science Publishing Ltd., London.

Danson, M.J. (1988) *Adv. Microbiol. Physiol.* **29**, 165 - 231.

Danson, M.J. (1989) *Can. J. Microbiol.* **35**, 58 - 64.

Danson, M.J. and Wood, P.A. (1984) *FEBS Lett.*, **172**, 289 - 293.

Darland, G., Brock, T.D., Samsonoff, W. and Conti, S.F. (1970) *Science (New York)* **170**, 1416 - 1418.

Davidson, B.E., Sajgò, M., Noller, H.F. and Harris, J.I. (1967) *Nature (London)* **216**, 1181 - 1185.

DeLange, R.J., Williams,L.C. and Searcy, D.G. (1981) *J. Biol. Chem.* **256**, 905 - 911.

De Rosa, M., De Rosa, S., Gambacorta, A., Minale, L., Thomson, R.H. and Worthington, R.D. (1977) *J. Chem. Soc. Perkin. Trans.* **1**, 653 - 657.

De Rosa, M., Gambacorta, A. and Bu'Lock, J.D. (1976) *Phytochem.* **15**, 143 - 145.

De Rosa, M., Gambacorta, A., Nicolaus, B., Giardina, P., Poerio, E. and Buonocore, V. (1984) *Biochem. J.* **224**, 407 - 414.

Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acid Res.* **12**, 387 - 395.

Dignam, J.D. (1990) *Met. in Enzymol.* **182**, 194 - 203.

Dunn, R., McCoy, J., Simsek, M., Majumdar, A., Chang, S.H., RajBhandary, U.L. and Khorana, H.G. (1981) *Proc. Nat. Acad. Sci. USA* **78**, 6744 - 6748.

Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, L, Boiwe, T. and Brändén, C. -I. (1974) *FEBS Lett.*, **44**, 200 - 204.

Entner, N. and Doudoroff, M. (1952) *J. Biol. Chem.* **196**, 853 - 862.

Evans, J.N.S., Raleigh, D.P., Tolman, C.J. and Roberts, M.F. (1986) *J. Biol. Chem.* **261**, 16323 - 16331.

Evans, J.N.S., Tolman, C.J., Kanodia, S. and Roberts, M.F. (1985) *Biochemistry* **24**, 5693 - 5698.

Feeney, R., Clarke, A.R. and Holbrook, J.J. (1990) *Biochem. Biophys. Res. Commun.* **166**, 667 -672.

Giardina, P., De Biasi, M. -G., De Rosa, M., Gambacorta, A. and Buonocore, V. (1986) *Biochem. J.* **239**, 517 - 522.

Green, G.R., Searcy,D.G. and DeLange, R.J. (1983) *Biochim. Biophys. Acta* **741**, 251 - 257.

Grossebeuter, W., Hartl, T. Gorisch, H. and Stezowski, J.J. (1986) *Biol. Chem. Hoppe-Seyler* **367**, 457 - 470.

Gupta, R. and Woese, C.R. (1980) *Curr. Microbiol.* **4**, 245 - 249.

Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557 - 580.

Hensel, R., Zwickl, P., Fabry, S., Lang, J. and Palm, P. (1989) *Can. J. Microbiol.* **35**, 81 - 85.

Hough, D.W. and Danson, M.J. (1989) *Lett. in Appl. Microbiol.* **9**, 33 - 39.

Huet, J., Schnabel, R., Sentenac, A. and Zillig, W. (1983) *EMBO J.* **2**, 1291 - 1294.

Jameson, B. and Wolf, H. (1988) *CABIOS* **4**, 181 - 186.

Jany, K.-D., Ulmer, W., Fröschle, M. and Pfleiderer, G. (1984) *FEBS Lett.* **165**, 6 - 10.

Jörnvall, H. (1970) *Eur. J. Biochem.* **16**, 25 - 40.

Kaine, B.P. (1987) *J. Mol. Evol.* **25**, 248 - 254.

Kaiser, K. and Murray N.E. (1985) 'The Use of Phage Lambda Replacement Vectors in the Construction of Representative Genomic DNA Libraries', pp. 1 - 48 in 'DNA Cloning Vol.1 - a Practical Approach', ed. D.M. Glover, IRL Press Ltd., Oxford.

Kilpatrick, M.W. and Walker, R.T. (1981a) *Biochem. Soc. Trans.* **9**, 387 - 388.

Kilpatrick, M.W. and Walker, R.T. (1981b) *Nucleic Acid Res.* **9**, 4387 - 3490.

Kilpatrick, M.W. and Walker, R.T. (1982) *Zentralb. Bakteriol., Mikrobiol. Hyg., Abt. 1, Orig. C*, **3**, 79 - 89.

Kimura, M., Kimura, J. and Zierer, R. (1984a) *FEBS Lett.* **175**, 208 - 212.

Kimura, M., Kimura, J., Davie, P., Reinhardt, R. and Dijk, J. (1984b) *FEBS Lett.* **176**, 176 - 178.

Kimura, M. and Wilson, K.S. (1983) *J. Biol. Chem.* **258**, 4007 - 4011.

Klenk, H.P., Haas, B., Schwass, V. and Zillig, W. (1986) *J. Mol. Evol.* **24**, 167 - 173.

Klibanov, A.M. and Ahern, T.J. (1987) 'Thermal Stability of Proteins', pp. 213 - 218, in 'Protein Engineering', ed. D.L. Oxender and C.F. Fox. Liss, New York.

Kobayashi, Y. and Horikoshi, K. (1980) *Agric. Biol. Chem.* **44**, 2261 - 2269.

Krauth-Siegel, R.L., Blatterspiel, R., Saleh, M., Schilz, E., Schirmer, R.H. and Untucht-Grau, R. (1982) *Eur. J. Biochem.* **121**, 259 -267.

Kuchino, Y., Ihara, M., Yabusaki, Y. and Nishimura, S. (1982) *Nature (London)* **298**, 684 - 685.

Kushner, D.J. (1985) 'The *Halobacteriaceae*', pp. 171 - 214, in 'The Bacteria, Vol.8: The Archaebacteria', ed. C.R. Woese and R.S. Wolfe. Academic Press Inc., New York.

Laemmli, U.K. (1970) *Nature (London)* **227**, 680 - 685.

Laine, B., Chartier, F., Imbert, M., Lewis, R. and Sautiere, P. (1986) *Eur J. Biochem.* **161**, 681 - 687.

Lake, J.A. (1988) *Nature (London)* **331**, 184 - 186.

Lake, J.A. (1989) *Can. J. Microbiol.* **35**, 109 - 118.

Lake, J.A. (1991) *TIBS* **16**, 46 - 50.

Lake, J.A., Henderson, E., Oakes, M. and Clark, M.W. (1984) *Proc. Nat. Acad. Sci. USA* **81**, 3786 - 3790.

Lampel, K.A., Uratani, B., Chaudhry, G.R., Ramaley,R. and Rudikoff, S. (1986) *J. Bacteriol.* **166**, 238 - 243.

Lane, D.J., Pace, B., Olsen, G.J. Stahl, D.A., Sogin, M.L. and Pace, N.R. (1985) *Proc. Nat. Acad. Sci. USA* **82**, 6955 - 6959.

Langworthy, T.A. (1977) *Biochim. Biophys. Acta* **487**, 37 - 50.

Langworthy, T.A., Smith, P.F. and Mayberry, W.R. (1972) *J.Bacteriol.* **112**, 1193 - 1200.

Larsen, H. (1973) *Antonie van Leeuwanhoek J. Microbiol. Serol.* **39**, 383 - 396.

Levy, H.R. (1979) *Adv. Enzymol.* **48**, 97 - 192.

Lewin, B. (1990) 'Genes IV', Oxford University Press.

Lobet, Y., Peacock,M. and Cieplak, W.Jr. (1989) *Nucleic Acid Res.* **17**, 4897.

Londei, P., Altamura, S., Cammarano, P. and Petrucci, L. (1986) *Eur. J. Biochem.* **157**, 455 - 462.

Luehrsen, K.R., Fox, G.E., Kilpatrick, M.W., Walker, R.T., Domdey, H., Krupp, G. and Gross, H.J. (1981) *Nucleic Acid Res.* **9**, 965 - 970.

Mackness, R. (1991) 'The Expression and Crystallisation of Glucose Dehydrogenase from *Thermoplasma acidophilum*', Project Report, Bath University Library.

Mahoney, W.C., Hogg, R.W. and Hermodson, M.A. (1981) *J.Biol. Chem.* **256**, 4350 - 4356.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) 'Molecular Cloning - a Laboratory Manual', Cold Spring Harbour Laboratory, USA.

Matsushita, K., Ohno, Y., Sinagawa, E., Adachi, O. and Ameyana, M. (1980) *Agric. Biol. Chem.* **44**, 1505 - 1512.

Matsumura, M., Katakura, Y., Imanaka, T. and Aiba, S. (1984) *J. Bacteriol.* **160**, 413 - 420.

Matthews, D.A., Alden, R.A., Freer, S.T., Xuong, N. and Kraut, J. (1979) *J. Biol. Chem.* **254**, 4144 - 4151.

McGill, T.J., Jurka, J., Sobieski, J.M., Pickett, M.H., Woese, C.R. and Fox, G.E. (1986) *Syst. Appl. Microbiol.* **7**, 194 - 197.

McPherson, A. (1990) *Eur. J. Biochem.* **189**, 1 - 23.

Messing, J. (1983) *Met. in Enzymol.* **101**, 20 - 78.

Morris, D.C., Searcy, D.G. and Edwards, B.F.P. (1985) *J. Mol. Biol.* **186**, 213 - 214.

Ng, T.K. and Kenneally, W.F. (1986) pp. 197 - 216, from 'Thermophiles: General, Molecular and Applied Microbiology', Ed. T.D. Brock, John Wiley & Sons.

Noll, K.M., Rinehart Jr., K.L., Tanner, R.S. and Wolfe, R.S. (1986) *Proc. Nat. Acad. Sci. USA* **83**, 4238 - 4242.

Norrander, J., Kempe, T. and Messing, J. (1983) *Gene* **26**, 101 -106.

Ohlsson, I., Nordström, B. and Brändén, C. -I. (1974) *J. Mol. Biol.* **89**, 339 - 354.

Oliver, J.L., Sanz, J.L., Amils, R. and Marin, A. (1987) *J. Mol. Evol.* **24**, 281 - 288.

Olsen, G.F. and Woese, C.R. (1989) *Can. J. Microbiol.* **35**, 119 - 123.

Pai, E.F., Karplus, P.A. and Schulz, G.E. (1988) *Biochemistry* **27**, 4465 - 4474.

Pal, G.P., Jany, K. -D. and Saenger, W. (1987) *Eur. J. Biochem.* **167**, 123 - 124.

Payton, M.A. and Haddock, B.A. (1985) in 'Comprehensive Biotechnology' (ed. M. Moo-Young), Vol.1, pp.337 - 357, Pergamon Press, Oxford.

Perball, B. (1988) 'A Practical Guide to Molecular Cloning', John Wiley & Sons, Inc., New York.

Perry, L.J. and Wetzel, R. (1984) *Science* **226**, 555 -557.

Pribnow, D. (1975) *J. Mol. Biol.* **99**, 419 - 443.

Pülich, W.M., Van Baalen, C., Gibson, J.L. and Tabita, F.R. (1976) *Plant Physiol.* **58**, 393 - 397.

Ree, H.K., Cao, K., Thurlow, D.L. and Zimmerman, R.A. (1989) *Can. J. Microbiol.* **35**, 124 - 133.

Ree, H.K. and Zimmerman, R.A. (1990) *Nucleic Acids Res.* **18**, 4471 - 4478.

Reiter, W., Palm, P. and Zillig, W. (1988) *Nucleic Acids Res.* **16**, 2445 - 2459.

Rice, D.W., Schulz, G.E. and Guest, J.R. (1984) *J. Mol. Biol.* **174**, 483 - 496.

Rossmann, M.G., Moras, D. and Olsen, K.W. (1974) *Nature (London)* **250**, 194 - 199.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Nat. Acad. Sci. USA* **74**, 5463 - 5467.

Sanz, J.L., Altamura, S., Mazziotti, I., Amils, R., Cammarano, P. and Londei, P. (1987) *Mol. Gen. Genet.* **207**, 385 - 394.

Schleifer, K. -H., Steber, J. and Meyer, H. (1982) *Zentralb. Bakteriol. Parasitkend. Infektionskr. Hyg. Abt. 1 Orig. Reihe C.* **3**, 171 - 178.

Scrutton, N.S., Berry, A. and Perham, R.N. (1990) *Nature (London)* **343**, 38 - 43.

Searcy, D.G. (1975) *Biochim. Biophys. Acta* **395**, 535 - 547.

Searcy, D.G. (1986) *Syst. Appl. Microbiol.* **7**, 198 - 201.

Searcy, D.G. and Doyle, E.K. (1975) *Int. J. Syst. Bacteriol.* **25**, 286 - 289.

Searcy, D.G., Stein,D.B. and Green, G.R. (1978) *Biosystems* **10**, 19 - 28.

Segerer, A., Langworthy, T.A. and Stetter, K.O. (1988) *Syst. Appl. Microbiol.* **10**, 161 - 171.

Shine, J. and Dalgarno, L. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 1342 - 1346.

Smith, L.D. (1989) 'Studies of Dehydrogenases from Thermophilic Archaebacteria with a View to Cofactor Regeneration', Ph.D Thesis, University of Bath.

Smith, L.D., Budgen, N., Bungard, S.J., Danson, M.J. and Hough, D.W. (1989) *Biochem. J.* **261**, 973 - 977.

Smith, L.D., Bungard, S.J., Danson, M.J. and Hough, D.W. (1988) *Biochem. Soc. Trans.* **16**, 864.

Southern, E. (1975) *J. Mol. Biol.* **98**, 503 - 517.

Sprott, G.D., McKeller, R.C., Shaw, K.M., Giroux, J. and Martin, W.G. (1979) *Can. J. Microbiol.* **25**, 192 - 203.

Stein, D.B. and Searcy, D.G. (1978) *Science (New York)* **202**, 219 - 221.

Stetter, K.O., Konig, H. and Stackebrandt, E. (1983) *Syst. Appl. Microbiol.* **4**, 535 -551.

Stetter, K.O. and Zillig, W. (1985) pp. 85 - 170, in 'The Bacteria, Vol.8: The Archaebacteria', ed. C.R. Woese and R.S. Wolfe. Academic Press Inc., New York.

Stezowski, J.J., Englmaier, R., Galdiga, C., Hartl, T., Rommel, I., Dauter, Z., Goerisch, H., Grossebeuter, W., Wilson, K. and Musil, D. (1989) *J. Mol. Biol.* **208**, 507 - 508.

Sutherland, K.J. (1990) '*Thermoplasma acidophilum* Citrate Synthase - Cloning and Sequencing of the Gene', Ph.D. Thesis, Bath University Library.

Sutherland, K.J., Henneke, C.M., Towner, P., Hough, D.W. and Danson, M.J. (1991) *Eur. J. Biochem.* **194**, 839 - 844.

Taylor, S.S. (1977) *J. Biol. Chem.* **252**, 1799 - 1806.

Teichner, A. Londei, P. and Cammarano, P. (1986) *J. Mol. Evol.* **23**, 343 - 353.

Tesch, A. and Klink, F. (1990) *FEMS Lett.* **71**, 293 - 298.

Thatcher, D.R. and Sawyer, L. (1980) *Biochem. J.* **187**, 884 - 886.

Thomm, M. and Wich, G. (1988) *Nucleic Acid Res.* **16**, 151 - 163.

Tomlinson, G.A., Koch, T.K. and Hochstein, L.I. (1974) *Can. J. Microbiol.* **20**, 1085 - 1091.

Towner, P. (1991a) 'Purification of DNA', pp. 47 - 68, in 'Essential Molecular Biology Vol.1' (ed. T.Brown), Oxford University Press, Oxford.

Towner, P. (1991b) 'Recovery of DNA from Electrophoresis Gels', pp. 127 - 141, in 'Essential Molecular Biology - Vol.1' see Towner (1991a).

Tsai, C.S., Senior, D.J. and Al-Kassim, L.S. (1989) *Comp. Biochem. Physiol.* **94B**, 655 - 659.

Tu, J. and Zillig, W. (1982) *Nucleic Acids Res.* **10**, 7231 - 7245.

Wakabayashi, S., Wada, K., Matsubara, H., Kerscher, L. and Oesterhelt, D. (1985) *J. Mol. Biol.* **186**, 481 - 482.

Walker, R.T. (1983) *Yale J. Biol. Med.* **56**, 367 - 372.

Wang, S.W. and King, C.K. (1979) *Adv. Biochem. Eng.* **12**, 119 - 146.

Weijer, W.J., Hofsteenge, J., Vereijken, J.M., Jekel, P.K. and Beintema, J.J. (1982) *Biochim. Biophys. Acta* **704**, 385 - 388.

Whitman, W.B. (1985) 'Methanogenic Bacteria', pp. 1 - 84, in 'The Bacteria, Vol.8: The Archaebacteria', ed. C.R. Woese and R.S. Wolfe. Academic Press Inc., New York.

Wierenga, R.K., DeMaeyer, M.C.H. and Hol, W.G.J. (1985) *Biochemistry* **24**, 1346 - 1357.

Woese, C.R. (1987) *Microbiol. Revs.* **51**, 221 -271.

Woese, C.R. and Fox, C.E. (1977) *Proc. Nat. Acad. Sci. USA* **74**, 5088 - 5090.

Woese, C.R. and Olsen, G.J. (1986) *Syst. Appl. Microbiol.* **7**, 161 - 177.

Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) *Proc. Nat. Acad. Sci. USA* **87**, 4576 - 4579.

Yang, D., Kaine, B.P. and Woese, C.R. (1985) *Syst. Appl. Microbiol.* **6**, 251 - 256.

Yannisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene* **33**, 103 - 109.

Zillig, W., Palm, P., Reiter, W. -D., Gropp, F., Pühler, G. and Klenk, H. -P. (1988) *Eur. J. Biochem.* **173**, 473 - 482.

Zwickl, P., Lottspeich, F., Dahlmann,B. and Baumeister, W. (1990) *FEBS Lett.* **278**, 217 - 221.