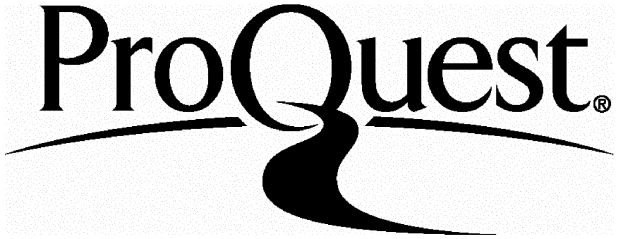
Accessing Genes from Environmental DNA Libraries

A thesis submitted to the University of London for the Degree of Doctor of Philosophy by

Dianna E. Wilkinson

Department of Biochemistry and Molecular Biology
University College London
Gower Street
London WC1E 6BT

2001



Abstract

The classical approach for isolating enzymes from environmental samples is to enrich, isolate and screen a variety of microorganisms for the desired enzyme activity. The majority of the physiological diversity is excluded with this approach because it is estimated that >99% of the microorganisms observed in the environment cannot be cultivated. An alternative to this cultivation-dependent approach is to clone DNA which has been extracted directly from microbial biomass present in water, soil and sediment. Using an appropriate host system, enzyme activity can subsequently be detected by screening for heterologous gene expression.

Geothermal regions are sources of thermophilic microbial diversity. This study sought to investigate the methods for extracting and cloning DNA from geothermal sediments for the purpose of detecting thermostable enzyme activities.

Methods for extracting and purifying DNA directly from soil and sediment were evaluated based on DNA yield, purity and quality. Purified environmental DNA was then used to evaluate cloning protocols based on cloning efficiency, recombination efficiency and total number of recombinants generated per ligation reaction.

Subsequently, two environmental gene libraries were constructed using a TA-cloning method with DNA directly extracted from sediments that were collected from Iceland geothermal sites. The environmental library designated as ICE16 was derived from biomass present in sediment at ~74°C, pH 7.4, while the DNA used to construct library ICE22 was derived from biomass present in sediment at ~58°C, pH 4.3. These libraries were screened for thermostable amylase, lipase, protease and phosphatase activities using established assays in both microtitre plate and indicator agar-plate formats. One transformant possessing phosphatase activity at 60°C (Phos22) and two transformants showing phenotypic differences on starch agar plates at 50°C (5ICE16, 6ICE16) were recovered from the ICE22 and ICE16 DNA libraries, respectively. These clones were selected for further evaluation including sequencing and expression studies.

Acknowledgements

Firstly, I would like to thank Dr. Don Cowan for providing me with the opportunity to conduct research in his laboratory and for his advice and encouragement throughout the course of this work. I am also grateful to Dr. John Ward for his advice.

A special acknowledgement goes to Dr. Thomas Jaenicke, a capable and knowledgeable colleague and friend who helped with the day to day challenges of scientific research. Acknowledgement is also due to Dr. Haitham Hussein for investigating recombination efficiencies of environmental DNA libraries employing alkaline phosphatases obtained from various commercial sources. I am also grateful to Mr. Bilal Jredah for his assistance with phosphatase screening and for isolation of the phosphatase-positive clones.

I thank my husband Ian for always being there with love, encouragement, support, patience and humour. I would also like to thank my family on both sides of the Atlantic for putting up with my ramblings and reminding me that there are more important things to life than work.

I thank everyone from the laboratories on the ground floor for providing light entertainment along the way.

Finally, my research was supported in part by a major research scholarship awarded by the Graduate School at University College London, as well as by funding provided by the Overseas Research Scholarship scheme.

In Memory of Doug Campbell

Table of Contents

Table of Figures

Chapter 1: Introduction

Chapter 2: Materials & Methods

Chapter 3: Direct Extraction of Environmental DNA from Geothermal Sediments

Chapter 4: Evaluation of Cloning Protocols and Preparation of Environmental DNA Libraries

Chapter 5: Screening Environmental Libraries for Thermostable Enzyme Activities

Chapter 6: Sequence and Expression Studies

Chapter 7: Discussion

Abbreviations & Symbols

Accession ..... GenBank sequence accession number
dATP ..... deoxyadenosine triphosphate
dCTP ..... deoxycytosine triphosphate
dGTP ..... deoxyguanosine triphosphate
dTTP ..... deoxythymidine triphosphate
EDTA ..... ethylenediaminetetraacetic acid
g ..... gram(s) or gravity
SDS ..... sodium dodecyl sulphate
UV ..... ultraviolet light
v/v ..... volume per volume
w/v ..... weight per volume

The Amino Acid Codes

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

The Nucleotide Codes

A	Adenine
C	Cytosine
G	Guanine
T	Thymine

Chapter 1
Introduction

1.1: Microbial Life in High-Temperature Environments

Thermophiles are generally defined as those organisms that thrive at elevated temperatures. Most thermophilic organisms known are termed 'moderate', with growth temperature optima ranging between ~55°C and ~80°C (Kristjansson & Stetter, 1992). Moderate thermophiles have representatives from many taxonomic groups including protozoa, algae, fungi, bacilli, streptomycetes, clostridia, cyanobacteria, purple bacteria, green bacteria, methanogens and halophiles. The prokaryotic thermophiles predominate in this category since the temperature maxima for eukaryotic thermophiles typically is below ~60°C (Brock & Madigan, 1991). Hyperthermophiles constitute a subgroup of thermophilic prokaryotes that thrive at temperatures greater than 80°C with a maximum growth temperature of ~113°C (Blochl *et al.*, 1997). In contrast with moderate thermophiles, hyperthermophiles typically cannot grow below ~60°C. Almost all hyperthermophiles are archaeal. The only bacterial hyperthermophilic representatives are *Thermotoga* and *Aquifex* (Stetter, 1998). Throughout this work, "thermophile" will be a collective term encompassing both moderate and hyperthermophilic microorganisms.

Common habitats for thermophilic organisms are natural geothermal areas such as terrestrial fresh-water hot springs, geysers, solfataric fields consisting of heated soils, bubbling mud pools and steam vents, in addition to submarine hydrothermal systems and abyssal hot vents (black smokers). Non-geothermal sources of high-temperature biotopes include burning coal refuse piles, domestic hot water tanks, self-heating compost piles and heated industrial waters (Kristjansson & Stetter,1992).

Chemical compounds found in geothermal fluids that can be used by thermophiles as carbon and energy sources include elemental hydrogen, hydrogen sulphide, sulphur dioxide, ammonia, methane, carbon dioxide, carbon monoxide, elemental sulphur (S^0) and dissolved mineral salts. Oxygen availability is often limited in thermal waters since its solubility decreases as temperature increases. Organic material can be derived allochthonously from plants and animals as well as from microbial biomass and

processes (Aragno, 1992). Because of the high concentration of salts such as sulphates or carbonates, high-temperature biotopes can be either acidic or basic.

1.2: Phylogeny of thermophilic Prokaryotes

Based on 16S and 18S rRNA sequences the universal phylogenetic tree shows a tripartite division of the living world consisting of the domains *Bacteria*, *Archaea* and *Eucarya* (Figure 1.1) (Woese & Fox, 1977; Pace *et al.*, 1997). The archaeal kingdom, with its high proportion of moderate thermophiles and hyperthermophiles, is placed nearest to the root of the phylogenetic tree. Among the bacteria, hyperthermophilic *Aquifex* and *Thermotoga* as well as moderately thermophilic *Thermus* are also deeply rooted. Based on these findings, the last common ancestor (the root of the tree) may have been a hyperthermophile.

1.3: Diversity of Thermophilic Prokaryotes

Tables 1.1 and 1.2 respectively list some representative thermophilic bacteria and archaea. Known moderately thermophilic prokaryotes include aerobic heterotrophs, oxygenic and anoxygenic phototrophs, aerobic chemolithoautotrophs that oxidise hydrogen, sulphur, nitrogen-containing compounds or iron, anaerobic fermenters, nitrate reducers and denitrifiers (Kristjansson & Stetter, 1992). Known hyperthermophiles include strictly anaerobic, S^0-reducing, autotrophs and heterotrophs, S^0-independent fermenters, microaerophiles, methanogens, sulphate reducers, iron oxidisers and nitrate reducers (Stetter, 1998; Adams, 1999).

It is inevitable that knowledge of the physiological diversity of thermophiles will continue to increase as more organisms are isolated and characterised. Our understanding of thermophilic genetic diversity, however, is incomplete because only a small fraction of naturally occurring microorganisms are routinely cultivatable in laboratory studies (Pace, 1997). Molecular phylogenetic analyses using nucleic acids obtained from uncultured organisms present in geothermal environments have revealed an incredible genetic diversity of microorganisms that surpasses that implied by cultured organisms (Ward *et al.*, 1990, 1992; Weller *et al.*, 1991; Barns *et al.*, 1994; Barns *et al.*, 1996; Hugenholtz *et al.*, 1998). Such diversity implies a largely untapped resource of novel biomolecules including enzymes.

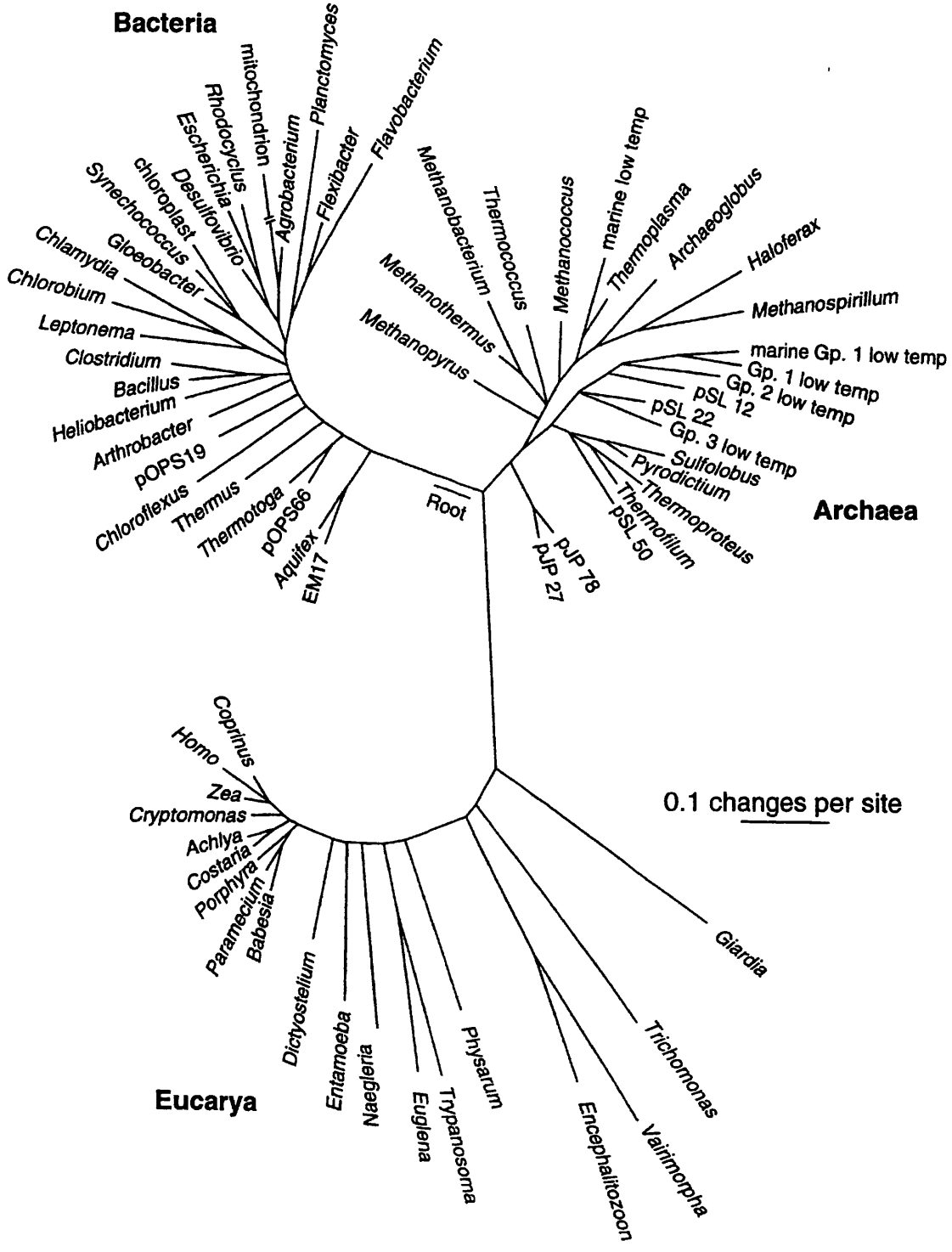


Figure 1.1: Universal phylogenetic tree based on rRNA sequences. Taken from Pace, 1997.

Organism	Temp. opt/max (°C)	pH (opt)*	Metabolism	Reference
Aquifex pyrophilus	85/95	5.4-7.5	Facultatively aerobic lithoautotroph	Huber *et al.* (1992)
Bacillus acidocaldarius	58/65	(2.0)	Aerobic heterotroph	Claus & Berkeley (1986)
Bacillus stearothermophilus	60/75	(7.0)	Facultatively aerobic heterotroph	Claus & Berkeley (1986)
Chlorobium tepidum	50/57	(6.9)	Obligate phototroph	Madigan (1986)
Chloroflexus auranticus	56/70	(8.0)	Anoxygenic phototroph/ Chemoheterotroph	Pierson & Castenholz (1974)
Clostridium thermoautotrophicum	58/68	(5.7)	Anaerobic chemothithoautotroph	Wiegel *et al.* (1981)
Dyctioglomus thermophilum	78/80	(7.0)	Anaerobic Chemoorganotroph	Saiki *et al.* (1985)
Rhodothermus marinus	65/77	(7.0)	Aerobic, slightly halophilic heterotroph	Alfredsson *et al.* (1988)
Synechococcus lividus	65/73	(8.0)	Oxygenic photoautotroph	Meeks & Castenholz (1971)
Thermus aquaticus	70/80	(7.7)	Aerobic obligate heterotroph	Brock & Freeze (1969)
Thermotoga maritima	80/90	5.5-9.0	Anaerobic heterotroph	Huber *et al.* (1986)

Table 1.1: Some thermophilic bacteria. *The value given in parentheses is the optimal pH. Table 1.1 was compiled from Kristjansson & Stetter (1992) and Stetter (1998).

Organism	Temp. opt/max (°C)	pH (opt)*	Metabolism	Reference
Acidianus infernus	88/95	1.5-5	Facultatively anaerobic heterotroph	Segerer *et al.* (1986)
Aeropyrum pernix	95/110	(7.0)	Obligately aerobic heterotroph	Sako *et al.* (1996)
Archaeoglobus fulgidus	83/92	5.5-7.5	Obligately anaerobic heterotroph	Stetter (1988)
Desulfurococcus mobilis	85/95	4.5-7.0	Obligately anaerobic heterotroph	Zillig *et al.* (1982)
Metallosphaera sedula	75/80	1.0-4.5	Obligately aerobic facultative lithoautotroph	Huber, G. *et al.* (1989)
Methanococcus jannaschii	85/86	3.0-6.5	Methanogen	Jones *et al.* (1983)
Methanohalobium evestigatum	50/60	5.5-7.0	Halophilic methanogen	Zhilina & Zavarzin, (1987)
Methanopyrus kandleri	98/110	5.5-7.0	Methanogen	Huber, R. *et al.* (1989)
Pyrococcus furiosus	100/103	5.0-9.0	Obligately anaerobic heterotroph	Fiala & Stetter (1986)
Pyrobaculum aerophilum	100/104	5.8-9.0	Facultatively anaerobic heterotroph	Volkl *et al.* (1993)
Pyrodictium abyssi	105/110	4.7-7.5	Obligately anaerobic heterotroph	Pley *et al.* (1991)
Pyrolobus fumarii	106/113	4.0-6.5	Facultatively aerobic autotroph	Blochl *et al.* (1997)
Staphylothermus marinus	92/98	4.5-8.5	Obligately anaerobic heterotroph	Fiala *et al.* (1986)
Stygiolobus azoricus	80/89	1.0-5.5	Obligately anaerobic lithoautotroph	Segerer *et al.* (1991)
Sulfolobus acidocaldarius	75/85	1-5	Obligately aerobic facultative lithoautotroph	Brock *et al.* (1972)
Thermococcus celer	87-93	4-7	Obligately anaerobic heterotroph	Zillig *et al.* (1983)
Thermodiscus maritimus	88/98	5-7	Obligately anaerobic heterotroph	Stetter (1986)
Thermoproteus neutrophilus	88/97	5.5-7.5	Obligately anaerobic lithoautotroph	Stetter (1986)

Table 1.2: Some thermophilic archaea. *The value given in parentheses is the optimal pH. Table 1.2 was compiled from Kristjansson & Stetter (1992), Baross & Holden (1996) and Stetter (1998).

1.4: Thermostable Proteins

Due to the high biodiversity of thermophiles and their ability to produce a variety of novel thermostable proteins, enzymes produced by thermophilic organisms have attracted the attention of both academia and industry. Proteins from thermophiles are typically thermostable and active at physiological temperatures and many are also resistant to organic solvents, detergents, proteolytic agents and pH extremes (Fontana *et al.*, 1991).

Thermophilic enzymes generally retain their thermal properties when they are expressed in mesophilic hosts indicating that their unique thermal properties are due to the primary amino acid sequence. Numerous studies have been undertaken to identify the universal mechanisms for promoting thermostability of proteins and it is generally accepted that high stability of thermophilic proteins, as compared to their mesophilic counterparts, is predominantly a result of differences in specific amino acid sequences (reviewed by Daniel, 1996; Vieille & Zeikus, 1996). Studies have shown that addition of hydrogen bonds, salt bridges, hydrophobic interactions or the release of local conformational strains can increase the stability of a protein by increasing its free energy of stabilisation ($\Delta G_{stab.}$) (Matthews *et al.*, 1987). Deduced from correlation studies, the mechanisms thought to play a role in contributing to the thermal adaptation of proteins include enhanced hydrophobicity and packing efficiency, the presence of salt bridges, reduced surface area, reduction of conformation strain, loop stabilisation and resistance to covalent destruction (reviewed by Jaenicke *et al.*, 1996; Vieille & Zeikus, 1996; Ladenstein & Antranikian, 1998) It should be stressed, however, that no universal stabilising mechanism has been identified for thermophilic proteins and that, in general terms, proteins derived from thermophiles are stabilised by the same types of intramolecular forces that stabilise mesophilic proteins.

Extrinsic factors that may contribute to the thermostability of a protein include an increased turnover, the presence of compatible solutes, specific ions, metabolites, cofactors and molecular chaperones. (Jaenicke *et al.*, 1996).

1.5: Biotechnological Significance of Enzymes from Thermophiles

Table 1.3 lists some examples of biotechnologically relevant enzymes from thermophilic prokaryotes. The main reason for selecting thermophilic enzymes for

biotechnological applications is their thermostability. Thermostable enzymes offer some advantages for biotechnological processes many of which run at high temperatures. Higher-temperature enzymic reactions may also eliminate or reduce cooling costs. Other advantages that thermophilic enzymes may have over their mesophilic counterparts include an increased resistance to chemicals and other potential denaturants; a longer shelf-life and less chance of contamination if operated at sufficiently high temperatures (Kristjansson & Stetter, 1992). The prospect of implementing thermophilic enzymes in industrial processes, however, is tempered somewhat by a reluctance of companies to invest in the development of new enzyme processes.

Enzyme	Function	Application
α-Amylase	Hydrolysis of endo-α-1,4-glycosidic bonds of starch and related compounds	High-temperature liquefaction and saccharification of starch for the production of sugars and sweeteners
Pullulanase	Hydrolysis of α-1,6-glycosidic bonds of amylopectin	
Glucoamylase	Hydrolysis of both α(1→4) and α(1→6) terminal linkages of starch	
Glucose isomerase (Xylose isomerase)	Isomerisation of glucose to fructose	
Cyclodextrin glycosyltransferase	Formation of non-reducing cyclodextrins from starch	Stabilisation/solubilisation of organic and inorganic compounds
Cellulases	Hydrolysis of cellulosic material	Production of chemical feed stock and fuel Detergent additive
Xylanases	Hydrolysis of xylan, a major component of hemicellulose	Paper bleaching
Proteolytic enzymes	Conversion of proteins to peptides and amino acids	Detergent additive Leather processing Amino acid production
Lipases	Hydrolysis of insoluble long-chain fatty acid esters	Detergent additive Food processing
DNA polymerases	DNA Replication	Polymerase Chain Reaction DNA sequencing
Restriction endonucleases	Cleavage of DNA at specific sequences	Molecular biology

Table 1.3: Examples of commercially relevant enzymes from thermophiles. Table 1.3 was compiled using data from Fogarty (1983), Edwards (1990), Sharp *et al.* (1990) and Ladenstein & Antranikian (1998).

1.6: Enzyme Discovery

The traditional route to finding novel enzymes from environmental samples is to enrich, isolate and screen a wide variety of microorganisms for the desired activity. Searching

for any one enzyme activity requires cultivating the samples under a range of enrichment conditions. In order to access as many microbes as possible, samples should be incubated at different pH, temperature and salinity ranges using media containing various carbon and energy sources under both aerobic and anaerobic conditions. Conditions should also be set up so as to eliminate known or undesirable enzyme activities. Upon identifying the microbe possessing the desired activity, the enzyme and corresponding gene is then recovered from the isolated organism. Although this method is a common route to enzyme discovery, a large fraction of the protein diversity could be missed due to difficulties in enriching and isolating environmental microorganisms in pure culture.

Molecular ecological studies of microbial assemblages such as those found in marine (Giovannoni *et al.*, 1990; Britschgi & Giovannoni, 1991), thermobiotic (e.g. Ward *et al.*, 1990, 1992; Weller *et al.*, 1991; Barns *et al.*, 1994; Barns *et al.*, 1996; Hugenholtz *et al.*, 1998) and terrestrial environments (e.g. Liesack & Stackebrandt, 1992) corroborate the widespread view that the prokaryote species so far cultured constitute only a small fraction of the actual microbial population observed in nature (Torsvik *et al.*,1990; Amann *et al.*,1995). The genomes of the total microbiota found in nature, termed the metagenome by Rondon *et al.* (2000), are therefore considered to contain considerably more genetic information than the culturable subset.

An alternative approach to conventional enzyme screening is to access this extensive microbial molecular diversity by isolating DNA without culturing the organisms present. Routine molecular biological techniques can then be applied to the environmental DNA in order to access and functionally characterise the target molecules. Figure 1.2 outlines some of the techniques used by investigators for recovering functional proteins encoded by the metagenome. Using a polymerase chain reaction (PCR)-based strategy, Seow *et al.* (1997) cloned β-ketoacyl synthase (KSβ) genes from multigenomic DNA derived from uncultivated soil bacteria. This was achieved by using degenerate primers corresponding to conserved regions flanking known KSβ genes. The cloned PCR products were subsequently sequenced, sub-cloned,

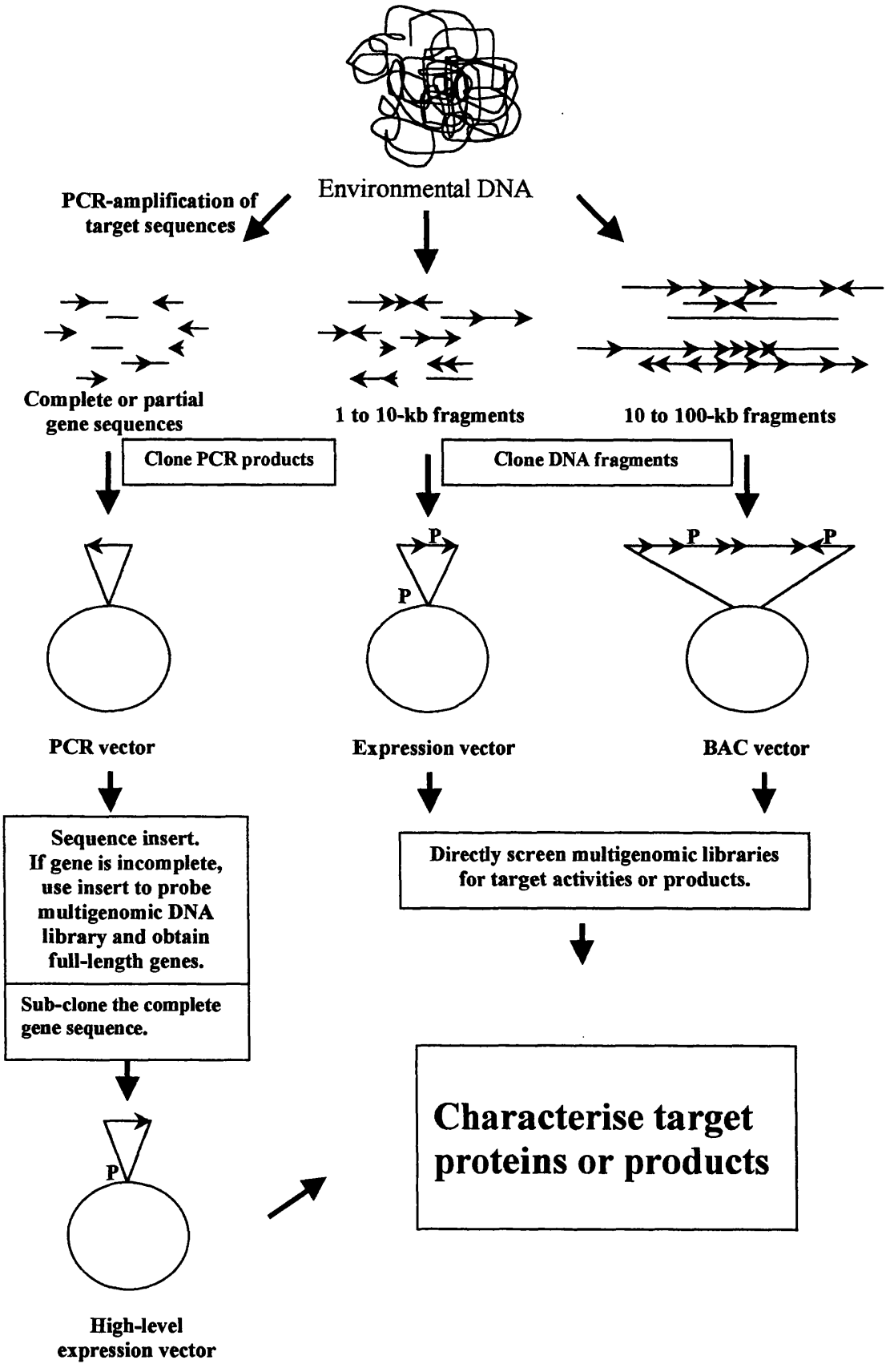


Figure 1.2: Methods for capturing functional gene products encoded by environmental DNA. Genes are represented by horizontal arrows, promoters by the letter P.

expressed and characterised in *Escherichia coli*, thus demonstrating that functional gene products can be obtained from the uncultivated metagenome.

An alternative means of accessing functional proteins from uncultivated microbes is to clone and directly express the metagenomic DNA (Short, 1997). In the first reported case of this kind, Cottrell *et al.* (1999) screened metagenomic libraries derived from uncultured marine microorganisms and directly identified twenty-two different chitinase-expressing clones. This was achieved by extracting DNA from filter-concentrated plankton and cloning 2 to 10-kb DNA fragments into a λ phagemid vector. Recombinant *E. coli* clones were screened for chitinase activity using a fluorogenic analogue of chitin. Chitinase-positive clones were detected within the library by their fluorescence when exposed to UV light.

Positive clones were subsequently isolated and the chitinase activities were functionally characterised directly without sub-cloning. The investigators also directly screened a marine environmental DNA library for cellulase activity but no cellulase-expressing clones were detected.

In separate investigations, environmental DNA libraries derived from uncultivated soil microorganisms were screened for the presence of genes conferring utilisation of 4-hydroxybutyrate (Henne *et al.*, 1999) or lipolytic activity (Henne *et al.*, 2000). Using various agar-plate assays, these investigators identified five different clones utilising 4-hydroxybutyrate, three different clones possessing lipase activity and one clone possessing esterase activity.

Rondon *et al.* (2000) used a bacterial artificial chromosome (BAC) vector to construct two environmental DNA libraries derived from the soil metagenome. Possessing insert sizes greater than 20 kb, these libraries catalogued more than 1 Gbp of DNA predicted to encode more than one million genes. Using a variety of agar-plate screening assays, the functional diversity of the libraries was demonstrated by detecting clones that were heterologously expressing DNase (one clone), antibacterial (one clone), lipase (two clones) and amylase (eight clones) activities. Clones expressing cellulase, chitinase, esterase, keritinase, protease and haemolytic activity or siderophore production were also screened but not detected. The investigators also identified a gene cluster in one sequenced DNA insert demonstrating the potential of BAC clones to contain complete operons. Such libraries could therefore be used to screen for secondary metabolites or

other similar products whose expression is typically encoded by genes and regulatory sequences clustered together in one contiguous segment on the chromosome.

Each of the above examples demonstrated that functional gene products can be obtained from uncultivated microorganisms. The PCR-based method is limited in accessing novel enzyme activities. Since the PCR-based approach used primers that were designed with conserved sequences, detection was restricted to genes adhering to the same consensus. Also, since the primers were designed using the DNA sequence information of previously cultivated organisms, the PCR-based method for enzyme discovery is not culture-independent. Entirely new analogous genes present in the uncultivated DNA would therefore be missed. Another limitation of the PCR-based approach to enzyme discovery is that functional expression of the genes can be achieved only after sequencing and sub-cloning the amplification products. General screening methods cannot be applied easily to such a specific approach.

Prior knowledge of DNA sequence is not required for the expression-cloning approach. Unlike the PCR approach, direct expression of metagenomic DNA is truly culture-independent. The generic application of expression cloning was also demonstrated by using different assays to identify various activities within a single DNA library. Furthermore, functional expression of metagenomic libraries was achieved directly without having to sequence the cloned DNA and sub-clone the target gene. Preliminary investigations of candidate clones could therefore be achieved without a significant investment in time and resource.

1.7: Accessing Enzymes Encoded by the Metagenome *via* Expression-cloning

There are three main phases to accessing functional proteins encoded by environmental DNA: (1) the isolation DNA from the environment without prior cultivation of the organisms present and preparation of the recovered DNA fragments for cloning; (2) the selection of an appropriate vector and host system for application in functional genetic expression and (3) the implementation of a screening strategy that allows for the detection of target heterologous proteins which are encoded by the cloned DNA and expressed by the host organism.

1.7.1: Extraction and purification of DNA from the environment

Many DNA extraction techniques for environmental samples are derived from protocols available in molecular biology manuals (e.g. Ausubel *et al.*, 1989; Sambrook *et al.*, 1989), which have been adapted to suit the particular needs of environmental samples. There are two main approaches to obtaining nucleic acids from environmental samples such as water, soil or sediment; (1) the cell extraction method, which is an *ex situ* approach, whereby microbial cells are separated from the environment prior to cell lysis and isolation of DNA or (2) the direct extraction method, whereby microbes are lysed directly in a slurry of the environmental matrix from which the DNA is isolated.

1.7.1.1: The cell extraction method

For uncultured water-borne microorganisms, DNA is best recovered *via* the cell extraction method because microbes are easily captured and concentrated using filtration technology (Paul & Pichard, 1995). Although this approach is not specific to prokaryotes, larger eukaryotic cells such as phytoplankton can be eliminated through prefiltration. The cell extraction technique for obtaining DNA from uncultivated terrestrial microbes was first published by Torsvik (1980) who modified a cell fractionation procedure (Faegri *et al.*, 1977) and combined it with his own lysis technique. There are two main features that are common to all procedures for obtaining the bacterial fraction from soil and sediment: (1) dispersal of soil/sediment aggregates and (2) centrifugation in which cells and soil particles are separated according to size, buoyant density or both.

Dispersion of soil/sediment aggregates can be achieved through physical techniques (e.g. vigorous shaking or homogenisation in a Waring blender) and/or by chemical means (e.g. chelating agents, ion exchange resins, non-specific detergents) in which the attractive forces between soil particle and cell surface are disrupted. After dispersion, the slurry is centrifuged and the bacterial cells separated from the environmental matrix according to sedimentation velocities and/or buoyant density (Bakken & Lindahl, 1995).

For uncultivated microorganisms obtained *via* the cell-extraction approach, routine procedures exist for cell lysis and DNA extraction (e.g. Sambrook *et al.*, 1989). For those circumstances in which the recovered DNA is highly contaminated with co-

extracted humic substances, more extensive purification is required as described in section 1.7.1.2.

1.7.1.2: The direct lysis method

Most direct extraction methods include a soil/sediment dispersal step followed by *in situ* cell lysis, DNA extraction and purification. These steps do not always proceed in succession and the extraction and purification steps may need to be repeated in order to obtain DNA that is of sufficient quantity and quality for most molecular biological protocols. Various investigators use different approaches to different steps. The dispersal step, whereby soil/sediment aggregates are broken down, releases microbial cells that may be entrapped within the environmental matrix and exposes them to lytic treatment(s). Dispersal of aggregates involves one or more of the following treatments: vigorous shaking (Selenska & Klingmuller, 1991a,b), vortexing (Saano & Lindstrom, 1995), grinding (Flemming *et al.*, 1994), sonication (Ogram *et al.*, 1987) and bead beating (Smalla *et al.*, 1993). These treatments may also cause cell lysis which can be enhanced by addition of lysozyme (Tsai & Olson, 1991), proteinase K plus SDS (Ogram *et al.*, 1987), and/or freeze-thaw treatments (Picard *et al.*, 1992).

After cell lysis, DNA needs to be extracted away from soil and cellular debris. This may be achieved through repeated centrifugations (Trevors *et al.*, 1992), phenol (Smalla *et al.*, 1993), phenol-chloroform (Tsai & Olson, 1991) and/or phenol-chloroform-isoamyl alcohol extractions (van Elsas & Smalla, 1995) or cetyl-trimethyl ammonium bromide (CTAB) and chloroform treatments (Saano & Lindstrom, 1995). DNA is then recovered by precipitation in solutions such as ethanol (Smalla *et al.*, 1993), polyethylene glycol (Ogram *et al.*, 1987) or isopropanol (Tsai & Olson, 1991).

The extraction and precipitation steps may also purify the DNA to a certain degree, however, humic substances and other soil constituents often co-precipitate with the recovered DNA. Humic compounds that persist in DNA extracts create problems for restriction analysis (Tsai & Olson, 1991), PCR amplification (Tsai & Olson,1992a,b), transformation efficiency (Tebbe & Vahjen, 1993) and cloning efficiency (Henne *et al.*, 1999). Additional purification steps are therefore usually required for efficient removal of humic substances. No single procedure effectively purifies DNA from all soil or sediment types. This leads to many different purification strategies. These may include CsCl-EtBr equilibrium density gradient ultracentrifugation (Ogram *et al.*, 1987),

hydroxyapatite chromatography (Ogram *et al.*, 1987; Steffan *et al.*, 1988), gel filtration with Sephadex columns (Tsai & Olson, 1992b), ion-exchange chromatography (Tebbe & Vahjen, 1993; More *et al.*, 1994), glassmilk-purification (Smalla *et al.*, 1993), CsCl-precipitation (Trevors *et al.*, 1992) preparative agarose gel electrophoresis (Knaebel & Crawford, 1995; Myrold *et al.*, 1995), electroelution (Chandler *et al.*, 1997) CTAB-treatment (Zhou *et al.*, 1996) and polyvinylpolypyrrolidone (PVPP)-treatment (Holben *et al.*, 1988; Steffan *et al.*, 1988).

1.7.1.3: Evaluating environmental DNA

All methods for isolating DNA from environmental samples aim at getting a high yield of DNA without unduly shearing it. The recovered DNA must also be sufficiently pure for use in molecular biological techniques such as restriction endonuclease digestion, PCR and cloning. The following describes some of the ways in which environmental DNA is assessed for use in molecular biological techniques.

For pure solutions, DNA yield can be determined spectrophotometrically where one absorbance unit at 260nm is equal to 50μg dsDNA/ml (Sambrook *et al.*, 1989). This method, however, is not suitable for impure DNA solutions due to interference from UV-absorbing contaminants such as proteins and humic acids or other phenolic compounds that co-purify with DNA. A colourimetric method using diphenylamine reagent containing acetaldehyde (Richard, 1974) is an alternative assay for determining the DNA content of highly contaminated DNA solutions (Nesme *et al.*, 1995); however, this method is time-consuming. Alternatively, DNA can be quantified by fluorometric analysis after its specific reaction with either Hoechst 33258 dye, ethidium bromide (Ausubel *et al.*, 1989) or PicoGreen dsDNA quantification reagent (Molecular Probes, Eugene, OR, USA). The fluorometric methods for DNA quantification are more sensitive than the spectrometric methods and require smaller sample volumes for measurement. DNA yield may also be determined from the intensity of ethidium bromide-stained DNA fragments in agarose gel and comparison with standards of known concentration (Smalla, *et al.*, 1993).

Because coextracted contaminants such as humic substances have been demonstrated to inhibit restriction enzymes (Tsai & Olson, 1991), PCR amplification (Tsai & Olson,1992a,b), transformation efficiency (Tebbe & Vahjen, 1993) and cloning efficiency (Henne *et al.*, 1999), the purity of environmental DNA must be assessed prior

to use in cloning protocols. The purity of a DNA solution can be determined *via* spectral analysis by measuring the absorbance of the DNA solution at 230nm (A_{230nm}), 260nm (A_{260nm}) and 280nm (A_{280nm}). A DNA solution is considered pure if the A_{260nm} to A_{230nm} ratio is between 1.8 and 2.3 and the A_{260nm} to A_{260nm} ratio is between 1.5 and 2.0. (Marmur, 1963). Another means of assessing the purity of a DNA sample is to test its amenability to general molecular manipulations such as restriction digestion or PCR amplification (e.g. Clegg *et al.*, 1997).

The molecular weight of DNA extracted from environmental samples is a criterion for assessing the harshness of a given extraction procedure and/or the suitability of the isolated DNA for cloning. The assessment of DNA fragment size distribution is achieved easily by visualisation of ethidium bromide-stained DNA fragments in agarose gel and comparison with standards of known fragment size (Sambrook, 1989).

Provided that the DNA obtained from an environmental source is of sufficient quantity, purity and quality, it should be suitable for use in any appropriate procedure for genomic cloning (e.g. Ausubel *et al.*, 1989; Sambrook *et al.*, 1989). The selection of an appropriate cloning protocol, however, may require an empirical approach.

1.7.2: Prokaryotic systems for expression-cloning

This section describes the general aspects of expressing heterologous genes under the control of prokaryotic transcriptional, translational and protein machinery. The use of eukaryotic systems for expressing heterologous sequences will be discussed in section 1.7.4.

1.7.2.1: Heterologous gene expression in *E. coli*

E. coli is the most frequently used prokaryotic system for heterologous gene expression. It is favoured by many investigators because it is the most studied organism in terms of biochemistry and genetics. *E. coli* is also easily manipulated experimentally and grows rapidly in relatively simple media. There are several manuals available that describe the techniques for heterologous expression in *E. coli* (e.g. Goeddel, 1990; Glover & Hames, 1995; Tuan, 1997). Most protocols for producing heterologous proteins in *E. coli* are based on the assumption that the starting material is either a single, well characterised gene, an isolated open reading frame (ORF) or a population of cDNA molecules. These

procedures typically rely on vectors that have been designed specifically for the large-scale, high-level expression of heterologous proteins. Such high-level expression vectors are generally not required (or even appropriate) for expression-based cloning of metagenomic DNA, which is essentially a shotgun approach. In this work, the terms "expression-cloning" and "functional genetic expression" refer to the means by which functional proteins are accessed from (meta)genomic DNA. This is in contrast to "high-level expression" which refers to the means by which functional proteins are produced in large amounts by (sub)cloning specific, generally well characterised sequences. Many of the same principles for high-level expression, however, may also be applied to expression cloning approaches and many of the factors that affect production of protein in a high-level *E. coli* expression system can also affect the functional expression of (meta)genomic DNA.

Vectors for expression-cloning in *E. coli*.: Environmental DNA expression libraries have been constructed using plasmids (Henne *et al.*, 1999; Henne *et al.*, 2000), bacteriophage λ- and λ phagemid-based vectors (Cottrell *et al.* 1999) and BAC cloning vectors (Rondon *et al.*, 2000). Theoretically, however, any other type of *E. coli* cloning vector may be used for expressing environmental DNA provided that it can be replicated and maintained stably in the host cell.

The salient features of a prokaryotic system for expression cloning include a set of appropriately configured genetic elements that effect both transcriptional and translational aspects of protein production. For heterologous expression of genes encoded by environmental DNA, these genetic elements may be vector-encoded and/or encoded by foreign sequences contained within the cloned DNA fragment. The cloning vector should also possess a selectable marker, such as an antibiotic-resistance gene in order to facilitate phenotypic selection of the vector, an origin of replication (*ori*) that determines vector copy number and a multiple cloning site to accommodate the cloned insert. If expression of the cloned gene is to be driven by vector-encoded sequences then the multiple cloning site must be positioned downstream of these control sequences (reviewed by Balbas & Bolivar, 1990; Hannig & Makrides, 1998).

Factors that may affect functional expression of cloned (meta)genomic DNA in *E. coli*.:

Transcription-related factors: In *E. coli*, there is a good correlation between promoter strength and the degree to which the –35 and –10 DNA elements agree with the consensus sequence, (TTGACA and TATAAT, respectively, for $E\sigma^{70}$ promoters) (Gross *et al.*, 1992). Transcription initiation can also be modulated by a variety of mechanisms, namely the interaction of one or more regulatory proteins that are provided *in trans* and that react with specific sequences (e.g. operators and activator sites) in the vicinity of the promoter (reviewed by Gralla, 1990). If heterologous expression relies on transcription being initiated from a vector-encoded promoter, then the foreign gene must be inserted downstream of and in the proper orientation to this promoter. Conversely, if expression is to be initiated from a cloned promoter then it must be recognised by the host RNA polymerase holoenzyme.

Regulation of transcription is achieved through trans-acting effector molecules which can be either gene-specific or under global control (Gralla, 1990). Although many promoters encoded by vectors may be easily repressed or induced as required (e.g. isopropyl-β-D-thiogalactopyranoside (IPTG) for induction of *lac*-based promoters), there can be no targeted regulation of transcription from unknown foreign promoters contained on a cloned DNA fragment.

Another transcription-related factor that may affect heterologous gene expression is the presence of transcription terminators. Appropriately placed transcription terminators act as a barrier to RNA elongation. They minimise sequestration of RNA polymerase that might otherwise be re-initiating transcription. Legitimate transcription termination also avoids unnecessary transcription, which may exert a metabolic strain to the host. Non-specific and premature transcription termination may also occur within any transcript due to terminator-like secondary structures forming within the transcript (Balbas & Bolivar, 1990).

Translation-related factors: A set of general rules for giving maximal efficiency of translation has been compiled by Stormo (1986), they include: (1) the preferential initiation codon is AUG, although GUG, UUG, AUU and AUA are not uncommon; (2) the Shine-Dalgarno (SD) sequence within the ribosome-binding site (RBS) should have at least four nucleotides taken from the consensus sequence, AGGAGG; (3) the spacing

between the SD site and the start codon should be 9±3 nucleotides; (4) besides the SD site, the sequence upstream of the start codon should be A/U-rich; (5) the region around the initiation site should be free of secondary structures.

Although many ribosome-binding sites do not meet each of these requirements, they do work adequately (Stormo *et al.*, 1982).

The preference with which a specific organism uses a particular degenerate codon for a particular amino acid is referred to as its codon usage. The codon usage is another factor that may affect protein expression. Ikemura (1982) discovered that highly expressed *E. coli* genes preferentially use codons recognised by abundant tRNA species. He proposed that the availability of charged tRNAs may be a rate-limiting factor of protein synthesis. Heterologous genes that contain a substantial number of codons that are rarely used in *E. coli* may thus be expressed inefficiently. One strategy to minimise the effects of preferential codon usage in *E. coli* is to use host strains that have an expanded intracellular tRNA pool (e.g. BL21-CodonPlus *E. coli* strains produced by Stratagene, La Jolla, CA, USA).

Another factor that affects protein expression is the stop codon which is an indispensable signal for termination of mRNA translation. The three stop codons differ in the efficiency of translation termination with a strong bias in favour of UAA in highly expressed genes (Sharp & Bulmer, 1988). Inappropriate termination of translation may result in inactive polypeptide.

Protein-related factors. To become mature and active proteins, polypeptides must fold into their native conformations, their disulphide bonds, if any, must form and, in the case of multi-subunit proteins, the subunits must be properly combined. Moreover, many proteins require post-translational modification such as proteolytic cleavage, glycosylation, phosphorylation, acetylation, addition of lipids to specific amino acid side chains and incorporation of prosthetic groups. The failure of the host cell to exert any of these processes will likely result in an incorrectly folded and/or inactive protein.

Sometimes cloned proteins are recognised as abnormal and are degraded by the host proteolytic system (Itakura *et al.*, 1977). One way that protein degradation can be minimised is to utilise *E. coli* strains that are defective in proteolysis (e.g. strains containing *lon*, *htpR* and/or *clpA* mutations) (Gottesman, 1990). It should be kept in

mind, however, that no one protease mutant will suffice to stabilise all foreign proteins. The extent of protein degradation is also greatly affected by the culture conditions used with the recombinant strain. Stress factors known to increase the rate of proteolysis include nutrient starvation and conditions that favour the heatshock response (Enfors, 1992). By minimising these factors, protein degradation may be slowed enough so as to allow for some accumulation and detection of active gene product.

Another factor that may affect the activity of an expressed protein is improper compartmentalisation. In *E. coli*, secreted proteins may be released into the periplasm or integrated into or transported across the outer membrane. *E. coli* does not normally secrete its proteins into the culture medium. Heterologous bacterial proteins that are secreted out of the cytoplasm when expressed in *E. coli* utilise the components of the general targeting pathway (reviewed by Stader & Silhavy, 1990). A foreign protein that is normally secreted in the homologous host but is not recognised by the secretion pathways of *E. coli*, may assemble poorly if at all in the cytoplasm of *E. coli*. For example, disulphide bond formation, which is required for activity in certain extracellular proteins, does not seem to be favoured in the cytoplasm of *E. coli* (Pollitt & Zalkin, 1983).The rapid expression of heterologous protein in *E. coli* may also lead to the formation of inclusion bodies due to aggregation of partially folded and inactive protein (Williams *et al.*, 1982). Incubation temperature and growth rate are two parameters that can be adjusted in order to reduce inclusion-body formation. There are a number of reports which show that lowering the temperature of growth reduces inclusion-body formation (Schein & Noteborn, 1988; Takagi *et al.*, 1988; Schein, 1991). Media composition and pH values are additional factors that affect the yield of correctly folded, soluble proteins (Schein, 1991).

A gene product that compromises its host is an other factor that may affect heterologous expression. Certain foreign proteins may be toxic to the host. A recombinant that is metabolically compromised by a cytotoxic product is more likely to be selected against and overgrown by cells that do not possess the activity.

***Host-related factors*:** Host-cell physiology can also affect the level of protein expression. Factors that may be important include the choice of nutrients and environmental factors such as temperature and dissolved oxygen. Another important component of an expression system which determines the final amount and activity of a heterologous protein is the genetic background of the host strain. An appropriate host

must possess the genetic traits necessary to work in conjunction with the expression signals of the system and/or increases mRNA or protein stability. While it may relatively be straightforward to select an appropriate host suitable to vector-encoded sequences, the metagenome-encoded sequence is an undefined variable. One way to increase the chance of expressing genes under the control of sequences encoded by the metagenome is to use the recombinant library to transform/transfect different *E. coli* strains or even different species of host organisms. This could be achieved by cloning the DNA into, say, a broad host-range vector and "shuttling" the recombinant library between hosts.

1.7.2.2: Heterologous expression in bacteria other than *E. coli*

E. coli is somewhat promiscuous in its ability to recognise foreign bacterial transcriptional and translational signals. There are numerous investigations reporting heterologous bacterial gene expression independent of vector-encoded sequences; however there are also numerous reports in which *E. coli* did not express bacterial genes due to its failure to recognise transcriptional, translational and/or post-translational control sequences. One reason for using alternative bacterial expression systems is that any expression control sequences that are "missed" by *E. coli*, may be recognised by a different host species. Another reason for selecting a different bacterial species as host for heterologous expression is that *E. coli* may not possess the necessary biochemical pathways that are required for phenotypic expression of a certain function, for example, degradation of xenobiotic compounds or photosynthesis. Using various bacterial hosts with different genetic and metabolic backgrounds for expression screening may increase the repertoire of functional gene products that are encoded by the environmental DNA.

Bacterial cloning systems that have the potential for use in expression cloning include, *Bacillus subtilis* (Harwood & Cutting, 1990). *Streptomyces* (Hopwood *et al.*, 1987), cyanobacteria (Kuhlemeier & van Arkel, 1987) and *Pseudomonas* sp. (Brunschwig & Darzins, 1992). It must be stressed however that the utility of these and other bacterial systems have yet to be demonstrated for functional expression of environmental DNA.

1.7.2.3: Heterologous gene expression of non-bacterial genes in *E. coli*

Functional expression of eukaryotic DNA in *E. coli*: In order to achieve functional expression of eukaryotic DNA in *E. coli*, the gene in question must be placed under the

control of vector-encoded sequences that are recognised by the host. In other words, the gene must be cloned downstream of an *E. coli* promoter, in the correct reading frame and with the start codon properly positioned with respect to a vector-encoded Shine-Dalgarno sequence. Furthermore, in order to produce an active protein, the heterologous transcript should not require any processing that cannot be achieved in *E. coli*, such as alternative splicing. Similarly, the produced polypepetide must not require post-translational modifications that cannot be carried out in *E. coli*.

Functional expression of archaeal DNA in *E. coli*: Due to fundamental differences in transcriptional and translational machinery (reviewed by Bell & Jackson, 1998), it would be intuitive to believe that functional genetic expression of archaeal genomic DNA would not be possible in *E. coli*. There are however several documented cases in which archaeal genes were expressed independent of vector-encoded expression control sequences. It has been demonstrated that DNA from *Methanococcus voltae* (Wood *et al.*, 1983) and *Methanococcus vannielii* (Meile & Reeve, 1985) could complement *hisA* mutants of *E. coli*. Because the mechanism by which archaeal ribosomes recognise the start codon is analogous to that in bacteria (Dennis, 1997), it is not surprising that translation of archaeal sequences could be achieved in *E. coli*. The archaeal transcription initiation machinery, however, is most closely related to that of the *Eukarya* (Bell & Jackson, 1998), and it seems unlikely that these elements could have been recognised by the *E. coli* host. Subsequent to these initial observations, Cue *et al.* (1985) identified, in each of the methanococcal DNAs, bacterial-like promoters possessing –35 and –10 elements upstream of each of the *hisA*-complementing genes. The investigators also identified putative SD sequences appropriately spaced upstream of translational start codons. Because limited information about archaeal transcription was available at the time, the investigators did not search for archaeal-like promoter elements. The authors predicted that, although the archaeal translation initiation signals were recognised by *E. coli*, it seemed possible that the bacterial-like promoters were fortuitously recognised by the host.

There are several other documented cases of functional expression of archaeal DNA in *E. coli*: for example, *Sulfolobus solfotaricus* DNA was able to confer thermostable α-amylase and glycosyltransferase activities to *E. coli* host cells (Kobayashi *et al.*, 1996); an ether-linked lipid biosynthetic gene encoded by *Sulfolobus acidocaldarius* genomic DNA was detected in *E. coli via* expression screening (Ohnuma, 1994); carbon

monoxide dehydrogenase subunits (Eggen *et al.*, 1991b) as well as an acetyl coenzyme A synthase (Eggen *et al.*,1991a), encoded by *Methanothrix soehngenii* DNA, were detected in *E. coli* through immunological screenings; and finally, DNA cloned from *Pyrococcus furiosus* was shown to confer thermostable esterase (Ikeda & Clark, 1998) and β-glucosidase (Voorhorst *et al.*, 1995) activities to its *E. coli* hosts. In each of these cases, the expression was achieved regardless of the orientation of the cloned DNA fragment suggesting that the archaeal DNA harboured sequences that were recognised as signals for fortuitous transcription as well as for translation. These findings indicate that *E. coli* may be employed as a host for expressing DNA derived from *Archaea* without the need for specifically designed vectors.

1.7.3: Archaeal systems for heterologous genetic expression

Molecular genetic techniques have been developed for halophilic archaea (Holmes & Dyall-Smith, 1990; Patenge *et al.*, 2000) and partly for thermophilic archaea (Aravalli & Garret, 1997) and methanogens (reviewed by Tumbula & Whitman, 1999). Archaeal vector-host systems provide an alternative route to heterologous expression of DNA from uncultivated archaea; however their utility for constructing environmental DNA libraries have yet to be demonstrated.

1.7.4: Eukaryal systems for heterologous genetic expression

The cloning of a eukaryotic gene in *E. coli*, does not in general lead to an efficient synthesis of the corresponding protein. This is because the eukaryotic DNA lacks the specific sequences necessary for it to be recognised by the host transcriptional, translational and protein processing machinery. In cases where the gene of interest is eukaryotic and incompatible with bacterial expression systems, yeast represents the next best expression system. Other microbial eukaryotic cloning systems include *Pichia pastoris* (Romanos *et al.*, 1991; Clare *et al.*, 1991), *Kluyveromyces* sp. (Fleer *et al.*, 1991) and *Aspergillus* sp. (Christiensen *et al.*, 1988).

Typically, heterologous expression in eukaryotic systems such as yeast involves placing a cDNA or subcloned coding sequences under the control of a strong eukaryotic promoter that will direct the synthesis and (over)production of the encoded mRNA; however, the utility of expression-screening using environmental mRNA has yet to be demonstrated.

1.8: Screening Systems

The screening of expression libraries is achieved through phenotypic selection using either a solid phase, such as agar plates or filter paper, or *via* microtitre well plates. Screening can be based on, for example, complementation of auxotrophy, ligand binding, antibody or other protein binding, resistance to cytotoxic substances such as antibiotics or by testing for catalytic activity. There are many factors that may be present during a screening assay that may affect whether a given protein activity is detected. The presence and concentration of constituents; such as substrate, product, cofactors, coenzymes and metal ions, will determine whether a particular enzyme activity will be detected. Assay conditions such as ionic strength, temperature and pH, or the presence of constituents such as inhibitors, will also affect whether a target enzyme is detected. Performing several screening assays under different conditions may increase the likelihood of detecting the functional target protein.

1.9: Aims

Environmental DNA libraries for expression cloning have been reported for uncultured marine and soil metagenomes. To date, no investigation describing expression libraries derived from thermobiotic habitats has been reported. The main aim of this work was to investigate the potential of expression-cloning in detecting activities encoded by DNA that was extracted from uncultivated thermophilic microorganisms. Other goals of this study included: (1) analysing, with respect to DNA yield, shearing and purity, the suitability of extraction procedures for obtaining clonable DNA from geothermal sediments; (2) investigating, using the DNA derived from geothermal sediment, various cloning protocols for generating stable, representative environmental libraries; (3) demonstrating the applicability of environmental libraries by performing various screening assays for detecting different functional thermostable proteins; (4) performing molecular and expression analyses on isolated recombinants in order to identify the molecular determinants responsible for the observed phenotype(s).

Chapter 2
Materials & Methods

2.1: Reagent Sources

Unless otherwise stated, all chemicals used were from BDH Ltd. (Poole, Dorset) or Sigma Chemical Co. (Poole, Dorset). and were of the highest grade available. Nutrient broth, and other ingredients for microbiological media were obtained from Oxoid Ltd. (Basingstoke, Hants.). Tryptose phosphate medium was obtained from Difco Laboratories (East Molesey, Surrey).

The antibiotics ampicillin, kanamycin, chloramphenicol, streptomycin and tetracycline were obtained from Sigma.

2.2: Bacterial Strains, Plasmids and λ Phagemid

Bacterial strains, plasmids and λ phagemid used in this work are listed in Table 2.1. Stock cultures of bacterial strains were maintained as cell suspensions in 25% (v/v) glycerol at -80°C (Sambrook *et al.*, 1989).

2.3: Sterilisation Conditions

Unless otherwise stated, all buffers, reagents, culture media and durable labware, were sterilised by autoclaving at 121°C for 20 minutes.

Strain or plasmid	Genotype or relevant description	Reference or source
***E. coli* strains**		
JM107	e14$^-$ (McrA$^-$) *end*A1 *gyr*A96 (Nal^r) *thi*-1 *hsd*R17 ($r_K^-m_K^+$) *sup*E44 *rel*A1 Δ(*lac-pro*AB) [F′ *tra*D36 *pro*AB *lac*I^{q}Z ΔM15]	Yanisch–Perron *et al.* (1985)
TOP10	F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *deo*R *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^r) *end*A1 *nup*G	Invitrogen Corp. (Carlsbad, CA, USA)
XL1-Blue	*rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* [F′ *pro*AB *lac*I^q ZΔM15 Tn10 (Tet^r)]	Stratagene Ltd. (La Jolla, CA, USA)
XL2-Blue MRF'	Δ(*mrc*A)183 Δ(*mcr*CB-*hsd*SMR-*mrr*)173 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* [F′ *pro*AB *lac*I^q ZΔM15 Tn10 (Tet^r) Amy (Cam^r)]	Stratagene
NovaBlue	*end*A1 *hsd*R17 ($r_{k12}^-m_{k12}^+$) *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac* [F′ *pro*A$^+$B$^+$ *lac*I^q ZΔM15 Tn10 (Tet^r)	Novagen (Madison, WI, USA)
BM25.8	*sup*E44, *thi* Δ(*lac–pro*AB) [F' *tra*D36, *pro*AB *lac*I^{q}Z ΔM15] λimm434 (Kan^r)P1 Cam^r hsdR ($r_{k12}^-m_{k12}^-$) Host for pTriplEx excision from phagemid λTriplEx (see below)	Stratagene
SASX41B	*hem*A41 *rel*A1 *spo*T1 *met*B1 *rrn*B-2 *mcr*B1 *cre*C510	*E. coli* genetic stock center (New Haven, CT, USA)
Phagemid		
λTriplEx	*lac*Z' Ap^r *lox*P pTriplEx	Clontech Laboratories Inc. (Palo Alto, CA, USA)
Plasmids		
pUC18/19	*lac*Z' Ap^r	Yanisch–Perron *et al.* (1985)
pT7Blue	*lac*Z' Ap^r	Novagen
PCR-XL-TOPO	*lac*Z' *ccd*B Km^r	Invitrogen
pTriplEx	*lac*Z' Ap^r *lox*P pTriplEx is excised from λTriplEx by a *Cre-lox* recombinase-mediated reaction in an appropriate host (e.g. *E. coli* BM25.8)	Clontech
pCR–	pCR-XL-TOPO carrying a short (0.5kb) DNA fragment and used in *E. coli* TOP10 cells as negative control for screening and protein analysis	This work
p5ICE16	2.9 kb fragment derived from Iceland geothermal sediment site 16 cloned into pCR-XL-TOPO	This work
p6ICE16	4.7 kb fragment derived from Iceland geothermal sediment site 16 cloned into pCR-XL-TOPO	This work
pPHOS22	1.7 kb fragment derived from Iceland geothermal sediment site 22 cloned into pCR-XL-TOPO	This work
pQR126	*lac*Zα Km^r.Amylase expression plasmid	J. Ward (unpublished)

Table 2.1: Bacterial strains, plasmids and λ phagemid used in this work. Genetic markers and phenotypes are presented using standard nomenclature (Demerec *et al.*, 1966; Bachmann, 1990).

2.4: Culture Media

E. coli strains were routinely grown either in nutrient broth No. 2 or on nutrient agar prepared according to the manufacturer's instructions. Culture media used in special protocols are listed below.

2.4.1: SOC media

SOC medium was used for the recovery and outgrowth of newly transformed *E. coli* strains (Sections 2.8.10 & 2.8.11).

SOC Medium

Constituent	**L^{-1}**
Tryptone	20g
Yeast extract	5g
1M NaCl	10ml
1M KCl	2.5ml

The pH was adjusted to 7.0 with 5M NaOH prior to sterilisation. After allowing the broth to cool to ~55°C, the following filter-sterilised ingredients were added aseptically.

Constituent	**L^{-1}**
1M $MgCl_2$	10ml
1M $MgSO_4$	10ml
2M Glucose	10ml

2.4.2: Luria-Bertani (LB) broth and agar plates

LB media were used for *E. coli* transformation (Sections 2.8.10 & 2.8.11) and transfection protocols (Sections 2.8.13 & 2.8.14).

LB broth

Constituent	L^{-1}
Tryptone	10g
Yeast extract	5g
NaCl	5g

The pH of LB broth was adjusted to 7.0 with 5M NaOH prior to sterilisation.

LB agar

Constituent	L^{-1}
LB broth	as above
Agar	15g

LB top agar

Constituent	L^{-1}
LB broth	as above
Agar	7g

The pH of LB broth was adjusted to 7.0 with 5M NaOH prior to addition of agar and sterilisation.

2.4.3: Addition of antibiotics and δ-aminolaevulinic acid

After autoclaving and allowing the media to cool to ~55°C, filter-sterilised antibiotic solution was added to the medium as necessary. Antibiotic final concentrations were: ampicillin, 50μg/ml broth or 100μg/ml agar; kanamycin, 50μg/ml; tetracycline, 15μg/ml; chloramphenicol, 35μg/ml and streptomycin, 20μg/ml.

For growth of the *E. coli hem*A mutant strain SASX41B, δ-aminolaevulinic acid was added to LB media at a final concentration of 20μM (Avissar & Beale, 1989).

2.5: Sampling Geothermal Sediments

Volumes ranging from 5ml to 500ml wet sediment were aseptically collected from various geothermal sites located within New Zealand and Iceland. In the laboratory, sediments were aliquoted into 50-ml polypropylene tubes and stored at -80°C.

New Zealand geothermal sediments (designated as Tok) were collected from the Tokaanu thermal region, Central North Island, New Zealand (D. Cowan, personal communication).

Icelandic geothermal sediments (designated as Ice) were collected from several thermal sites located in the area of south-west Iceland (This work).

Temperature and pH values of geothermal sites were determined with a Solomat 520C digital thermometer/pH meter (Norwalk, CT, USA).

2.6: Direct Extraction of DNA from Geothermal Sediments

2.6.1: Mortar and pestle method

Direct DNA extraction using a mortar and pestle plus SDS (MPS method) used in this work is a modification of that developed by Saano & Lindstrom (1995).

Homogenisation: In replicate, 5ml sodium phosphate-EDTA (NE) buffer (120mM sodium phosphate buffer (pH 8.0), 50mM EDTA) was added to a sterile mortar containing 10g wet geothermal sediment plus ~0.5g sterile quartz sand. The sample was ground with a sterile pestle until homogenous and aseptically transferred to a 50-ml polypropylene tube. Both mortar and pestle were rinsed with 5ml NE buffer which was then added to the homogenised sediment.

Cell lysis: Lysozyme was added to the sample at a final concentration of 20mg/ml. The sample was incubated with periodic shaking at 37°C for 30 minutes. After addition of sodium dodecyl sulphate (SDS) (1% (w/v) final concentration) and proteinase K

(100mg/ml final concentration), the sample was incubated with shaking at 65°C for 1 hour. The sample was then passed through three freeze-thaw cycles by alternating between incubating in a dry ice/ethanol bath for 10 min and thawing in a 65°C water bath for 20 min. The sediment slurry was then adjusted to 0.7M NaCl before adding 1/10 volume cetyl-trimethyl ammonium bromide (CTAB) solution (10% (w/v) CTAB in 0.7M NaCl, pH 8.0). The sample was gently vortexed and incubated at 65°C for 20 minutes.

Extraction and precipitation of crude DNA: After adding an equal volume of chloroform, the sediment sample was vortexed gently before transferring to a 30-ml centrifuge tube and centrifuging at 4°C for 15 minutes at 9000x g. The upper aqueous phase was retained and transferred to a fresh centrifuge tube. 2ml SNE buffer (0.7M NaCl, 120mM sodium phosphate buffer (pH 8), 50mM EDTA) were added to the sediment-containing organic phase. The sediment was gently vortexed and the back-extracted DNA recovered by centrifuging as described above. The back-extracted aqueous phase was transferred to a separate centrifuge tube and the sediment-containing organic phase was discarded. Crude DNA was precipitated from the retained aqueous phases by adding an equal volume of isopropanol and incubating at -20°C for at least one hour. The DNA was recovered by centrifugation at 4°C for 30 minutes at 10000x g. The crude DNA pellets were briefly dried under vacuum and dissolved in 2.5ml Tris-EDTA (TE) buffer (10mM Tris-Cl (pH 8.0), 1mM EDTA).

DNA purification: In preparation for anion-exchange chromatography, primary and corresponding back-extracted DNA samples were combined and adjusted by adding 10 volumes of equilibration buffer (50mM MOPS (pH 7.0), 750mM NaCl, 15% (v/v) isopropanol, 0.15% (v/v) Triton X-100). The DNA was applied to a 500G column (Qiagen Ltd., Crawley, West Sussex) and purified according to the manufacturer's recommendations for genomic DNA. Precipitated and dried DNA was dissolved in 200μl TE buffer (pH 8). All DNA samples were stored at -20°C.

2.6.2: Bead beating method

Environmental DNA was extracted directly from pre-weighed geothermal sediments using the FastDNA spin kit for soil (Bio101 Inc., Vista, CA, USA) and Mini-Bead Beater-8 (BioSpect Products Ltd., Bartlesville, OK, USA). The procedure was conducted as described in the manufacturer's instructions for soil DNA extraction,

except that the bead beater was set at ~1500 rpm for 1 minute. The DNA was eluted from each column with 200μl distilled H_2O.

2.7: Analytical Procedures

2.7.1: Dry sediment weights

Dry weights of sediment were determined using a HG53 halogen moisture analyser (Mettler-Toledo Ltd., Beaumont Leys, Leicester). Aliquots of wet sediment were pre-weighed then dried to a constant weight at 80°C.

2.7.2: Spectrophotometry

Routine DNA samples were quantified by spectrophotometric analysis using a Beckman DU7500 spectrophotometer (High Wycombe, Bucks.) according to the method of Sambrook *et al.* (1989) in which one absorbance unit at 260nm is equal to 50μg dsDNA/ml. A DNA solution was considered pure if the A_{260nm} to A_{230nm} ratio was between 1.8 and 2.3 and the A_{260nm} to A_{280nm} ratio was between 1.5 and 2.0. (Marmur, 1963).

2.7.3:PicoGreen assay for dsDNA

When measuring impure or limited amounts of DNA, concentration was determined fluorometrically using the PicoGreen dsDNA quantitation reagent and kit (Molecular Probes, Eugene, OR, USA). The procedure was conducted as described in the manufacturer's instructions. Readings were carried out using a TD-700 Laboratory Fluorometer (Turner Designs, Sunnyvale, CA, USA), (excitation ~480nm, emission ~570nm). Known concentrations of λDNA were used to produce a standard curve for each batch of assays.

2.7.4: Agarose gel electrophoresis

DNA fragments were routinely separated by gel electrophoresis in 1% (w/v) agarose gel as described by Sambrook *et al.* (1989). The electrophoresis buffer used was Tris-acetate-EDTA (TAE) (40mM Tris-base, 20mM sodium acetate, 2mM EDTA, pH 8.0). Following electrophoresis, gels were stained in TAE buffer containing 0.05μg/ml ethidium bromide. After destaining in TAE buffer, DNA bands within the gel were

visualised and photographed under UV light. In some cases DNA was quantified by visual analysis on agarose gels using a known concentration of standard DNA as a reference.

2.7.5: SDS-polyacrylamide gel electrophoresis

Proteins were analysed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) according to Ausubel *et al.* (1995).

Gels: Stock acrylamide/bis-acrylamide monomer solution was obtained from National Diagnostics (Hessle, East Riding of Yorkshire). Protein resolving gels consisted of 15% acrylamide/0.4% bis-acrylamide in 375mM Tris-Cl (pH 8.3) and 0.1% (w/v) SDS. Protein stacking gels consisted of 6% acrylamide/0.16% bis-acrylamide in 125mM Tris-Cl (pH 6.8) and 0.1% (w/v) SDS. Gels were polymerised by adding 0.8mg/ml ammonium persulphate and 0.05% (v/v) N,N,N',N'-tetramethylenediamine (TEMED).

Sample preparation: An appropriate volume of *E. coli* culture was centrifuged at maximum speed for 2 minutes in a 1.5-ml microfuge tube and re-suspended in 200μl sample buffer (125mM Tris-Cl (pH 6.8), 2% (w/v) SDS, 20% (v/v) glycerol, 0.005% (w/v) bromophenol blue, 20μl mercaptoethanol). Samples were frozen at -20°C for two hours and then boiled for 10 minutes before loading 20μl of the protein sample for electrophoresis. Protein molecular mass markers (Calbiochem, Cambridge, MA, USA) were electrophoresed on each gel as standards.

Electrophoresis: Protein gels were electrophoresed overnight at constant voltage (50-75V) in Tris-glycine-SDS buffer (25mM Tris base, 200mM glycine, 0.1% (w/v) SDS).

Detection of protein bands: Protein gels were stained in Coomassie Blue solution (0.2% (w/v) Coomassie Brilliant Blue R, 50% (v/v) methanol, 10% (v/v) glacial acetic acid) on an orbital shaker for 30 minutes. Gels were destained overnight in a 10% (v/v) methanol plus 10% (v/v) acetic acid solution.

2.8: Molecular Biology Techniques

2.8.1: DNA precipitation

In those cases when DNA solutions required a desalting, buffer exchange, enzyme removal and/or concentration step, DNA was ethanol-precipitated according to Ausubel

et al. (1995). When working with volumes less than 50μl, 2μl pellet paint (Novagen) were added to the DNA solution before precipitation. This step coloured the DNA pellet pink, making it easier to visualise for recovery. The precipitated DNA pellet was dried under vacuum, re-suspended in the appropriate diluent and store at -20°C.

2.8.2: Restriction digests

Restriction endonucleases and reaction buffers were obtained from New England Biolabs (Beverly, MA, USA). Digests were carried out at the manufacturer's recommended temperature according to the method of Ausubel *et al.* (1995). Restriction reactions were stopped and prepared for agarose gel electrophoresis by adding 1/10 volume of 10x DNA gel loading buffer (20% (w/v) Ficoll 400, 0.1M EDTA, 1.0% (w/v) SDS, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol). Alternatively, if restricted DNA was to be used directly in an enzymic reaction, the digest was stopped either by heating at 65°C for 10 minutes or by purification using QIAex II gel extraction kit (Qiagen).

2.8.3: Preparative gel electrophoresis

To avoid exposure to both ethidium bromide and UV light, whenever possible, environmental DNA was size-fractionated for cloning *via* a modified agarose gel electrophoresis protocol (Section 2.7.4). 0.5μg λDNA (Gibco Life Technologies, Gaithersburg, MD, USA) cleaved with the restriction endonuclease *Hin*d-III (λ*Hin*d III) (Section 2.8.2) in 10μl of 1x DNA gel loading buffer was placed into the two flanking wells of a 1% (w/v) agarose gel. Purified environmental DNA in 1x DNA gel loading buffer was placed into the inner well(s) of the gel. The DNA was electrophoresed according to Section 2.7.4 except that after electrophoresis, a strip from each side of the gel was excised and stained in ethidium bromide. To localise the DNA, the two excised lanes were placed under UV light and, using the λDNA as a reference, DNA fragments of the appropriate size within the gel were marked using a razor blade. The two stained strips were then repositioned against the unstained gel and, using the razor marks as a guide, the gel region containing the environmental DNA fragments was excised.

DNA fragments were extracted from agarose gel using the QIAex II gel extraction kit (Qiagen). The procedure was conducted according to the manufacturer's

recommendations except that 60μl of size-fractionated DNA was eluted with $T_{low}E$ buffer (10mM Tris-Cl (pH 8), 0.1mM EDTA).

2.8.4: Blunt-end polishing of DNA ends

Single-stranded 3' and 5'-termini were removed from DNA using Vent DNA polymerase (New England Biolabs). Blunt-end polishing was carried out in 100-μl reactions in the presence of 200μM dNTPs (50μM each of dATP, dCTP, dGTP & dTTP) and 1x ThermoPol buffer (20mM Tris-Cl (pH 8.8 at 24°C), 10mM KCl, 10mM $(NH_4)_2SO_4$, 2mM $MgSO_2$, 0.1% (v/v) Triton X-100). 1U Vent DNA polymerase was added and the mixture incubated at 55°C for 30 minutes. The reaction conditions allowed for primer extension without strand displacement of the DNA template. The reaction was quenched on ice before purifying the DNA using the QIAquick PCR purification kit (Qiagen).

2.8.5: Addition of deoxyadenosine to the 3' termini of blunt-ended DNA

Up to 5μg blunt-ended DNA was adjusted to 95μl with distilled H_2O. 11μl of 10x ThermoPol buffer, 5μl of 2mM dATP and 1U Taq DNA polymerase (Gibco Life Technologies) were added and incubated at 72°C for 20 minutes. The reaction mixture was purified using the QIAquick PCR purification kit according to the manufacturer's instructions except that 100μl TE buffer (pH 8) was used to elute the DNA. The eluted DNA was immediately used for cloning (Section 2.8.9).

2.8.6: Dephosphorylation of DNA

Unless otherwise stated, the removal of phosphate residues from the 5'-termini of DNA was performed according to Ausubel *et al.* (1995), using calf intestine alkaline phosphatase (CIAP) (Gibco Life Technologies). DNA in 20μl of 1x CIAP buffer (20mM Tris-Cl (pH 8), 1mM $MgCl_2$, 1mM $ZnCl_2$) was incubated with CIAP (1U/pmol overhanging DNA ends or 5U/pmol blunt DNA ends) at 37°C for 30 minutes. CIAP was inactivated by incubating the reaction mixture at 65°C for 15 minutes.

2.8.7: Ligation of DNA

When not performed as part of a kit protocol, ligation reactions were carried out using T4 DNA ligase (Boehringer-Mannheim, Lewes). Up to 5μg DNA in 50μl of 1x ligation buffer (66mM Tris-Cl (pH 7.5 at 20°C), 5mM $MgCl_2$, 1mM dithioerythritol, 1mM ATP) were incubated with 1-5U T4 DNA ligase at 16°C for 18-24 hours. When used to transform *E. coli*, ligation mixtures were diluted 5-fold prior to adding to competent cells.

2.8.8: Perfectly Blunt cloning kit

The Perfectly Blunt Cloning Kit (Novagen) was used for the blunt-end ligation of environmental DNA into the dephosphorylated *Eco*R V-cloning site of pT7Blue. The procedure was carried out essentially as described in the manufacturer's instructions as outlined below.

Assuming an average insert size of 5kb, ~200ng environmental DNA was added to 5.0μl of End Conversion Mix plus distilled H_2O to give a final volume of 10μl. The proprietary End Conversion Mix contains all of the constituents that are required for blunt-end polishing of DNA fragments. The end conversion reaction was incubated at 22°C for 15 minutes. After heat inactivation at 75°C for 5 minutes, the reaction was briefly cooled on ice before proceeding.

50ng dephosphorylated *Eco*R V-digested pT7Blue and 4U T4 DNA ligase (each supplied in the kit) were added to the mixture and incubated at 22°C for 2 hours. The ligation reaction was subsequently used for the transformation of *E. coli* NovaBlue cells as described in Section 2.8.10.

2.8.9: pCR-XL-TOPO cloning kit

The pCR-XL-TOPO cloning kit (Invitrogen) was used for the cloning of 3'A-DNA into the 3'T-DNA cloning site of pCR-XL-TOPO. The procedure was carried out essentially as described in the manufacturer's instructions as outlined below.

3'A-DNA (Section 2.8.5) was precipitated, dried under vacuum (Section 2.8.1) and re-suspended in 10μl of 1x shrimp alkaline phosphatase (SAP) buffer (50mM Tris-Cl (pH 8.5 at 20°C), 5mM $MgCl_2$). 1U SAP (Boehringer-Mannheim) was added and the

mixture incubated at 37°C for 30 minutes. SAP was inactivated by incubating the reaction mixture at 65°C for 15 minutes.

pCR-XL-TOPO is supplied as a linearised plasmid with 3'thymine overhangs. Covalently bound to the vector is topoisomerase I, which catalyses the ligation of plasmid and input DNA (Shuman, 1994). The TOPO-cloning reaction was initiated by adding 10ng pCR-XL-TOPO to 4μl of 5'dephosphorylated, 3'A-DNA. The mixture was incubated at 22°C for exactly 5 minutes. To quench the cloning reaction the sample was quickly placed on ice before immediately being used to transform electrocompetent *E. coli* TOP10 cells (Section 2.8.11).

2.8.10: Chemical transformation of *E coli*

Preparation of chemically competent cells: *E. coli* strains were made chemically competent according to the frozen storage buffer (FSB)-based transformation protocol of Hanahan *et al.* (1995):

i. 5ml of a fresh over-night culture of *E. coli* strain was inoculated into 100ml LB broth contained in a 1-L baffled flask. The cells were incubated with shaking at 37°C until the OD_{550nm} reached ~0.6 units (~3 hours).

ii. The cells were harvested by centrifugation at 4°C for 15 minutes at 1000x g.

iii. After thoroughly decanting the broth, the cell pellet was re-suspended in 35ml ice-cold FSB (10mM potassium acetate, 10% (w/v) glycerol, 100mM KCl, 45mM $MnCl_2$, 10mM $CaCl_2$, 3mM hexamine cobalt trichloride) and incubated on ice for 15 minutes.

iv. The cells were harvested as above then re-suspended in 8ml ice-cold FSB. 280μl dimethylsulphoxide (DMSO) were added and the cell suspension gently mixed. After incubating on ice for 5 minutes, an additional 280μl DMSO were added and the cell suspension incubated on ice for a further 15 minutes.

v. 210μl cells were aliquoted into cryovials and flash frozen in an ethanol/dry ice bath before storing at -80°C.

Transformation of frozen chemically competent cells:

i. After thawing on ice, 200µl cells were transferred to a 15-ml conical polypropylene tube. 10-1000ng DNA in <10µl diluent was added to the cells and incubated on ice for 30 minutes.

ii. The cells were heat shocked at 42°C for 30 seconds then immediately chilled on ice.

iii. 800µl SOC medium was added immediately to the cells and incubated at 37°C for 30 minutes before plating (Section 2.8.12).

Alternatively, commercially available *E. coli* NovaBlue competent cells (Novagen) or *E. coli* XL2-Blue-MRF' ultracompetent cells (Strategene) were used in transformation reactions. The transformation of commercially available competent cells was essentially as described above except that 1-50ng DNA in 1 µl diluent were used to transform 20µl *E. coli* NovaBlue cells and 0.1-50ng DNA in 1µl diluent were used to transform 100µl *E. coli* XL2-Blue MRF' in the presence of 2µl mercaptoethanol.

2.8.11: Transformation of *E. coli* cells by electroporation

An Equibio Easyject Plus electroporator (Flowgen Instruments Ltd., Sellingbourne, Kent) was used for electroshock transformation of *E. coli* TOP10 electrocompetent cells (Invitrogen).

On ice, 1-2µl DNA (10-100ng) were added to 50µl electrocompetent cells and transferred to a chilled, 0.1cm *E. coli* pulser cuvette (Bio-Rad Laboratories, Hercules, CA, USA). The cells were electroporated at an electrical field strength of 18.0 kV/cm. Immediately following electroporation, 450µl SOC medium prewarmed to room temperature were added to the cuvette. The cells were transferred to a 15-ml conical tube and incubated with shaking at 37°C for 1 hour.

2.8.12: Plating of transformants and selection of recombinants

10-300µl transformed cells were plated onto LB plates containing the appropriate antibiotic. For those transformations relying on blue/white screening of recombinants, isopropyl-β-D-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) were added to the LB plates at final concentrations of 70µg/ml and 80mM, respectively. All plates were incubated overnight at 37°C.

2.8.13: Phagemid λTriplEx cloning

2.8.13.1: Preparation of blunt-ended environmental DNA for cloning into λTriplEx (Clontech)

Size-fractionated (Section 2.8.3), blunt-ended (Section 2.8.4) DNA was precipitated, dried under vacuum (Section 2.8.1) and dephosphorylated using SAP as described in Section 2.8.9.

10μg *Eco*R I adapter-DNA (Gibco Life Technologies), containing a phosphorylated blunt end and a non-phosphorylated *Eco*R I half-site (Figure 2.1), were ligated to the dephosphorylated blunt-ended environmental DNA as described in Section 2.8.7.

The *Eco*R I half sites of the adapted DNA were phosphorylated by incubating the ligation mixture with 30U T4 polynucleotide kinase (Promega, Madison, WI, USA) at 37°C for 30 minutes (Sambrook *et al.*, 1989).

Adapter dimers and unligated adapters were removed using cDNA size fractionation columns (Gibco Life Technologies) according to the manufacturer's protocol. Precipitated and dried DNA was resuspended in 10μl distilled H_2O.

1-3μl adapted DNA were used in ligation reactions with 0.5μg dephosphorylated *Eco*R I-digested λTriplEx DNA essentially as described in Section 2.8.7, except that reaction volumes were scaled down to 5μl.

```
5'-pGTCGACGCGGCCGCG
   CAGCTGCGCCGGCGCTTAA- OH-5'
```

Figure 2.1: DNA sequence of the *Eco*R I adapter which was ligated to dephosphorylated blunt-ended DNA for cloning into λTriplEx. In addition to the *Eco*R I half-site (-AATTC), the adapter also contains the recognition sequences for *Not* I (GCGGCCGC) and *Sal* I (GTCGAC). Insert DNA can therefore be recovered from the vector by digesting with either *Eco*R I or one of the rare-cutting enzymes, *Not* I or *Sal* I.

2.8.13.2: λ packaging reaction

DNA from each ligation reaction was separately packaged into λ phage particles using Gigapack III Gold packaging extract (Stratagene) according to the manufacturer's instructions as outlined below.

i. Packaging extracts were kept at less than 0°C until immediately before adding ligated DNA.
ii. 4μl ligated DNA (0.1-1.0μg) was added to the packaging extract and incubated at 22°C for 2 hours.
iii. 500μl SM buffer (50mM Tris-Cl (pH 7.5), 100mM NaCl, 8mM $MgSO_4$, 0.1%(w/v) gelatine) were added to the packaging reaction followed by 20μl chloroform.
iv. To remove debris, the reaction mixture was spun briefly in a bench top centrifuge.
v. 5 to 20-fold dilutions of packaged DNA in 10μl SM buffer were used for plating the unamplified library.

2.8.13.3: Titring the unamplified λ library

The λ packaging reaction was titred according to the manufacturer's instructions as outlined below.

i. *E. coli* XL1-Blue was inoculated into 15ml LB broth supplemented with $MgSO_4$ (10mM final concentration) and maltose (0.2% (w/v) final concentration).
ii. The cells were grown overnight, with shaking, at 30°C then centrifuged for 15 minutes at 500x g. and re-suspended in 7ml of 10mM $MgSO_4$.
iii. 1μl of appropriately diluted packaging reaction was added to 200μl cells. To allow the phage to attach to the cells, the mixture was incubated at 37°C for 15 minutes.
iv. 2ml molten LB top agar (48°C), supplemented with IPTG (70μg/ml final concentration) and X-gal (80mM final concentration), were added to the cells, mixed and quickly poured onto prewarmed LB agar plates.
v. After cooling at room temperature for at least 15 minutes, the plates were incubated at 37°C for 10 to 16 hours to allow formation of plaques.

vi. The titre was determined as plaque forming units (pfu)/ml. Packaging efficiencies were determined as pfu/μg vector DNA.

2.8.13.4: Converting the phage λTriplEx to the plasmid pTriplEx

The conversion of a λTriplEx clone to a pTriplEx clone involves *in vivo* excision and circularisation of a complete plasmid from recombinant phage.

i. *E.coli* BM25.8 was inoculated into 10ml LB broth supplemented with $MgSO_4$ (10mM final concentration) and incubated overnight at 37°C with shaking.

ii. 100μl of 1M $MgCl_2$ was added to the overnight culture.

iii. Using the sterile tip of a Pasteur pipette, an agar plug from a well-isolated plaque was retrieved and placed into 350μl TB buffer. To elute the phage, the sample was vigorously vortexed for several minutes and incubated at 37°C for 4 hours.

iv. For transfection, 200μl *E. coli* BM25.8 cells and 150μl eluted phage were combined and incubated, without shaking, at 31°C for 30 minutes.

v. 400μl LB broth was added and the sample incubated, with shaking, at 31°C for 1 hour.

vi. pTriplEx stocks were stored at -80°C until needed.

2.8.14: Preparation of plasmid DNA from *E. coli*

Plasmid DNA was prepared from *E. coli* using QIAprep plasmid miniprep or Qiagen plasmid midiprep purification kits (Qiagen) as described in the manufacturer's instructions.

Alternatively, for the direct use in electroporation (Section 2.8.11), plasmid DNA was recovered from individual colonies of *E. coli* using a modification of the QIAprep minprep protocol. Instead of inoculating into 5ml nutrient broth for overnight growth, the colony was added directly to buffer P1 and plasmid prepared according to the manufacturer's instructions.

2.9: Expression Screening

Two approaches were used to screen environmental DNA libraries. The first involved a microtitre plate format and the second employed indicator agar plates.

2.9.1: Screening libraries for enzyme activities using microtitre plates

The strategy used for screening libraries in microtitre plates is outlined in Figure 2.2.

Step 1: Amplified plasmid libraries were aliquoted into master microtitre plates (master plates) at a concentration of ~500 cfu/well. This was achieved by appropriately diluting each library into nutrient broth plus antibiotic and pipetting 200μl (~2.5 cfu/μl) into each master-plate well (88 wells per master plate). One row (eight wells) per master plate were reserved for controls.

Step 2: After incubating over night at 37°C, an appropriate volume from each master-plate well was aliquoted for assay by replica-plating into fresh microtitre plates (assay plates). Glycerol (20% (v/v) final concentration) was added to each master well and the master plates stored at -80°C until required for further investigation.

Amylase, phosphatase and lipase microtitre plate assays are described in Section 2.9.2.

Step 3: To isolate the target recombinant, 10-fold serial dilutions of positive wells (diluted master wells) were incubated overnight at 37°C.

Step 4: Cultures of serial dilutions were replica-plated for re-assay and stored at -80°C as described in step 2.

Step 5: For each re-assayed dilution series, the positive well with the highest dilution factor was selected for further investigation. Cells from the corresponding diluted master well were plated onto nutrient agar and grown overnight at 37°C.

Step 6: Individual colonies were transferred to a unique well of a fresh microtitre plate (sub-master plate) and grown overnight at 37°C.

Step 7: Sub-master plates containing individual colonies were aliquoted for re-assay and stored as described in step 2.

Step 8: Positive clones were isolated for further analyses including insert DNA sequencing (Section 2.10) and protein characterisation (Section 2.7.5).

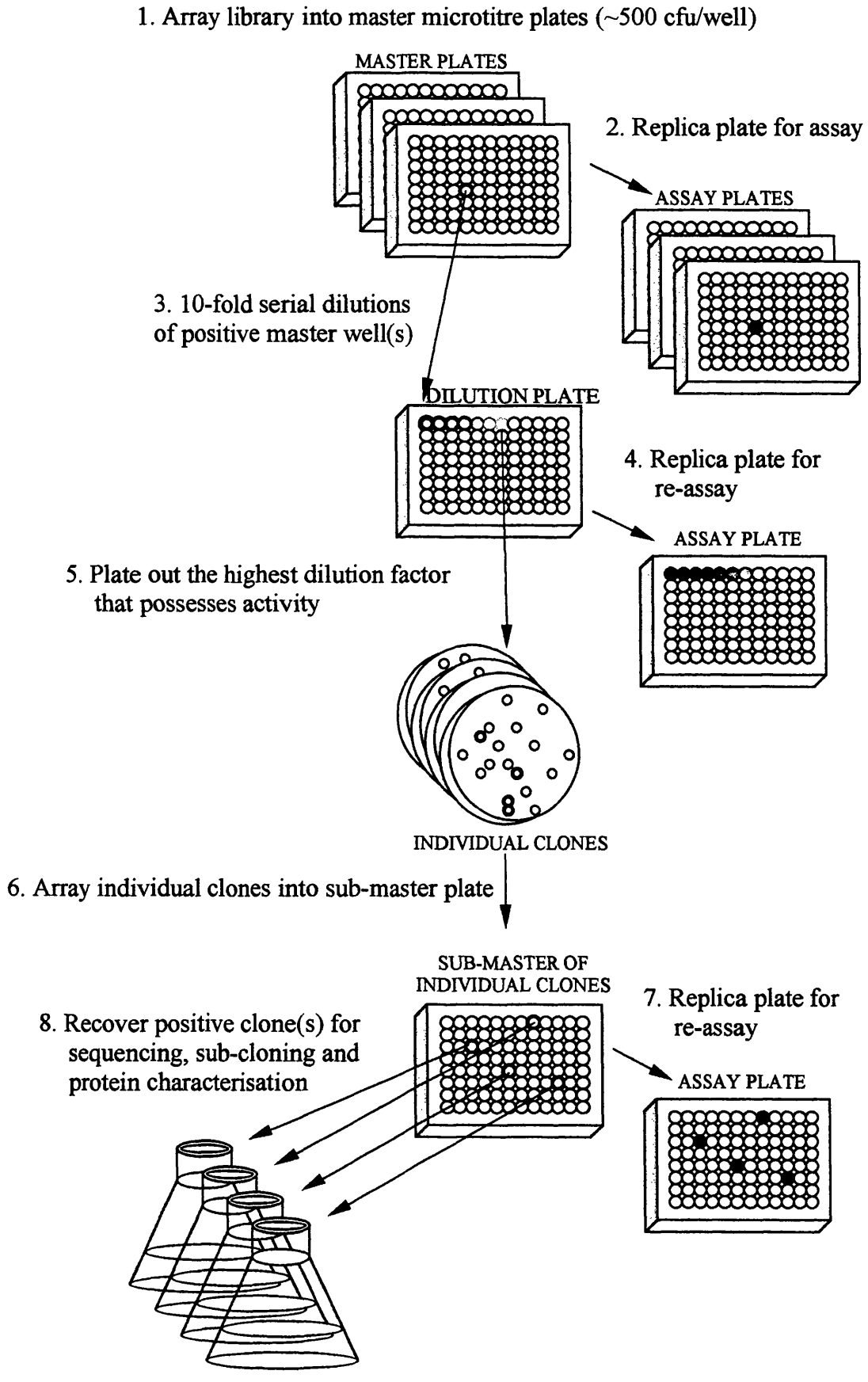


Figure 2.2: Flowchart depicting the strategy for screening environmental DNA libraries in microtitre plates.

2.9.1.1: Microtitre-plate screening for α-amylase activity

The microtitre plate amylase assay is a modified version of the assay of Blanchin-Roland & Masson (1989). 150-µl aliquots from each master plate well were replica-plated to assay wells containing 55µl of 1% (w/v) soluble starch in 30mM sodium phosphate buffer (pH 8.0). Lids were sealed to the assay plates with Parafilm and the plates incubated overnight at 50°C.

Starch utilisation was detected by transferring 12.5µl of the overnight incubation to fresh microtitre plate wells containing 250µl iodine solution (freshly prepared by adding 200µl 2.2% I_2 /4.4% KI (w/v) to 100ml 2% (w/v) KI solution). Microtitre plates were read at 570nm using a Dynatech MR7000 microtitre plate reader (Dynex Technologies, Ashford, Middlesex).

E. coli TOP10/pCR– and *E. coli* JM107/pQR126 were used as negative and positive controls, respectively. Control wells containing no added starch were used as blanks.

2.9.1.2 Microtitre-plate screening for lipase activity

Libraries were screened for lipase activity in microtitre plates using the synthetic chromogenic substrate, *p*-nitrophenyl palmitate (PNP-palmitate).

70µl aliquots from each master plate well were replica-plated to assay wells containing 15µl of 0.5M Tris-Cl (pH 8.0). Lids were sealed to the assay plates with Parafilm and the plates incubated at 80°C for 15 minutes. 15µl of 1.5mM PNP-palmitate in ethanol were added to the assay wells and incubated at 60°C. Plates were inspected periodically for up to 8 hours. Lipase activity was indicated by the development of a yellow colour due to the production of *p*-nitrophenolate.

E. coli TOP10/pCR– and lipase (Sigma) were used as negative and positive controls, respectively.

2.9.1.3: Microtitre-plate screening for phosphatase activity

Libraries were screened for phosphatase activity in microtitre plates using the synthetic chromogenic substrate, *p*-nitrophenyl phosphate (PNP-phosphate).

75-μl aliquots from each master plate well were replica-plated to assay wells containing 15μl of 0.5M Tris-Cl (pH 8.0). Lids were sealed to the assay plates with Parafilm and the plates incubated at 80°C for 15 minutes. To start the reaction, 10μl of 30mM PNP-phosphate in distilled H_2O was added to the assay wells and the plates incubated at 60°C. Plates were inspected periodically for up to 8 hours. Phosphatase activity was indicated by the development of a yellow colour which corresponded to the production of *p*-nitrophenolate.

E. coli TOP10/pCR– and phosphatase (Sigma) were used as negative and positive controls, respectively.

2.9.2: Screening libraries for enzyme activities using indicator agar plates

Environmental DNA libraries were screened for α-amylase, lipase, phosphatase and protease activities using indicator agar plates prepared as described below. Libraries were plated out at a titre of 2000 cfu per 140mm plate and incubated overnight at 37°C. Prior to enzyme detection, indicator plates were placed into sealed plastic bags, transferred to 50°C and incubated overnight.

Because incubating at 50°C kills *E. coli*, plasmids were prepared from these clones using a modified mini-preparation protocol for plasmid isolation (Section 2.8.14). Recovered plasmids were re-introduced into *E. coli* by electroporation (Section 2.8.11).

The controls used in the microtitre plate assays were also used in the indicator plate assays. Protease (Sigma) was used as a positive control for the skim milk indicator plates. The enzymes were used by spotting into the agar of the appropriate indicator plate.

2.9.2.1: Starch indicator plates for α-amylase activity (Gerhardt *et al.*, 1981)

Indicator plates were prepared by supplementing nutrient agar with soluble starch at a final concentration of 0.25% (w/v). After the 50°C-incubation step, the starch plates

were flooded with iodine solution which was prepared by adding 1ml 2.2% I_2 /4.4% KI solution to 500ml 2% (w/v) KI solution. The iodine solution was decanted as each plate became saturated with the purple-black starch/iodine complex. A halo around the colony indicated amylase activity.

2.9.2.2: Tween-80 indicator plates for lipase activity (Gerhardt *et al.*, 1981)

Indicator plates were prepared by supplementing nutrient agar with $CaCl_2$ (0.01% (w/v) final concentration). After autoclaving and cooling to 50°C, Tween-80 was added to the medium at a final concentration of 1% (v/v). After the 50°C incubation step, plates were inspected for lipase activity as indicated by the occurrence of opaque haloes around lipase-positive colonies.

2.9.2.3: TPMG (tryptose phosphate-methyl green) indicator plates for phosphatase activity (Riccio *et al.*, 1997)

Indicator plates were prepared by supplementing tryptose phosphate agar with phenolphthalein diphosphate (1mg/ml final concentration), and methyl green (50μg/ml final concentration). Phosphatase activity is indicated by the formation of green halos or the green staining of phosphatase-positive colonies.

2.9.2.4: Skim milk indicator plates for protease activity

Indicator plates were prepared by supplementing nutrient agar with 1% (w/v) powdered skim milk. Protease activity is indicated by the occurrence of clear zones around protease-positive colonies.

2.10: DNA Sequencing and Analysis

DNA sequences were determined by Oswel DNA services (Northampton). Nucleotide and deduced amino acid sequences obtained from environmental genomic libraries were compared with each other or with entries in the nonredundant nucleic acid and protein databases of the National Center for Biotechnology Information (NCBI) server using BLASTN or BLASTX (Altschul *et al.*, 1990; Gish & States, 1993).

Open reading frames were located within DNA sequences using the ORF finder at NCBI. Multiple protein sequence alignments were achieved using CLUSTAL W server at Baylor College of Medicine (Thompson *et al.*, 1994). Protein motifs within deduced

amino acid sequences were identified using FingerPRINTscan software located on the server at the University of Manchester Bioinformatics Unit (Scordis *et al.*, 1999). Putative bacterial promoter sequences were screened by either visually scanning regions upstream of ORFs for consensus –35 (TTGACA) and –10 (TATAAT) hexamers (Gross *et al.*, 1992) or by using online software for bacterial promoter prediction at www.fruitfly.org/seq_tools/promoter.html (Reese *et al.*, 1996). Upstream sequences were also visually scanned for archaeal consensus promoter elements (Soppa, 1999). tRNA gene sequences were scanned using the tRNAscan-SE search sever at Washington University, St. Louis (Lowe & Eddy, 1997).

2.11: Thermostable Enzyme Activities in *E. coli* Cell Extracts

Enzymic activities of selected clones were assayed using cell extracts prepared from recombinant *E. coli*. Cultures were grown to stationary phase and diluted in nutrient broth to give an OD_{600nm} reading of ~1.5. Cells (50ml) were harvested (10 min at 4°C and 8000x g) and resuspended in 5ml nutrient broth. The cells (1ml) were disrupted on ice using a MSE Soniprep 150 ultrasonic disintegrator (SANYO Gallenkamp Plc., Leicester). Sonication was performed at an amplitude of 8-μm for 2 min at 10-sec intervals. The extract was cleared by centrifugation at top speed in a microfuge for 10 min at 4°C. Aliquots of cell extracts were incubated at different temperatures for 15 minutes before being assayed for enzymic activities using microtitre plates or 1.5-ml microfuge tubes essentially as described in Section 2.9.2.

Chapter 3

Direct Extraction of Environmental DNA from Geothermal Sediments

3.1: Aim

This chapter describes the investigation of two different procedures for isolating environmental DNA from geothermal sediment. A mortar and pestle grinding method and a bead beating method for DNA extraction were compared with respect to DNA yield, shearing and purity. A survey of DNA recovered from New Zealand and Iceland sediments was also performed using these extraction protocols.

3.2: Background

The methods for isolating DNA from soils and sediments are based either on the recovery of bacterial cells from the environmental sample prior to cell lysis and DNA isolation (pioneered by Holben *et al.* 1988) or on the direct lysis of microbial cells present within the environmental matrix followed by DNA extraction (pioneered by Ogram *et al.*, 1987).

Although specific for bacterial DNA, the cell extraction method for soils and sediments is not commonly used for molecular studies of bacterial communities because the procedure is time consuming and DNA yield, purity and fragment size are not necessarily improved over those of the direct lysis approach. (Steffan *et al.*, 1988; Krsek & Wellington, 1999). Based on this information, the direct lysis approach to DNA extraction was selected for isolating DNA from Iceland and New Zealand geothermal sediments.

In this work, two different procedures with contrasting scales of operation, as well as, different physical, enzymic and chemical elements for extracting and purifying DNA directly from sediment were evaluated in terms of the quantity and quality of DNA recovered. A mortar and pestle method and a bead beating method for DNA extraction were chosen for comparison studies because they are representative alternative approaches to obtaining DNA from soils and sediments.

3.3: Comparison of Methods for Direct Extraction of DNA from Geothermal Sediment

Described in detail in Section 2.6, the mortar and pestle plus SDS (MPS) protocol (modified from Saano & Lindstrom, 1995) and a proprietary bead beating (BB) method (Bio101) for the recovery of DNA from sediments are briefly outlined in Tables 3.1 and 3.2, respectively. Sediment (Ice22) collected from a runoff stream (~58°C, pH 4.3) of the main geyser at Krysuvik-Seltun, Hverir, Iceland was selected as the test material because it contained macroscopic filamentous biomass material and was thus likely to yield sufficient quantities of DNA for comparative analysis as well as for library preparation.

3.3.1: DNA yield

The yield, purity and fragment size of Ice22-DNA recovered *via* MPS and BB methods are listed in Table 3.3. For calculating yields, purified Ice22-DNA concentrations were determined spectrofluorometrically using PicoGreen, a nucleic acid stain specific for dsDNA (Section 2.7.3). Although each of the replicates was treated in the same manner, the yield of DNA recovered *via* the MPS method was not reproducible. The reason for this variability may be explained by sample heterogeneity due to the filamentous biomass which was matted throughout the sediment. An additional homogenisation step using, say, a Waring blender prior to the grinding step may improve the reproducibility of DNA yields using the MPS extraction procedure. The difficulty in obtaining reproducible DNA yields using the MPS method may also be due to extensive sample handling. The MPS method involved several manual processing steps (e.g. grinding, sample transfers, extractions), any of which could have introduced error into the process. The BB method on the other hand required very little processing. The physical disruption of the cells *via* the BB method was mechanised and the extraction and purification of DNA required fewer steps than that of the MPS method. Taking into consideration the difficulty in obtaining reproducible results, the MPS method yielded approximately nine-fold greater amount of DNA per gram dry sediment than that of the BB method. Although not tested here, it has been shown that yields of DNA recovered through bead beating can be increased almost two-fold, without appreciable shearing of DNA, by increasing homogenisation speed from ~1700 rpm to 2510 rpm (Miller *et al.*, 1999). In this work, homogenisation was carried out at 1500 rpm in the same type of

bead beater as that used in Miller's lab (Mini Bead Beater-8, BioSpec Products). This suggests that DNA yields *via* the BB method could be improved by increasing the homogenisation speed to 2510 rpm.

Step		Purpose
1.	10g sediment + ~0.5g sterile quartz sand 5ml NE buffer: Grind using M&P until homogenised	Disperse cells
2.	Add lysozyme and incubate with periodic shaking at 37°C for 30 minutes.	Digest cell walls
3.	Add SDS and proteinase K and incubate with shaking at 65°C for 60 minutes.	Lyse cells
4.	Freeze-thaw (3X)	Lyse cells
5.	Add CTAB at high salt concentration. Vortex and incubate at 65°C for 20 minutes.	Bind proteins, carbohydrates, humic substances and other contaminants
6.	Chloroform-extraction	Remove CTAB, its complexes and other contaminants
7	Back-extraction using NE buffer	
8	Isopropanol precipitation	Purify and concentrate DNA
9	Anion-exchange chromatography (500G column)	purify DNA

Table 3.1: Mortar and pestle plus SDS (MPS) procedure for extracting and purifying DNA from soil and sediment (Adapted from Saano & Lindstrom, 1995).

Step		Purpose
1.	0.5g sediment + ceramic and silica particles + homogenisation and protein solubilisation buffers (FastDNA fit for soil): Bead-beat (Mini-Beadbeater-8)	Disperse and lyse cells
2.	Centrifuge	Remove sediment and other particles
3.	Add protein-precipitating solution and centrifuge	Remove proteins
4.	Recover DNA using silica matrix and guanidine thiocyanate/ethanol wash (FastDNA spin kit for soil)	Remove humic substances and other contaminants
5.	Isopropanol precipitation	Purify and concentrate DNA

Table 3.2: Bead beating plus spin column purification (BB method) for isolating DNA from soil and sediment.

Method	Yield (μg DNA/gws)	Yield (μg DNA/gds)	A260/A230	A260/A280	DNA fragment size (kb)
MPS (n=5)	3.0±2.3	13.2±9.6	2.2±0.43	1.8±0.15	<0.5 to >23
BB (n=3)	0.35±0.03	1.5±0.12	0.06±0.004	1.9±0.19	~0.5 to ~10

Table 3.3: Comparison of mortar and pestle plus SDS (MPS) and bead beating (BB) methods for isolating DNA from Ice22 sediment. Yields and absorbance ratios are the mean ± standard deviation; gws, g wet sediment; gds, g dry sediment.

3.3.2: DNA purity

The purity of extracted DNA was determined spectrophotometrically by calculating the A_{260nm} to A_{230nm} and A_{260nm} to A_{280nm} ratios for humic acid and protein contamination, respectively (Section 2.7.2). DNA solutions were considered pure if the A_{260nm} to A_{230nm} ratio was between 1.8 and 2.3 and the A_{260nm} to A_{280nm} ratio was between 1.5 and 2.0. (Marmur, 1963).

The A_{260nm} to A_{230nm} and A_{260nm} to A_{280nm} purity ratios of Ice22-DNA recovered *via* the MPS method were 2.2 ± 0.43 and 1.87 ± 0.15, respectively, indicating that the DNA was acceptably pure. The A_{260nm} to A_{280nm} ratio of DNA recovered by the BB procedure was 1.9 ± 0.19 indicating that the sample was essentially free from contaminating proteins; however, the A_{260nm} to A_{230nm} ratio was only 0.06 ± 0.004, indicating that the sample may have been contaminated with humic substances. The purified DNA solution obtained *via* the BB method was not discoloured in any way suggesting that humic substances were not the absorbing species. To determine whether the silica matrix used in the minicolumns contained UV-absorbing substances, distilled H_2O was processed in the same manner as DNA solutions for purifying DNA extracted using the FastDNA spin kit for soil (Bio101) (Steps 4 and 5 in Table 3.2). The eluted distilled H_2O was found to absorb at 230nm but not at 260nm or 280nm indicating that the silica used in the purification columns interfered with spectrophotometric measurement of DNA.

An alternative approach for assessing the purity of DNA was to investigate its accessibility to molecular techniques such as those used for library construction. TA-cloning efficiencies of Ice22-DNA extracted by the bead beating method were similar to those of Ice22-DNA obtained *via* mortar and pestle (Section 4.5.2). Indeed, DNA recovered from Iceland sediments *via* the BB method was used in the construction of environmental DNA libraries (Section 4.7).

3.3.3: DNA fragment size

The size distribution of extracted DNA fragments was assessed by agarose gel electrophoresis (Section 2.7.4). The MPS method recovered Ice22-DNA ranging in size from <0.5 kb to >23 kb (Figure 3.1), while the BB method sheared the DNA to <10 kb (Figure 3.3).

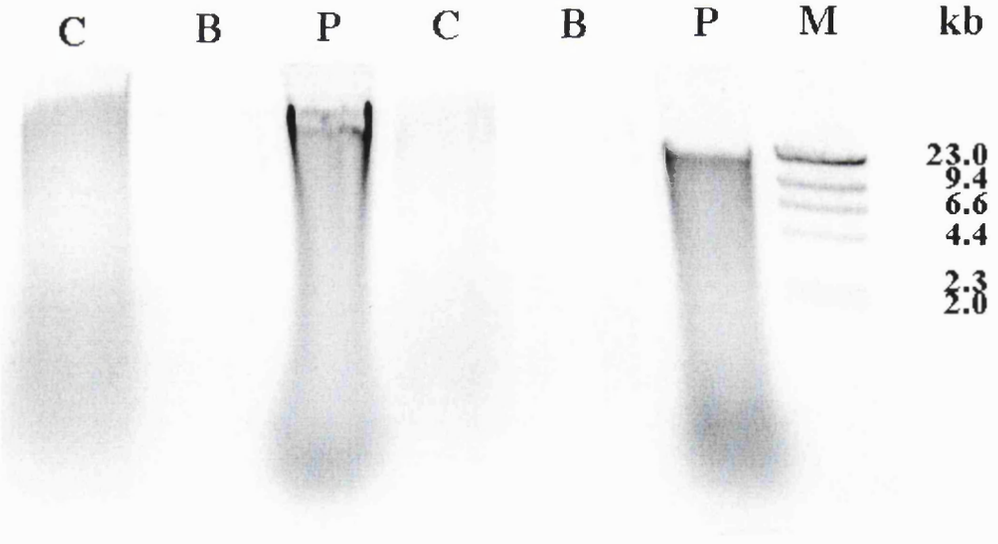


Figure 3.1: 1% agarose gel of environmental DNA extracted directly from Ice22 geothermal sediment *via* the MPS method. 20μl of DNA from replicate extractions using 10g wet sediment was electrophoresed in 1% agarose gel. M is λ-*Hin*d III DNA marker; C, crude DNA obtained after chloroform extraction and isopropanol precipitation; B, back-extracted DNA obtained from solvent-containing sediment using NE buffer and isopropanol precipitation; P, purified DNA obtained *via* anion exchange chromatography. Yields cannot be directly compared because crude, back-extracted and purified DNA each have different ethidium bromide staining efficiency.

3.4: Survey of DNA Extracted from New Zealand and Iceland Geothermal Sediments

3.4.1: New Zealand geothermal sediments

For the construction of environmental DNA libraries, New Zealand (Tok) geothermal sediments were collected from the Tokaanu thermal region, Central North Island, New Zealand (D. Cowan, personal communication). The descriptions of Tok sediments are given in Table 3.4. During the course of cloning procedures (Chapter 4), the DNA yields (μg DNA/g dry sediment) of Tok sediments extracted *via* the MPS method were quite low when compared to that of Ice22 sediment extracted in the same manner. With the exception of DNA extracted from TokC, all Tok DNAs were fragmented to <10kb (Table 3.5). While high molecular weight DNA is usually required for general cloning procedures, DNA that was fragmented to between 2 and 10kb was not ruled out as starting material for blunt-end cloning protocols. This is because prior cleavage of blunted DNA fragments with restriction endonuclease(s) is not required for ligation to the cloning vector.

Tok Sample	Description	Temp. °C	pH
A	Mineral-deposited sediment	84	6.6
B	Fine sandy sediment from bubbling vent	91	6.1
C	Mineral-deposited sediment	73	6.6
D	Mud from bubbling pool	76	ND
E	Mineral-deposited sediment	56	5.5

Table 3.4: New Zealand geothermal sediments (Tok). ND, not done.

Yield \ Sample	Tok A	Tok B	Tok C	Tok D	Tok E
μg DNA/gws	0.20	0.14	0.057	0.015	0.028
μg DNA/gds	0.33	0.028	1.1	0.024	0.056
DNA size (kb)	6-10	0.5-10	<0.5-20	0.5-4	0.3-3

Table 3.5: Yields and fragment size distributions of DNA extracted from New Zealand (Tok) geothermal sediments *via* the MPS method. gws, g wet sediment; gds, g dry sediment.

3.4.2: Iceland geothermal sediments

Iceland (Ice) sediments were collected from several thermal sites located in the area of south-west Iceland. The descriptions of Ice sediments are given in Table 3.6. A major disadvantage of the MPS method is that it was not reproducible and it involved time-consuming manual grinding, incubation, extraction and purification steps. The processing time for the extraction of four samples using the MPS method was 2 to 3 days. The BB method was faster and accommodated up to eight 0.5-g samples per procedure. For this reason, the BB method was selected for extracting the Iceland sediments. With the BB method, twenty-two sediments were easily surveyed by extracting simultaneously eight samples and visualising the recovered DNA on agarose gels (Figures 3.2 and 3.3). The time required to extract all of the Iceland sediments was reduced from a predicted two weeks, using the MPS method, down to one day with the BB method.

Yields of DNA extracted from Ice sediments *via* the BB method are given Tables 3.7 and 3.8. With the exception of Ice5 and Ice22, all sediments that contained macroscopic amounts of biomass possessed higher amounts of DNA than either muddy, sandy or mineral-deposited sediments. Because it contained macroscopic amounts of biomass, it was surprising that Ice22 sediment produced relatively low DNA yields with the BB treatment, especially since Ice22 was selected for initial studies based on its macroscopic appearance.

Yields of less than ~0.03μg DNA per g wet sediment could not be visualised on agarose gel when 1/10 of the extracted volume was electrophoresed. Relative yields and fragment sizes of recovered DNA were easily estimated with this survey making this approach a useful tool for selecting sediments for library construction. Once a sediment is selected using this survey approach, DNA isolation could be scaled up simply by increasing the number of 0.5-g extractions or by extracting 10g of the chosen sediment using the MPS method.

Ice Sample	Description	Location	Temp. °C	pH
1	Light-grey mud	Bubbling mud pool located at Skalafell	70	2.1
2	Steam-heated soil	2 metres from site 1	62	ND
3	Green algal biomass	Downstream of Blue Lagoon	21	5
4	Grey silicious sediment	Bubbling vent Krysuvik-Seltun, Hverir	65	5.1
5	Grey silicious biomass	Heated water downstream of main geyser, at Krysuvik-Seltun, Hverir	~85	5.5
6	Stream sediment	Freshly dug ditch downstream of bore hole, Laugaluatin, Brun	78	8.8
7	Mineral-deposited sediment	Edge of Blesi, a deep pool located at Geysir	76	9.4
8	Fine sandy sediment	Edge of deep pool located adjacent to Blesi	76	9.8
9	Brown sandy mud	1.5 metre diameter pool located in mound north of site 8	79	8.0
10	Grey mud	50cm bubbling mud pool 100 metres uphill to site 9	~92	10.2
11	Red mud	50cm mud pool 50 metres down stream of site 9	67	6.0
12	grey mud	50cm still pool near site 11	83	3.1
13	Mineral-deposited sediment	Piped runoff stream emptying into the Hveregirdi River	70-85	7.9
14	Grey silicious biomass	Piped runoff stream emptying into the Hveregirdi River	50-60	8.7
15	Grey silicious biomass	Different section of same runoff stream at site 14	60-75	9.6
16	Grey silicious biomass	Different section of same runoff stream at site 14	69-79	9.5
17	Sandy mud	Bubbling vent on ridge ~500 metres north-east of site 14	62	9.6
18	Grey silicious biomass	Piped runoff 5 meters downstream of site 17	50-60	9.5
19	Green-orange biomass	Piped runoff 5 meters downstream of site 18	52	9.1
20	Reddish-brown mud	Shallow still pool north of Hveregirdi River	69	9.5
21	Black mineral-deposited sand	Capped steam vent downstream of site 20	70	9.7
22	Grey silicious filamentous biomass	Runoff from main geyser at Krysuvik-Seltun, Hverir	57-59	4.3
23	Light-grey mud	Boiling mud pool near site 22	97	4.0
24	Light-grey mud	Boiling mud pool adjacent to site 23	98	3.7

Table 3.6: Iceland geothermal sediments (Ice). ND, not done.

Yield \ Sample	Ice1	Ice3	Ice4	Ice5	Ice6	Ice7
μg DNA/gws	0.014	6.2	3.2	3.4	0.45	0.028
μg DNA/gds	0.027	25.5	5.8	6.0	1.3	0.050

Yield \ Sample	Ice8	Ice8d	Ice9	Ice10	Ice11	Ice12
μg DNA/gws	0.026	ND	0.018	0.50	5.81	0.033
μg DNA/gds	0.048	ND	0.039	1.0	19.9	0.18

Table 3.7: DNA yields of Iceland geothermal sediments (Ice1 and Ice3 through Ice12) extracted *via* the BB method. gws, g wet sediment; gds, g dry sediment.

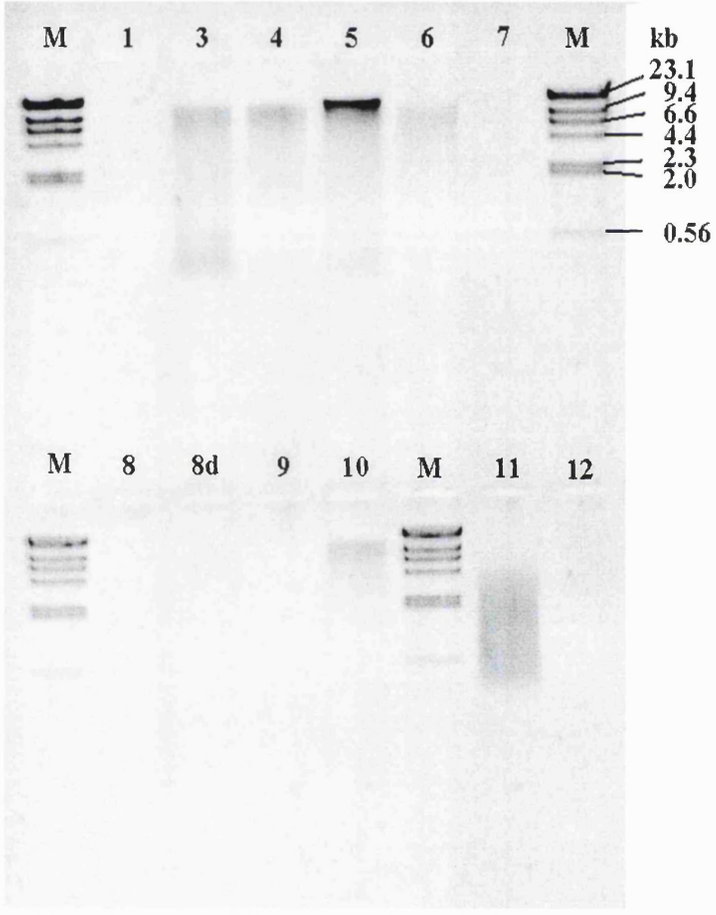


Figure 3.2: Fragment size distributions and relative yields of DNA extracted from Iceland geothermal sediments (Ice1 and Ice3 through Ice12) *via* the BB method. Except for Ice8d (8d), all sediments were wet prior to bead beating. Ice8d is the same as Ice8 (8) except that the sample was dried at 60°C overnight prior to bead beating. 1/10 volumes of extracted DNA were loaded with lane number corresponding to sample name. M is a λ*Hin*d III DNA weight marker.

Yield \ Sample	Ice13	Ice15	Ice16	Ice17	Ice18	Ice19
μg DNA/gws	1.42	49.4	9.73 (0.50)	3.54	27.5	26.1
μg DNA/gds	2.19	224	17.7 (0.96)	10.0	193	186

Yield \ Sample	Ice20	Ice21	Ice22	Ice23	Ice24
μg DNA/gws	2.4	0.016	0.350 (0.030)	0.153	0.031
μg DNA/gds	22.5	0.020	1.53 (0.117)	0.233	0.49

Table 3.8: DNA yields of Iceland geothermal sediments (Ice13 and Ice15 through Ice24) extracted *via* the BB method. gws, g wet sediment; gds, g dry sediment. Values in parentheses are standard deviations of triplicate extractions using Ice22 sediment.

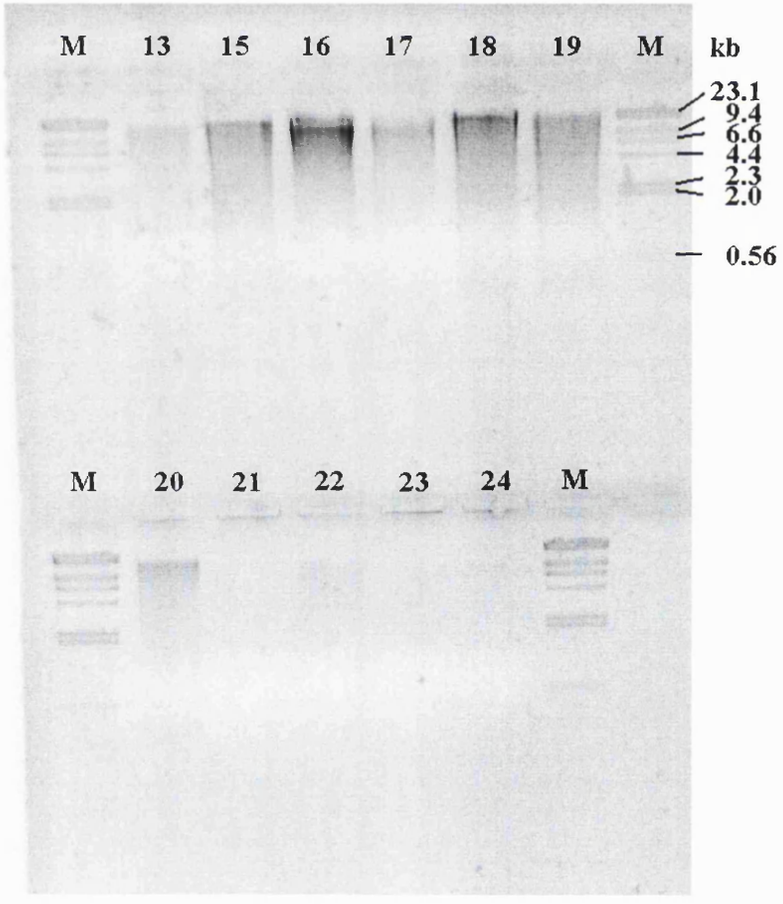


Figure 3.3: Fragment size distributions and relative yields of DNA extracted from Ice geothermal sediments *via* the bead-beating method. 1/10 volumes of extracted Iceland DNA (samples 13 and 15 through 24) were loaded with lane number corresponding to sample name. M is a λ*Hin*d III DNA weight marker. gws, g wet sediment; gds, g dry sediment.

3.5: Summary

Two different protocols for isolating DNA directly from geothermal sediment were compared with respect to DNA yield and quality. Ice22 geothermal sediment was selected for initial studies because it contained macroscopic amounts of biomass and was predicted to produce sufficient quantities of DNA for analysis. The mortar and pestle plus SDS method recovered ~9-fold more DNA from Ice22 sediment than bead beating. The mortar and pestle plus SDS approach, however, was more laborious and not as reproducible as the bead beating method.

A_{260nm} to A_{230nm} and A_{260nm} to A_{280nm} ratios indicated that DNA obtained from Ice22 sediment *via* mortar and pestle was essentially free of contaminating substances. Ice22-DNA recovered *via* bead beating was relatively pure with respect to the A_{260nm} to A_{280nm} ratio; however, A_{230nm} -absorbing substances present within the silica-based purification matrix prevented accurate measurement of the A_{260nm} to A_{230nm} ratio. The mortar and pestle plus SDS method was less shearing to Ice22-DNA than the bead beating method: however, the fragment size of Ice22-DNA recovered by both methods was suitable for cloning procedures. Ice22-DNA recovered by both extraction methods had similar cloning efficiencies as described in Chapter 4.

The mortar and pestle method was used to extract DNA from five New Zealand sediments for use in subsequent cloning protocols. The bead beating method provided a quick and simple means of surveying a collection of twenty-two Iceland sediments. With this approach it was discovered that Ice22 sediment produced relatively low DNA yields compared with other sediments that possessed macroscopic amounts of biomass. By providing information in relative terms of DNA yield and size, this approach could be used for selecting quickly those sediments that are suitable for use in library preparation.

Chapter 4
Evaluation of Cloning Protocols and Preparation of Environmental DNA Libraries

4.1: Aims

The main aim of the work described in this chapter was to evaluate various cloning protocols for generating stable environmental libraries using DNA extracted from Iceland (Ice) and New Zealand (Tok) geothermal sediments (Chapter 3). These environmental DNA libraries would subsequently be screened for enzymic activities as described in Chapter 5.

Preliminary environmental libraries were constructed and evaluated based on cloning efficiency (cfu/μg vector), recombination efficiency (% of transformants containing recombinant vector) and the number of recombinant transformants produced per ligation reaction. Library construction was facilitated by using standard molecular biology equipment and kits. In order to best represent the microbial community, the protocol chosen for constructing environmental DNA libraries needed to have a high efficiency of cloning. In order to minimise the number of non-recombinant transformants, the chosen protocol also had to generate greater than 70% recombinants. Because screening of libraries relied upon heterologous expression, DNA inserts needed to be at least 1kb in length in order to accommodate complete ORFs and upstream control sequences such as promoter(s) and ribosome-binding site(s). It was also advantageous for the vector to possess promoter and ribosome-binding sequences immediately upstream of the cloning site in case the native genomic control sequences could not be recognised by the expression host. In such cases however, proper alignment of the cloned gene would be required for expression.

4.2: Background

E. coli was chosen as the host for the environmental libraries because it is one of the best genetically characterised bacteria and many of the established cloning protocols are based on this organism. *E. coli* pUC-based cloning vectors were selected for constructing environmental libraries because they offer all the features desirable in a

general cloning vector. pUC-based vectors are maintained stably at high copy number and relatively high expression can be achieved through gene dosage. Most pUC-based vectors contain a multiple cloning site located immediately downstream of the *lac* promoter and ribosome-binding site for sequences encoding the β-galactosidase α peptide (*lacZ*α) (Vieira & Messing, 1982; Yanisch-Perron *et al.*, 1985). This feature is especially important when environmental promoter sequences are not included in the insert or are not recognised by the host. This is because readthrough transcription driven by the *lac* promoter may provide an alternative route to heterologous expression of cloned DNA. Expression of environmentally derived DNA as an α peptide-gene fusion may also be also possible if the gene is cloned in the same reading frame as the α peptide.

The following plasmid vectors were used in this work: pUC19 (Vieira & Messing, 1982; Yanisch-Perron *et al.*, 1985) was used for cloning restriction-enzyme-digested DNA; pT7Blue (Novagen) was used for cloning blunt-ended DNA and pCR-XL-TOPO (Invitrogen) was used for cloning 3'-adenylated DNA (3'A-DNA). Also evaluated was the multifunctional phagemid vector, λTriplEx (Clontech). Physical maps and salient features of these vectors are shown in figures 4.1, 4.2, 4.3 and 4.4, respectively. Expression vectors designed for heterologous (over)expression were not considered a feasible option for constructing environmental libraries because they are specifically designed to express cDNA or subcloned ORFs. Such vectors are not suitable for shotgun-cloning of genomic DNA because the former relies on mRNA as starting material and the later assumes that the DNA sequence of the ORF is known. Constructing new vectors for expression screening was not within the scope of this project.

4.3: An Initial Cloning Experiment Using Soil DNA

Preliminary work for evaluating cloning protocols was conducted using DNA that was extracted directly from readily available local soil (loamy sand) (Table 4.1). Only after establishing this ground work was the DNA extracted from geothermal sediment used to construct and evaluate environmental libraries.

The mortar and pestle plus SDS (MPS) method (Section 2.6.1) was used to extract DNA directly from common garden soil (GS-DNA) collected on campus at University College London. GS-DNA was prepared for cloning as described in Section 2.8 and

outlined in Figure 4.5. Briefly, GS-DNA was partially digested with *Sau*3AI (Figure 4.6) and DNA fragments ranging from 1 to 10 kb were recovered by preparative gel electrophoresis. Dephosphorylated, *Bam*HI-digested pUC19 (pUC19-*Bam*HI) isolated *via* preparative gel electrophoresis was ligated to *Sau*3AI-digested GS-DNA fragments and transformed into *E. coli* XL2-Blue MRF' ultracompetent cells. Recombinants were visualised by blue/white screening.

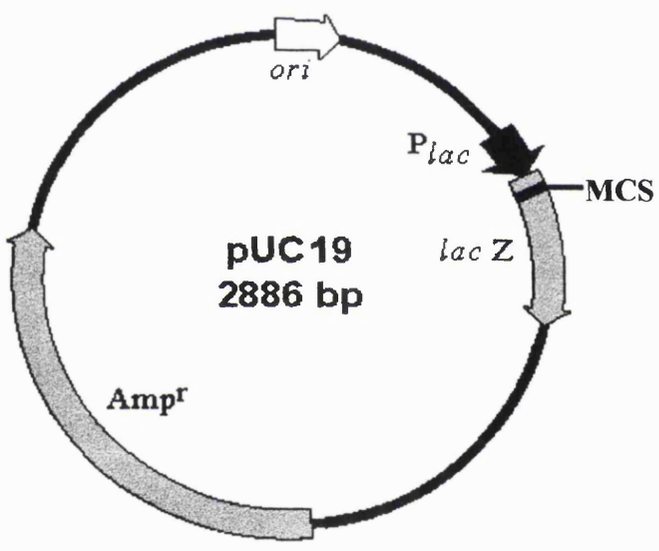


Figure 4.1: pUC19 (Vieira & Messing, 1982; Yanisch-Perron *et al.*, 1985) vector map demonstrating the salient details of vector design. *ori*, origin of replication; Amp^r, ampicillin resistance ORF; P *lac*, inducible *lac* promoter; *lacZ*, α region of the β-galactosidase gene. In this study, environmental DNA was cloned into the unique *Bam*HI-site within the multiple cloning site (MCS). Insertional inactivation of *lacZ* allowed for blue/white selection of recombinants.

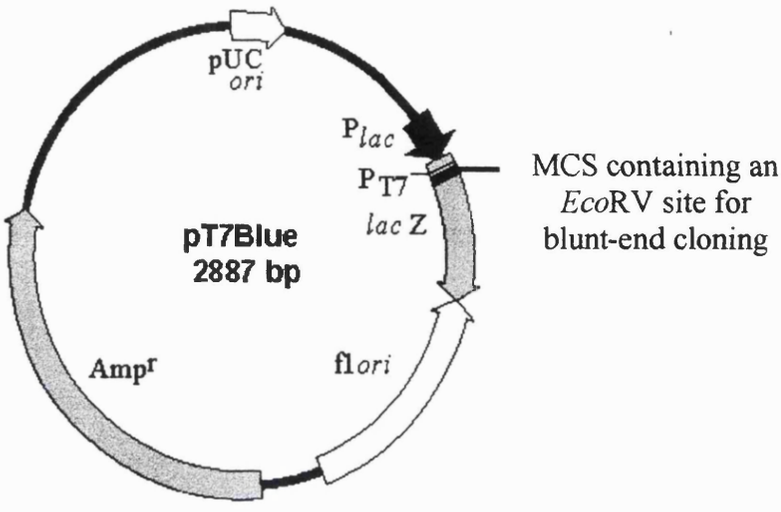


Figure 4.2: pT7Blue vector supplied with the Perfectly Blunt cloning kit (Novagen). pUC *ori*, pUC-derived origin of replication; f1 *ori*, f1 phage origin of replication; PT7, promoter for T7 RNA polymerase. The vector was supplied linearised and nonphosphorylated at the unique *Eco*RV. Insertional inactivation of *lacZ* allowed for blue/white selection of recombinants.

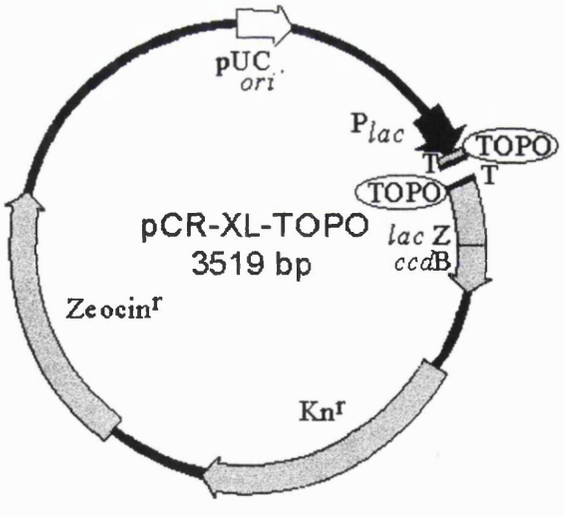


Figure 4.3: pCR-XL-TOPO vector supplied with the TOPO-XL-PCR cloning kit (Invitrogen). *ccd*B, lethal gene ORF fused to *lacZ*; Zeocin^r, Zeocin resistance ORF; Kn^r, kanamycin resistance ORF. The vector was supplied linearised at the T- cloning site located within the multiple cloning site. Ligation of insert DNA is achieved by exploiting the activity of topoisomerase which is covalently bound to vector DNA termini (TOPO). The multiple cloning site is located in the 5' end of the *lacZ-ccd*B gene fusion. Ligation of a DNA fragment into the T-cloning site disrupts expression of the lethal fusion permitting growth of only recombinants upon transformation. Figure 4.3 was modified from TOPO XL PCR Cloning Kit user manual, version A.

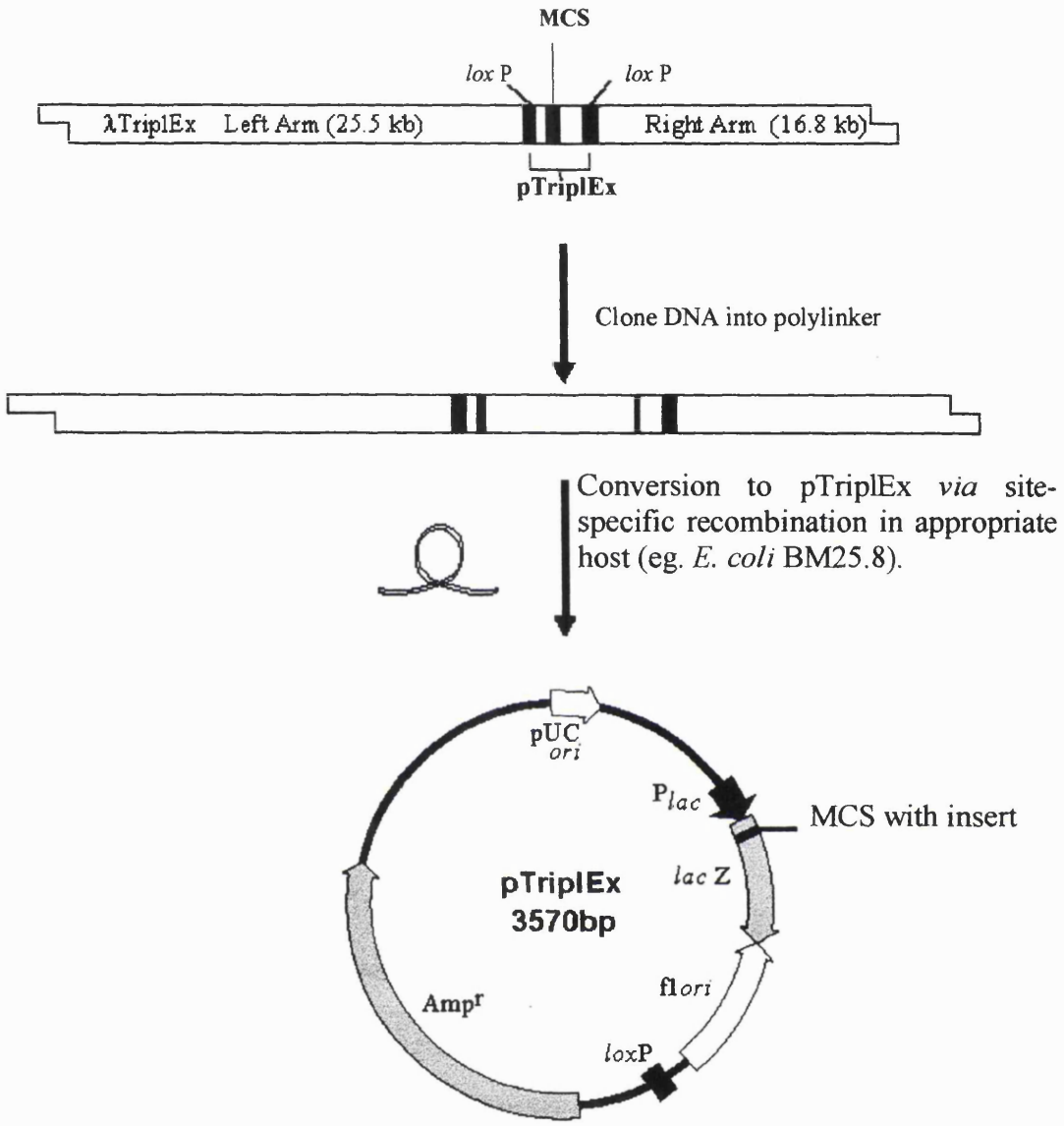


Figure 4.4: Phagemid λTriplEx (Clontech). The λTriplEx DNA arms were supplied predigested and dephosphorylated at the unique *Eco*R I site within the multiple cloning site. Insertional inactivation of *lacZ* allowed for blue/white selection of recombinants. The MCS is located within a pUC-based plasmid (pTriplEx) which is flanked by *lox*P, sequences recognised by *Cre* recombinase for excision and circularisation of pTriplEx (Elledge *et al.*, 1991). With this system, high titre libraries can be constructed using λTriplEx DNA. Subsequently, recombinant phage can be easily converted to recombinant pTriplEx for further analyses. Figure 4.4 was modified from λTriplEx library user manual version PT303-1.

Donor DNA	Description of collection site	Yield of purified DNA (μg DNA/gws)	DNA recovered for cloning (μg)
Garden Soil (GS)-DNA	Loamy sand	6.7 (MPS)	67+
Tokaanu geothermal sediment (Tok)-DNA		MPS	
TokA-DNA	Runoff stream 76°C, pH 6.6	0.20	8.1
TokB-DNA	Runoff stream 91°C, pH 6.1	0.14	0.43
TokC-DNA	Runoff stream 73°C, pH6.6	0.057	1.5
TokD-DNA	Mud pool 76°C	0.015	0.15
TokE-DNA	Runoff stream 56°C, pH 5.5	0.028	1.0
Iceland geothermal sediment (Ice)-DNA			
Ice22-DNA	Runoff stream 57-59°C, pH 4.3	3.0 (MPS) 0.35 (BB)	270+ 2.8+
Ice16-DNA	Runoff stream 69-79°C, pH 9.5	9.9 (BB)	110+

Table 4.1: Description of soil and sediments and DNA yields of samples used for the construction of environmental libraries. gws, g wet sediment; MPS, mortar and pestle plus SDS method of DNA extraction; BB, bead beating method of DNA extraction. (+) indicates that more DNA could be obtained by repeating the extraction protocol. In other words, unlike the Tok sediments, the Ice sediment samples were not depleted during this investigation.

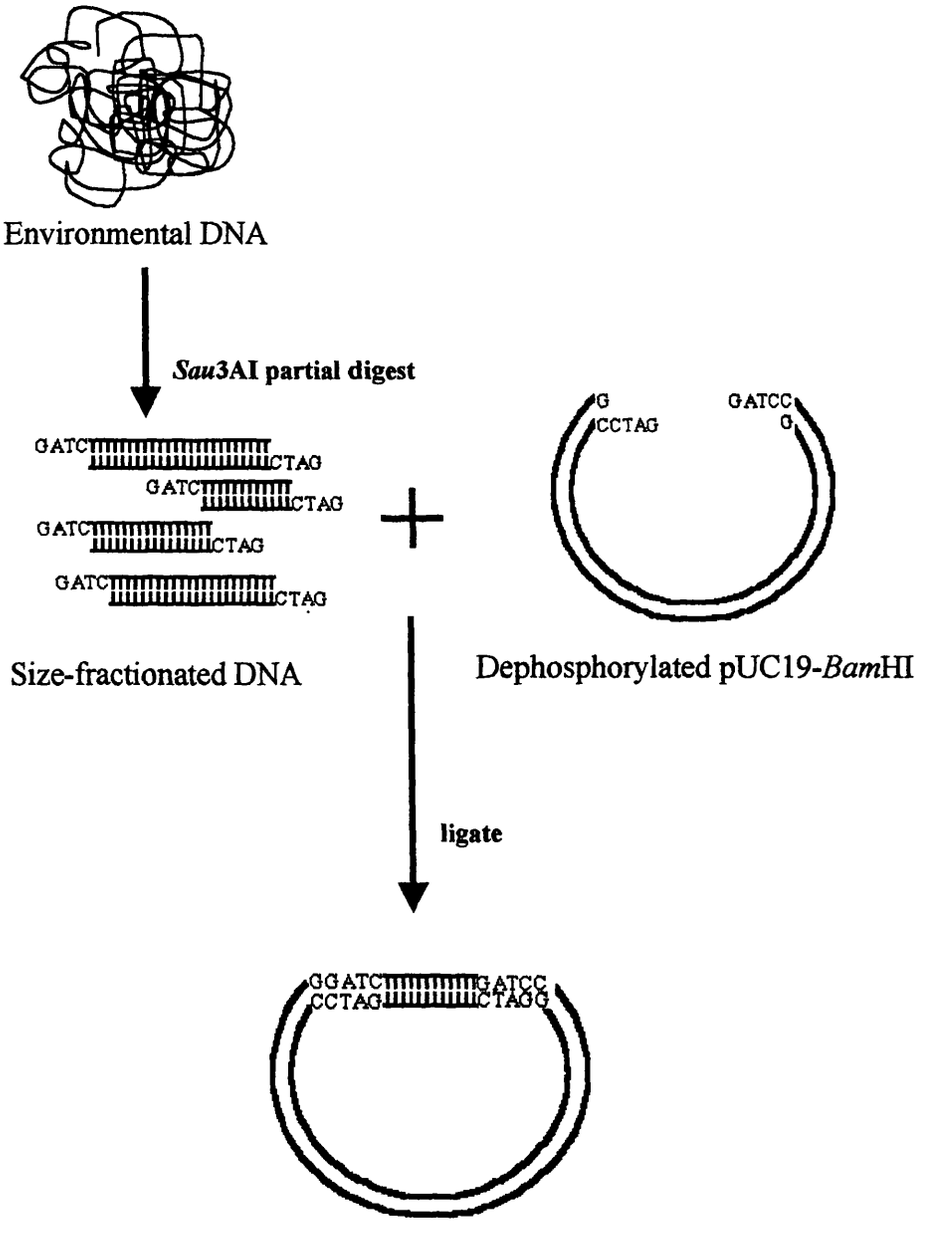


Figure 4.5: Overview of environmental library construction using pUC19.

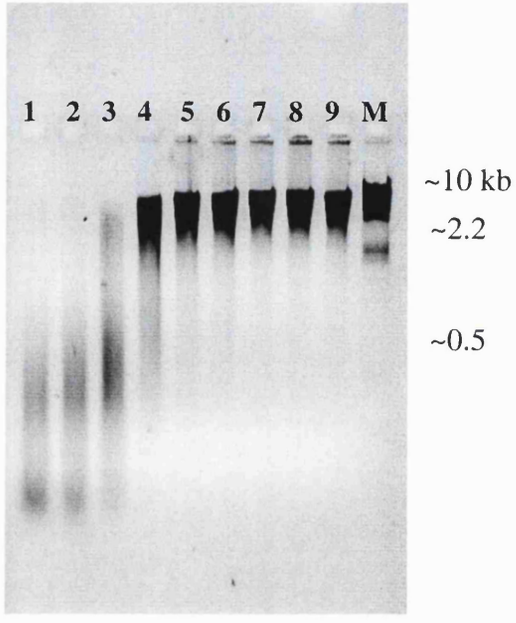


Figure 4.6: 1% agarose gel of GS-DNA partially digested with *Sau*3AI. 2-fold serial dilutions of *Sau*3AI were incubated with 0.5 μg GS-DNA in 15μl at 37°C for 1 hour. Lane 1, 2.5 U *Sau*3AI/μg GS-DNA; lane 2 1.25U/μg; lane3, 0.625U/μg; lane 4, 0.312U/μg; lane 5, 0.16U/μg; lane 6, 0.08U/μg; lane 7, 0.04U/μg, lane 8, 0.02U/μg, lane 9, 0U/μg. M is a λ*Hin*d III DNA weight marker. 30 to 50 μg GS-DNA was cleaved in subsequent large scale reactions, one third under the conditions of lane 3, one third under the conditions of lane 4 and one third under the conditions between lanes 3 and 4. Note that a portion of the DNA appeared refractory to cleavage by *Sau*3AI, and thus may not have been cloned into pUC19-*Bam*HI.

The cloning and recombination efficiency for GS-DNA ligated into pUC19-*Bam*HI were 8.2 x 10^4 cfu/μg DNA and 52% recombinant transformants, respectively (Table 4.2). The transformation efficiencies of an uncut pUC19 control and a control insert ligated into pUC19-*Bam*HI were 3.6 x 10^8 cfu/μg vector and 2.0 x 10^4cfu/μg vector, respectively, indicating that transformation and ligation reactions were functioning properly.

A likely cause for the low cloning efficiency of GS-DNA was the quality of DNA. GS-DNA was considered pure, however, because it was successfully digested with *Sau*3AI. Difficulty may still be encountered with the preparation of pure environmental DNA because some genomic DNA fragments may be methylated and resistant to digestion by restriction enzymes (Arber & Dussoix, 1962). A portion of GS-DNA did appear refractory to restriction by *Sau*3AI, suggesting that a subpopulation of GS-DNA could not be cloned. For the portion of GS-DNA fragments that could be digested, cloning may still have been problematic due to the heterogeneity of DNA termini. This is because DNA fragments extracted directly from sediments vary greatly in length (ranging from ~0.5kb to ~15kb in the case of GS-DNA). Complete or even partial digestion of these relatively short GS-DNA fragments may have generated a mixed population of DNA ends with some fragments possessing cohesive termini, and others containing only one or no clonable end(s).

Low recombination frequencies (less than 70%) of GS-DNA may have been due to incomplete cleavage or incomplete dephosphorylation of plasmid DNA prior to ligation. Purifying the linearised pUC19 by preparative gel electrophoresis minimised the former suggesting that in this case, incomplete dephosphorylation was the cause for low recombination efficiencies. Alkaline phosphatase from different commercial sources, however, did not significantly improve recombination efficiencies of environmental libraries (H. Hussein, unpublished results).

Because GS-DNA had such low cloning and recombination efficiencies, the protocol using pUC19-*Bam*HI was not suitable for generating environmental libraries. To see if cloning efficiencies with GS-DNA could be improved, a cloning protocol that did not rely on cohesive-end ligation was investigated.

Ligation		*n*	cloning efficiency (X 10^6cfu or pfu/ μg vector)	% recombination[a]	recombinants per ligation[b]
vector +	input DNA				
pUC19-*Bam*HI +					
GS-DNA-*Sau*3AI		3	0.082 (0.0096)	52	1100
pT7Blue-*Eco*RV +					
GS-DNA-*Sau*3AI		1	0.040	81	1600
GS-DNA		1	0.041	65	1300
TokA-DNA		2	0.022-0.026	90	1100
TokB-DNA		1	0.12	76	4600
TokC-DNA		1	0.0079	86	340
TokD-DNA		1	0.11	67	3700
TokE-DNA		0	Not done. Purified DNA was degraded to <2kb.		
Ice22-DNA (mortar & pestle + SDS)		1	0.082	89	3600
pCR-XL-TOPO-3'T					
Ice22-DNA (mortar & pestle + SDS)		5	0.78 (0.39)	88	6900
Ice22-DNA (bead beating)		2	0.85-1.0	75	7000
Ice16-DNA (bead beating)		3	1.4 (0.24)	84	12000
λTriplEx-*Bam*HI +					
Ice16-DNA (bead beating)		3	0	0	0

Table 4.2: Comparison of cloning protocols using environmental DNA. Evaluations were based on cloning efficiency, recombination efficiency and the number of recombinants recovered per ligation. Standard deviations are given in brackets. *n* is the number of ligations performed per protocol. **a:** For pUC19, pT7Blue and λTriplEx, recombination efficiency was determined by blue/white screening for recombinants on IPTG and X-gal. For pCR-XL-TOPO, recombination efficiency was determined by restriction analysis of isolated plasmids. **b:** The number of recombinants generated per ligation reaction = (cloning efficiency) X (amount of vector used per ligation) X (% recombination).

4.4: Blunt-End Cloning of Environmental DNA

pT7Blue Perfectly Blunt cloning kit (Novagen) is designed for cloning of any DNA fragment regardless of whether the termini are blunt ends or possess 5'-overhangs or 3'-overhangs. This feature of the kit made it a good candidate for constructing environmental libraries because the DNA, whether prepared by restriction digests or through mechanical shearing, could be used as input DNA.

To prevent self ligation and improve recombination efficiencies, the kit came supplied with pT7Blue vector that had been cleaved and dephosphorylated at its *Eco*RV-cloning site (pT7Blue-*Eco*RV) (Figure 4.2). As outlined in Figure 4.7, the termini of input DNA were converted to blunt ends by incubating with the kit's proprietary blunt-end conversion enzyme mix and ligated into pT7Blue-*Eco*RV (Section 2.8.8). Subsequent transformation into *E. coli* NovaBlue Competent Cells generated recombinant colonies that were visualised by blue/white screening (Section 2.8.12).

4.4.1: Blunt-end cloning of soil DNA

The blunt-end cloning and recombination efficiencies for *Sau*3AI-digested GS-DNA were 4.0 X 10^4 cfu/μg vector and 81% recombinant transformants, respectively. Compared with pUC19-*Bam*HI (Table 4.2), the cloning efficiency of pT7Blue-*Eco*RV with GS-DNA decreased by 51%. However, since the recombination efficiency increased by 56%, the blunt-end cloning method actually increased the overall number of recombinants by ~500 cfu/ligation. In a similar procedure GS-DNA was size-fractionated but not digested with a restriction endonuclease prior to blunt-end cloning. The efficiencies with this donor DNA were 4.1 X 10^4 cfu/μg vector and 65% recombinant transformants, respectively. Although direct comparisons cannot be made between the two donor DNAs, this demonstrates that input DNA with ragged termini can be used for cloning with the Perfectly Blunt cloning kit. The cloning efficiency of an insert control was ~10 fold greater than those of GS-DNA.

4.4.2: Blunt-end cloning of DNA from New Zealand geothermal sediments

The blunt-end cloning protocol using pT7Blue-*Eco*RV was considered a viable option for cloning DNAs extracted from New Zealand and Iceland geothermal sediments because it eliminated the requirement for cohesive DNA termini. This was considered

an important feature because DNA extracted directly from geothermal sediments were already fragmented without previous digestion (Figures 3.4 & 3.5; Table 3.5).

DNA was extracted from Tok geothermal sediments (Tok-DNA) (Table 4.1) using mortar and pestle as described in Section 2.6.1. DNA fragments ranging between 1 and 10 kb were recovered by preparative gel electrophoresis and cloned using the blunt-end DNA protocol as outlined in Figure 4.7.

Cloning efficiencies were highly variable between different Tok-DNA samples (Table 4.2). This is not surprising since cloning efficiencies are affected by many variables including inefficiencies of blunt-end ligation and unknown variables that may be present in DNA solutions extracted from different sediments. The conclusion was, however, that the efficiencies obtained with Tok-DNA using the pT7Blue Perfectly Blunt cloning protocol were not sufficient for generating environmental libraries. Furthermore, insufficient amounts of source material (Table 4.1) prevented scale-up of the cloning protocol. All Tok-DNAs were depleted before libraries of sufficient size could be constructed.

4.4.3: Blunt-end cloning of DNA from Iceland geothermal sediments

Although it was possible to generate libraries from less than 10μg Tok-DNA, it became clear that, greater amounts of DNA were required to generate environmental libraries that were suitably large enough for screening. 100μg DNA is generally required as starting material for general cloning purposes (Kaiser *et al.*, 1995). 10 to 1000 fold greater volumes of sediment was therefore collected from Iceland in order to extract sufficient DNA for cloning.

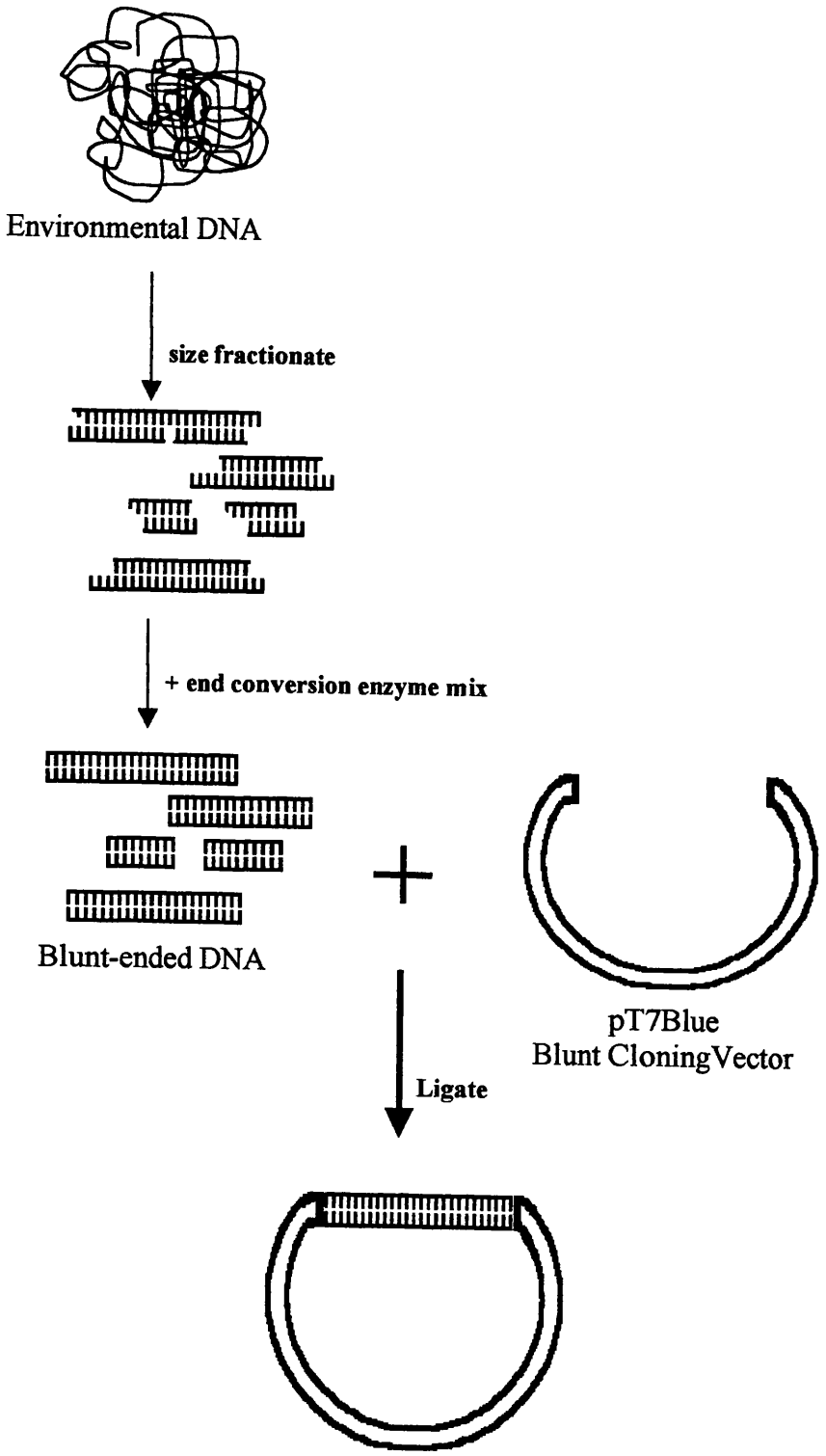


Figure 4.7: Overview of environmental library construction using pT7Blue.

DNA was extracted from Ice geothermal site 22 (Ice22-DNA) using the mortar and pestle plus SDS method as described in Section 2.6.1. Ice22-DNA was prepared for cloning into pT7Blue as described for Tok-DNA except that, after size fractionation, the DNA was incubated with Vent DNA polymerase, purified by spin column and resuspended in the appropriate volume of distilled H_2O (Section 2.8.4). An appropriate amount of resuspended DNA was then added to the blunt-ending reaction for cloning into pT7Blue. Incubation with Vent DNA polymerase is an integral step for TA-cloning and was included in the blunt-end cloning protocol so that the same type of input DNA could be used for comparison purposes.

The efficiencies of Ice22-DNA in pT7Blue were 8.2 X 10^4 cfu/μg vector and 89% recombinant transformants. Although not directly comparable, these efficiencies are similar to those obtained for TokD-DNA (Table 4.2). Because at least 600-fold more Ice22-DNA than TokD-DNA could be obtained, larger libraries could be prepared by scaling up the number of ligation and transformation reactions. Before deciding whether to scale up the blunt-end cloning procedure, a TA-cloning procedure was evaluated to see if the efficiencies of cloning could be further improved with Ice22-DNA.

4.5: TOPO TA-Cloning

The TA-cloning protocol using pCR-XL-TOPO was selected to determine whether the cloning efficiency of Ice22-DNA extracted by mortar and pestle could be improved over those obtained with the blunt-end cloning protocol tested in Section 4.4.3.

pCR-XL-TOPO was supplied linearised with 3' thymidine overhangs and topoisomerase covalently bound to the plasmid (Shuman, 1994) (Figure 4.3). The vector contains the lethal *ccd*B gene adjacent to upstream-sequences encoding the C-terminus of the *LacZ*α fragment. Ligation of insert into pCR-XL-TOPO disrupts expression of the *lacZ*α-*ccd*B gene fusion permitting growth of only recombinant transformants. Non-recombinants are killed upon plating. Blue/white screening was, therefore, not required (Bernard *et al.*, 1994).

Environmental DNA was prepared for TA-cloning as outline in Figure 4.8. The termini of input DNA were converted to blunt ends by Vent DNA polymerase before adding adenosines to the 3'-ends. The input DNA was then dephosphorylated before being ligated to the T-cloning site of pCR-XL-TOPO (Section 2.8). Termed TOPO-TA

cloning, ligation was achieved by exploiting the activity of topoisomerase supplied covalently bound to the vector (Shuman, 1994) (Figure 4.3).

4.5.1: TOPO-TA cloning of Ice22-DNA extracted *via* mortar and pestle is more efficient than the corresponding blunt-end cloning protocol

The TOPO-TA-cloned Ice22-DNA was electroporated into *E. coli* TOP10 cells as described in Section 2.8.11. Nearly a 10-fold increase in cloning efficiency was observed compared to the blunt-end DNA cloning protocol for the same input DNA (i.e. DNA extracted from Ice22 sediment *via* mortar and pestle) (Table 4.2). The TOPO-TA cloning protocol generated nearly twice as many recombinants per ligation compared to the blunt-end protocol.

4.5.2: TOPO-TA cloning of Ice22-DNA extracted *via* mortar and pestle is comparable to that of Ice22-DNA obtained through bead beating

To compare the cloning efficiencies of Ice22-DNA extracted by two different means, Ice22-DNA extracted by bead beating (Section 2.6.2) was cloned into pCR-XL-TOPO as described above for Ice22-DNA extracted by mortar and pestle. The cloning efficiency of ICE22-DNA extracted by the bead beating method increased by 15%. With a slightly lower recombination frequency, the total number of recombinants produced per ligation decreased by ~1%. The efficiencies of Ice22-DNA extracted by the two different methods could therefore be considered similar.

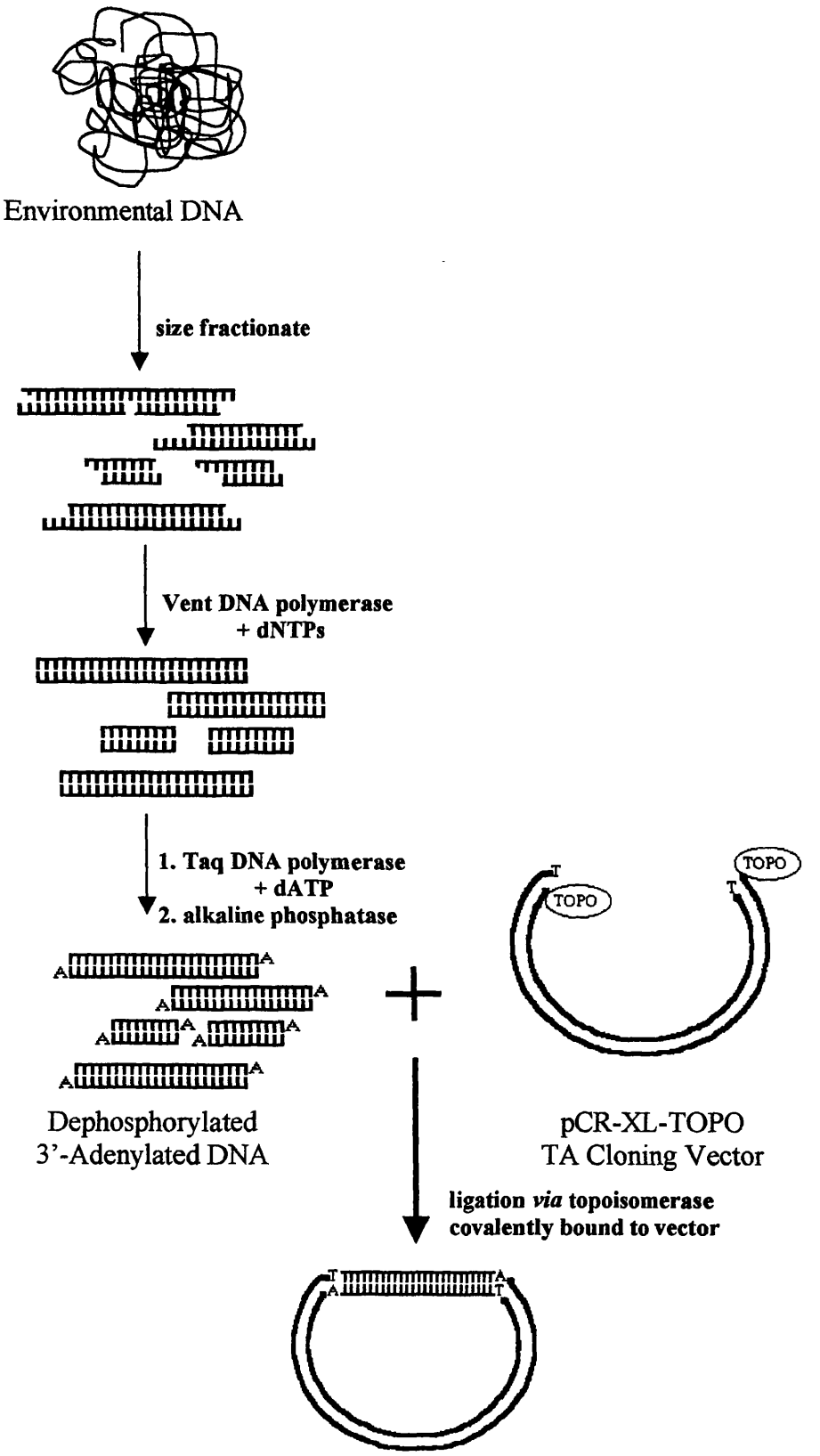


Figure 4.8: Overview of environmental library construction using pCR-XL-TOPO. TOPO is topoisomerase covalently bound to vector ends.

4.5.3: TOPO-TA cloning of DNA extracted directly from Ice16 geothermal sediment (Ice16-DNA)

Because agarose gel electrophoresis indicated high yields per extraction (Figure 4.9), DNA extracted from Ice16 geothermal sediment (Ice16-DNA) was selected for TOPO-TA-cloning. Ice16-DNA was extracted by the bead beating method (Section 2.6.2) and prepared for cloning into pCR-XL-TOPO essentially as described for Ice22-DNA (Section 4.5.3).

The TOPO-cloning efficiencies of Ice16-DNA were 1.5-times higher than those obtained with Ice22-DNA prepared in the same manner (Table 4.2). 75% more recombinants were generated per ligation with Ice16-DNA than with Ice22-DNA.

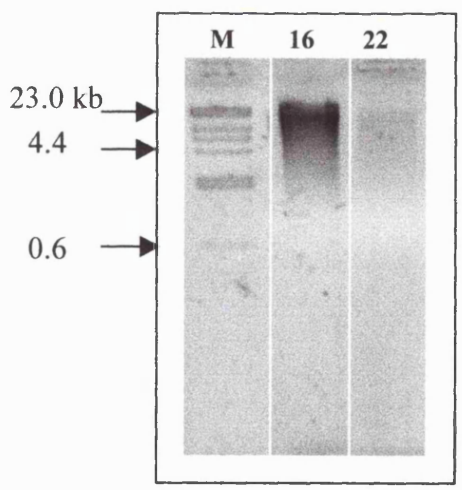


Figure 4.9: Comparison of Ice16-DNA (16) and Ice22-DNA (22) extracted from geothermal sediment *via* the bead beating method. 1/10 volume of purified DNA extracted from ~0.5g wet sediment was electrophoresed on 1% agarose gel. (M) λ-*Hin*d III DNA marker. For a detailed discussion about DNA yields see Section 3.4.2.

4.6: λTripleEx

The phagemid vector, λTriplEx was the next vector investigated to see whether cloning efficiencies could be further improved for Ice16-DNA. λTriplEx was chosen because it offered the high efficiencies of packaging and transduction of λ phage into *E. coli* and could accommodate inserts up to 10kb in size. Another feature that made this phagemid a candidate for environmental library construction was that it possesses two ribosomal binding sites and one translational slip site located immediately downstream of the *lacZ* promoter (Figure 4.10). Any heterologous expression driven from the *LacZ* promoter could therefore be permitted in all three reading frames.

Ice16-DNA was prepared for cloning into λTriplEx as described in Section 2.8.13 and outlined in Figure 4.11. Briefly, *Eco*R I adapter was ligated to the dephosphorylated blunt termini of Ice16-DNA. After adding 5'phosphate groups to the *Eco*R I half sites, the adapted DNA molecules were size fractionated to remove any small DNA fragments including adapter dimers which preferentially clone. The adapted DNA was then ligated to the λTriplEx DNA arms and the recombinant phage was packaged for transfecting *E. coli* XL1-Blue cells.

The efficiencies obtained with Ice16-DNA cloned in λTriplEx were 4.5 X 10^7 pfu/μg vector and 80% recombinants, as determined by blue/white screening. 19 white plaques were selected for *Cre-lox*-mediated phagemid conversion to pTriplEx (Section 2.8.14 and Figure 4.4). None of the selected plaques contained inserts when analysed by restriction analysis using *Hin*d III (Figure 4.12).

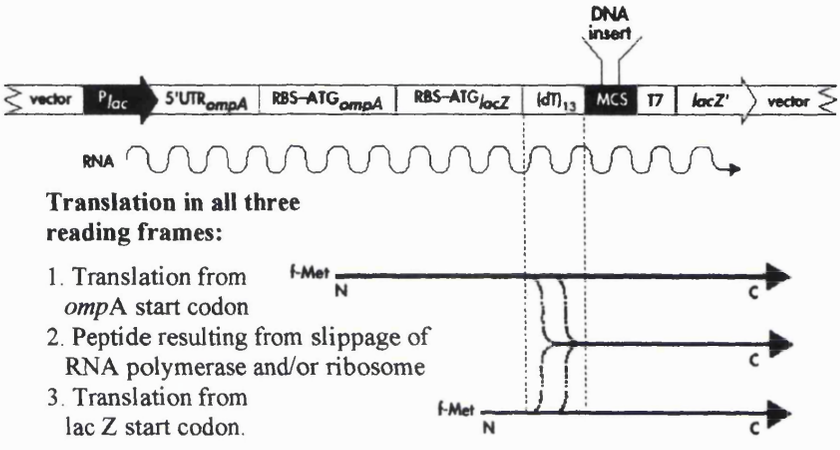


Figure 4.10: Generation of polypeptides from all three reading frames in a single recombinant λTriplEx clone. λTriplEx contains two translation start sites (i.e. two sets of ribosome-binding sites(RBS) and ATG start codons) in different reading frames and a slip site (a stretch of dTs) that can cause ribosomes to shift frames between the regulated *lac* promoter (P*lac*) and the multiple cloning site (MCS) (Atkins, *et al.*, 1990). RNA polymerase may also slip during transcription of the (dT)13 region (Wagner *et al.*, 1990). By the time the ribosomes begin translating the insert, roughly one-third will be in each of the three reading frames. 5'UTR *ompA* indicates the 5'untranslated region of the *ompA* gene in *E. coli*. These UTR sequences help stabilise mRNA for increased expression (Emory *et al.*, 1992). Figure 4.10 was modified from λTriplEx library user manual version PT3003-1.

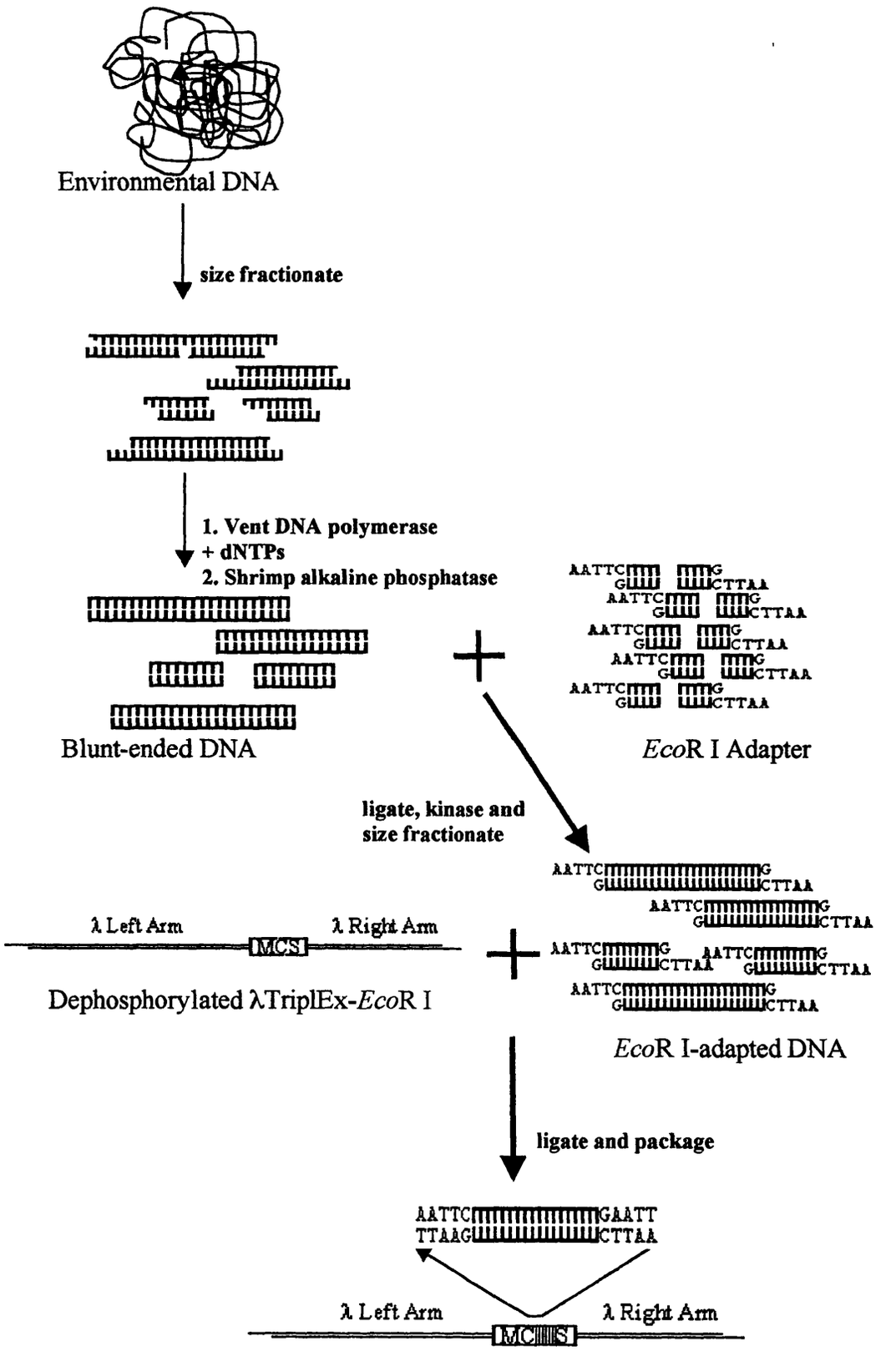


Figure 4.11: Overview of environmental DNA library construction using λTriplEx. MCS is multiple cloning site.

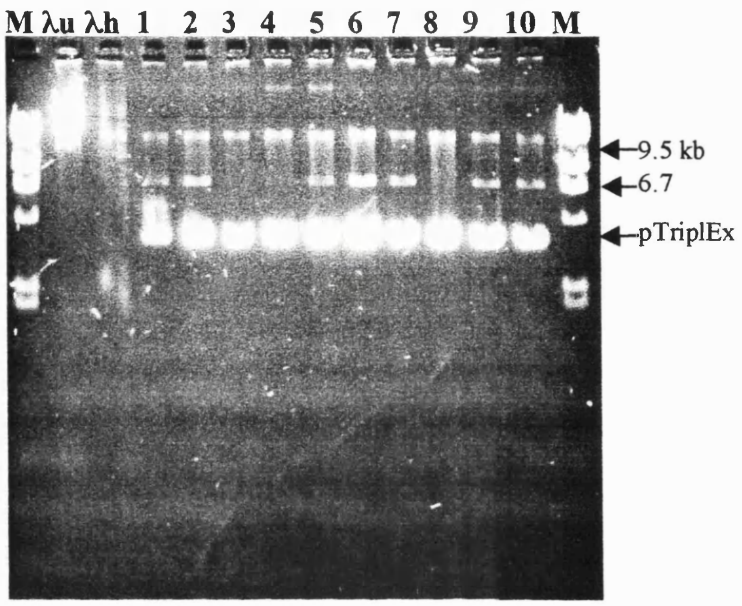


Figure 4.12: 1% agarose gel of Ice16 pTriplEx clones digested with *Hin*d III. Lanes 1-10 are 10 of the clones randomly selected from the Ice16 λTriplEx library. None of the 19 Ice16 pTriplEx clones analysed contained inserts. λu and λh are λDNA undigested and digested with *Hin*d III, respectively. M is λ*Hin*d III DNA marker. pTriplEx is the 3.6 kb linear plasmid vector. The bands corresponding to 9.5 kb and 6.7 kb are λDNA which has been co-purified with the plasmid DNA.

One reason for the low cloning efficiencies of Ice16-DNA with λTriplEx could be due to religation of vector arms; however, this was avoided by using dephosphorylated λTriplEx arms. Preferential insertion of adapter dimers or small DNA fragments into λ TriplEx is another cause for low recombination efficiencies. This was avoided, however, by removing small input DNA fragments *via* preparative gel electrophoresis. Furthermore, adapter dimers were removed by size exclusion chromatography prior to ligating with λ-DNA arms. Low efficiencies could also arise through difficulties of ligating blunt-ended DNA to adapter DNA. Although this could not be checked retrospectively, Vent DNA polymerase and shrimp alkaline phosphatase were not considered suspect because these reactions were used to successfully clone Ice16-DNA into pCR-XL-TOPO. T4-DNA ligase was not suspect because the control insert ligation was successful (4.9 X 10^6 pfu/μg vector). Finally, if the T4 polynucleotide kinase

reaction failed then the adapted DNA could not be ligated to the vector and therefore no plaques or only blue, non-recombinant plaques would develop. Until further investigations of this protocol are carried out, the reasons for unsuccessfully cloning Ice16-DNA into λTriplEx are inconclusive.

4.7: Construction of Environmental DNA Libraries

The TA-cloning protocol using pCR-XL-TOPO (Section 4.5) was selected for scale-up production of environmental DNA libraries because it generated more recombinants per ligation (TOPO-cloning) reaction (Table 4.2). The TA protocol also accommodated DNA fragments regardless of the type of termini. Conversely, the protocol using pUC19-*Bam*HI (Section 4.3) required input DNA having termini that were cohesive to the cloning site. Digesting relatively low molecular weight environmental DNA was thought to generate fragments possessing heterogeneous and thus unclonable ends. The inability to clone such heterogeneous ends may be reflected in the relatively low cloning efficiencies of the pUC19-*Bam*HI cloning protocol (Table 4.2).

The blunt-end cloning protocol using pT7Blue (Section 4.3) also permitted the cloning of heterogeneous DNA termini; however, it was not selected for large scale library preparation because cloning efficiencies were lower than those of the TA-cloning protocol (Table 4.2). λTriplEx was not considered for library construction because no recombinants could be generated with Ice16 DNA. (Section 4.6)

Designated ICE16 and ICE22, two environmental libraries were constructed using DNA extracted from Ice16 and Ice22 geothermal sediments, respectively. Although a portion of ICE22 was constructed initially using DNA extracted *via* mortar and pestle (Section 4.5.2), the scale-up of both ICE16 and ICE22 was performed only using DNA extracted *via* bead-beating (Sections 4.5.2 & 4.5.3). Scale-up of library construction was achieved by increasing the number rather than the volume of ligation (TOPO-TA-cloning) reactions. For each library, transformants were amplified, pooled and stored at-80°C until used in activity screens as described in Chapter 5.

Unamplified ICE16 and ICE22 environmental libraries possessed 44000 and 39000 transformants, respectively. For both libraries, restriction analysis of randomly selected clones indicated that 84% of plasmids contained an insert. ICE16 and ICE22 DNA libraries, therefore, contained 37000 and 33000 independent recombinant clones,

respectively. Insert sizes of recombinants ranged from ~1 kb to 10 kb with a median insert size of 4.7kb for both libraries (Figure 4.13). This represents 170 Mbp of cloned DNA for the ICE16 and 160 Mbp of cloned DNA for ICE22. Given an average of 1kb per gene, these libraries might contain 170000 and 160000 genes, respectively

Library Name	sediment	Cloning Efficiency (X 10^6cfu/μg vector)	% Recombinants	Insert Size (kb)	Independent Clones[a]
ICE16	~70°C, pH 9.5	1.4	84	~1-10	37000
ICE22	~55°C, pH 4.3	0.78	84	~1-10	33000

Table 4.3: Characterisation of the two environmental DNA libraries derived from Iceland geothermal sediments using pCR-XL-TOPO as cloning vector. **a:** Defined as the number of recombinants present in a library prior to amplification.

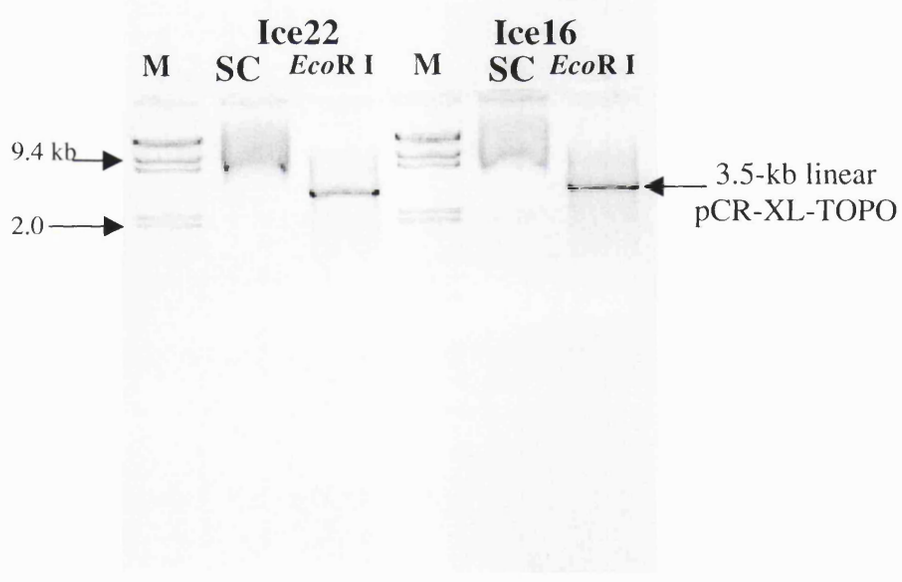


Figure 4.13: 1% agarose gel of supercoiled plasmid DNA and *Eco*R I-digested plasmids recovered *en masse* from ICE16 and ICE22 environmental libraries. Plasmid DNA was recovered by minipreparation from 3ml of amplified library. M is λ-*Hin*d III DNA marker; SC, supercoiled plasmid population; *Eco*R I, Plasmid DNA digested with *Eco*R I. *Eco*R I-digested DNA fragments from each library range between ~1 to 10Kb.

4.8: Summary

Four protocols for constructing environmental DNA libraries were evaluated, based on cloning efficiency, recombination efficiency and total number of recombinants generated per ligation. pUC19-*Bam*HI was the plasmid vector investigated for cohesive-end cloning of GS-DNA, while pT7Blue-*Eco*R V, used for blunt-end cloning, was investigated using the same input DNA. The blunt-end cloning protocol generated ~45% more recombinants than the corresponding cohesive-end cloning protocol. Environmental DNA libraries could not be produced using pT7Blue and DNA extracted from New Zealand sediments due to limiting amounts of source material.

pCR-XL-TOPO, supplied in the TOPO-XL cloning kit was used to investigate the efficiency of TA-cloning. The TOPO-cloning protocol generated ~92% more recombinants than the corresponding blunt-end protocol.

A phagemid protocol, using λTriplEx, failed to generate any recombinants.

Based on these findings, the TA-cloning protocol was selected for constructing environmental DNA libraries. ICE16 and ICE22 environmental libraries were constructed using DNA directly extracted from Iceland geothermal sediments. The two libraries contained 37000 and 33000 independent clones, respectively with insert sizes ranging between ~1 to 10kb. Both libraries were amplified and stored at -80°C (Section 2.2) until used in expression-screening procedures as described in Chapter 5.

Chapter 5
Screening Environmental Libraries for Thermostable Enzyme Activities

5.1: Aims

In Chapter 4, two environmental DNA libraries (ICE16 and ICE22) were constructed using DNA extracted directly from different geothermal sediments. The main aim of the work described in this chapter was to demonstrate the applicability of the environmental libraries by performing various screening assays for the detection of different functional thermostable proteins. This was done by adopting a library screening strategy that could accommodate various assays for detecting different thermostable enzymes that were encoded by the metagenome. Although enzymes of potential biotechnological interest were selected as target molecules for expression screening, this exercise was conducted primarily to demonstrate that heterologous sequences can be expressed in ICE16 and ICE22 environmental DNA libraries.

5.2: Background

ICE16 DNA library consisted of ~44000 independent transformants and was derived from Icelandic geothermal sediment collected from a piped runoff stream (~74°C, pH 9.5) that emptied into the Hveregirdi River. ICE22 DNA library consisted of ~39000 independent transformants and was derived from sediment collected from a stream (~50°C, pH 4.3) that ran from the main geyser at Krysuvik Seltun, Hverir, Iceland. Both libraries contained ~84% recombinant transformants with an average insert size of ~5kb. As described in the following sections, these multigenomic libraries were screened for thermostable α-amylase, lipase, phosphatase and protease activities using both microtitre plate and indicator agar plate platforms.

5.3: Library Pooling Strategy

The Clarke and Carbon (1976) equation for calculating the required size of DNA libraries was adapted in this work for estimating the number of recombinants which would need to be screened in order to have a certain probability of assaying all independent clones (Equation 5.1).

$$N = \frac{\ln(1-p)}{\ln\left(1-\frac{x}{y}\right)} \quad \text{(Equation 5.1)}$$

where, N= the number of agar plates or microtitre wells to be screened
p = the probability of screening all independent transformants within an amplified library
x = the cell density per agar plate or microtitre well
y = the number of independent transformants present within an amplified library.

In this work, x = 2000 cfu per agar plate or 500 cfu per well
y = 44000 for ICE16 and 39000 for ICE22.

For ICE16, 400 microtitre wells (at a density of 500 cfu/well), would need to be assayed in order to have, say, a 99% probability of screening each its 44000 independent transformants. In other words, for each functional screen performed on ICE16, 200000 clones should be assayed. Similarly for ICE22, 180000 clones (arrayed in 292 microtitre wells) should be assayed in order to provide a 99% probability of screening each of its 39000 independent transformants. These estimations are based on the assumptions that the growth characteristics of transformants are unaffected by sequence content and that cloned DNA inserts amplify at the same rate.

The strategy used for screening libraries in microtitre plates is presented in Figure 2.2. Amplified ICE16 and ICE22 DNA libraries were arrayed into master microtitre plates (master plates) as outlined in section 2.9.1. After incubating overnight in nutrient broth at 37°C, an aliquot from each master well was taken for assay. Glycerol (20% (v/v) final concentration) was then added to each master well for storage at -80°C. Alternatively, indicator agar plates were set up by plating ICE16 and ICE22-DNA libraries at a titre of 2000cfu/140mm-diameter plate. Approximately 50 indicator plates were screened per library, per activity. This provided a ~90% probability of assaying all independent transformants within each amplified library per screen. Increasing this probability by simply increasing the number of indicator plates was impracticable due to the large number of plates involved in handling and storage of the libraries. Screening ICE16 and ICE22 for thermostable α-amylase, protease, lipase and phosphatase is described in the following sections.

5.4: Amylase Screening

Starch, a polymer of glucose, is one of the most widely available plant polysaccharides and represents a major source of organic raw material. α-Amylase (EC 3.2.1.1), the enzyme that hydrolyses the α-1,4 glucosidic bonds of starch, is produced in many bacteria and fungi. α-Amylase is used in industrial processes such as starch liquefaction and the production of glucose syrup. High hydrolysis temperatures are desirable in the starch industry because starch granules cannot be attacked by α-amylases unless they have been ruptured by heat (Fogarty, 1983). Since enzymes from thermophilic sources are generally heat-stable, we decided to screen ICE16 and ICE22 for thermostable α-amylase activity.

5.4.1: Amylolytic activity screen using microtitre-plate assay

As described in Section 2.9.1.1, the ICE16 DNA library was screened for thermostable α-amylase activity by replica-plating 150μl from each of 264 master plate wells into assay wells containing 55μl of 1% (w/v) soluble starch in 30mM sodium phosphate buffer (pH 8.0) plus 0.1% SDS and 0.02% (v/v) Triton X-100. After sealing the plates with Parafilm and incubating overnight at 50°C, starch utilisation was detected by transferring 12.5μl of the overnight incubation to fresh microtitre plate wells containing 250μl iodine solution. Microtitre plates were read at 570nm.

0.27 % (w/v) starch was present in each sample and control well prior to the 50°C incubation step. Amylase activity was measured in ICE16 after this incubation step by comparing the amount of starch measured in each well to that of negative control wells. 0.27% (S.D. 0.029%, n=16) starch was measured in the wells that were incubated with *E. coli* TOP10/pCR– (Section 2.2), indicating that no starch was utilised by the negative control. After the 50°C incubation, all positive control wells (n=16) had less than the lower cut-off for starch (0.0039%) remaining. This indicated that essentially all of the starch was utilised by *E. coli* JM107/pQR126 (Section 2.2). For ICE16, any master well that, when assayed, fell outside two standard deviations of the negative well was re-assayed for amylase activity. No re-assayed master well was positive for starch hydrolysis, indicating that amylase activity was not detected in the ICE16 environmental DNA library.

ICE22 library was screened for amylase activity as described above except that each assay plate was visually inspected for starch hydrolysis rather than *via* a plate reader. For ICE22, any microtitre plate assay well that displayed a starch/iodine colour different from that of the negative well was re-assayed. No re-assayed master plate well was positive for starch hydrolysis, indicating that amylase activity was not detected in ICE22 environmental DNA library.

5.4.2: Amylolytic activity screen using indicator agar-plate assay

ICE16 and ICE22 environmental DNA libraries were also screened at pH 7 using starch indicator agar plates as described in Section 2.9.2.1. After growth at 37°C, the plated recombinants were transferred to 50°C and incubated again overnight. Although host cell proteins heat-denature relatively quickly, the long duration of the heat inactivation was not considered deleterious because zones of starch hydrolysis would indicate amylase activity whether or not any heterologous thermostable amylase remained at the time of detection. This high-temperature incubation also appeared to lyse the *E. coli* cells allowing for the release of any heterologously expressed proteins to come into contact with the starch agar. For detection of amylase activity, the starch plates were flooded with iodine solution, which was then poured off as each plate became saturated with the purple-black starch/iodine complex.

Flooding the plates with iodine solution revealed no zones of hydrolysed starch indicating that no amylase activity was detected in either ICE16 or ICE22 library. The positive control, *E. coli* JM107/pQR126, when plated onto the starch plates, produced large zones of clearing which were evident prior to addition of iodine. This demonstrated that starch hydrolysis can be detected with this assay system.

Although no zones of clearing were present, five dark coloured transformants were noted amongst the general population of creamy-white transformants present in ICE16. These five clones were selected to determine whether their cloned inserts were associated with the observed phenotypic change.

Because the plated library was heat inactivated and flooded with iodine, the *E. coli* host could not be grown for plasmid preparation. Each dark transformant was therefore used directly for plasmid mini-preparation as described in Section 2.8.14. The plasmid DNA

recovered from each dark-coloured transformant was reintroduced into *E. coli* TOP10 cells *via* electroporation (Section 2.8.11).

Plasmid DNA recovered from one dark-coloured transformant failed to re-transform *E. coli* TOP10 cells. This is most likely due to insufficient amount of plasmid isolated from the dead colony. Plasmids recovered from four of the dark transformants did re-transform *E. coli*; however, when assayed on starch plates, one re-transformant failed to produce dark-coloured colonies. The three re-transformants that did produce dark colonies when assayed on starch plates were designated 5ICE16, 6ICE16 and 12ICE16.

Restriction digests of plasmids from these three transformants (p5ICE16, p6ICE16 and p12ICE16, respectively), indicated that p6ICE16 and p12ICE22 are derived from the same clone (Figure 5.1). This was confirmed by comparing the DNA-insert sequence of p6ICE16 with the partial DNA-insert sequence of p12ICE16 (results not shown). Further experimental work was conducted on 5ICE16 and 6ICE16 to investigate the observed phenotype of these transformants. These experiments are described and discussed in Chapter 6.

5.5: Protease Screening

Proteases are used primarily in the detergent industry and in the dairy industry and are produced commercially from both bacterial and fungal sources. For use as detergent enzymes, proteases must be stable at high temperatures (Fogarty, 1983). Having been derived from geothermal sources, ICE16 and ICE22 libraries make good candidates for the screening of thermostable protease activities.

5.5.1: Proteolytic activity screen using indicator agar-plate assay

ICE16 and ICE22 DNA libraries were plated onto skim milk indicator plates and screened for thermostable protease activity as described in Section 2.9.2.4. *E. coli* TOP10/pCR– cells plated in this manner were protease-negative (i.e. the surrounding agar was not clear). Protease solution spotted onto the protein indicator plates resulted in the formation of clear spots. This demonstrated that the indicator plates could be used to detect protease activity. Out of approximately 100,000 recombinants screened per library, no clone possessing thermostable protease activity was detected.

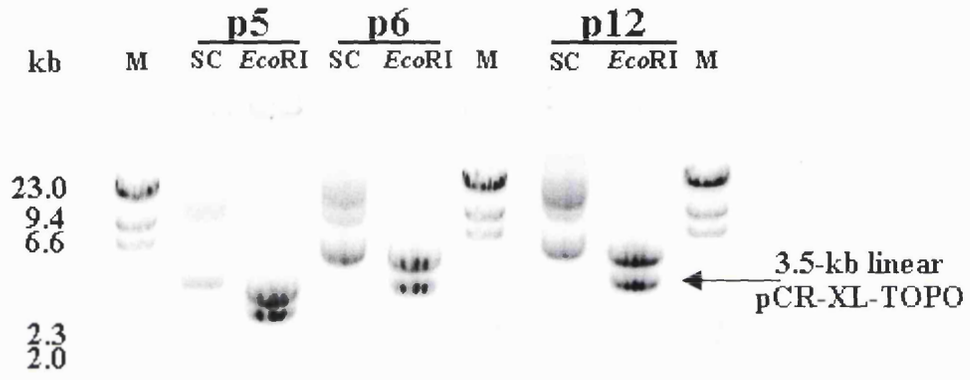


Figure 5.1: 1% agarose gel of plasmids (p5, p6 and p12) isolated from *E. coli* TOP10/pCR-XL-TOPO clones p5ICE16, p6ICE16 and p12ICE16, respectively. These recombinants displayed a dark-coloured phenotype when assayed on starch indicator plates as described in Section 5.4.2. M is λ-*Hin*d III DNA marker; SC, supercoiled plasmid DNA; *Eco*R I, plasmid DNA digested with *Eco*R I.

5.6: Lipase Screening

Lipases (glycerol ester hydrolases) (EC 3.1.1.3) hydrolyse fats into di or monoglycerides and fatty acids which are readily metabolised by the cell (Fogarty, 1983). Because lipases derived from thermophilic sources make good candidates for lipid modification in industrial processes, ICE16 and ICE22 DNA libraries were screened for lipolytic activity using both microtitre plate and indicator agar plate formats.

5.6.1: Lipolytic activity screen using microtitre-plate assay

As described in Section 2.9.1.2, ICE16 and ICE22 DNA libraries were screened for thermostable lipase activity by replica-plating 75μl from each master well to fresh assay wells containing 15μl of 0.5M Tris-Cl (pH8.0). After heat-inactivating the host-encoded proteins, 15μl of 1.5mM PNP-palmitate was added to each assay well and the microtitre plates incubated at 60°C. The plates were visually inspected periodically for the development of a yellow colour indicating lipase activity. The colour of the negative control wells containing *E. coli* TOPO10/pCR–, remained unchanged throughout the assay indicating little or no non-enzymatic hydrolysis of substrate. Positive control wells, containing lipase which was added after the heat inactivation step, resulted in the immediate formation of a yellow colour demonstrating that the substrate was able to be hydrolysed.

Out of 264 master wells screened (500 clones per well) for each library, no thermostable lipase activity was detected in either ICE22 or ICE16 master plates.

5.6.2: Lipolytic activity screen using indicator agar-plate assay

ICE16 and ICE22 environmental DNA libraries were also screened at pH 7 using Tween80 (oleic acid ester) indicator agar plates as described in Section 2.9.3. After growth at 37°C, the plated recombinants were transferred to 50°C and incubated overnight. *E. coli* TOP10/pCR– cells plated in this manner were lipase-negative (i.e. the surrounding agar was not opaque). Lipase solution spotted onto the Tween-80 indicator plates resulted in the formation of opaque spots demonstrating that the substrate was able to be hydrolysed. Out of approximately 100,000 recombinants screened per library, no clone possessing thermostable lipase activity was detected using this method.

5.7: Phosphatase Screening

Phosphatases are a group of enzymes that hydrolyse esters or anhydrides of phosphoric acid. Phosphomonoesterases differ in their substrate specificity and pH optimum and a common 'artificial' substrate for detecting their activity is *p*-nitrophenyl phosphate (PNP-phosphate).

5.7.1: Phosphatase activity screen using microtitre-plate assay

As described in section 2.9.1.3, ICE16 and ICE22 DNA libraries were screened for thermostable alkaline phosphatase activity by replica-plating 75μl from each master well to fresh assay wells containing 15μl of 0.5M Tris-Cl (pH8.0). After heat-inactivating the host-encoded proteins, 10μl of 30mM PNP-phosphate was added to each assay well and the microtitre plates incubated at 60°C. The plates were visually inspected periodically for the development of a yellow colour indicating phosphatase activity. Assays were stopped when the colour of the negative control well, containing *E. coli* TOPO10/pCR–, turned bright yellow indicating the non-enzymatic hydrolysis of the substrate. Positive control wells, containing phosphatase solution added to the positive control well after the heat inactivation step, resulted in the immediate formation of a yellow colour demonstrating that the substrate was able to be hydrolysed.

Out of 264 master wells screened (500 clones per well) for each library, alkaline phosphatase activity was detected in three different wells for ICE22 but in no wells for ICE16. The three phosphatase-positive wells were each serially diluted and reassayed for alkaline phosphatase. For each serial dilution, the phosphatase-positive well with the highest dilution factor was plated onto nutrient agar for isolation of individual recombinants. These isolated clones were then reassayed (Figure 2.2). Three phosphatase-positive clones, each derived from a separate master well, were isolated for further investigation (B Jredah, unpublished results). Restriction digests (Figure 5.2) and partial sequence comparisons indicated that the three phosphatase-positive clones isolated from ICE22 were monoclonal. The plasmid isolated from this clone was designated pPhos22. The experimental work conducted to investigate the phosphatase activity of this clone is described and discussed in Chapter 6.

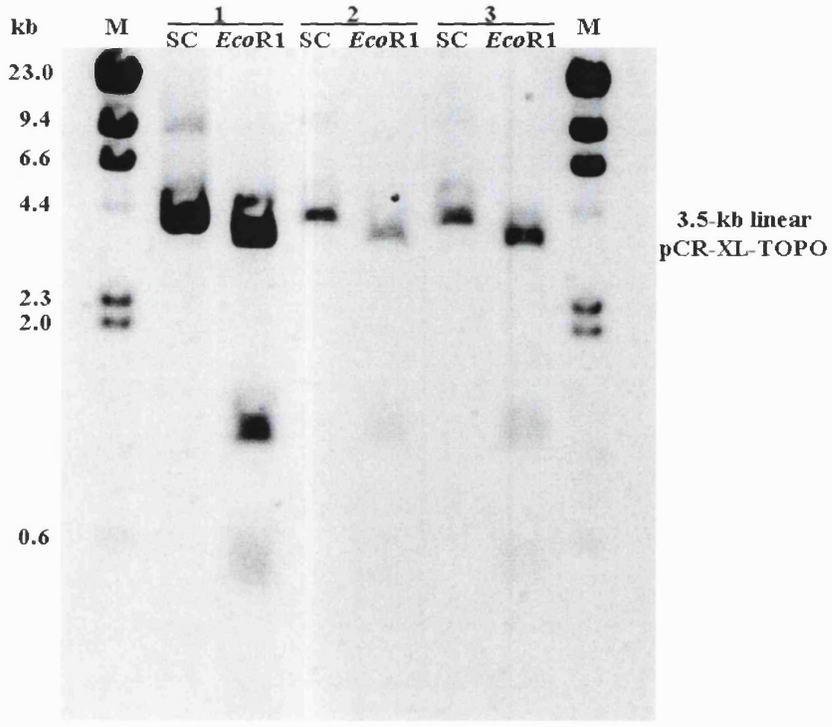


Figure 5.2: 1% agarose gel of plasmids prepared from three *E. coli* TOP10/pCR-XL-TOPO clones detected in ICE22 environmental library displaying thermostable alkaline phosphatase activity. As described in Section 5.7.2, these monoclonal recombinants were independently isolated from three different master wells of the arrayed library. M is λ-*Hin*d III DNA marker; SC, supercoiled plasmid DNA; *Eco*R I, plasmid DNA digested with the restriction enzyme, *Eco*R I.

5.7.2: Phosphatase activity screen using indicator agar-plate assay

The tryptose phosphate methyl green (TPMG)-based expression cloning system (Section 2.9.2.3) allows for the generical detection of phosphatase activity. Using phenolphthalein diphosphate (PDP) as substrate, this system has been used to detect a Class A non-specific acid phosphatase, an acid hexose phosphatase and an alkaline phosphatase expressed in *E. coli* genomic libraries derived from *Providencia stuartii*, *Providencia rettgeri* and a Gram-negative environmental strain LR-23, respectively (Riccio *et al.*, 1997). As an alternative means for detecting thermostable phosphatase activity, ICE16 and ICE22 libraries were screened at pH 7.2 using TMPG + PDP

indicator plates. After growth at 37°C, the plated recombinants were transferred to 50°C and incubated again overnight. *E. coli* TOP10/pCR– cells plated in this manner were unstained (i.e. the clones showed a phosphatase-negative phenotype). Phosphatase solution spotted onto the TMPG + PDP plates resulted in the formation of green-stained spots demonstrating that phosphatase activity could be detected with this system. Out of approximately 100,000 clones screened for each library, no phosphatase-positive colonies were detected using this method.

5.8:Summary

To allow for high-throughput screening of enzyme activities, ICE16 and ICE22 environmental libraries were arrayed into master microtitre plates at a titre of ~500 cfu per well. The master plates were then used as a source of recombinants for screening thermostable α-amylase, lipase and phosphatase activities. Approximately 132000 recombinants were screened per assay per library using the microtitre plate format. This provided a ~95% probability of assaying all independent clones within each amplified library per assay. Alternative assays for screening for thermostable α-amylase, lipase, phosphatase as well as protease activities were conducted using indicator agar plates. Approximately 100000 recombinants were screened per library per assay. This provided a ~90% probability of assaying all independent clones within each amplified. Increasing this probability by simply increasing the number of indicator plates screened was impracticable.

Although amylase activity was not detected in the environmental libraries, two recombinants (*E. coli* TOP10/p5ICE16 and *E. coli* TOP10/p6ICE16) displaying a dark-coloured phenotype on starch plates at 50°C were isolated from ICE16. Also, one recombinant (*E. coli* TOP10/pPhos22) possessing thermostable phosphatase activity was isolated from ICE22 during microtitre plate screening using PNP-phosphate as substrate. This result demonstrates that heterologous sequences, encoded by DNA derived from geothermal sediment and cloned into pCR-XL-TOPO, can be expressed in *E. coli* at detectable levels. No other enzyme activity was detected in either library using the methods described here. *E. coli* TOP10 strains, p5ICE16, p6ICE16 and pPhos22 were selected for further analyses as described and discussed in Chapter 6.

Chapter 6

Sequence and Expression Studies

6.1: Aim

The aim of the work described in this chapter was to characterise the recombinant *E. coli* strains identified during expression screening of environmental DNA libraries. In order to identify the molecular determinants responsible for the observed phenotypes, molecular and expression analyses were conducted on these isolated clones.

6.2: Background

Environmental DNA libraries from two different geothermal sediments were constructed (Chapter 4) and screened for thermostable enzyme activities (Chapter 5). Plasmids (p5ICE16 and p6ICE16) from two clones isolated from ICE16 environmental library conferred an atypical phenotype to *E. coli* hosts when propagated on starch plates. After heat inactivation and addition of iodine solution, these transformants turned dark while the remainder of the library retained the creamy white phenotype typical of *E. coli*. The plasmid (pPhos22) of another clone recovered from environmental DNA library ICE22 conferred thermostable alkaline phosphatase activity to its *E. coli* host.

In order to identify the gene(s) conferring these phenotypes, the DNA inserts of p5ICE16, p6ICE16 and pPhos22 were sequenced and compared to the sequences in the NCBI databases (Section 2.10). Finally SDS-PAGE of crude cell extracts (Section 2.7.5) and complementation assays were performed on selected clones in order to identify heterologously expressed proteins.

6.3: p5ICE16

Recombinant plasmid p5ICE16, which conferred a dark phenotype to *E. coli* TOP10 cells, was obtained from ICE16 environmental DNA library (Section 5.4.2). The DNA used for ICE16 library construction was derived from a biomass-containing sediment; ~70°C, pH 9.5 (Section 4.7).

6.3.1: p5ICE16 contains a 2.9-kb insert

The physical map of the p5ICE16-DNA insert is shown in Figure 6.1. Three ORFs were identified within this 2878-bp fragment (Section 2.10). The ATG translational start at bp 112 was assigned to p5*orf*1 which was apparently truncated 112-bp downstream at the T/A-cloning site (Figure 6.2). A TAG stop codon was identified in the vector sequence 12 codons downstream of the truncation. This p5*orf*1-vector fusion encodes a 48-aa peptide. A possible Shine-Dalgarno (SD) sequence (119GAGGT115) but no promoter-like sequences were identified in the region preceding p5*orf*1 (Section 2.10). Located on the opposite DNA strand, p5*orf*2 was shown to encode a 251-aa protein with a predicted molecular mass of 28.9 kDa (Section 2.10). The ATG codon at bp 216 was assigned as the translational start for p5*orf*2 which ends with a TGA stop codon at bp 971. p5*orf*2 is preceded by putative promoters (bp 59-108 and 174TATAAT179) and SD sequences (203GGAG206) (Figure 6.2). p5*orf*3, whose coding strand is in the same orientation as that of p5*orf*1, is located downstream of p5*orf*2 and encodes a 612-aa protein with a predicted molecular mass of 66.8 kDa (Figure 6.3). The GTG translation initiation codon at bp 2817 and the TAA stop codon ending at bp 981 were assigned to p5*orf*3 which is preceded by putative −10 promoter (2868TATAAT2863) and SD sequences (2827GGAG2824). No archaeal promoter-like sequences were identified in regions upstream of these ORFs (Section 2.10).

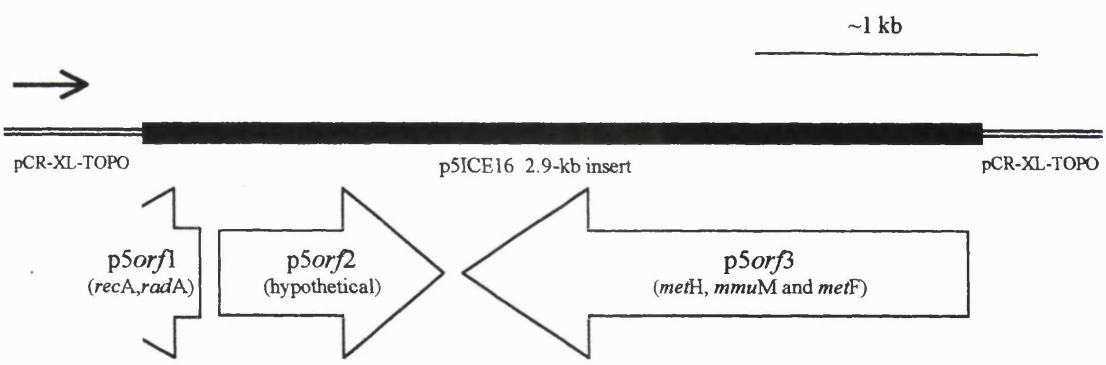


Figure 6.1: Physical map of the 2.9-kb Iceland16 DNA fragment (black bar) cloned into the TA vector, pCR-XL-TOPO (double line) and recovered from *E. coli* TOP10/p5ICE16. Block arrows show the relative localisation of each gene and the orientation of the open reading frames (p5*orf*1, p5*orf*2 and p5*orf*3). p5*orf*1 is truncated at the cloning site as indicated by an incomplete arrowhead. The genes encoding homologous sequences are represented in parentheses. The black arrow indicates the direction of transcription from the *lac* promoter.

The p5ICE16-DNA fragment appeared to be of bacterial origin, given the homology of potential genes on the insert to sequences found in the databases.

6.3.2: The incomplete p5*orf*1 gene product has homology to RecA-like proteins

BLAST searches (Section 2.10) showed that the deduced amino acid sequence of the incomplete p5*orf*1 has highest homology to the N-terminal sequence of a putative RadA homologue from *Aquifex aeolicus* (55% identity over 37 aa) (Accession O66827) (Deckert *et al.*, 1998). Archaeal *rad*A (Seitz *et al.*, 1998), bacterial *rec*A (Kowalczykowski *et al.*, 1994) and eukaryotic *rad*51 (Ogawa *et al.*, 1993) are analogous genes whose product catalyses the synaptic event in homologous recombination. RecA-like proteins are recombinases that facilitate the alignment of homologous DNA sequences and catalyse the exchange of DNA strands (Kowalczykowski, 1991). Since p5*orf*1 was truncated after only 37 codons, there was insufficient nucleotide sequence information to further investigate its gene product. Cloning and nucleotide sequence analysis of intact p5*orf*1 would reveal whether it is homologous to *rec*A.

6.3.3: p5*orf*2 gene product is a conserved hypothetical protein with limited homology to *N*-acetyltransferase

BLAST searches using the deduced p5*Orf*2 sequence revealed highest homologies across the entire protein to the N-terminal half of conserved hypothetical proteins from *Bacillus subtilis* (24% identity over 215 aa) and *Sphingomonas yanoikuyae* (24% identity over 176 aa) (Accessions CAA65708 and AAC43641, respectively). The N-terminal third of p5*Orf*2 also showed limited homology across the entire length of putative Ard1-like N-acetyltransferases from *Pyroccocus abyssii* and *Pyroccocus horikoshi* (~20% identity over 172 aa) (Accessions CAB50578 and BAA29368, respectively). Ard1 is a eukaryotic *N*-acetyltransferase that co-translationally modifies specific proteins (Arnold *et al.*, 1999; Polevoda, *et al.*, 1999). No genetic, biochemical or structural information pertaining to these proteins identified in the BLAST search could be found in the literature. p5*orf*2 was therefore characterised as coding for a conserved hypothetical protein. Further investigation into p5*Orf*2 may reveal whether it is analogous to Ard-1.

Figure 6.2: p5*orf*1 and p5*orf*2

```
1       GCGTTTCCTCTTGCAGTGTGTTCCACTCCCCACATTCGGAGCAGCGCCCC      50
        CGCAAAGGAGAACGTCACACAAGGTGAGGGGTGTAAGCCTCGTCGCGGGG
37 trunc. T  E  E  Q  L  T  N  W  E  G  C  E  S  C  R  G      22
                P1 ..........................................
51      ACCCAGCGCGGCGTTGTATAGCCGCAAACAGAGCAAACATAGTGCGTGCG     100
        TGGGTCGCGCCGCAACATATCGGCGTTTGTCTCGTTTGTATCACGCACGC
21       V  W  R  P  T  T  Y  G  C  V  S  C  V  Y  H  T  R      5
         ..........>
101     TACCCTGGGCATTGACCTCACGAAATTATACCTGTTTCTGAAACCGCGAA     150
        ATGGGACCCGTAACTGGAGTGCTTTAATATGGACAAAGACTTTGGCGCTT
4         V  R  P  M<p5orf1|                                    1

151     GCAGGAATTGCGCAAAGGTGGTGTATAATTATGCCCACAGATTACTCGCG     200
                                P2 ------->
201     ACGGAGAGCTGACCCATGTCGGCTCTCCGTTTTTGCGTCTTATGGCTCGT     250
1             |p5orf2>M  S  A  L  R  F  C  V  L  W  L  V      12

251     TCCCAACCAGGGAAAGGCAGGCACAGCAATGATGACGGTAGGGCTATTCC     300
13       P  N  Q  G  K  A  G  T  A  M  M  T  V  G  L  F  Q      29

301     AGGTGCGCCCTGCAGAGGCGCGTGATCTGGATCGGGTCGAGGAGCTGGAT     350
30       V  R  P  A  E  A  R  D  L  D  R  V  E  E  L  D       45

351     CAGTTGTTTGGCAACGCCGCGCTTTCGCGCGACTACTTCGAGGCGTGGCT     400
46      Q  L  F  G  N  A  A  L  S  R  D  Y  F  E  A  W  L      62

401     GCAGAACCATCCGCTGGGGTTTCTGGTCGCCGAGTTCGACGGGCGCATTT     450
63       Q  N  H  P  L  G  F  L  V  A  E  F  D  G  R  I  Y      79

451     ACGCCTATAGCATGGTGATTTACCTCCATCGCCACCAGATTCGCGACAAC     500
80       A  Y  S  M  V  I  Y  L  H  R  H  Q  I  R  D  N        95

501     TGGTACCACGACACAGCGGGGGGCACTTGCGCCAGCCATAGCCCGCTGGG     550
96      W  Y  H  D  T  A  G  G  T  C  A  S  H  S  P  L  G      112

551     CGAATATCTTTACGCCGTGAGCATCGCCTCACAAAAGCAGGAGGCGGCAC     600
113      E  Y  L  Y  A  V  S  I  A  S  Q  K  Q  E  A  A  R     129

601     GGGCGCTGTTCATCGCCAGCCGGCGGCTGTTCATTCGCTCTCATGTGTAT     650
130      A  L  F  I  A  S  R  R  L  F  I  R  S  H  V  Y       145

651     CAGACGCTCGTCTACGGCAGACTCCCTCGCTTCCGCAAATGGGTCCTCGA     701
146     Q  T  L  V  Y  G  R  L  P  R  F  R  K  W  V  L  E      162

701     GCAGGGCTACGACCCCGAAACGCTCACCCCTGAACAGAAGCAACGCCTGC     750
163      Q  G  Y  D  P  E  T  L  T  P  E  Q  K  Q  R  L  L     179

751     TGAACCTTTATGTCAACCTGCTTTATGACCCCTATCAGGTGTTTTACGAA     800
180      N  L  Y  V  N  L  L  Y  D  P  Y  Q  V  F  Y  E       195

801     GGGCTGAGCTTTATGCCCGAAGGCGGAGTCGTGGACTATCTGGAGGGGGA     850
196     G  L  S  F  M  P  E  G  G  V  V  D  Y  L  E  G  D      212

851     CGACGAGTCGCTCAACTGTGCGCTGAAGCTGGTATGGCGCAACCCCTATT     900
213      D  E  S  L  N  C  A  L  K  L  V  W  R  N  P  Y  Y     229

901     ATCGCCCGCTGACGACGGATCGCGCCGTGGTGGAAGCGCCTGAAGTAGCG     950
230      R  P  L  T  T  D  R  A  V  V  E  A  P  E  V  A       245

951     GTGACGACTCCGAAGCAGTGAGAAGGGATTACCGTATTGCCTCAATCACC    1000
246     V  T  T  P  K  Q  *251      AATGGCATAACGGAGTTAGTGG
                                p5orf3end*  R  I  A  E  I  V cont.
```

Figure 6.2: Nucleotide sequence of the region of p5ICE16 containing the truncated p5*orf*1, p5*orf*2 and the 5'-end of p5*orf*3 and deduced aa sequences of the C-terminal region of p5*Orf*1, p5*Orf*2 and N-terminal region of p5*Orf*3. The nucleotide sequence has been numbered with respect to the orientation of the insert with bp 1 of the insert positioned closest to the plasmid-encoded *lac* promoter. Both DNA strands are shown for p5*orf*1 and p5*orf*3, whose coding strand is opposite to that of p5*orf*2. The putative promoters (P1 and P2) upstream of p5*orf*2 are indicated by arrows. The transcriptional start, as predicted by promoterscan (Section 2.10) is highlighted in bold. Possible ribosome-binding sites for p5*orf*1 and p5*orf*2 are indicated by double-underlines. Stop codons are indicated by asterisks. Nucleotide positions 1001 through 2878 containing the remainder of p5*orf*3, are not shown.

Figure 6.3: p5*orf*3

```
2878   GGCTTCGTGGTATAATCCCTCCGTGAATCGCGAGCAGTTTCGGGCGCGTC   2829
                P3 ------>
2828   TGGAGCAGGGGGTGCTGATTGCCGATGGCGCCATCGGCACGATGCTGGCA   2779
1        |p5orf3>V  L  I  A  D  G  A  I  G  T  M  L  A      13

2778   TTGCGCGGGGTGCCAACCCCCTATGAACTTGCGAACCTGCTTTATCCCGA   2729
14     L  R  G  V  P  T  P  Y  E  L  A  N  L  L  Y  P  D    30

2728   TACCGTGCGGGCGCTCCATCGGGAGTATTACGAGGCGGGCGCACGCCTGA   2679
31      T  V  R  A  L  H  R  E  Y  Y  E  A  G  A  R  L  I    47

2678   TAGAGACCAACACCTACACTGCCAATCGGGTACGCCTGTTTAATCTGCCT   2629
48       E  T  N  T  Y  T  A  N  R  V  R  L  F  N  L  P      63

2628   GAACGCGGCAGCGAAGCGCCCCCCACCTATAGCCTGCTGGAACAGTTCGG   2579
64     E  R  G  S  E  A  P  P  T  Y  S  L  L  E  Q  F  G    80

2578   CTCGCCGGAAGAGCTGGTACGGCGCATCAATCAGGAGGCGGTGCGCCTGG   2529
81      S  P  E  E  L  V  R  R  I  N  Q  E  A  V  R  L  A    97

2528   CACGCGAGGCGGTGGGCGCAGACGCGCTCGTGTTCGGCTCAGTGGGACCT   2479
98       R  E  A  V  G  A  D  A  L  V  F  G  S  V  G  P      113

2478   GTGGGCAAGCCGCTGGAACCCATCGGCGAAACCCGACTGGACGAAGCCGA   2429
114    V  G  K  P  L  E  P  I  G  E  T  R  L  D  E  A  E    130

2428   GGGGGCGTTCCGCGAGCAGATGCAGGCGCTCCTTGAGGCGGGCGTGGACG   2379
131     G  A  F  R  E  Q  M  Q  A  L  L  E  A  G  V  D  G    147

2378   GACTGATTCTGGAAACTTTCATAGACCCCCGCGAACTGGAGCTGGCGATT   2329
148      L  I  L  E  T  F  I  D  P  R  E  L  E  L  A  I      163

2328   CGGGTCGCTCGCGAGCTCGCCCCTGATCTGCCCCTGATTGCCTCCAAAGG   2279
164    R  V  A  R  E  L  A  P  D  L  P  L  I  A  S  K  G    180

2278   CTTTGTGGAAGACGGCGAGACGCTCATGGAAGGGCTACCCGAACGCTTCG   2229
181     F  V  E  D  G  E  T  L  M  E  G  L  P  E  R  F  A    197

2228   CGCACACCGTCAGCGCGCTGGGTGTGGACGCCGTCGGAGCGAACTGCGTG   2179
198      H  T  V  S  A  L  G  V  D  A  V  G  A  N  C  V      213

2178   GTGGGACCTCAGCGCATGCTGGACATCGTGCGCATGATGGCAACGGGCAC   2129
214    V  G  P  Q  R  M  L  D  I  V  R  M  M  A  T  G  T    230

2128   CGAATTGCCCCTTTCCTCGATGCCGACGCCCGGCTTGCCGCAACTGGTGC   2079
231    E  L  P  L  S  S  M  P  T  P  G  L  P  Q  L  V  R    247

2078   GCGGGCAGGTGGTGTACGACATTCACCCCGACTATTTCGGGCGGTATGCA   2029
248      G  Q  V  V  Y  D  I  H  P  D  Y  F  G  R  Y  A      263

2028   GTGCGCCTTGTGGAGGCAGGCGCCCAGATTGTGGGCGGCTGCTGCGGTAC   1979
264    V  R  L  V  E  A  G  A  Q  I  V  G  G  C  C  G  T    280

1978   GACGCCCGACCATATCCGCGCGGTGGCGCAGGCAGTATCCCGAACGCCTG   1929
281     T  P  D  H  I  R  A  V  A  Q  A  V  S  R  T  P  V    297

1928   TCAAGCGCAGGGCAGGGGGCATCCGTGCAGTGGTGCGTGAGCGCAAGGAA   1879
298      K  R  R  A  G  G  I  R  A  V  V  R  E  R  K  E      313

1878   GAAGAGCTGCCCCTTGCGGAGCCGTCGCGCCTCTCTCAGATACTGGGCAA   1829
314    E  E  L  P  L  A  E  P  S  R  L  S  Q  I  L  G  K    330

1828   AGAGCGTGTAATCGCAGTGGAATTAGACCTGCCCCGCGGCTTGAAAGTGC   1779
331     E  R  V  I  A  V  E  L  D  L  P  R  G  L  K  V  Q    347

1778   AAAAGGTTATCGAAGGCGCACGCCTGCTGAAAGAGCATGGGGTGCATGTG   1729
348      K  V  I  E  G  A  R  L  L  K  E  H  G  V  H  V      363
```

Figure 6.3: Continued

Figure 6.3: p5*orf3* continued

```
1728   ATTGACATCTCTGACGGCGCGCGCGCCCGCCTGCGCATGAATGTGATTGC   1679
364    I  D  I  S  D  G  A  R  A  R  L  R  M  N  V  I  A    380

1678   CATCTCCCATCTGGTACAGCGCGAGGCGGGGATTGAGGTGATGATGCACT   1629
381     I  S  H  L  V  Q  R  E  A  G  I  E  V  M  M  H  F    397

1628   TCGCCTGTCGCGACCGCAACCTGCTGGCAATTCAGGCAGATTTGTTAGGC   1579
398      A  C  R  D  R  N  L  L  A  I  Q  A  D  L  L  G     413

1578   GCGCACGCCCTGGGCATCCGCAATGTGCTGGCAATCACGGGCGACCCCGC   1529
414    A  H  A  L  G  I  R  N  V  L  A  I  T  G  D  P  A    430

1528   CCAGATTGGCGACTACCCCACTGCCACCAGCGTGTTCGATGTGGACGCGA   1479
431     Q  I  G  D  Y  P  T  A  T  S  V  F  D  V  D  A  I    447

1478   TCGGGCTGGTGCGCATCCTGCGGCGCTTCAACGAGGGGCGCGACCTCGCA   1429
448      G  L  V  R  I  L  R  R  F  N  E  G  R  D  L  A     463

1428   GGCAACACGATTGGCGTGCGGGCGAACTTCACGATTGCTGTCGCCTACAA   1379
464    G  N  T  I  G  V  R  A  N  F  T  I  A  V  A  Y  N    480

1378   CCCGCTCGCGCCCGACCCCGAAACCGAGCGCGACCGCCTGCGCAAAAAGA   1329
481     P  L  A  P  D  P  E  T  E  R  D  R  L  R  K  K  I    497

1328   TAGAGGAGGGCGCGCATCTGGTCTACACCCAGCCGATTTTTGAGATGCGC   1279
498      E  E  G  A  H  L  V  Y  T  Q  P  I  F  E  M  R     513

1278   GTCGTGGAAGAGACCGCCGAGTTGATGAACCGACTCGGCGTGCCGTGGCT   1229
514    V  V  E  E  T  A  E  L  M  N  R  L  G  V  P  W  L    530

1228   GGTGGGGGTGCTACCGCTGCGAAGCGCACGCCACGCCGAGTTTATGCATA   1179
531     V  G  V  L  P  L  R  S  A  R  H  A  E  F  M  H  N    547

1178   ATGAGGTGCCGGGCGTGTCCATTCCCGAGCCGATTTTGCGGCGGATGGCG   1129
548      E  V  P  G  V  S  I  P  E  P  I  L  R  R  M  A     563

1128   GAGGCGCCCGAGGAGGACGCGCTCGCCGTGGGGCTGGAGATCGCCCGTCG   1079
564    E  A  P  E  E  D  A  L  A  V  G  L  E  I  A  R  R    580

1078   ATTCGTGGCTGAGGCGGCGCCCTATGCGCAGGGGGTCTACCTGATGCCAC   1029
581     F  V  A  E  A  A  P  Y  A  Q  G  V  Y  L  M  P  P    597

1028   CTGCAGGCAGCGCCCAGATTGCCCTGCAGGTGATTGAGGCAATACGGTAA    979
598      A  G  S  A  Q  I  A  L  Q  V  I  E  A  I  R  *     712

978    TCCCTTCTCACTGCTTCGGAGTCGTCACCGCTACTTCAGGCGCTTCCACC    929
```

Figure 6.3: Nucleotide sequence of the region of p5ICE16 that contains the nt and deduced aa sequence of p5*orf3*. The DNA sequence shown is complementary to the strand encoding p5*orf2*. The nucleotide sequence has been numbered with respect to the orientation of the insert. Base pair 1 of the insert (not shown) is positioned closest to the plasmid-encoded *lac* promoter. The putative –10 promoter element (P3) and ribosome-binding site upstream of p5*orf3* are indicated by an arrow and a double-underline, respectively. The stop codon is indicated by an asterisk.

6.3.4: The amino acid sequence deduced from p5*orf3* possesses homologies to proteins involved in methionine biosynthesis

BLAST searches using the 612-aa sequence derived from p5*orf3* revealed that *B. subtilis* possesses a gene (*yit*J) whose derived product (Accession CAA70665) shares 36% sequence identity with p5*orf3* gene product. Although the function of *Yit*J is unknown, transcript analysis of *yit*J upstream regulatory sequences revealed that *yit*J forms part of the S Box regulon for methionine and cysteine biosynthetic genes in *B subtilis* (Grundy & Henkin, 1998). The N-terminal half of p5*orf3* was also shown to be 33% identical to the N-terminal third of a putative cobalamin-dependent methionine synthase (*Met*H) from *Thermotoga maritima* (Nelson *et al.*, 1999) and 27% identical to the N-terminal third of the *Met*H proteins isolated and characterised from *E. coli* (Banerjee *et al.*, 1989; Old *et al.*, 1990; Blattner *et al.*, 1993) and *Homo sapiens* (Leclerc *et al.*, 1996; Li *et al.*, 1996; Chen *et al.*, 1997). The N-terminal half of the p5*orf3* protein was also shown to be 30% identical to the 312-aa homocysteine *S*-methyltransferase (*Mmu*M), also a methionine synthase, isolated from *E. coli* (Neuhierl *et al.*, 1999). Hypothetical proteins from *Pyrococcus horikoshi* (Accession BAA30190) and *Pyrococcus abyssi* (Accession CAB49818) were the only archaeal sequences shown to possess any homology (~26% identity over ~300 aa) to the N-terminus of p5*orf3* gene product.

Homologies of p5*orf3* gene product with *Met*H, *Mmu*M and *Yit*J suggested that it may be involved in the synthesis of methionine. Based on this evidence, p5*orf3* was tentatively identified as coding for a methionine biosynthetic enzyme and is hereinafter referred to as p5*met*X.

While the N-terminal half of p5*Met*X was shown to be homologous to *Met*H proteins, amino acid sequence alignments indicated that the C-terminal half of p5*Met*X has no homology to cobalamin-dependent methionine synthase (results not shown). BLAST searches however revealed that the C-terminal region of p5*Met*X is ~30% identical to hypothetical proteins from *T. maritima* (Nelson *et al.*, 1999) and *Synechosystis* sp. (Kaneko *et al.*,1996). These three proteins, in turn, possess homologies to the central region of methylenetetrahydrofolate reductases (MTHFRs) from *H. sapiens* (31% identity over 111 aa for p5*Met*X) (Goyette *et al.*, 1994) and *E. coli* (22% identity over 132 aa for p5*Met*X) (Saint-Girons *et al.*, 1983; Plunkett *et al.*, 1993; Guenther *et al.*,

1999). No archaeal sequences were shown to possess homologies to the C-terminus of p5*orf3* gene product.

In *E. coli*, methylenetetrahydrofolate reductase (*Met*F) forms part of the folate branch of the methionine biosynthetic pathway (Saint-Girons *et al.*, 1988). Homology with *Met*F provided further evidence that p5*Met*X may be involved in methionine synthesis.

6.3.4.1: The N-terminal half of p5*Met*X contains homologous sequences for homocysteine binding in *E. coli Met*H and *Mmu*M

In *E. coli*, *Met*H, *Mmu*M and *Met*F are each involved in the biosynthesis of methionine (Figure 6.4). In the previous section, BLAST searches revealed that p5*Met*X possesses regions of homology with each of these proteins suggesting that p5*Met*X may also be involved the synthesis of methionine. Multiple sequence alignments of p5*Met*X against these proteins was performed to see if any determinants involved in the methionine biosynthetic pathway could be identified in p5*Met*X.

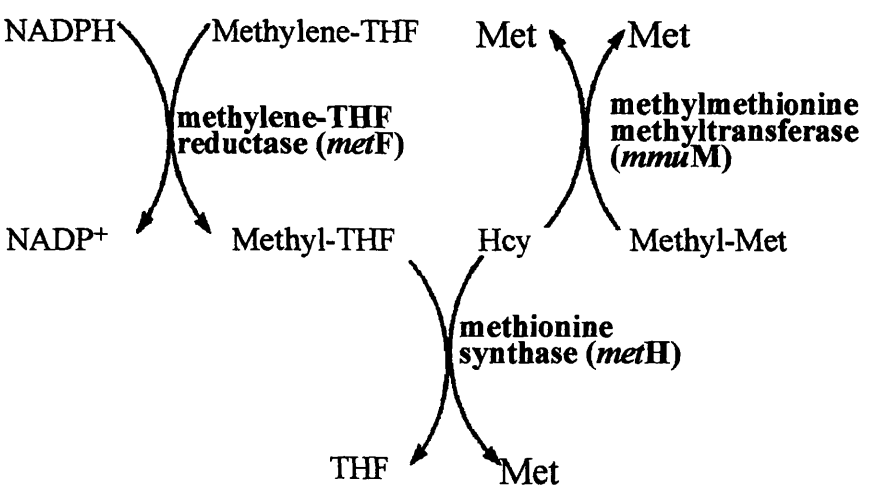


Figure 6.4: Methionine (Met) biosynthetic pathways from homocysteine (Hcy) and methylenetetrahydrofolate (methylene-THF) or S-methylmethionine (Methyl-Met). The *E. coli* genes encoding the proteins possessing regions of homology to p5*met*X gene product are indicated in parentheses. Adapted from Matthews *et al.* (1998) and Thanbichler *et al.* (1999).

In *E. coli*, *Met*H catalyses the transfer of a methyl group from methyltetrahydrofolate (Methyl-THF) to enzyme-bound cob(I)alamin then to homocysteine to produce tetrahydrofolate (THF) and methionine (Banerjee *et al.*, 1990a). Occasionally, the cobalamin cofactor is oxidised to inactive cob(II)alamin (Frasca *et al.*, 1988; Drummond *et al.*, 1993) The enzyme reductively methylates cob(II)alamin using reduced flavodoxin (Fujii *et al.*, 1977) and S-adenosylmethionine (Ado-Met) as electron and methyl donor, respectively (Banerjee *et al.*, 1990b). *E. coli Met*H has been characterised as a modular protein with four distinct regions for binding homocysteine (Goulding *et al.*, 1997), Methyl-THF (Roberts *et al.*, 1994), cobalamin cofactor (Banerjee *et al.*, 1989) and Ado-Met (Drummond *et al.*, 1993), respectively (Figure 6.5).

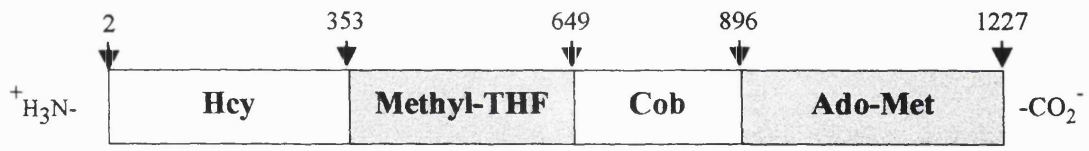


Figure 6.5: Schematic showing the four distinct regions of the mature *E. coli Met*H protein. Hcy, homocysteine-binding and methylation module (aa 2-353); Methyl-THF, methyltetrahydrofolate-binding module (aa 354-649); Cob, cobalamin cofactor-binding module (aa 650-896); Ado-Met, *S*-adenosylmethionine-binding module (aa 897-1227). The N-terminal half of p5*Met*X displayed homology to the Hcy-binding and methylation region only. Relative locations of amino acid residues are indicated by arrows.

A truncated peptide of the *E. coli Met*H protein, consisting of amino acid residues 2-353, is able to catalyse the methyl transfer from exogenous methylcobalamin to homocysteine indicating that it must contain the homocysteine binding and methylation site(s) (Goulding *et al.*, 1997). This homocysteine-binding module of *Met*H has a conserved motif (GGCCGTXPXHI) found in all cobalamin-dependent methionine synthases and which has been implicated in homocysteine binding and activation (Goulding *et al.*, 1997). A very similar sequence (GGCCRTXPXDI) is also found in *Mmu*M which is a methionine synthase that methylates homocysteine using methylmethionine as methyl donor (Neuhierl *et al.*, 1999). In *E. coli Met*H, Cys310 and Cys311 within this motif, as well as Cys247 have been identified as putative ligands for the zinc cofactor involved in homocysteine activation (Goulding & Matthews, 1997; Peariso *et al.*, 1998).

A multiple sequence alignment of the N-terminal half of p5*Met*X against the N-terminal amino acids of *Yit*J, as well as the amino acid sequence of the homocysteine-binding module of (putative) *Met*Hs from *T. maritima*, *E. coli* and *Homo sapiens* and the full-length *Mmu*M from *E. coli* is shown in Figure 6.6. This alignment demonstrated that the N-terminal half of p5*Met*X possesses the homocysteine-binding motif as well as each of the three highly conserved cysteines found in zinc-dependent methyltransferases. These findings indicated that p5*Met*X may possess zinc-dependent methyltransferase activity using a catalytic mechanism similar to that of the homocysteine-binding module of *Met*H. Database searches indicated that p5*Met*X possesses no residues displaying homology to the other modules of *Met*H, suggesting that p5*Met*X contains no determinants for binding either Methyl-THF, cobalamin cofactor or Ado-Met as characterised in *Met*H.

```
p5METX     1 ----------------------------VLIADGAIGTMLALR--------G---VPTP
BSUBT      1 -------------------MGLLEDLQRQVLIGDGAMGTLLYSY--------G--IDRCF
TMARIT     1 MRNRREVS---------------KLLSERVLLLDGAYGTEFMKY--------G--YDDLP
ECOLI      1 MSSKVEQLR--------------AQLNERILVLDGGMGTMIQSYRLNEADFRGERFADWP
HUMAN      1 MSPALQDLSQPEGLKKTLRDEINAILQKRIMVLDGGMGTMIQREKLNEEHFRGQEFKDHA
MMUM_EC    1 MSQNNPLRALLDKQD----------------ILLLDGAMATELEA------RG---CNLA

p5METX    21 YELAN------LLYPDTVRALHREYYEAGARLIETNTYTANRVRLFNLPERGSEAPPTYS
BSUBT     32 EELN-------ISKPEEIQRIHKAYVEAGANIIQTNTYGANYIKLSRH---GLEDDIKKM
TMARIT    36 EELN-------IKAPDVVLKVHRSYIESGSDVILTNTFGATRMKLRKH---GLEDKLDPI
ECOLI     47 CDLKGNNDLLVLSKPEVIAAIHNAYFEAGADIIETNTFNSTTIAMADY---QMESLSAEI
HUMAN     61 RPLKGNNDILSITQPDVIYQIHKEYLLAGADIIETNTFSSTSIAQADY---GLEHLAYRM
MMUM_EC   36 DSLWSAKVL--VENPELIREVHLDYYRAGAQCAITASYQATPAGFAAR---GLDEAQSKA

p5METX    75 LLEQFGSPEELVRRINQEAVRLAREAVGADALVFGSVGPVG--------KPLEPIGETRL
BSUBT     82 NQEAVK----IARASAGD------------AYVLGTMGGIR--------TFNK--NAYSL
TMARIT    86 VRNAVR----IARRAAGE-------K-----LVFGDIGPTG-------ELPYP-LGSTLF
ECOLI    104 NFAAAK----LARRCADEWTARTPEKP---RYVAGVLGPTNRTASISPDVNDPAFRNITF
HUMAN    118 NMCSAG----VARKAAEEVTLQTGIK----RFVAGALGPTNKTLSVSPSVERPDYRNITF
MMUM_EC   91 LIGKS---VELARKAREAYLAENPQAGT--LLVAGSVGPYGAYLADGSEYRGDYHC--SV

p5METX   127 DEAEGAFREQMQALLEAGVDGLILETFIDPRELELAIRVAR-----ELAPDLPLIASKGF
BSUBT    116 DEIKRSFREQLYLLLHEEPDGLLLETYYDLEEAREVLKIAR-----K-ETDLPIMLNVSM
TMARIT   122 EEFYENFRETVEIMVEEGVDGIIFETFSDILELKAAVLAAR-----EVSRDVFLIAHMTF
ECOLI    157 DGLVAAYRESTKALVEGGADLILIETVFDTLNAKAAVFAVKTEFEALGVELPIMISGTIT
HUMAN    170 DELVEAYQEQAKGLLDGGVDILLIETIFDTANAKAALFALQNLFEEKYAPRPIFISGTIV
MMUM_EC  144 EAFQAFHRPRVEALLDAGADLLACETLPNFSEIEALAELLT-----AYPRARAWFSFTLR

p5METX   182 VEDGETLMEGLPERFAH-TVSALGVDAVGANCVVGPQRMLDIVRMMATGTELPLSSMPTP
BSUBT    170 HEQG-VLQDGTPLSDALRSIADLGADIVGINCRLGPYHMIEALSEVPIFDDVFLSVYPNS
TMARIT   177 DEKGRSLTGTDPANFAI-TFDELDIDALGINCSLGPEEILPIFQELSQYTDKFLVVEPNA
ECOLI    217 DASGRTLSGQTTEAFYN-SLRHAEALTFGLNCALGPDELRQYVQELSRIAECYVTAHPNA
HUMAN    230 DKSGRTLSGQTGEGFVI-SVSHGEPLCIGLNCALGAAEMRPFIEIIGKCTTAYVLCYPNA
MMUM_EC  199 DSEHLSDGTPLRDVVAL-LAGYPQVVALGINCIALEN-TTAALQHLHGLTVLPLVVYPNS

p5METX   241 GLPQL--VRGQVVYDIHPDYFGRYAVRLVEAG-AQIVGGCCGTTPDHIRAVAQAVS-RTP
BSUBT    229 SLPSL--EEGRLVYETDDTYFQNSASEFRKQG-ARIIGGCCGTTPNHIRAMAEAVGGLAP
TMARIT   236 GKPIV--ENGKTVYPLKPHDFAVHIDSYYELG-VNIFGGCCGTTPEHVKLFRKVLGNRKP
ECOLI    276 GLPNA---FG--EYDLDADTMAKQIREWAQAGFLNIVGGCCGTTPQHIAAMSRAVEGLAP
HUMAN    289 GLPNT---FG--DYDETPSMMAKHLKDFAMDGLVNIVGGCCGSTPDHIREIAEAVKNCKP
MMUM_EC  257 GEHYDAVSKTWHHHGEHCAQLADYLPQWQAAG-ARLIGGCCRTTPADIAALKARS 310

p5METX   299 RRAGGIRAVVRERKEEELPLAEPSRLSQI327→612
BSUBT    288 EKEVKTRAKEFIS--VHHERTEPG-LDEI313→612
TMARIT   293 LQRKKKR---------IFAVSSPSKLVTF312→768
ECOLI    333 LPEIPVA--C-------RLSGLEPLNIGED353→1227
HUMAN    346 PPATAFEGHM------LLSGLEPFRIGPY368→1265
MMUM_EC
```

Figure 6.6: Comparison of the deduced N-terminal amino acid sequence of p5*orf3* (p5METX) from the 2.9-kb p5ICE16 insert against the N-terminal half of *Yit*J from *B. subtilis* (BSUBT) (Accession CAA70665), the homocysteine-binding region of (putative) cobalamin-dependent methionine synthases from *T. maritima* (TMARIT) (Accession AAD35357), *E. coli* (ECOLI) (Accession P13009) and *H. sapiens* (HUMAN) (Accession Q99707) and the full-length homocysteine *S*-methyltransferase from *E. coli* (MMUM_EC) (Accession Q47690). Red residues indicate conservation in at least three of the proteins while the blue residues represent a conservative substitution. The motif implicated in homocysteine binding and which is found in zinc-dependent methyltransferases is indicated by a green bar. The cysteine residues identified as putative ligands of the zinc cofactor in *E. coli Met*H are highlighted in yellow.

6.3.4.2: The C-terminal half of p5*Met*X aligned with the Flavin Adenine Dinucleotide- (FAD)-binding region in *E. coli Met*F

E. coli methylenetetrahydrofolate reductase (*Met*F) is a 296-aa flavoprotein that catalyses the transfer of reducing equivalents from NADH to methylenetetrahydrofolate (methylene-THF) to form NAD^+ and methyl-THF (Hatch *et al.*, 1961; Sheppard *et al.*, 1999). In *E. coli*, this reaction is the penultimate step in *de novo* methionine biosynthesis and provides Methyl-THF as substrate for *Met*H (Saint-Girons *et al.*, 1988).

The X-ray structure of *E. coli Met*F has been determined and the regions involved in the noncovalent binding of FAD (aa 60-183) and interaction with methylene-THF (aa 271-290) have been assigned (Guenther *et al.*, 1999). The region(s) of *Met*F that interacts with NADH are not known.

A multiple sequence alignment using the C-terminal half of p5*Met*X against the C-terminal half of *Yit*J from *B. subtilis* and the full-length hypothetical proteins from *T maritima* and *Synechosystis* sp., as well as the C-terminal region of *Met*F proteins from *Homo sapiens* and *E. coli* is shown in Figure 6.7. The C-terminal half of p5*Met*X aligned with the region in the *E. coli Met*F that is involved in binding FAD. This suggests that p5*Met*X may be a flavoprotein. Furthermore, p5*Met*X has the conserved alanine that when mutated (A177V; *E. coli* numbering) results in the propensity of the *E. coli* enzyme to lose its flavin cofactor (Guenther *et al.*, 1999). p5*Met*X did not align well with the C-terminal half of *E. coli Met*F (results not shown) suggesting that p5*Met*X lacks the determinant(s) for binding methylene-THF as characterised in *Met*F.

Based on the comparisons with *E. coli Met*H, *Mmu*M and *Met*F, it was proposed that p5*met*X may encode a zinc-dependent methyltransferase with novel methionine synthase activity that uses noncovalently bound FAD as cofactor and uses homocysteine and, perhaps, methylmethionine but not methyl(ene)-THF as reactants. It will be of interest to further examine p5*met*X and identify the biological role of this putative methionine synthase gene.

```
p5METX  230 TELPLSSMPTPGLPQLVRGQVVYDIHPDYFGRYAVRLVEAGAQIVGGCCGTTPDHIRAVA
BSUBT   218 DDVFLSVYPNSSLPSLEEGRLVYETDDTYFQNSASEFRKQGARIIGGCCGTTPNHIRAMA
TMARIT    1 ------------------------------------------------------------
SYNECH    1 ---------------------------------------METGPQVWAMTKPANRKNSVKIS
HUMAN     1 ------------------------------------------------------------
ECOLI     1 ------------------------------------------------------------

p5METX  290 QAVS-RTPVKRRAGGIRAVVRERKEEELPLAEPSRLSQILGKERVIA-VELDLPRGLKVQ
BSUBT   278 EAVGGLAPITEKEVKTRAKEFISVHHERTEP---GLDEIAAKKRSII-VELDPPKKLSFD
TMARIT    1 --------------------------------MRKSLEIGKCVVL---EVLTPRGTNLG
SYNECH   24 CQR--------GSHRSQLKPPLSPLDVMTIS---RFRQAAEAKEFLITAEVTPPKGGNPE
HUMAN     1 KDS--------SRCSTPGLDPERHERLREKM---RRRLESGDKWFS--LEFFPPRTAEGA
ECOLI     1 -----------------MSFFHASQRDALN---QSLAEVQGQINVS-FEFFPPRTSEME

p5METX  348 KVIEGARLLKEHG-------VHVIDISLGARARL----RMNVIAISHLVQREAGIEVMMH
BSUBT   334 KFLSAAAELKEAG-------IDALTLADNSLATP----RISNVACGALVKQQLDMRSLVH
TMARIT   25 RFMEFTEEAWKTG-------IDAFTVTDMPMGRV----RMAPWAASHLLV-ESGKDVLMH
SYNECH   73 RMLAVAAKLRGRV--------HGVNVTDGSRAVL----RMSSIAACVLLQ-QRGIEAICQ
HUMAN    75 VNLISRFDRMAAG------GPLYIDVTWHPAGDPGSDKETSSMMIASTAVNYCGLETILH
ECOLI    39 QTLWNSIDRLSSLK------PKFVSVTYGANSGE----RDRTHSIIKGIKDRTGLEAAPH

p5METX  397 FACRDRNLLAIQADLLGAHALGIRNVLAITGDPAQIGDYPTATSVFDVDAIGLVRILRRF
BSUBT   383 ITCRDRNIIGLQSHLMGLDTLGLNDVLAITGDPSKIGDFPGATSVYDLTSFDLIRLIKQF
TMARIT   73 FTRNTRNMIRIQSDLLGCHALGVKNLLLLSGDNPSHGDYPETTSVNDIDILDLIRLTKLM
SYNECH  120 FTRNTRNMIRIQSDLLGCHALGVKNLLLLSGDNPSHGDYPETTSVNDIDILDLIRLTKLM
HUMAN   129 MTCCRQRLEEITGHLHKAKQLGLKNIMALRGDPI--GDQWEEEEGGFNYAVGLVKHIRSE
ECOLI    89 LTCIDATPDELRTIARDYWNNGIRHIVALRGDLP----PGSGKP--EMYASDLVTLLKEV

p5METX  457 NEGRDLAGNT-IGVRANFTIAVAYNPLAPDPETERDRLRKKIEEGAHLVYTQPIFEMR-V
BSUBT   443 NEGLSLSGKP-LGKKTNFSVAAAFNPNVRHLDKAVKRLEKKIDCGADYFVSQPVYSEQ-Q
TMARIT  133 NEGTDLAGNK-IYGKTDFFVGGALNP-FSDKD--IKRAKQKIEAGVDFLVTQPLFNSE-V
SYNECH  180 NQGLDFNQAPLVDGRLDLFPGAAVDPQLKSWSGLQSRFERKLTAGAQFFQSQLITDFD-R
HUMAN   187 F-GDYFDICV-AGYPKG-------HPEAGSFEADLKHLKEKVSAGADFIITQLFFEADTF
ECOLI   143 A---DFDISV-AAYPEV-------HPEAKSAQADLLNLKRKVDAGANRAITQFFFDVESY

p5METX  515 VEETAELMNR-LGVPWLVGVLPLRSARHAEFMHNEVPGVSIPEPILRRMAEAPEEDAL--
BSUBT   501 LVDIHNETKH-LKTPIYIGIMPLTSSRNAEFIHNEIPGIKLSDTIREKMAHAGEDKEKQK
TMARIT  188 AKKIKEELN----TKILASIVFFENAKQMSYFS-SVPGIEIPDEIVNSTEKG-DDYLK--
SYNECH  239 LDKFMTQVAVGCGKPILAGIFLFKSAKNATFIKRVVPGVNIPDALIDRLARAEDPLK---
HUMAN   238→341
ECOLI   192→290

p5METX  572 AVGLEIARRFVAEAAPYAQGVYLMPPAGSAQIALQVIEAIR 612
BSUBT   560 AEGLAIARSLLDTACELFNGIYLITPFLRSDLTAELTSYIQQKDEQRQNIFLH 612
TMARIT  240 EKSFESVLRFVEETKGVLDGFYIVAIVKDLEKIRKVVEIAKS 281
SYNECH  296 -EGVAIAAEQVKLAKELCQGVHLMAIKKE-ELIPEILDQAGIAPLEN 340
HUMAN
ECOLI
```

Figure 6.7: Comparison of the deduced C-terminal amino acid sequence of p5*orf*3 (p5METX) from the 2.9-kb p5ICE16 insert against the C-terminal half of *Yit*J from *B. subtilis* (BACSU) (Accession CAA70665), and full-length hypothetical proteins from *T. maritima* (TMARIT) (Accession AAD3559) and *Synechocystis* sp. (SYNECH) (Accession BAA17723). These proteins also show homology to the central region of Methylenetetrahydrofolate reductases from *H. sapiens* (HUMAN) and *E. coli* (ECOLI) (Accessions S461154 and P00394, respectively). Red residues indicate conservation in at least three of the proteins while the blue residues represent a conservative substitution. The region of the *E. coli Met*F that is involved in noncovalent binding of FAD is indicated by double-underlines. The alanine residue that, when mutated to valine, results in the propensity of the *E. coli* enzyme to lose its FAD cofactor is highlighted in yellow. As a reference, the homocytsteine-binding motif identified in the N-terminal half of p5*Met* is indicated by a green bar.

6.3.5: Prospects for heterologous expression of p5ICE16-insert sequences in *E. coli* TOP10 cells

The orientation of p5*orf*1-vector fusion and p5*orf*3 (p5*met*X), with respect to the vector-encoded *lac* promoter, indicated that heterologous expression would only be permitted *via* their own promoter(s). No bacterial promoter elements, could be identified in p5*orf*1-upstream sequences. Expression of the p5*orf*1-vector fusion from insert-encoded sequences was, therefore, thought unlikely in *E. coli* TOP10 cells.

Putative bacterial –10 promoter element and ribosome-binding sequences were assigned to p5*met*X (p5*orf*3) (Figure 6.3); however, because any putative –35 element was truncated during the cloning procedure, p5*met*X could not have be expressed as cloned in *E. coli* TOP10 host cells.

Bacterial promoter-like sequences were identified upstream of p5*orf*2. Assuming that *E. coli* RNA polymerase holoenzyme was able to recognise these promoter sequences, transcription of p5*orf*2 from insert-encoded sequences was possible. The presence of a moderate SD sequence 9 bp upstream of the assigned ATG start site suggested that translation of p5*orf*2 was also possible. Because the biological role of p5*orf*2 gene product could not be inferred from its deduced amino acid sequence, the activity of p5O*rf*2 could not be employed as a means to characterising the clone. To see if a protein corresponding to p5*Orf*2 was produced in *E. coli* p5ICE16, SDS-PAGE was performed on crude cell extracts of the clone (Section 2.7.5). No protein band of the expected size (28.9 kDa) was observed for *E. coli* TOP10/p5ICE16. Indeed, no protein band could be attributed to the heterologous expression of the cloned DNA in *E. coli* TOP10 cells (results not shown). Despite the fact that cloned Ice16-DNA was shown to confer the dark phenotype to *E. coli* host cells (Section 5.4.2), no gene or protein encoded by the DNA insert of p5ICE16 could be identified as the determinant responsible for the observed phenotype.

6.4: p6ICE16

Recombinant plasmid p6ICE16, which conferred a dark phenotype to *E. coli* TOP10 cells, was obtained from environmental DNA library ICE16 (Section 5.4.2). The DNA used to construct ICE16 was derived from a biomass-containing sediment; ~70°C, pH 9.5 (Section 4.7).

6.4.1: p6ICE16 contains a 4.7-kb insert

The physical map of the p6ICE16-DNA insert is shown in Figure 6.8. Five open reading frames and a putative tRNA gene were identified within this 4686-bp fragment.

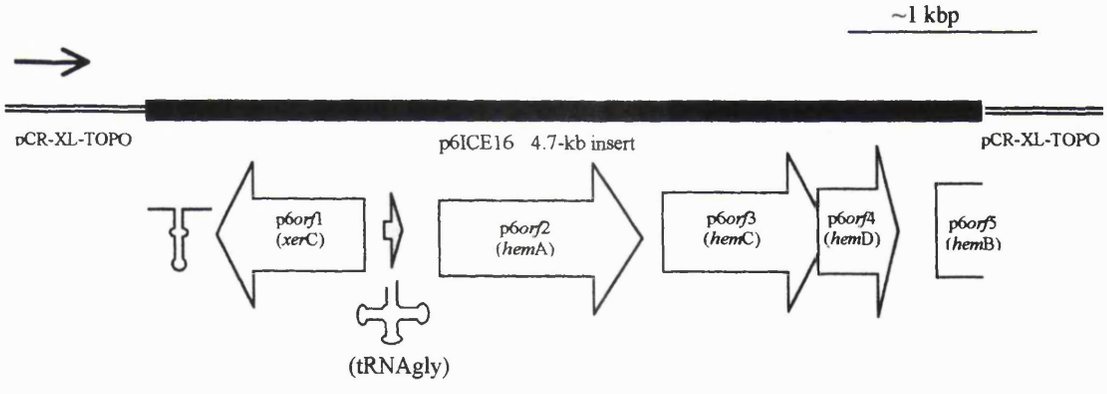


Figure 6.8: Physical map of the 4.7-kb Iceland16 DNA fragment (black bar) cloned into the TA vector, pCR-XL-TOPO (double line) and recovered from *E. coli* TOP10/p6ICE16. Block arrows show the relative localisation of each gene and the orientation of coding sequences found. Open reading frames (p6*orf*1 to p6*orf*5) are represented by white block arrows. p6*orf*5 is truncated at the cloning site as shown by a block with no arrowhead. The shaded block arrow upstream of p6*orf*2 represents the gene coding for a tRNA. The symbol downstream of p6*orf*1 indicates the presence of a palindromic structure. The genes encoding homologous sequences are shown in parentheses. The black arrow indicates the direction of transcription from the *lac* promoter.

The p6ICE16-DNA fragment appeared to be of bacterial origin, given the homology of potential genes on the insert to sequences found in the databases.

6.4.2: p6*orf*1 encodes a probable *xer* site-specific recombinase

p6*orf*1, which is positioned in an orientation opposite to that of the other putative coding sequences, encodes a protein of 312 amino acids, corresponding to a molecular mass of 35.9 kDa.. This ORF has a possible SD sequence, 1412AGGAGGT1406, located 12 nucleotides upstream of an assumed ATG start codon. Downstream of the TGA stop codon at bp 463 is a series of imperfect direct and inverted repeats (Figure 6.9).

BLAST searches using the amino acid sequence as predicted for p6*orf*1 indicated high homology to known bacterial proteins involved in the integration/recombination process. The highest homologies found were with the *xer*C gene product of *Haemophilus influenzae* (27% identity over 288 aa) (Fleischmann *et al.*,1995) and with the putative integrase-recombinase gene product of the archaeon *Methanobacterium thermoautotrophicum* (41% identity over 159 aa) (Smith *et al.*,1997). These proteins also have similarity to *Xer*D of *E. coli* (Subramanya *et al.*, 1997).

In *E. coli, xer* site-specific recombination occurs at specific DNA sequences and requires the activity of two related recombinases, *Xer*C and *Xer*D, each of which catalyses a specific pair of strand cutting, exchanging and rejoining reactions (Colloms *et al.*, 1990; Blakely *et al.*, 1993, Blakely & Sherratt, 1994). *Xer* proteins are members of the phage integrase family of recombinases. *Xer* recombinases function in the stable inheritance of prokaryotic chromosomes and plasmids containing the target site sequence by converting circular multimers to monomers thereby allowing DNA segregation during cell division (Nash, 1996). Homologues of *Xer*C and *Xer*D have been found in many prokaryotes.

Integrase family proteins possess two motifs which are crucial for the conservation of structure (Abremski & Hoess, 1992; Blakely & Sherrett, 1996). These two motifs contain four completely conserved amino acids: Arg in motif I and Arg, His and Tyr in motif II. In *Xer*C and *Xer*D, the invariant tyrosine residue in motif II acts as a nucleophile during strand cleavage (Pargellis *et al.*, 1988; Parsons *et al.*, 1988; Evans *et al.*, 1990) while the arginine in motif I, and the histidine and arginine in motif II, are

involved in activation of the scissile phosphodiester and phosphotyrosyl during strand cleavage and rejoining, respectively (Parson *et al.*, 1988; Pan & Sadowski, 1992).

Xer family members possess a third motif which contains a highly conserved lysine also implicated in catalysis (Cao *et al.*, 1997; Cao & Hayes, 1999). Multiple sequence alignments of p6*Orf*1 against Xer proteins indicated that p6*Orf*1 possesses each of the five highly conserved amino acid residues found in phage integrase-recombinases (Figure 6.10). Based on this evidence, p6*orf*1 was tentatively identified as coding for a phage integrase-recombinase.

Downstream of the p6*orf*1 stop codon are a series of imperfect direct and inverted repeats (Figure 6.9). The presence of palindromic structures downstream of p6*orf*1 suggested that the RNA transcript would terminate within this region.

Figure 6.9: p6*orf*1

```
1482                                          ATGTCCTGGACGCAAAAACCCG   1461

1460   CATTGACCCCCCTCAAAACTGAGCGTTAGCGCGTGTGAGTTTTTGCCCAG             1411

1410   GAGGTAGCCTATGCCAAAGCCCCTCAAACTGTCCAGACCCACCCCCAACC             1361
1          |p6orf1>M  P  K  P  L  K  L  S  R  P  T  P  N  L            14

1360   TTGACGCCCTGCTCGGGCAATATCTGCGGGAGCGGGAGGCTCACGGCAAA             1311
15       D  A  L  L  G  Q  Y  L  R  E  R  E  A  H  G  K                30

1310   AGCCCCGCTACGGTGCACTGGCAGGAGACCGCCGCGCGGCTGTTGATGGA             1261
31     S  P  A  T  V  H  W  Q  E  T  A  A  R  L  L  M  E                47

1260   GGAGACGGGGGCGCGCGACTTGGCCGAGCTGACCCCGGCCATTTTGCTGG             1211
48       E  T  G  A  R  D  L  A  E  L  T  P  A  I  L  L  D              64

1210   ATTGGTTTGCCGCTAAGCGCTCGCAAGGCGGGCGGGGTCGGCGCATTTCC             1161
65       W  F  A  A  K  R  S  Q  G  G  R  G  R  R  I  S                80

1160   GTCTCCATCAGCACCCTCTCGCAGTACGAGCGGGCGTTGCGGCCCTTCCT             1111
81     V  S  I  S  T  L  S  Q  Y  E  R  A  L  R  P  F  L                97

1110   CAAATGGCTCTATCAGCGCGGCTACACCCAGACCGACCTGTCGCTGGAGC             1061
98       K  W  L  Y  Q  R  G  Y  T  Q  T  D  L  S  L  E  L              114

1060   TGCCCCACTACCGCCCCCCCAAACAGGTTGTCCAACCGTTCACCCCGGAC             1011
115      P  H  Y  R  P  P  K  Q  V  V  Q  P  F  T  P  D                130

1010   GAGCTACGGGCTTTTTTCGCGGCGGCCTCAGAGCCCCCCAACGCCCGGCG              961
131    E  L  R  A  F  F  A  A  A  S  E  P  P  N  A  R  R               147

960    CAAACTGGCCTTGTATCGGCTTCTGCTCGACACCGGCATCCGCCGGGGTG              911
148      K  L  A  L  Y  R  L  L  L  D  T  G  I  R  R  G  E              164

910    AGGCAGTCTCGCTCCAGCTTGATGCGATTTACTGGCGGGAGCGCACCATG              861
165      A  V  S  L  Q  L  D  A  I  Y  W  R  E  R  T  M                180

860    CGGGTGGAGGGGAAGACCGGCGGGCGCATCGTCTATTTCTCCGAGCGCTC              811
181    R  V  E  G  K  T  G  G  R  I  V  Y  F  S  E  R  S               197

810    GCTCAGGGCCATCAACGAGTACCTCAACAACGAGCGCCGTCCCCGTAGGC              761
198      L  R  A  I  N  E  Y  L  N  N  E  R  R  P  R  R  P              214

760    CGGGTGAGGCCACGCTTTTCCTCGACCGCCACGGCGACCCTTTGCGGGCC              711
215      G  E  A  T  L  F  L  D  R  H  G  D  P  L  R  A                230

710    GGGGAGATTACCCAGGAAACCATTCGCATCGCCCGTCGTGCAGGGATTAT              661
231    G  E  I  T  Q  E  T  I  R  I  A  R  R  A  G  I  M               247

660    GCGCGACCACGTGGGGCCGCACACCTTCCGCCACACGTTTGCGGTGGAGT              611
248      R  D  H  V  G  P  H  T  F  R  H  T  F  A  V  E  Y              264

610    ACCTGAAAGCTGGGGGCGACCTGCGGAGCCTACAGGTGTTGCTTGGACAC              561
265      L  K  A  G  G  D  L  R  S  L  Q  V  L  L  G  H                280
```

Figure 6.9: Continued

Figure 6.9: p6*orf*1 continued

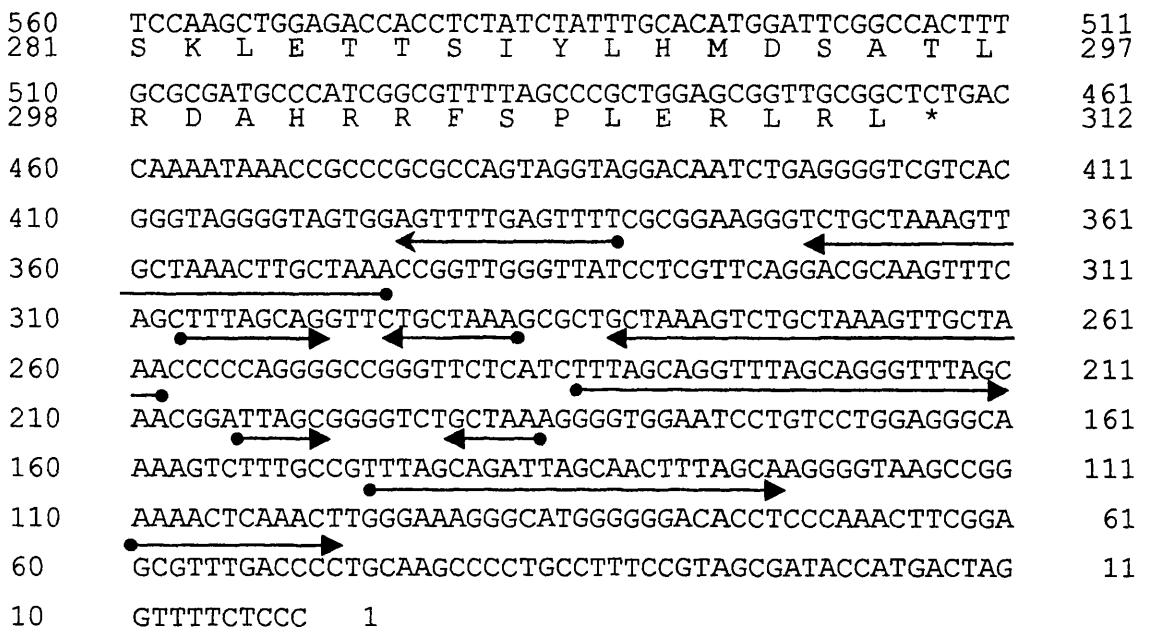


Figure 6.9: Nucleotide and deduced amino acid sequence of the first 1482 bases of p6ICE16 which contains ORF1 (p6*orf*1). p6*orf*1 was subsequently identified as encoding a putative phage integrase-recombinase (Section 6.4.2). The DNA sequence shown is complementary to the strand encoding the tRNA gene and p6*orf*s 2 through 5. The nucleotide sequence has been numbered with respect to the orientation of the insert. Base pair 1 is positioned closest to the plasmid-encoded *lac* promoter. A possible ribosome-binding site for p6*orf*1 is double-underlined and the stop codon is indicated by an asterisk. Imperfect direct and inverted repeated sequences are shown by horizontal arrows.

```
p6ORF1    1 ----MPKPLKLSRPTPNLDALLGQYLRE--------REAHGKSPATVHWQETAARLLMEE
HAEIN     1 -------------MLTALNRYWDYLRIE--------RQMSPHTITNYQHQLDATIKILAQ
METTH     1 MVDTMNMKRESEERQGLLERYNFPELIEDYLIELEIRNYSPNTIKTYKSIVKNFYEFLMN
ExerD     1 ------------QDLARIEQFLDALWLE--------KNLAENTLNAYRRDLSMMVEWLHH
ExerC     1 ----------MTDLHTDVERYLRYLSVE--------RQLSPITLLNYQRQLEAIINFASE

p6ORF1   49 TG-ARDLAELTPAILLDWFAAKRSQGGRGRRISVSISTLSQYERALRPFLKWLYQRGYTQ
HAEIN    40 QD-IHSWTQVTPSVVRFILAESKKQG-------LKEKSLALRLSALRRFLSFLVQQGELK
METTH    61 EDDLYDDRRVLRSFKRYIQYLKRDKK-------VTQNYIYLVTVVVKKFFEFSGIDCLEE
ExerD    41 RGLTLATAQSD--DLQALLAERLEGG-------YKATSSARLLSAVRRLFQYLYREKFRE
ExerC    43 NG-LQSWQQCDVTMVRNFAVRSRRKG-------LGAASLALRLSALRSFFDWLVSQNELK

                                                                      I
p6ORF1  108 T-DLSLELPHYRPPKQVVQPFTPDELRAFFAAASE-PPNARRKLALYRLLLDTGIRRGEA
HAEIN    92 V-NPATGISAPKQGRHLPKNMDGEQVQQLLANDSK-EPIDIRDRAILELMYSSGLRLSEL
METTH   114 VKAPKRTKSLPKSLNEDEVKSLINAVEVADDGSVIRRFIKTRDRLILSLLYSSGLRVSEL
ExerD    92 D-DPSAHLASPKLPQRLPKDLSEAQVERLLQAPLIDQPLELRDKAMLEVLYATGLRVSEL
ExerC    95 A-NPAKGVSAPKAPRHLPKNIDVDDMNRLLDIDIN-DPLAVRDRAMLEVMYGAGLRLSEL

                                  III
p6ORF1  166 VSLQLDAIYWRERTMRVEGKTGG-RIVYFSERSLRAINEYLNNERRPRRPGE---ATLFL
HAEIN   150 QGLDLNSINTRVREVRVIGKGNKERVVPFGRYASHAIQEWLKV-R-ALFNPKD--EALFV
METTH   174 VSLRINDIDPDERTIRIRGKGDKDRIVLFDENTRDLLMDYLKR-R---IHES---EYLFL
ExerD   151 VGLTMSDISLRQGVVRVIGKGNKERLVPLGEEAVYWLETYLEHGRPWLLNGV-SIDVLFP
ExerC   153 VGLDIKHLDLESGEVWVMGKGSKERRLPIGRKPVAWIEHWLDL-R-DLFGSED--DALFL

                                                          II
p6ORF1  222 DRHGDPLRAGEITQETIRIARRAGIMRDHVGPHTFRHTFAVEYLKAGGDLRSLQVLLGHS
HAEIN   206 SQLGNRISHRAIQKRLETWGIRQGLNSH-LNPHKLRHSFATHMLEASSDLRAVQELLGHS
METTH   227 NRFGDPLTPRYVQMMIKNYARKAGIKKK-VTPHILRHSFATHLLKNGVDIRAIQQLLGHS
ExerD   210 SQRAQQMTRQTFWHRIKHYAVLAGIDSEKLSPHVLRHAFATHLLNHGADLRVVQMLLGHS
ExerC   209 SKLGKRISARNVQKRFAEWGIKQGLNNH-VHPHKLRHSFATHMLESSGDLRGVQELLGHA

p6ORF1  282 KLETTSIYLHMDSATLRDAHRRFSPLERLRL
HAEIN   265 NLSTTQIYTHLNFQHLAEVYDQAHPRAKRKK
METTH   286 NLSTTQIYTSVDMQTLKNVYDRARLL-----
ExerD   270 DLSTTQIYTHVATERLRQLHQ----------
ExerC   268 NLSTTQIYTHLDFQHLASVYDAAHPRAKRGK
```

Figure 6.10: Comparison of the deduced amino acid sequence of p6*orf*1 against the putative *Xer*C proteins of *H. influenzae* (HAEIN) (Accession P44818) and *M. thermoautotrophicum* (METTH) (Accession AE000865) as well as against the *Xer*D (ExerD) and *Xer*C (ExerC) proteins of *E. coli*. (Accessions P21891 and P22885, respectively) Red residues represent conservation in at least three of the proteins while the blue residues represent a conservative substitution. Motifs I and II, which represent structural conservation in integrase family proteins, and motif III which is conserved in *Xer*C and *Xer*D, are indicated by green lines. The invariant (RRHY) residues found in all phage integrase family members and the highly conserved K found in *Xer* recombinases are highlighted in yellow.

6.4.3: p6*Orf*2, p6*Orf*3, p6*Orf*4 and p6*Orf*5 are homologous to enzymes involved in the tetrapyrrole biosynthetic pathway

The deduced amino acid sequences of p6*orf*2, p6*orf*3, p6*orf*4 and p6*orf*5 are shown in Figure 6.11.

The ATG start codon, at bp 1740, and a TGA stop codon, at bp 2871, were assigned to p6*orf*2. This ORF is preceded by putative promoter (1670TATGTT1675) and SD sequences (1732AG1733) and encodes a 377-aa protein with a predicted molecular mass of 41.1 kDa. p6*orf*3, which encodes a 299-aa protein corresponding to a molecular mass of 32.5 kDa, starts with the GTG codon at bp 2902 and ends with a TGA stop codon at bp 3801. A putative SD sequence (2893AGG2895) was identified upstream of p6*orf*3.

The ATG translational start at bp 2781 and TGA stop at bp 4215 were assigned to p6*orf*4 which overlaps the 3'-end of p6*orf*3 by 14 bp. A putative SD sequence (3777GGAGGT3782) was assigned 9 bp upstream of p6*orf*4. p6*orf*5 is preceded by a potential SD sequence (4449GAGGT4453) 5 bp upstream of a GTG start codon. p6*orf*5 has no stop codon and encodes a 76-aa peptide before being truncated at the T/A-cloning site.

BLAST searches using the deduced protein sequences encoded by these four ORFs revealed homologies to enzymes involved in the early steps of tetrapyrrole biosynthesis (Figure 6.12).

Figure 6.11: p6*trna*, p6*orfs*2-5

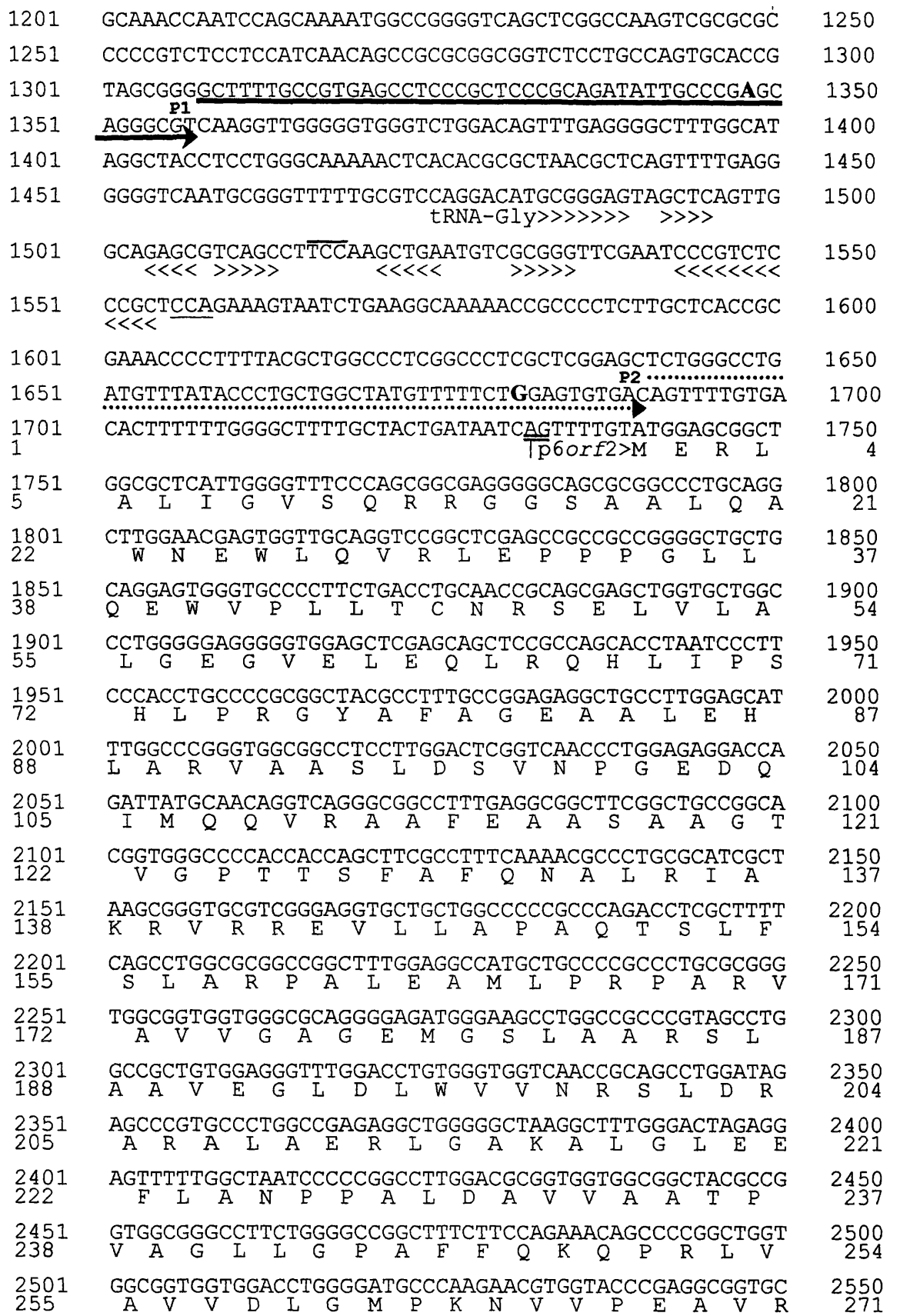


Figure 6.11: Continued

Figure 6.11: p6*trna*, p6*orfs*2-5 continued

```
2551   GTGGGGCGGTGCTCTTTGACCTGGAGCGCCTACAGCGCCTTGGGGAGGAG   2600
272      G  A  V  L  F  D  L  E  R  L  Q  R  L  G  E  E     287

2601   CGCCGCGCGCGCTTGCAGGCTGATCTGGCCCGGGCCGAGCGGATTGTGCT   2650
288    R  R  A  R  L  Q  A  D  L  A  R  A  E  R  I  V  L    304

2651   GGAGGAGGTGGAGCAGGTGGTGATGGAGTGGGCGGAGCGCTCAATGGCCC   2700
305    E  E  V  E  Q  V  V  M  E  W  A  E  R  S  M  A  P    321

2701   CGGCCATCGCCCGGATGCGCGAGGTCTACCGCCGCACCTTGGATGAGCTG   2750
322      A  I  A  R  M  R  E  V  Y  R  R  T  L  D  E  L     337

2751   GTGGGTGAGCTGGTGGGGCCGGAGATGGTGGAGCGGCTGGCCCACCGCTT   2800
338    V  G  E  L  V  G  P  E  M  V  E  R  L  A  H  R  F    354

2801   CGCCCACTTTCCGGTTAAGGGCCTGAGGGGGCTGGCCCGGTGGCATGGGG   2850
355    A  H  F  P  V  K  G  L  R  G  L  A  R  W  H  G  A    371

2851   CCGAGGTAGCCCAGACCTTCTGAAGGAAGGCTGGGTTGTGGAAGGCGGGG   2900
372      E  V  A  Q  T  F  *                               377

2901   CGTGGCTAGGGTGCGCCTGGCCACCCGTGGAAGCCGCCTGGCCTTGTGGC   2950
|p6orf3>V  A  R  V  R  L  A  T  R  G  S  R  L  A  L  W  Q    17

2951   AGGCCGAGTGGGTGGCCAAGCAGCTCGTTCAGCAGGGGGCCGAGGTGGAG   3000
18       A  E  W  V  A  K  Q  L  V  Q  Q  G  A  E  V  E      33

3001   CTGGTGGTGGTGGAGACCCAGGGCGACCGCGAAAAGCGCCCCTTTGCGCA   3050
34     L  V  V  V  E  T  Q  G  D  R  E  K  R  P  F  A  Q     50

3051   GATGCAGGGCCAGGGCTTTTTCACCAAAGCAGTTCAGGAAGCGGTGCTGG   3100
51     M  Q  G  Q  G  F  F  T  K  A  V  Q  E  A  V  L  E     67

3101   AGGGACGCGCCGACTTTGCGGTGCACTCCTACAAAGACCTCCCGAGCGCC   3150
68       G  R  A  D  F  A  V  H  S  Y  K  D  L  P  S  A      83

3151   CGTCCGGCGGGTTTGGAGATTGCCGCGGTGCCGCCTCGAGAGGACCCCCG   3200
84     R  P  A  G  L  E  I  A  A  V  P  P  R  E  D  P  R    100

3201   TGAGCTCCTCTTGGTACGGCTGCAAGCGGTGGACCAGGTAGCCCCTGGCC   3250
101    E  L  L  L  V  R  L  Q  A  V  D  Q  V  A  P  G  L    117

3251   TGCCCCTGCGGGCTGGGGCCCGGGTGGGGAGCAGCGCTGCCCGGCGGCAG   3300
118      P  L  R  A  G  A  R  V  G  S  S  A  A  R  R  Q     133

3301   GCCCAGCTCGCCCACCTGCGCCCCGACCTTTCCCTTTTGGAGCTGCGGGG   3350
134    A  Q  L  A  H  L  R  P  D  L  S  L  L  E  L  R  G    150

3351   CAACGTGCTCACCCGGGTGGAAAAGCTTCGGCAGGGCGAGTACGACGCGG   3400
151    N  V  L  T  R  V  E  K  L  R  Q  G  E  Y  D  A  V    167

3401   TTCTTTTAGCCTACGCGGGGGTGCGACGGCTAGGGCTGGACTTAAGTCCC   3450
168      L  L  A  Y  A  G  V  R  R  L  G  L  D  L  S  P     183

3451   TTCCACTGGCAGGTGCTTCCGCCTACGCTTTTGGTGCCAGCCCCGGCCCA   3500
184    F  H  W  Q  V  L  P  P  T  L  L  V  P  A  P  A  Q    200

3501   AGGGGCGTTGGCCCTGGAGTGCCGCCAGGACGACAAGCGTCTGAGGCCCT   3550
201      G  A  L  A  L  E  C  R  Q  D  D  K  R  L  R  P  L   217

3551   TGCTGGAGCCTTTGGACGACCCCTCGGCCGGCGTGCGGTGCGGCGGAGCC   3600
218      L  E  P  L  D  D  P  S  A  G  V  R  C  G  G  A     233

3601   GGTCTTATGGCCCGCTTGGCCGGGGGTTGCCAGCTTGCCTTGGGGGCGCT   3650
234    G  L  M  A  R  L  A  G  G  C  Q  L  A  L  G  A  L    250

3651   GGCCCAGGaGACGCCGGAGGGACTCCAGCTCCTGGCCTGGTACCGGGGCC   3700
251    A  Q  E  T  P  E  G  L  Q  L  L  A  W  Y  R  G  R    267

3701   GAAGCTACCAGGCCCGAGGGTCGGACCCAGAGGCGGTGGCGGAGGCGGTT   3750
268      S  Y  Q  A  R  G  S  D  P  E  A  V  A  E  A  V     283
```

Figure 6.11: Continued

Figure 6.11: p6*trna*, p6*orfs*2-5 continued

```
3751   TTCAAGGAGATTTGCAGCGAGTACCCGGAGGTGGTATGCGAATCGCCCTG   3800
284    F  K  E  I  C  S  E  Y  P  E  V  V  C  E  S  P  *.    299
1                                |p6orf4>M  R  I  A  L        5

3801   ACCCAGTCTGAGGGCCGCTTGGCGGGTTTGCAGGAGGCCTTGGAGGCTAT   3850
6      T  Q  S  E  G  R  L  A  G  L  Q  E  A  L  E  A  M     22

3851   GGGCCTCGAGGTTTGGCGGGTTCCCCTGGTGCAGACCCGCTTTCTGCCTG   3900
23      G  L  E  V  W  R  V  P  L  V  Q  T  R  F  L  P  A    39

3901   CTGACCTGACCCCCATCCAGGACTGTCGTTGGTGGCTTTTTACCAGCGTC   3950
40       D  L  T  P  I  Q  D  C  R  W  W  L  F  T  S  V      55

3951   GCGGCGGTGCGGGCGGTGCAGGCTTTGGGGGCCAGCCTCGAGGGGCGGAG   4000
56     A  A  V  R  A  V  Q  A  L  G  A  S  L  E  G  R  R     72

4001   GCTGGGTGCGGTGGGCCCTGCGACCCAGCGGGCTTTGGAGGAGGCGGGGG   4050
73      L  G  A  V  G  P  A  T  Q  R  A  L  E  E  A  G  G    89

4051   GGGTTGTGGAGCTGGTGGCCCCGGAGGGGAACGCGCTGAGCCTGGCCAAG   4100
90       V  V  E  L  V  A  P  E  G  N  A  L  S  L  A  K      105

4101   GCTTTTCTGGCCCGGCGCCCCTTCGGCTTTGTGGGCCTGCCCCAGGGCAA   4150
106    A  F  L  A  R  R  P  F  G  F  V  G  L  P  Q  G  N     122

4151   CCGGGCTCAGCCCCACCTGGCCCGGGTGTTTGCGGGAAGGCGGGGCTACA   4200
123     R  A  Q  P  H  L  A  R  V  F  A  G  R  R  G  Y  T    139

4201   CCCGTGGAAGGCGGTGACGGTTTACGAAACCCTGGTTCGCCCCTGGCCGC   4250
140      R  G  R  R  *                                      143

4251   AGGGTCTGGAGGCCCCCGAGCTGCTCTTGCTGGCCAGCCCCTCTGCGGTG   4300

4301   GAGGCCCTGCCCGAAGCGGTGGGAGGAAGGGCTCACAGCCTGGCCTTGGG   4350

4351   CCCGAGCACCGCGGCGGCTTTGGGGGAGCGGGGCTGGTCCTACACCCTTC   4400

4401   TGGCCAGTCCCAGTGTGGAGGCCGTTTTGCAGGCCATTCGGGATATCCGA   4450

4451   GGTGAACTGTGCTTGACTTGAAAGAACCCAAGACCCTGCCGCTGGATGCC   4500
1      |p6orf5>V  L  D  L  K  E  P  K  T  L  P  L  D  A     14

4501   CCCTTTCGGCTGACCCAGCGGCCCCGGCGGCTGCGGGCCACGGCGGCTCT   4550
15     P  F  R  L  T  Q  R  P  R  R  L  R  A  T  A  A  L     31

4551   GCGGGAGAGCGTTGCGGAAACCCACCTTCGGCCTACCGATTTCATTGCGC   4600
32      R  E  S  V  A  E  T  H  L  R  P  T  D  F  I  A  P    48

4601   CCTTCTTTGTGCTGCCCGGACGGGGTCCTTCTGAGGCTATCCCGGCCCTG   4650
49       F  F  V  L  P  G  R  G  P  S  E  A  I  P  A  L      63

4651   CCTGGGGTTTACCGCCACAGCGTGGAGGGGTTTTTG                 4686
64     P  G  V  Y  R  H  S  V  E  G  F  L  truncated         76
```

Figure 6.11: Nucleotide and deduced amino acid sequence of the region of p6ICE16 containing p6*orf2*, p6*orf3*, p6*orf4*, p6*orf5* as well as the tRNA-Gly gene (Section 6.4.8) identified upstream of p6*orf2*. p6*orfs* 2-5 were subsequently identified as putative tetrapyrrole biosynthetic genes (*hem*ACDB, respectively) (Section 6.4). The putative promoter-like sequences (P) and ribosome-binding sites are indicated by arrows and double-underlines, respectively. The transcriptional starts, as predicted by PromoterScan (Section 2.10) are highlighted in bold. Stop codons are indicated by asterisks. The triplet encoding the glycine anticodon is overlined and the CCA triplet found at the 3'-end of mature tRNAs is underlined.

6.4.4: p6*orf2* encodes a glutamyl-tRNA reductase (*Hem*A)

BLAST searches using the deduced protein sequence of p6*orf2* showed highest homologies to putative *Hem*A proteins from *Archaeoglobus fulgidus* (32% identity over 303 aa) (Klenk *et al.*, 1997), *Deinococcus radiodurans* (37% identity over 263 aa) (White *et al.*, 1999), *Shewanella putrefaciens* (33% identity over 329 aa) (incomplete genome sequence), and *Aquifex aeolicus* (29% identity over 333 aa) (Deckert *et al.*, 1998). The *hem*A gene product is a glutamyl-tRNA reductase (EC 1.2.1.-), an enzyme that catalyses the second step of the C5 pathway for δ-aminolaevulinic acid (ALA) biosynthesis. This pathway for ALA biosynthesis is found in higher plants, algae, archaea and most bacteria (reviewed by Kannangara *et al.*, 1994).

A multiple sequence alignment (Figure 6.13) of the probable *hem*A gene product encoded by p6*orf2* (p6*Hem*A) against the putative *Hem*A from *A. fulgidus* as well as against glutamyl tRNA reductases possessing less significant homologies from *B. subtilis* (23% identity over 277 aa) (Petricek *et al.*, 1990, *E. coli* (25% identity over 308 aa) (Ikemi *et al.*, 1992) and cucumber (31% identity over 107 aa) (Tanaka *et al.*, 1996) revealed the characteristic fingerprint for the ADP-binding fold of NAD(P)H-dependent glutamyl tRNA reductases (Wierenga *et al.*, 1986). This alignment indicated that the ADP pocket starts at Arg170 and ends at Ser201 in the deduced amino acid sequence of p6*orf2* (hereinafter referred to as p6*hem*A) thus providing further evidence that it encodes a glutamyl tRNA reductase.

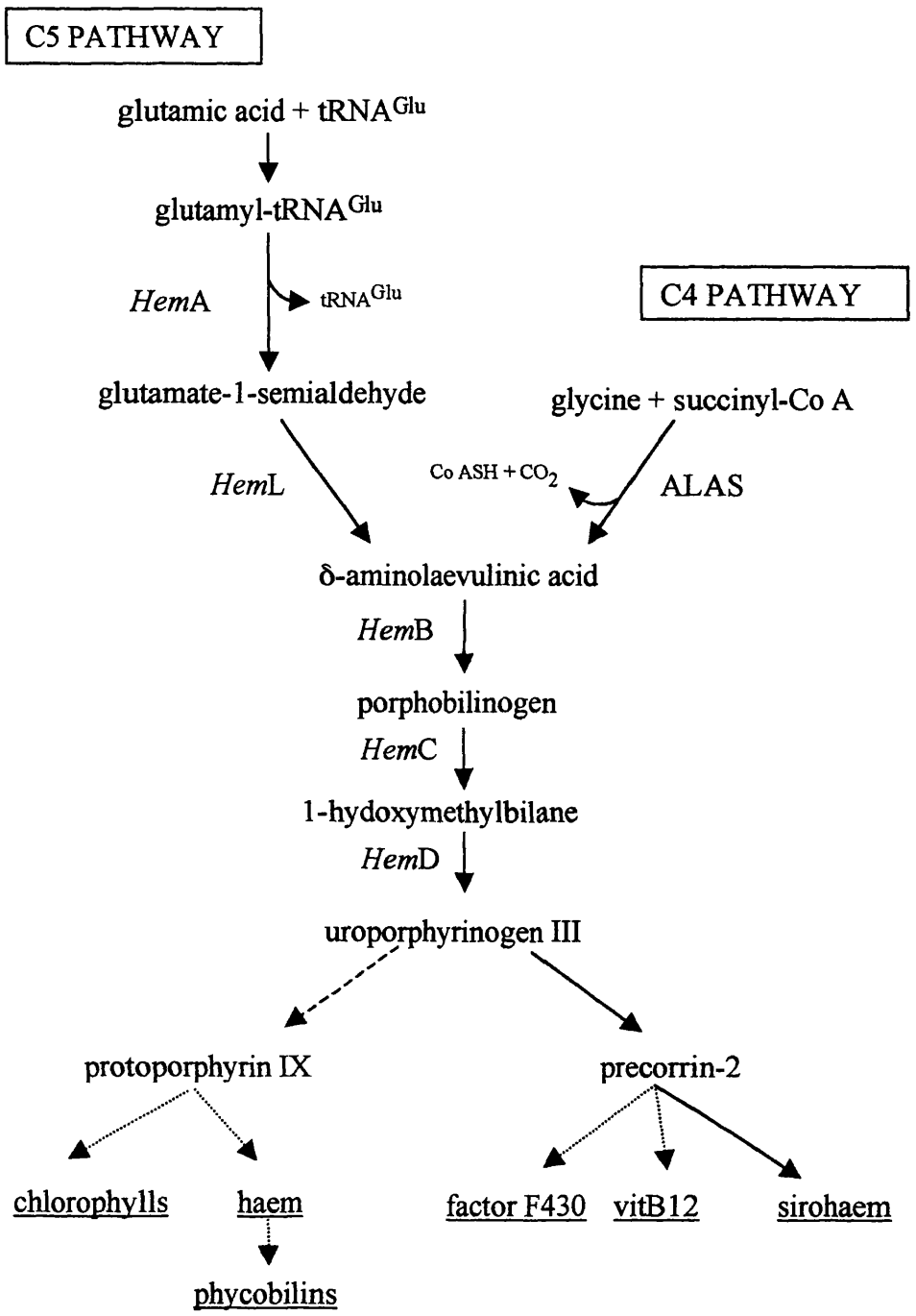


Figure 6.12: Overview of tetrapyrrole biosynthesis with gene products indicated for the formation of uroporphyrinogen III (UROIII) from glutamyl-tRNA. In animals, yeast, algae and proteobacteria of the α-subdivision, δ-aminolaevulinic acid (ALA) is synthesised *via* the C4 pathway which is catalysed by ALA synthase (ALAS). In plants, archaea and most bacteria, ALA is synthesised from glutamic acid *via* the C5 pathway. δ-ALA is the first committed step in tetrapyrrole biosynthesis whereas UROIII is the macrocyclic tetrapyrrole precursor for all natural prosthetic groups such as those underlined above. With the exception of *Hem*L, the enzymes of the C-5 pathway that catalyse the formation of UROIII from glutamyl-tRNA are homologous to putative gene products identified in p6ICE16. Solid and dashed arrows represent single and multiple biosynthetic steps, respectively. (Figure 6.12 was adapted from Johansson & Hederstedt, 1999).

Figure 6.13: Alignment of the predicted amino acid sequence of ORF2 from p6ICE16 (p6*hem*A) against the putative *Hem*A from *A. fulgidus* (ARCFU) (Accession AE000782) and the glutamyl tRNA reductases from *B. subtilis* (BACSU) (Accession P16618), *E. coli* (ECOLI) (Accession BVECHA) and cucumber (CUCSA) (Accession P49295). Red amino acid residues represent conservation in at least three of the proteins while the blue residues indicate conservative substitutions. The ADP-binding pocket, serving as a fingerprint for NAD(P)H-dependent glutamyl tRNA reductases, is indicated by a green line.

6.4.5: The incomplete p6*orf*5 gene product has homology to porphobilinogen synthase (*Hem*B)

The deduced amino acid sequence of the incomplete p6*orf*5 (designated p6*hem*B) was shown to possess similarities to the N-terminal sequence of several porphobilinogen (PBG) synthase (*Hem*B) proteins. The highest homology found was with the putative *hem*B gene product of *Thiobacillus ferrooxidans* (50% identical over 40 amino acids) (Accession TIGR_920). PBG synthase (EC 4.2.1.24) also called ALA dehydratase, catalyses the formation of porphobilinogen from two molecules of aminolaevulinic acid (Jordan & Seehra, 1980). PBG synthases are classified on the basis of their metal requirements. The truncation of p6*hem*B after only 78 codons did not reveal enough nucleotide sequence information for predicting the metal requirement of its gene product. Cloning and nucleotide sequence analysis of the complete sequence of p6*hem*B would reveal its similarity to *hem*B of other bacterial species.

6.4.6: p6*orf*3 and p6*orf*4 encode a porphobilinogen deaminase (*Hem*C) and uroporphyrinogen III synthase (*Hem*D), respectively

BLAST searches using the deduced amino acid sequences of p6ICE16 *orf*3 and *orf*4 (hereinafter referred to as p6*hem*C and p6*hem*D, respectively) revealed homologies to porphobilinogen (PBG) deaminase (*HemC*) (EC 4.3.1.8) and uroporphyrinogen (URO) III synthase (*Hem*D) (EC 4.2.1.75), respectively. *Hem*C and *Hem*D catalyse sequential reactions whereby URO III is formed from four PBG molecules. To ensure their co-expression, many bacterial *hem*C and *hem*D genes are either located adjacent to each other (e.g. *Clostridium josui* (Fujino *et al.*, 1995)) or with the stop codon of *hem*C overlapping the *hem*D translational start (e.g. *E. coli* (Jordan *et al.*, 1988), *B. subtilis* (Hansson *et al.*, 1991), *Bacillus sphaericus* and *Bacillus stearothermophilus* (Johansson & Hederstedt, 1999)). Consistent with these findings, the start of p6*hem*D overlaps the stop of p6*hem*C by 16 nucleotides (Figure 6.11).

Figure 6.14 shows the alignment of the conceptually translated sequence of p6*hem*C with the *Hem*C proteins from phylogenetically diverse species. PBG deaminase, also called hydroxymethylbilane synthase, catalyses the polymerisation of four molecules of PBG to produce hydroxymethylbilane (HMB). A five-element amino acid fingerprint was identified by Louie (1993) as a signature common to all PBG deaminases. A fingerPRINTScan (Section 2.10) showed that the putative p6*Hem*C possesses each of

the five elements of the PBG deaminase fingerprint. Furthermore, the invariant cysteine residue responsible for cofactor binding (Miller *et al.*, 1988) as well as invariant arginines shown to be required for activity in the *E coli Hem*C (Lander *et al.*, 1991) are also found in p6*Hem*C.

URO (III) synthase, also called cosynthase, catalyses the cyclisation of HMB to form URO III. While homology searches demonstrated that the p6*hem*D gene product is 40% identical across the entire protein to a putative *hem*D gene product from *D. radiodurans* (Accession AAF10363) (White *et al.*, 1999), no significant homology to other uroporphyrinogen III synthases in the database could be found. This finding is not unexpected because *Hem*D proteins are not highly conserved.

6.4.7: p6*hem* sequences form part of an apparent operon

The close spacing of p6*hem*A, p6*hem*C, p6*hem*D and truncated p6*hem*B suggested that they form part of an operon. The arrangement of the p6*hem*ACDB gene cluster was found to be similar to that the *hem* operon of *Bacillus subtilis* (*hem*AXCDBL) (Hansson *et al.*, 1991); however, the ORF corresponding to *hem*X was not found between p6*hem*A and p6*hem*C. Furthermore, a putative *hem*L gene encoding glutamate-1-semialdehyde aminotransferase (EC.5.4.3.8), which catalyses the synthesis of ALA *via* the C5 pathway, was not included in the cloned environmental DNA fragment. Other Gram-positive bacteria possess a similarly conserved organisation of the *hem* genes, for example, *Staphylococcus aureus* (Kafala & Sasarman, 1997), *Bacillus sphaericus*, *Bacillus strearothermophilus*, *Brevibacillus brevidus*, *Paenibacillus macerans* (Johansson & Hederstedt, 1999) and *Clostridium josui* (Fujino *et al.*, 1995). It is noteworthy that the Gram-positive radioresistant *D. radiodurans* displays no clustering of its putative *hem* genes as indicated by sequence searches of its genome (Section 2.10).

In Gram-negative bacteria such as *E. coli*, *hem*C and *hem*D are closely linked, whereas *hem*A, *hem*B and *hem*L are located at different positions on the chromosome (Jordan *et al.*, 1988; Bachmann, 1990).

Since *A. fulgidus* sequences were shown to possess homology with the p6*hem*A gene product, its whole genome sequence was searched in order to determine the organisation

of its putative *hem* genes. Two putative *hem* gene clusters (*hem*AB and *hem*LCD) were identified in different sections of this archaeal genome.

Whether the differences in the organisation on the chromosome of *hem* genes from different prokaryotes reflect evolutionary or regulatory differences is an interesting subject for further studies.

6.4.8: The gene upstream of *hem*A encodes a tRNA-Gly

Database searches using the p6ICE16-insert DNA sequence identified a putative tRNA gene (bp 1483-1558) located upstream of p6*orf2* (p6*hem*A) (Figure 6.11). The putative tRNA showed 96% identity across its entire coding sequence to tRNA-Gly of *Stigmatella aurantiaca* (Accession X82820). A tRNA-scan (Section 2.10) of the p6ICE16-DNA insert confirmed the glycine anticodon (5'UCC3') of this putative tRNA gene product (Figure 6.15).

```
p6hemC     1 ------------------------------VARVRLATRGSRLALWQAEWVAKQLVQQG-
AQUAE      1 ------------MYPVWGTEGSNPSPSAIPAMKIRIGTRKSKLALWQANYVKDFLEKHW-
ECOLI      1 ----------------------------MLDNVLRIATRQSPLALWQAHYVKDKLMASHP
EUGGR    121 TTKPAD--LQEVSGGRIWSLASTTGSNIGAGKTVRVATRKSPLAMWQAEFIQSELERLWP
ARATH     42 ISSPALGKCRRKQSSSGFVKACVAVEQKTRTAIIRIGTRGSPLALAQAYETREKLKKKHP
HUMAN      1 -----------MSGN--GNAAATAEENSPKMRVIRVGTRKSQLARIQTDSVVATLKASYP

                                  1                                2
p6HEMC    29 -----AEVELVVVETQGD-REKRPFAQMQGQGFFTKAVQEAVLEGRADFAVHSYKDLPSA
AQUAE     48 ----GVEVELVKITTTGDKITDVPLAKIGGKGLFVKEIEKALLEGSIDLAVHSLKDVPMV
ECOLI     33 ----GLVVELVPMVTRGDVILDTPLAKVGGKGLFVKELEVALLENRADIAVHSMKDVPVE
EUGGR    179 ----GITVELQPMSTRGDKILDSPLAKVGGKGLFVKELETALLENRSDIAVHSTKDVPME
ARATH    102 ELVEDGAIHIEIIKTTGDKILSQPLADIGGKGLFTKEIDEALINGHIDIAVHSMKDVPTY
HUMAN     48 ----GLQFEIIAMSTTGDKILDTALSKIGEKSLFTKELEHALEKNEVDLVVHSLKDLPTV

                                                          3
p6HEMC    83 RPAGLEIAAVPPREDPRELLLVRLQAVDQVAPGLPLRAGARVGSSAARRQAQLAHLRPDL
AQUAE     104 IPKGLKLGAITKRENPYDVLISR---SGKKLYELP--SGSVIGTSSLRRQVQIKKRRRDL
ECOLI      89 FPQGLGLVTICEREDPRDAFVSN---NYDSLDALP--AGSIVGTSSLRRQCQLAERRPDL
EUGGR     235 LPEGLVLGVICKRHDPCDAIVFPKGSNLKSLEDLP--HGARVGTSSLRRQCQLLLKRPDL
ARATH     162 LPEKTILPCNLPREDVRDAFICL---TAATLAELP--AGSVVGTASLRRKSQILHKYPAL
HUMAN     104 LPPGFTIGAICKRENPHDAVVFHPKFVGKTLETLP--EKSVVGTSSLRRAAQLQRKFPHL

                4
p6HEMC    147 SLLE-LRGNVLTRVEKLRQG-EYDAVLLAYAGVRRLGLDLSPFH--WQVLPPTLLVPAPA
AQUAE     159 KVEV-LRGNVDTRMRKLKEG-LYDAVILAYAGVKRMGYESEIT------EVLEDFIPAVG
ECOLI     144 IIRS-LRGNVGTRLSKLDNG-EYDAIILAVAGLKRLGLESRIR----AALPPEISLPAVG
EUGGR     293 KFLE-LRGNVNTRLAKLDSG-DYDAIILAAAGLKRLGFSDRVLPGETNIIDPNVMCPAAG
ARATH     217 HVEENFRGNVQTRLSKLQGG-KVQATLLALAGLKRLSMTENVA----SILSLDEMLPAVA
HUMAN     162 EFRS-IRGNLNTRLRKLDEQQEFSAIILATAGLQRMGWHNRVG----QILHPEECMYAVG

                                                    5
p6HEMC    199 QGALALECRQDDKRLRPLLEPLDD-PSAGVRCGGAGLMARLAGGCQLALGALAQETP---
AQUAE     211 QGSLAIEIREGDKRIEELIKPLNNEESFLCAIAERTFLRRLEGGCQVPVGAFAKIEN-GT
ECOLI     198 QGAVGIECRLDDSRTRELLAALNHHETALRVTAERAMNTRLEGGCQVPIGSYAELID-GE
EUGGR     351 QGALSIELRTNDPEIAALLEPLHHIPDAVTVACERAMNRRLNGGCQVPISGFAQLKD-GQ
ARATH     272 QGAIGIACRTDDDKMATYLASLNHEETRLAISCERAFLETLDGSCRTPIAGYASKDEEGN
HUMAN     217 QGALGVEVRAKDQDILDLVGVLHDPETLLRCIAERAFLRHLEGGCSVPVAVHTAMKD-GQ

p6HEMC    255 -EGLQLLAWYRG------------RSYQARGSDP----------------------EAVA
AQUAE     270 LKMKAFISDIEA----------ERYIEGYREGNP--------------------EEAEKLG
ECOLI     257 IWLRALVGAPDG----------SQIIRGERRGAP--------------------QDAEQMG
EUGGR     410 LRMEARVGSVTG-KGPLIIQSKTFRLPWSGRTWP----QLQ----K--------ESEALG
ARATH     332 CIFRGLVASPDG-----------TKVLETSRKGP----------------YVYEDMVKMG
HUMAN     276 LYLTGGVWSLDGSDSIQETMQATIHVPAQHEDGPEDDPQLVGITARNIPRGPQLAAQNLG

p6HEMC    280 EAVFKEICSEYPEVVCESP-----------
AQUAE     301 LSLAEELLKKGGEEILKEIYSSQ-------
ECOLI     288 ISLAEELLNNGAREILAEVYN----GDAPA
EUGGR     453 VEVADMLLADGAQAYLDEAYASRTLGWA--
ARATH     365 KDAGQELLSRAGPGFFGN------------
HUMAN     336 ISLANLLLSKGAKNILDVA---RQLNDAH-
```

Figure 6.14: Comparison of the deduced amino acid sequence of ORF3 from the 4.7-kb p6ICE16 insert (p6HEMC) against PBG deaminases from *A. aeolicus* (AQUAE) (Accession AE000681) (Deckert *et al.*, 1998), *E. coli* (ECOLI) (Accession P06983) (Jordan *et al.*, 1988), *Euglena gracilis* (EUGGR) (Accession P13446) (Weinstein & Beale, 1983), *Arabidopsis thaliana* (ARATH) (Accession Q43316) (Lim *et al.*, 1994) and *H. sapiens* (HUMAN) (Accession P08397) (Raich *et al.*, 1986; Grandchamp *et al.*, 1987). Red residues indicate conservation in at least three of the proteins while the blue residues represent a conservative substitution. The five elements making up the signature for PBG deaminase proteins are indicated by green bars. The invariant cysteine, responsible for cofactor-binding is highlighted in yellow. The invariant arginines that, when mutated, each result in complete loss of activity in the *E. coli* protein, are highlighted in grey.

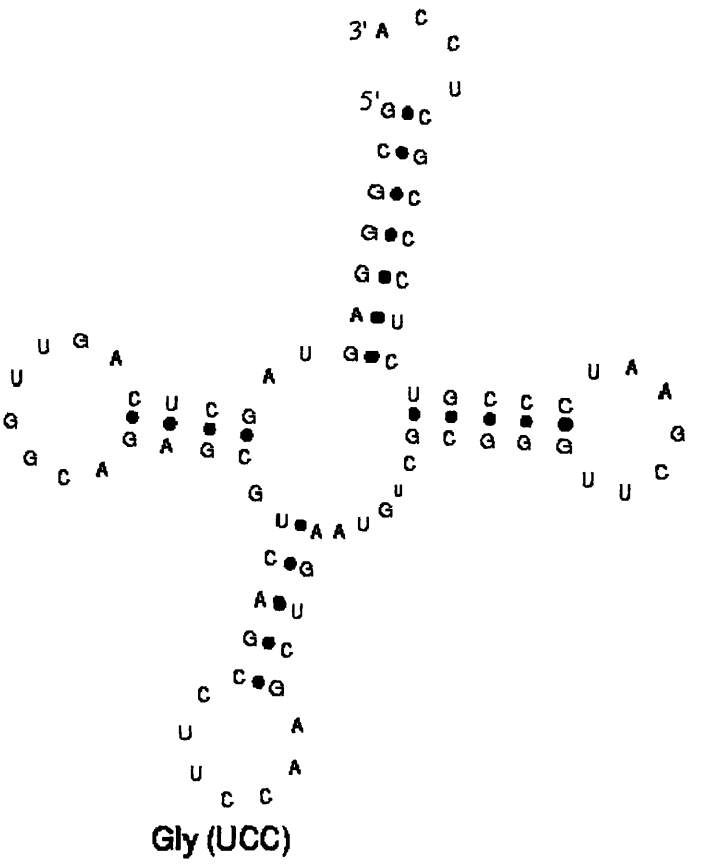


Figure 6.15: Deduced nucleotide sequence and structure of the transfer-RNA molecule encoded by p6ICE16-DNA insert (bp 1482-1558). This gene product has the characteristic CCA sequence at its 3'-end and the anticodon sequence, UCC, identifying it as a glycine specific tRNA.

6.4.9: Prospects for heterologous expression of p6ICE16 DNA sequences

The orientation of p6*orf*1, with respect to the vector-encoded *lac* promoter, indicated that heterologous expression would only be permitted *via* its own promoter. No bacterial promoter sequence could be identified in p6*orf*1-upstream sequences. Significant homologies of p6*orf*1 with sequences of *M. thermoautotrophicum* prompted a scan for consensus archaeal promoter elements. Again, no obvious promoter sequences could be identified. Based on this information it was thought unlikely that p6*Orf*1 could be expressed in *E. coli* TOP10.

Apparent p6*orf*1 transcription terminators (Figure 6.8) located immediately adjacent to the cloning site were thought to preclude heterologous expression of insert-encoded genes *via* the vector-encoded *lac* promoter. Any expression of the tRNA gene and/or p6*orf*s2-5 (p6*hem*ACDB) would therefore only have been permitted *via* their own promoter(s). Promoter scans of the p6ICE16-DNA insert identified two putative promoter sequences. One was located upstream of the tRNA-Gly gene (bp 1308-1357) and the other was located upstream of (p6*hem*A) (bp 1641-1690). No other bacterial or archaeal promoter sequences were found within the 4.7 kb insert. This suggested that the tRNA-Gly gene and p6*hem*ACDB cluster constitute separate transcription units and provided further evidence that the putative p6*hem* genes form part of an operon.

The accumulation of intermediate products of the C5 pathway for uroporphyrinogen III synthesis (Figure 6.12) was thought to be responsible for the dark phenotype observed with *E. coli* TOP10/p6ICE16. *E. coli* strains expressing cloned *hem*A or *hem*ACD genes of various species have a fluorescent pink phenotype due to the overproduction of tetrapyrrole pigments (Li *et al.*, 1989; Chen *et al.*, 1994; Fujino *et al.*, 1995). *E. coli* TOP10/p6ICE16 did not exhibit this pink fluorescent phenotype when exposed to long-wave UV light, suggesting that the tetrapyrrole pigments were not being overproduced. This possibility was further supported by the finding that plasmid p6ICE16 did not complement *E. coli hem*A mutant strain SASX41B. Finally, to see if any proteins were being heterologously expressed from the p6ICE16-DNA insert, SDS-PAGE was performed using *E. coli* crude cell extracts. No protein band typifying a heterologously expressed gene product was observed for *E. coli* TOP10/p6ICE16 (results not shown). Even though cloned Ice16-DNA was demonstrated to confer the observed phenotype to

E. coli TOP10 cells (Section 5.4.2), no gene or gene product imparting the atypical phenotype could be identified for p6ICE16.

6.5: pPhos22

Recombinant plasmid pPhos22, which conferred a thermostable phosphatase activity to *E. coli* TOP10 cells, was obtained from Ice22-environmental DNA library (Section 5.7.1). The DNA used for Ice22-DNA library construction was derived from a biomass-containing sediment; ~55°C, pH 4.3 (Section 4.7).

6.5.1: pPhos22 contains a 1.7-kb insert

The physical map of the pPhos22-DNA insert is shown in Figure 6.16. Three open reading frames were identified within this 1689-bp fragment.

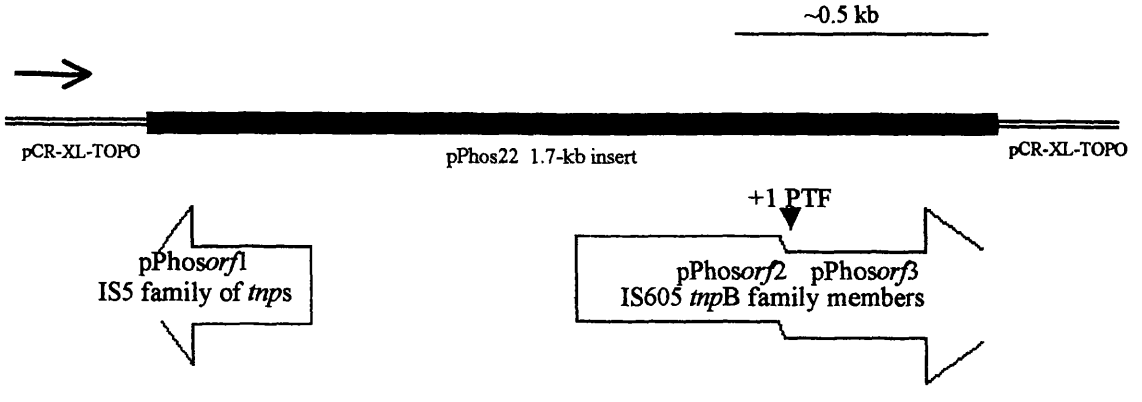


Figure 6.16: Physical map of the 1.7-kb DNA fragment (black bar) extracted from Ice22 sediment, cloned into the TA vector, pCR-XL-TOPO (double line) and recovered from the transformant designated pPhos22. Block arrows show the relative localisation of each gene and the orientation of coding sequences found. pPhos*orf*1 and pPhos*orf*3 have been truncated at the cloning site as indicated by incomplete arrows. A fusion of pPhos*orf*2 and pPhos*orf*3 can arise from a putative +1 programmed translation frameshift (+1 PTF). The genes having homologous sequences to the ORFs identified are shown in parentheses. The black arrow indicates the direction of transcription from the vector-encoded *lac* promoter.

The nucleotide sequence of the 1.7-kb insert of pPhos22 and deduced amino acid sequences of its ORFs (pPhos*orf*1, pPhos*orf*2 and pPhos*orf*3) are shown in Figure 6.17. pPhos*orf*1 (bp 342-1) is preceded by a potential SD sequence (351GAG349) 6 bp upstream of a CTG start codon. pPhos*orf*1 has no stop codon and encodes a 114-aa peptide before being truncated at the cloning site. Three possible bacterial promoter sequences were identified in the 5'-flanking region of pPhos*orf*1.

Two additional putative promoter sequences were identified on the opposite DNA strand upstream of a second ORF (pPhos*orf*2; bp 871-1302). pPhos*orf*2 is preceded by a possible SD sequence, 866AG867, located 5 nucleotides upstream of the assumed ATG start codon. pPhos*orf*2 encodes a 143-aa peptide before ending at a TGA stop codon. Overlapping the 3'-end of pPhos*orf*2 by 11 nucleotides, a third ORF (pPhos*orf*3) was assigned with an ATG translational start at bp 1291. pPhos*orf*3 encodes a 132-aa peptide before being truncated at the cloning site. The deduced amino acid sequence continues with vector-encoded sequences for another 114 amino acids before encountering a TAG translational stop. A fusion of pPhos*orf*3 and vector sequences would therefore encode a 246-aa peptide. A fusion of pPhos*orf*2 and pPhos*orf*3 (pPhos*orf*2-3) is also predicted to arise from a +1 frameshift identified at a stretch of nine adenosine residues located near the end of pPhos*orf*2 and just before the putative start codon of pPhos*orf*3. This frameshift was not considered the result of a sequencing error because the DNA sequence of this A9 motif was confirmed by re-sequencing the complementary strand encompassing this motif. For sequence analysis purposes, one fused pPhos*orf*2-3 has been considered.

6.5.2: pPhos*orf*1 and pPhos*orf*2-3 gene products each possess homologies to transposases encoded by different families of insertion sequences (IS) elements

BLAST searches revealed that pPhos22 DNA insert possesses no homology at the nucleotide level to sequences deposited in the databases; however, searches using the conceptually translated sequences of pPhos*orf*1 and pPhos*orf*2-3 indicated that each possessed homology to bacterial transposases encoded by different families of IS elements.

Figure 6.17: pPhos*orf*1, pPhos*orf*2 and pPhos*orf*3

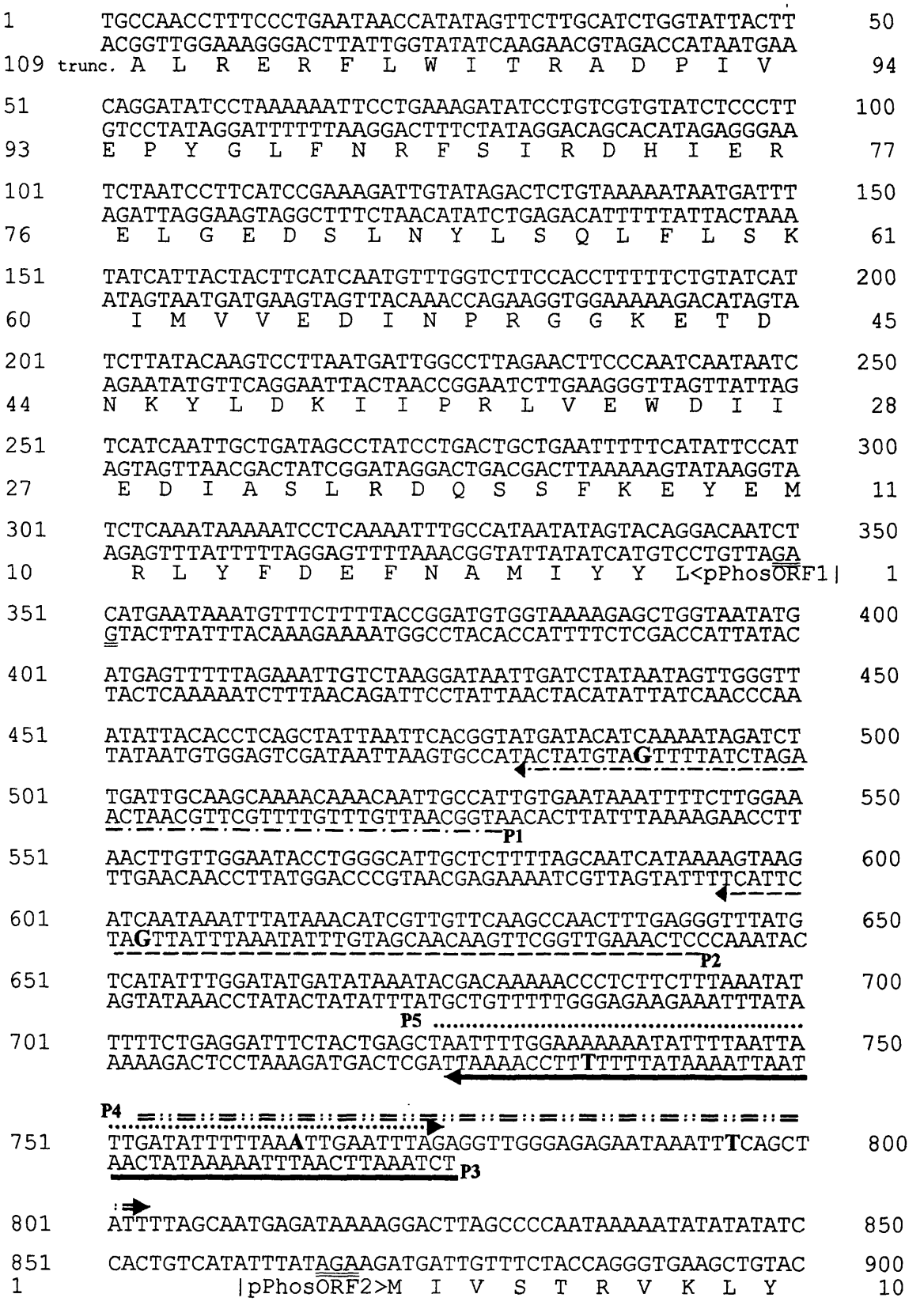


Figure 6.17: Continued.

Figure 6.17: pPhos*orf*1, pPhos*orf*2 and pPhos*orf*3 continued

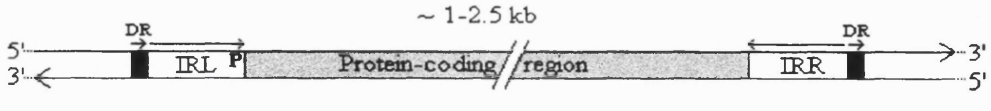
```
901    ACCAATGAAAAACAGAGAGTATTGCTGGAGAAGCACATTGGTAGCTGGCG    950
11      T  N  E  K  Q  R  V  L  L  E  K  H  I  G  S  W  R     27

951    ATCCTTTTACAACTACTTTCTTGAGAAGGGGAATGAATACTGCCTGAGAC   1000
28      S  F  Y  N  Y  F  L  E  K  G  N  E  Y  C  L  R  R     44

1001   GCAAGGAATCGAAGAAATGCTCACTGAGTTATATTGAAACTCAAAGCATG   1050
45      K  E  S  K  K  C  S  L  S  Y  I  E  T  Q  S  M      60

1051   GTGAGAGAACTCAGGAAGTATCAGTGGCTTTACGAAATGAATTCACAATC   1100
61      V  R  E  L  R  K  Y  Q  W  L  Y  E  M  N  S  Q  S     77

1101   GCTCCAGATGTCTTTACGCTACCTGGATAATGTGTTCAAGAAGTTCTTTA   1150
78      L  Q  M  S  L  R  Y  L  D  N  V  F  K  K  F  F  N     94

1151   ATAAGAATTCTGAATATTGCGGATTCAGGAAGAAGGGTAAAAATAACCAC   1200
95      K  N  S  E  Y  C  G  F  R  K  K  G  K  N  N  H      110

1201   TTCGCGTTTCCACAGCACATAAAGATAAAGGGGGATAGAATATGTTTTCC   1250
111     F  A  F  P  Q  H  I  K  I  K  G  D  R  I  C  F  P    127

1251   AGAGTTTTCAGAAGTCATATGTTTAAAGGCTCTAAAAAAAAATGTCATAT   1300
128     E  F  S  E  V  I  C  L  K  A  L  K  K  N  V  I  *
                    +1 FRAME SHIFT *  K  K  M  S  Y      143

1301   GAAATAAAGAGTATCAACCAGATAGTTATAACCAAGGAAGGGGGTGATTA   1350
144     E  I  K  S  I  N  Q  I  V  I  T  K  E  G  G  D  Y    160

1351   TTACTGCTCCATAATCTACGAAAATGGAGAGGAGCCTTTGGAGGGTGTGC   1400
161     Y  C  S  I  I  Y  E  N  G  E  E  P  L  E  G  V  P    177

1401   CAATGTCAGTAGAGAATTGGGTGGGCATAGATACGGGTGTTGAGAAGTTT   1450
178     M  S  V  E  N  W  V  G  I  D  T  G  V  E  K  F      193

1451   GCCACGCTCTCGGATGGTATTGGAATAGATAATCCACACTTCATCAATGA   1500
194     A  T  L  S  D  G  I  G  I  D  N  P  H  F  I  N  E    210

1501   GGTTGAGAATAAGATCAATAAGATTCAAAGAGAACTATCAAGAAAGCAAA   1550
211     V  E  N  K  I  N  K  I  Q  R  E  L  S  R  K  Q  K    227

1551   AGGGTTCGAATAACTGGCAGAAAACAATATTGAAGATTCAGAGGAATTAC   1600
228     G  S  N  N  W  Q  K  T  I  L  K  I  Q  R  N  Y      243

1601   AGGAAGCTAAGGAGGAAGAGGGATGAATTCCTTGCCAGGAATCGACCGTG   1650
244     R  K  L  R  R  K  R  D  E  F  L  A  R  E  S  T  V    260

1651   TATGGTCAAGCGGTATGATAGCATAGTTTTTGAATACCCA             1689
261     M  V  K  R  Y  D  S  I  V  F  E  Y  P RANSADIHHT    273

       GGRSSMHLEGPIRPIVSRITIHWPSFYNVVTGKTLALPNLIALQHIPLSP
       AGVIAKRPAPIALPNSCAAYTYGSLRFTPIKERAVIVCLWMYRVILLTRR
       GDGW*     386-aa fusion protein
```

Figure 6.17: Nucleotide sequence of pPhos22 and the translated peptide sequences of its OFRs (pPhos*orf*1, pPhos*orf*2 and pPhos*orf*3. The nucleotide sequence has been numbered with respect to the orientation of the insert with bp 1 positioned closest to the plasmid-encoded *lac* promoter. Both DNA strands are shown for pPhos*orf*1 and upstream sequences, whose coding strand is opposite to that of pPhos*orf*2-3. The putative promoters (P) upstream of pPhos*orf*1 and pPhos*orf*2 are indicated by arrows. The transcriptional starts, as predicted by promoterScan (Section 2.10) are highlighted in bold. Possible ribosome-binding sites are indicated by double-underlines. Stop codons are indicated by asterisks. The A9 motif encoding the putative +1 translational frameshift between pPhos*orf*s 2 and 3 is double overlined. The vector sequence forming part of the 386-aa peptide fusion with pPhos*Orf*2-3 is shown in red.

Mahillon and Chanler (1998) loosely define an IS element as a small (<2.5kb) segment of mobile DNA with a simple genetic organisation, encoding no functions other than those involved in its mobility. These functional components include *cis*-acting DNA sequences which define the ends of the element and a transposase which catalyses the transposition reaction. The *cis*-acting sequences are usually in the form of terminal inverted repeats which are recognised by the transposase (Figure 6.18).

Transposases are members of the polynucleotidyltransferase superfamily (Grindley & Leschziner, 1995; Rice *et al.*, 1996). They catalyse the endonucleolytic cleavage of the phosphodiester bonds at the ends of the IS element and transfers these ends to a target DNA site. It is this phosphoryl transferase activity which was thought to have been detected in *E. coli* TOP10/pPhos22 (Section 5.7.1). Transposases are generally encoded by one or two ORFs which take up nearly the entire length of the insertion sequence. In the case of two ORFs, a full-length transposase may be generated by fusing the two ORFs *via* a programmed translational frameshift (Chandler & Fayet, 1993; Farabaugh, 1996; Gesteland & Atkins, 1996).

IS elements may be classified according to similarities and differences in their genetic organisation, in the nucleotide sequence of their defined ends and target DNA, in the protein sequence of their encoded transposase and the mechanism of transposition (replicative, conservative or alternative). Because 3' sequences of pPhos*orf*1 and pPhos*orf*2-3 were each truncated during cloning, the IS elements encoding them were incomplete. The terminal inverted repeats and target DNA sequences for each putative IS-element could therefore not be identified. The type of IS element encoding pPhos*orf*1 and pPhos*orf*2-3 was assigned on the basis similarities found within their respective transposases.



not to scale

Figure 6.18: General structure of bacterial IS elements. IS elements contain a central region encoding protein(s) for transposition. The terminal inverted repeats (IRL and IRR) define the ends of the element and contain sequences recognised by the transposition proteins for sequence-specific binding and strand cleavage. The transposase promoter (P) is located partially in the IRL. The target DNA is usually duplicated resulting in the formation of directly repeating sequences (DR) which flank the IS element. Adapted from Mahillon & Chandler (1998).

6.5.3: pPhos*orf*1 gene product has similarities to IS5-like transposase

BLAST searches with the pPhos*orf*1-encoded partial protein showed significant homology with the N-terminal residues of putative transposases (*Tnp*s) from several prokaryotic sources. Highest homologies were with a putative IS5-like transposase from *Sphingomonas* Sp. LB126 (45% identical over 107 aa) (Accession CAB87573), an IS1194-encoded *tnp* gene product from *Streptococcus thermophilus* (44% identical over 101 aa) (Accession CAA73953.1) (Bourgoin *et al.*, 1998) and an IS1168-encoded *tnp* gene product from *Bacteroides vulgatus* (45% identical over 92 aa) (Accession I40597) (Haggoud *et al.*,1994). Both IS1194 and IS1168 are subgroups of the IS5 family. A major feature that defines the IS5 family of insertion sequences is the similarity between their (putative) transposases (Rezsohazy *et al.*, 1993; Mahillon & Chandler, 1998). Like transposases from many IS families and retroviral integrases, the IS5-encoded transposases possess a DDE motif (Fayet *et al.*, 1990; Katzman *et al.*, 1991; Kulkosky *et al.*, 1992). This carboxylate triad is generally located in the C-terminal half of the transposase and is thought to form part of a catalytic pocket which is necessary for phosphoryl cleavage and DNA strand transfer (Khan *et al.*, 1991). Although pPhos*orf*1 gene product has significant homology to the N-terminal sequences of IS5-like transposases, truncation after 114 amino acids deleted any potential DDE motif. Cloning and sequencing of the remainder of pPhos*orf*1 should reveal whether it is equivalent to transposases encoded by IS5 family of insertion sequences.

6.5.4: It is likely that the IS5-like *tnp* gene is not expressed in *E. coli* clone pPhos22

Three possible bacterial promoter sequences were identified upstream of pPhos*orf*1. While it is plausible that transcription can be driven by these sequences, expression of an active protein is most unlikely because the section of the predicted *tnp* gene product thought to contain the catalytic residues for phophoryltransfer had been deleted during cloning. pPhos*orf*1 would therefore not have been responsible for the thermostable phosphatase activity detected during the screening procedure. Based on this information, no further characterisation was carried out on pPhos*orf*1.

6.5.5: pPhos*orf*2-3 gene product has homologies to *Tnp*B encoded by IS605 family of insertion sequences

The conceptually translated sequence of pPhos*orf*2-3 revealed significant homologies to several bacterial transposases belonging to the IS605-*Tnp*B family. Highest homology

was found with several copies of IS605-*Tnp*B from *Helicobacter pylori* including that encoded by the plasmid pHPM186 (34% identical over 271aa) (Accession AAC28361.1). pPhos*orf*2-3 gene product also shared homology with the putative IS605/*Tnp*B-like transposases from *Thermotoga maritima* (32% identity over 262 aa) (Accession AAD36121) and the putative transposase encoded by the IS605 family member IS1341 of the thermophilic bacterium PS3 (30% identity over 217aa) (Murai *et al.*, 1995) (Accession BAA07634).

As mentioned in Section 6.5.3, the carboxylate triad (DDE) motif is highly conserved in many IS families. IS605 family members are exceptions in that they possess no well-defined DDE triad (Mahillon & Chandler, 1998). No IS605 transposase has been isolated and characterised to date and the active site of IS605/TnpB-like proteins has yet to be assigned. It was therefore not possible to reconcile sequence similarities to protein function for pPhos*Orf*2-3. A multiple sequence alignment of pPhos*orf*2-3 gene product against the IS605/*Tnp*B-like transposases of five other bacterial species (Figure 6.19) does however reveal regions of amino acid conservation which may be involved in the mechanistic or catalytic activity of these proteins. Since the pPhos*orf*2-3 was truncated at the cloning site, a fully functional transposase was unlikely to have been expressed in *E. coli* TOP10/pPhos22; however thermostable phosphatase activity was detected in this recombinant *E. coli* strain suggesting that the residues required for this activity were retained on the cloned pPhos22-DNA insert.

```
ORF          1 ---MIVSTRVKLYTNEKQRVLLEKHIGSWRSFYNYFLEKGNEYCLRRKESKKCSLSYIET
HPYLORI      1 ---MLNAIKFRIYPNAQQKELISKHFGCSRVVYNYFLDYRQKQYAKGIKET--YFTMQKV
TMARIT       1 MTKMLRTYKFRIYPTREQEEKLAKHFGHTRFVYNFFLNYANIIYRVMERPTY-YNEWASV
ECOLI        1 -MKRLQAFKFQLRPGGQQECEMRRFAGACRFVFNRALARQNENHEAGNKYIP-YGKMASW
DINODOS      1 ---MLKAYKYRIYPNSEQALLIEKHFGCSRFVFNWALALQKRYYAMFGKSLS-RTQIQSQ
PS-3         1 --MANKAYQFRLYPTKEQEQLLAKTFGCVRFVYNKMLEERIQMFEKFKDDQE-SLKQQTC

ORF         58 QSMVRELRKYQWLYEMNSQSLQMSLRYLDNVFKKFFNKNSEYCGFR-KKGKNNHFAFPQH
HPYLORI     56 LTQIKHQEKYHYLNECNSQSLQMALRQLVSAYDNFFSKRARYPKFKSKKNAKQSFAIPQN
TMARIT      60 LVKLKKTNKYSWLNEVNSQALQQSLKDLERAFKNFFKKQAGYPKFKKKFSRQTFRIPQH
ECOLI       59 LVEWKNATETQWLKDSPSQPLQQSLKDLERAYKNFFRKRAAFPRFK-KRGQNDAFRYPQG
DINODOS     57 LVKKKKTGKFAWLNEVNSQSLLNALLNVYTAFTNFFKGRTKFPRFKSKKILGRSYQCPQH
PS-3        58 PTPAKYKKEFPWLKEVDSLALANAQLNLQKAFQHFFSGRAGFPKFKNRKAKQSYTTNMVN

ORF        117 IKIKGDRIC-------FPEFSEVICLKALKKMSYEIKSINQIVITKEGGDYYCSIIYENG
HPYLORI    116 IEIKTETQT-----IALPKFKEGIKAKLHRELPKDSVIKQAFISCIAD-QYFCSISYETK
TMARIT     120 IQLYIKEDNPKYGCIFVPKFKEGIKVRLHRKLPKDGKIKQATFIKTATNKYYAAIVFEVQ
ECOLI      118 VKLDQENS------RIFLPKLGWMRYRNSRQVTG--VVKNVTVSQSCG-KWYISIQTESE
DINODOS    117 CTVSFEQG------IINLPKIEGIKTVFSREFVG--YIKTVTISKTATGKYYASVLIENS
PS-3       118 GNIKLSDG------YIKLPKLKWIKLKQHREIPAHHIIKSCTITKTKTGKYYISILTEYE

ORF        170 EEPLEGVPMSVENWVGIDTGVEKFATLSDGIGIDNPHFINEVENKINKIQRELSRKQKGS
HPYLORI    170 EPIPKPTIIKK--AVGLDMGLRTLIVTSDKIEYPHIRFYQKLEKKLTKAQRRLSKKVKGS
TMARIT     180 DAEVQNTSTG---ILGIDLGIKDTITLSDGKKYKMP-DLSKYERQIKRLHRRLSRKQRGS
ECOLI      169 VSTPVHPSAS---MVGLDAGVAKLATLSDGTVFEPVNSFQKNQKKLARLQRQLSRKVKFS
DINODOS    169 DILPTPTTVEPNLTVGIDLGINHLLNLSDGSKFDNPKYLANASKRLAVQQKIFARKQKQS
PS-3       172 HQPAPKEVQT---VVGLDFSMSTLYVDSEGKRANYPRFYRKALETLAKEQRKWSRKKKGS

ORF        230 NNWQKTILKIQRNYRKLRRKRDEFLARESTVMVKR-YDSIVFEYP 273
HPYLORI    228 NNRKKQAKKVARLHLACSNTRDDYLHKISNEITNQ--YDLIGVETLNVKGLMR-------
TMARIT     236 KNWEKARLCLAKLYEKIVNIKNDWLHKITHDLVSESQAGKIVVEDLNIKGMVQN------
ECOLI      226 NNWQKQKRKIQRLHSCIANIRRDYLHKVTTAVSKN--HAMIVIEDLKVSNMSKSAAGTVS
DINODOS    229 KNYQKQKLAVARIHEKVRQQRLDLHHKITHSLIYENQATSYALEDLAVKNMVKN------
PS-3        229 NRWHKQRLKVAKLHEKIANQRKDFLHKESHKLAKR--YDCVVIEDLNMKGMSQA------

ORF            ------------------------------------------------------------
HPYLORI    279 ---------TYHSKSLANASWGKFLTMLKYKAQRKAKTLLGIDRFFPSSQLCSYCGFNTG
TMARIT     290 ---------HRLARHIHMQSWRRFLELLEYKAKRCGIEVIKANRYYPSSQMCSECGYINK
ECOLI      284 QPGRNVRAKSGLNRSILDQGWYEMRRQLAYKQLWRGGQVLAVPPAYTS-QRCAYCGHTAK
DINODOS    283 ---------RKLAKAINDVGWGQFVTLLTYKATWYGKNILKVNRFFASSKICSHCHHKLD
PS-3       281 ---------LHFGQGVHDNGWGMFTTFLQYKLVEQGKKLIKIDKWFPSSKTCSCCGRVKE

ORF            ------------------------------------------------------------
HPYLORI    330 KKH-ENITKFTCPHCNITHHRDYNASVNIRNYALGMLDDRHKIKIDKSRVGIIRTDYAHY
TMARIT     341 EVKDLSVREWTCPVCGAHHDRDVNAAKNLVRYGLMLSIGREPSEFTPVDSALAAEPERGL
ECOLI      343 ENR-LSQSKFRCQVCGYTANADVNGARNILAAGHAVLACGEMVQSGRPLKQEPTEMIQAT
DINODOS    334 NLP-LSVRNWTCPSCQTHHDRDTNAASNIRQQALADVAGLATV-----------------
PS-3       332 SLS-LSERTFRC-ECGFESDRDVNAAINIKHEGMKRLAIV--------------------

ORF            ---------------------------------------
HPYLORI    389 TDERIKACGASSNGVISKYGNILDLASYGAMKQEKAQSL
TMARIT     401 RAITG-------------------------------
ECOLI      402 A---------------------------------------
DINODOS        ---------------------------------------
PS-3           ---------------------------------------
```

Figure 6.19: Sequence alignment of pPhos*Orf*2-3 (ORF2-3) and putative IS605-like transposases from *H. pylori* (HPYLORI) (Accession AAC28361.1), *T. maritima* (TMARIT) (Accession AAD36121.1), *E. coli* (ECOLI) (Accession AAC74514.1), *Dichelobacter nodosus* (DINODOS) (Accession AAB16749.1) and thermophilic bacterium PS3 (PS-3) (Accession BAA07634). Red residues indicate conservation in at least three of the proteins while the blue residues represent a conservative substitution. The junction of the pPhos*Orf*2-3 fusion is indicated by double underlining. The DNA insert-encoded protein ends at Pro273; however vector-encoded residues continue for another 114 aa (not shown).

6.5.6: *E. coli* TOP10/pPhos22 cells accumulate 46.0-kDa and 42.5-kDa proteins

To investigate whether any of the peptides encoded by pPhos22 were heterologously expressed in *E. coli* TOP10 cells, SDS-PAGE was performed using *E. coli* TOP10/pPhos22 crude cell extracts. The *E. coli* TOP10/pPhos22 cell extracts possessed two protein bands of increased intensity with molecular masses of 42.5 kDa and 26.0 kDa (Figure 6.20, lane 1) when compared to the negative control, *E. coli* TOP10/ pCR– (lane2). Of these two proteins expressed in pPhos22, the larger band corresponded to the size expected for a fusion protein resulting from the proposed programmed +1 translational frameshift occurring between pPhos*orf*2 and pPhos*orf*3-vector (45.2 kDa). The 26.0-kDa band could have represented the product of pPhos*orf*3-vector (28.2 kDa); however, the possibility of it representing a post-translational cleavage product of the larger protein was not ruled out. No protein band corresponding to that of the pPhos*orf*2 or pPhos*orf*1-vector gene product (17.3 and 14.7 kDa, respectively) was observed. These findings suggested that one or both of these protein bands represents the heterologously expressed gene product(s) that conferred thermostable phosphatase activity to *E. coli* TOP10 cells.

Programmed translational frameshifting is found in several (putative) transposases (e.g. IS1 (Sekine & Ohtsubo, 1989; Luthi *et al.*, 1990; Escoubas *et al.*, 1991), IS 3 family members (IS911, Polard *et al.*, 1991; IS150, Vogele *et al.*, 1991; Sekine *et al.*, 1994; IS2 (Hu *et al.*, 1996) and the IS605/*Tnp*B-like transposase from *Borrelia burgdorferi* (Accession AE000786). Programmed translational frameshifting is one of the ways that insertion elements control the activity of transposition. IS3 expresses a fusion protein (*Orf*AB) by programmed –1 translational frameshifting between two consecutive and partially overlapping ORFs (*orf*A and *orf*B). Evidence suggests that *Orf*AB is the active transposase. In the absence of ribosomal frameshifting, IS3 produces two other proteins, *Orf*A and *Orf*B, in which the *orf*B gene product is synthesised by coupling with translation of *orf*A (Sekine *et al.*, 1994). Active transposase (*Orf*AB) is similarly formed by a programmed –1 translational frameshift in IS911 (Polard *et al.*, 1991). The *orf*A gene product which is formed in the absence of ribosomal shifting (*Orf*A) was shown to strongly stimulate *Orf*AB-mediated intermolecular transposition of IS911 (Polard *et al.*, 1992).

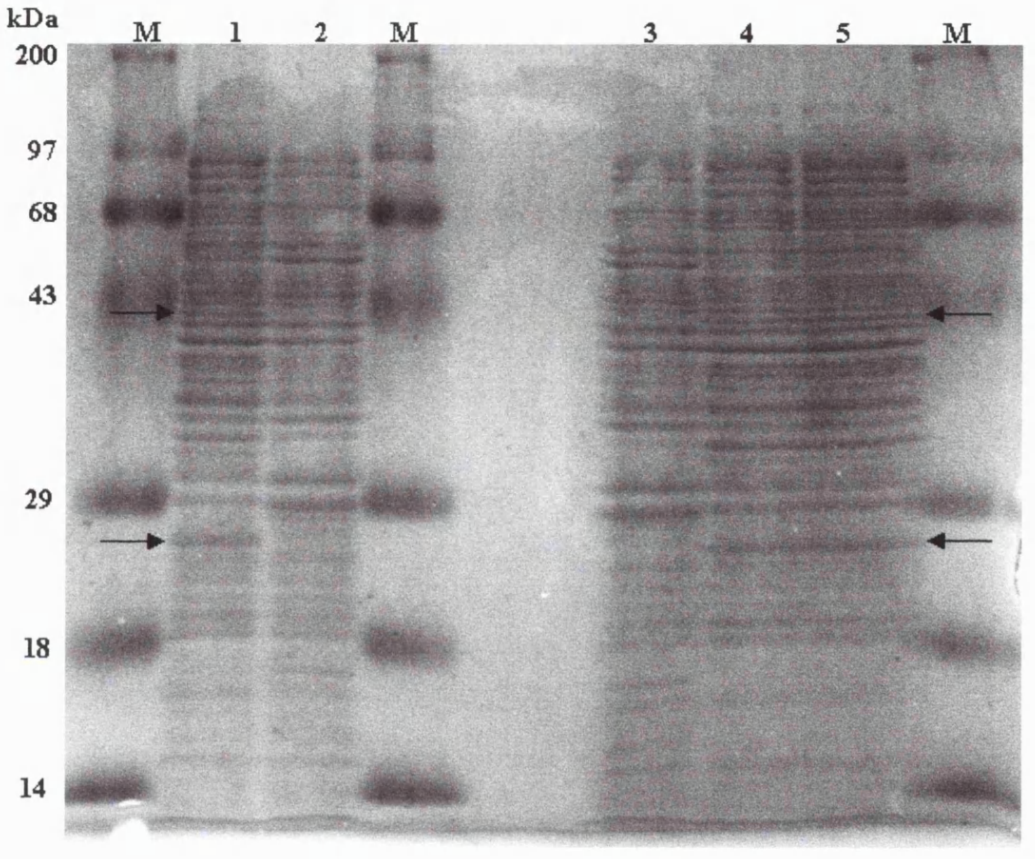


Figure 6.20: Coomassie-stained SDS-15% polyacrylamide gel showing the proteins of crude cell extracts of *E. coli* strains TOP10 pCR-XL-TOPO (lanes 2 & 3) and TOP10 pPhos22 (lanes 1, 4 & 5) Lane 1, 4,and 5 represent three independently isolated clones of pPhos22. Molecular mass standards (M) Sizes (kDa) are indicated on the left. The overexpressed protein bands corresponding to pPhosorfOrf2-3-vector fusion (42.5 kDa) and pPhosOrf3-vector fusion (26.0kDa)are indicated by arrows.

A similar arrangement is found in the unrelated IS1 except that no downstream product analogous to IS3-*Orf*B is produced and, unlike the IS911 protein, the product of the upstream frame (*Ins*A) binds specifically to the ends of the element and acts as a repressor of transposition (Machida *et al.*, 1984; Machida & Machida, 1989). The IS1-encoded fusion protein (*Ins*A*Ins*B) is thought to function as the transposase (Escoubas *et al.*, 1991). For IS3, IS911 and IS1-encoded transposase fusions, the putative phosphoryltransferase catalytic domain (DDE triad) is encoded by the downstream *orf*B (Mahillon & Chandler, 1998).

The apparent frameshift window located between pPhos*orf*2 and pPhos*orf*3, along with similarities of genetic organisation with known insertion elements that employ programmed translational frameshifting as a mechanism for controlling transposition activity, suggested that pPhos*orf*2-3 gene sequences may employ programmed translational frameshifting as a similar means of controlling transposition. It was proposed that the 46.0-kDa protein expressed in *E. coli* TOP10/pPhos22 had arisen from a +1 frameshift and that the pPhos*orf*2-3 sequences represented the transposase. As stated previously, IS605 family members do not possess the well-defined DDE triad found in most transposases. Because no mechanistic, biochemical or structural information on IS605/*Tnp*B-like transposases was available, the roles of pPhos*orf*2, truncated pPhos*orf*2-3 and truncated pPhos*orf*3 gene products could not be inferred.

6.5.7: Biochemical characterisation of thermostable phosphatase-positive *E. coli* TOP10/pPhos22

In an effort to characterise the phosphatase activity conferred upon *E. coli* TOP10 cells, a thermostability study was performed using *E. coli* TOP10/pPhos22 (Section 2.11). Phosphatase thermostability was assessed by incubating aliquots of crude *E. coli* extracts at various temperatures (60°C, 70°C, 80°C, 90°C and 100°C) for 15 minutes then transferring 75μl of each incubation to fresh tubes containing 15μl 0.5M Tris-Cl (pH 8), which had been pre-equilibrated to 60°C. To start the reaction 10μl of 30mM PNP-phosphate was added and the sample was incubated at 60°C. Incubations were inspected periodically for the development of a yellow colour indicating the presence of phosphatase activity. For each incubation the rate of colour change observed for *E. coli* TOP10/pPhos22 was identical to that for the negative control. This suggested that the release of *p*-nitrophenolate was due to nonspecific hydrolysis of PNP-phosphate and not

due the heterologously expressed proteins encoded by pPhos22. Repeating the procedure with serially diluted aliquots gave similar results. Repeating the microtitre plate assay on the three original clones which had been stored in sub-master plates at -80°C confirmed that the originals were active for phosphatase activity. Culturing these recovered clones however resulted in the loss of phosphatase activity as described above. Furthermore, even though these clones contained recombinant plasmids with DNA insert of the expected size (1.7 kb), repeating SDS-PAGE on these samples resulted in the loss of the heterologously expressed protein bands identified in these clones in Section 6.5.6. Several attempts at recovering an active clone failed and resulted in the depletion of the sub-master and master plates. Therefore no further characterisation of *E. coli* TOP10/pPhos22 could be performed.

6.6: Summary

As described in Chapter 5, three recombinant *E. coli* clones were isolated during expression-screening of environmental DNA libraries. Two of these clones, *E. coli* TOP10/p5ICE16 and *E. coli* TOP10/p6ICE16, displayed an atypical, dark phenotype during amylase detection on starch agar plates. The third clone, *E. coli* TOP10/pPhos22, was isolated during thermostable-alkaline-phosphatase screening. In an effort to identify the genes and proteins responsible for these phenotypes, the DNA inserts of p5ICE16, p6ICE16 and pPhos22 were sequenced and compared to nucleotide and protein databases at NCBI.

The 1.7-kb fragments from p5ICE16 possessed sequences encoding a partial putative RecA-like protein, a conserved hypothetical protein and an unknown gene product thought to be involved in methionine biosynthesis. The 4.7-kb fragments from p6ICE16 was shown to encode a putative prokaryotic site-specific recombinase, a tRNA-Gly gene and an apparent *hem* operon involved in the early steps of tetrapyrrole biosynthesis. Identification of the genes responsible for imparting the dark phenotype to *E. coli* TOP10/p5ICE16 and *E. coli* TOP10/p6ICE16 could not be inferred directly from database sequence annotations.

Since the gene encoding the conserved hypothetical protein possessed bacterial promoter-like elements in upstream sequences, it was considered a likely candidate for heterologous expression in *E. coli* TOP10/p5ICE16. Because the gene contained

homologies to conserved but uncharacterised proteins, biological activity could not be used as a means to characterise the clone further.

In *E. coli* TOP10/p6ICE16, the putative *hem* operon was considered to be the likely cause of the dark phenotype. It was preceded by a putative bacterial promoter and homologous *hem* sequences had been shown to be involved in the biosynthesis of macrocyclic pigments. It was thought that overproduction or accumulation of the pigment precursors resulted in the observed phenotype. An *E. coli hem*A auxotroph, however, could not be rescued when transformed with p6ICE16, suggesting that the putative *hem*A gene and, therefore, *hem* operon were not being expressed.

SDS-PAGE of *E. coli* TOP10/p5ICE16 and *E. coli* TOP10/p6ICE16 crude cell extracts could not identify any heterologously expressed proteins attributed to the cloned inserts.

Even though cloned Ice16-DNA was shown to confer a dark atypical phenotype to *E. coli* TOP10 cells (Section 5.4.2), no gene or gene product could be identified as the determinant(s) responsible for the observed phenotype of *E. coli* TOP10 strains p5ICE16 and p6ICE16.

Sequence analysis of the 1.7-kb fragment derived from phosphatase-positive *E. coli* TOP10/pPhos22 identified two putative transposases belonging to different families of prokaryotic insertion sequences. It was thought that the phosphoryl transferase activity of the putative IS605-tnpB-like transposase was what had been detected during library screening. SDS-PAGE of *E. coli* TOP10/pPhos22 crude cell extracts identified protein bands thought to correspond to alternatively translated products encoded by the IS605/tnpB-like gene. Studies were carried out to investigate the thermostability of the cloned phosphatase; however, all phosphatase activity was lost whenever *E. coli* TOP10/pPhos22 was scaled up for investigation. Furthermore, the heterologously expressed proteins previously identified by SDS-PAGE were lost from this recombinant. The originally isolated clone was reconfirmed as being phosphatase positive, however, activity was lost each time the recombinant was propagated.

Chapter 7
Discussion

7.1: Aims

The main aim of this research project was to investigate the potential of expression-cloning in detecting thermostable proteins encoded by DNA derived from uncultivated thermophilic microorganisms. This was achieved by: (1) analysing, with respect to DNA yield, shearing and purity, the suitability of extraction procedures for obtaining clonable DNA from geothermal sediments; (2) investigating various cloning protocols and generating stable representative environmental libraries; (3) demonstrating the applicability of the environmental DNA libraries by performing various screening assays for detecting heterologously expressed thermostable proteins; (4) performing molecular and expression analyses on isolated recombinants in order to identify the molecular determinants responsible for the phenotype(s) that were observed during expression-screening.

7.2: Direct Extraction of Environmental DNA from Geothermal Sediments

7.2.1: Comparison study

Chapter 3 describes the comparison study of two methods for the direct extraction of DNA from geothermal sediment (Ice22, ~58°C, pH 4.3). A mortar and pestle plus SDS (MPS) protocol (modified from Saano & Lindstrom, 1995) and a bead beating (BB) method (Bio101) for DNA extraction were compared with respect to DNA yield, purity and fragment size.

Table 7.1 summarises the findings of this study. The objective of this comparison was to select an extraction protocol that recovered sufficient amounts of DNA the quality of which was suitable for use in cloning procedures. In considering the merits of DNA extraction protocols the most important include the amount of effort involved in the process, the quality of the DNA recovered and the amount of DNA recovered.

One of the initial advantages of the MPS DNA extraction protocol was that it required only routine laboratory equipment and reagents The BB method on the other hand required a dedicated bead beater and proprietary DNA extraction kit and reagents. Because it involved more manual steps, the MPS method, however, was considerably more tedious than and not as reproducible as the BB method for extracting DNA.

Both extraction procedures recovered DNA that was sufficiently pure for cloning as determined by spectrophotometric analysis (Section 3.3.2). Although the BB method was more shearing to DNA than the MPS method (Section 3.3.3), both procedures recovered DNA whose fragment-size distribution was suitable for cloning. Although the MPS method could recover more DNA per g dry sediment (Section 3.3.1), the BB method was superior in terms of reproducibility. Because it could accommodate 8 X 0.5-g samples simultaneously, the BB method was easily scaled up and required less time to extract multiple samples than the MPS method.

Both DNA extraction procedures were considered suitable for cloning in terms of DNA yield, purity and fragment size. When selecting one of these protocols for recovering DNA for cloning purposes, one must balance the need for high yields against reproducibility and speed of operation. DNA recovered by both methods was used successfully as input DNA for cloning protocols as described in Chapter 4.

7.2.2: Impact of isolation techniques on diversity of recovered DNA

A primary assumption in this work was that greater DNA recovery reflected a more representative (diverse) sample of DNA from the microbial community. While DNA yield is not the best way to estimate diversity, the use of quantitative measures to estimate the extent and bias of cell lysis or to quantitate the level of genetic diversity of recovered DNA would have been very time consuming. Miller *et al.* (1999) reported that, although a lysozyme/SDS/freeze-thaw treatment recovered less DNA per g dry sediment than a SDS/bead-mill homogenisation treatment, the lysis efficiencies of the two procedures were essentially the same (~65%). Microscopic examination of suspensions of *Bacillus subtilis* endospores has indicated that SDS/bead-mill homogenisation, with 2% survival of endospores, was substantially more effective at endospore disruption than an SDS/lysozyme/freeze-thaw treatment (94% endospore survival) (More *et al.*, 1994).

The findings of these studies on quantitative cell lysis suggest that, although the lysis efficiencies of freeze/thaw and bead-beating methodologies may be the same, the freeze-thaw treatment may be biased against those cells that are not easily lysed.

Because bead-beating is more efficient at lysing endospores and, presumably, other resistant cell structures (More *et al.*, 1994), it is possible that the DNA recovered *via* the BB method (Section 2.6.2) may have been more representative of the microbial community than that of the MPS method (Section 2.6.1). This could be true even though the DNA yield using the BB method was lower than that of the MPS method. It should be noted however that bead-mill homogenisation is itself biased against a resistant fraction (~4%) of cells, which are small and coccoid, that are present within sediment (More *et al.*, 1994).

Method	MPS (*n*=5)	BB (*n*=3)
Sample size (g)	10	0.5
Time required to extract one sample	2-3 days	2-3hours
DNA yield (μg DNA/gws)	3.0±2.3	0.35±0.03
DNA yield (μg DNA/gds)	13.2±9.6	1.5±0.12
A260/A230[a]	2.2±0.43	0.06±0.004[b]
A260/A280[a]	1.8±0.15	1.9±0.19
DNA fragment size (kb)	<0.5 to >23	~0.5 to ~10

Table 7.1: Comparison of mortar and pestle plus SDS (MPS) and bead beating (BB) methods for isolating DNA from Ice22 sediment (Chapter3). Yields and absorbance ratios are the mean ± standard deviation; gws, g wet sediment; gds, g dry sediment; a) DNA solutions were considered pure if the A_{260nm} to A_{230nm} ratio was between 1.8 and 2.3 and the A_{260nm} to A_{280nm} ratio was between 1.5 and 2.0. (Marmur, 1963); b) the silica used in the purification columns interfered with spectrophotometric measurement of DNA (Section 3.3.2).

When obtaining environmental DNA for use in library construction, it must also be kept in mind that there is a remarkable complexity of soil and sediment types as well as microbial community structures. Such complex samples may introduce multiple factors that may affect the performance of a DNA extraction protocol. While the findings of Miller *et al.* (1999) and More *et al.* (1994), may have provided a guideline for estimating the diversity of the DNA recovered *via* MPS and BB methods, detailed analysis of cell lysis efficiencies must be conducted with these protocols in order to characterise the extent of any bias that may be occurring with these treatments.

Future investigations may include determining the lysis efficiencies of MPS and BB treatments. This may be achieved *via* direct microscopic counts of cells within geothermal sediments, obtained before and after extraction, using acridine orange (e.g. Cullen & Hirsch, 1998), 4:6-diamidino-2 phenylindol (DAPI) (e.g. Miller *et al.*, 1999) or 5-(4,6-dichlorotriazin-2-yl) amino fluorescein (DTAF) (e.g. Zhou *et al.*, 1996).

Denaturing/Temperature gradient gel electrophoresis (D/TGGE) of environmental DNA, that has been amplified with, say, universal, Domain- or Genus-specific PCR primers, is another means to determine whether the MPS and BB extraction treatments are selective (e.g. Krsek & Wellington, 1999). If any differences in the D/TGGE band pattern is observed between different extraction treatments, then the diversity of the extracted DNA can be thought to be affected by the different treatments.

There are other reasons to further explore the diversity of extracted DNA. For example, when a direct lysis method is used to extract bacterial DNA from soil and sediment, it is possible that DNA of eukaryotic organisms (e.g. fungi, algae, plants), might also be extracted. In this work it has been assumed that the DNA extracted from New Zealand and Iceland is predominantly prokaryotic. Two arguments support this assumption. First, with few exceptions (e.g. Chevaldonne *et al.*, 1992), thermophilic eukaryotes including fungi and algae cannot survive at temperatures greater than ~60°C. With the exception of Ice3, no sediment sample collected in the course of this work was obtained from a site that was below 50°C. In many cases, sediments were collected from biotopes at temperatures greater than ~60°C (Tables 3.4 & 3.6) suggesting that only prokaryotic DNA was sampled. Second, due to their small size, bacteria have the highest number of genomes per unit biomass. Even though bacterial genomes are smaller than those of fungi and algae, it is likely that in most soils, and presumably sediments, the

predominant amount of DNA is bacterial (Torsvik *et al.*, 1995). In future work, the relative amounts of prokaryotic DNA obtained by various extraction methods can be determined *via* quantitative hybridisation studies of the environmental DNA using prokaryote- and eukaryote-specific probes (e.g. Leff *et al.*, 1995).

7.3: Evaluation of Cloning Protocols and Preparation of Environmental DNA Libraries

Chapter 4 describes the evaluation of various cloning protocols for generating environmental libraries using DNA extracted directly from Iceland and New Zealand geothermal sediments. Cloning protocols were evaluated based on cloning efficiency (cfu/ μg vector), recombination efficiency (% of transformants containing recombinant vector) and the number of recombinants generated per ligation reaction. *E. coli* was selected as the host for environmental expression libraries because it is one of the most genetically characterised bacteria and many cloning protocols are based on this organism. The cloning vectors investigated in this work were pUC19 (for cohesive-end cloning) (Section 4.3), pT7Blue (blunt-end cloning) (Section 4.4), pCR-XL-TOPO (T/A cloning) (Section 4.5) and λTriplEx (phagemid cloning) (Section 4.6).

As a result of this study, the T/A cloning protocol was selected for constructing environmental libraries because the TOPO-TA-cloning reaction (Section 2.8.9) generated more recombinants per ligation reaction than any other cloning procedure investigated (Table 4.2). This protocol also accommodated DNA fragments regardless of the type of termini they possessed. This was important because DNAs extracted from Iceland and New Zealand geothermal sediments generally were sheared to ~10kb (Section 3.4). Preparing this relatively low-molecular-weight DNA for cohesive-end cloning by restriction digestion was thought to generate fragments possessing heterogeneous and thus unclonable ends.

pCR-XL-TOPO was considered an appropriate vector for expression cloning because it possesses many of the features found in vectors used for expression cloning. For example, like other pUC-based cloning vectors (Vieira & Messing, 1982; Yanisch-Perron *et al.*, 1985), pCR-XL-TOPO can be maintained stably at high copy number in the *E. coli* host and relatively high expression can be achieved through gene dosage.

pCR-XL-TOPO also encodes promoter and ribosome-binding sequences immediately upstream of the cloning site. This feature was considered advantageous for expression-cloning because read through transcription, and possibly translation, driven from vector-encoded control sequences may have provided an alternative route to heterologous gene expression.

Designated ICE16 and ICE22, two environmental libraries were constructed using DNA extracted from Ice16 and Ice22 geothermal sediments, respectively (Section 4.7). The quality of the two environmental library was assessed by determining the cloning efficiency (cfu/μg vector), the mean insert size and the percentage of transformants that contained cloned inserts (Table 7.2).

Library	Source	Cloning Efficiency (X 10^6cfu/μg vector)	% Recombinants	Insert Size (kb)	Independent Clones[a]
ICE16	Sediment (~70°C, pH 9.5)	1.4	84	~5	37000
ICE22	Sediment ~55°C, pH 4.3	0.78	84	~5	33000
pBluescript	Soil	0.017[b]	80	~6.5	930000

Table 7.2: Characterisation of the two environmental DNA libraries derived from Iceland geothermal sediments using pCR-XL-TOPO as cloning vector (Section 4.7). Also included is the description of an environmental expression library derived from uncultivated soil microorganisms using pBluescript as cloning vector (Henne *et al.*, 1999). **a:** Defined as the number of recombinants present prior to library amplification. **b:** estimated from the cloning information provided by the investigators and by assuming that, at an insert-to-vector molar ratio of 3:1, 50ng of vector was used per ligation.

There is limited information in the literature concerning the efficiencies of cloning protocols for the construction of environmental DNA libraries. When preparing expression libraries using DNA derived from soil, Henne *et al.* (1999) obtained ~2500 transformants per μg of isolated soil DNA. The soil metagenome libraries contained ~80% recombinant clones with an average insert size of ~6.5kb (Table 7.2). Assuming an insert-to-vector molar ratio of 3:1, this corresponds to ~17000 transformants per μg vector, a value that is in concordance with some of the efficiencies reported for protocols evaluated in this work (Chapter 4, Table 4.2). The cloning efficiencies of ICE16 and ICE22 DNA libraries were actually better than that estimated for the soil DNA library constructed by Henne *et al.* (1999). This finding provides supporting evidence that TOPO-TA cloning is a valid option for use in constructing environmental DNA libraries.

The soil DNA libraries constructed in Henne's laboratory contained 930000 unique recombinant clones, whereas ICE16 and ICE22 contained only 37000 and 33000 unique recombinants, respectively. This suggests that more recombinant clones were required for ICE16 and ICE22. Scaling the size of the Ice-DNA libraries up to, say, 930000 unique recombinant clones would have required a considerable amount of time and resources. For example, ~80 ligation reactions would have to have been conducted in order to scale the size of ICE16 up to 930000 unique clones. This calculation is based on a cloning efficiency of 1.4 X 10^6 cfu per μg vector and the library containing 84% recombinants, each possessing a 5-kb insert. Since pCR-XL-TOPO comes supplied in a proprietary cloning kit, such a scale-up would have been quite expensive and time consuming. It is not clear exactly how 930000 unique recombinant clones were generated for the soil DNA libraries prepared in Henne's laboratory (Henne *et al.*, 1999), but with such a low cloning efficiency (2000 recombinants per μg isolated soil DNA), a considerable amount of time and resource were probably also required. Environmental DNA libraries smaller than ICE16 and ICE22 have been successfully employed in detecting enzyme activities through expression screening (Table 7.3) (Ronan *et al.*, 20000). This suggests that ICE16 and ICE22 contained sufficient numbers of clones for use in expression screening and, although beneficial, further scale-up of library size was not essential.

7.4: Screening Environmental Libraries for Thermostable Enzyme Activities

As described in Chapter 5, the potential of ICE16 and ICE22 expression libraries was assessed by testing the accessibility of heterologous enzyme activities. ICE16 and ICE22 environmental libraries, derived from uncultivated thermophilic microorganisms (Section 4.7), were screened for thermostable α-amylase, lipase, phosphatase and protease activities using both microtitre plate and indicator plate platforms (Section 2.9).

Out of ~100000 to 132000 clones screened per library per assay (~1.4 million clones total) (Sections 5.4 to 5.7), thermostable phosphatase was detected in one clone (*E .coli* TOP10/ pPhos22) during microtitre plate screening of ICE22. This result demonstrates that heterologous sequences, encoded by DNA derived from geothermal sediment and cloned into pCR-XL-TOPO, can be expressed in *E. coli* at detectable levels. Although no other enzyme activity was detected in either library using the methods described (Section 2.9), two recombinant clones (*E. coli* TOP10/p5ICE16 and *E. coli* TOP10/ p6ICE16), displaying an atypical dark phenotype were isolated from ICE16 (Section 5.4.2).

Table 7.3 summarises ICE16 and ICE22 along with other environmental DNA libraries employed for expression screening. Expression libraries are compared here with respect to the type of cloning vector used, the number of 1-kb genes represented per library, the number of different activities detected per library and the frequency at which these unique activities occurred.

7.4.1: Vectors for expression-screening

pCR-XL-TOPO, the vector used to construct ICE16 and ICE22, is based on pUC19. pBluescript, which was used to construct environmental DNA libraries derived from uncultivated soil (Henne *et al.*, 1999; Henne *et al.*, 2000) and marine (Cottrell *et al.*,1999) microorganisms, is also based on pUC19. Other vectors that have been used to construct environmental DNA libraries include phage λZAP, for cloning DNA from uncultivated marine microorganisms (Cottrell *et al.*, 1999), and a bacterial artificial chromosome (BAC), for cloning soil metagenomic DNA (Rondon *et al.*, 2000). Any other *E. coli*-based vector (e.g. phage M13, F-factor-based cosmid) could theoretically

be used for expression screening of environmental DNA libraries provided that it can be replicated and stably maintained in the host. To date, there have been no reports in the literature describing alternative prokaryotic or eukaryotic cloning systems for heterologous expression of environmental DNA sequences. Using various prokaryotic hosts with different genetic and metabolic backgrounds for expression-cloning may increase the likelihood of detecting functional gene products encoded by metagenomic DNA. Future work may include cloning environmental DNA into species-specific or broad host-range vectors for transformation and screening of alternative host species.

Although a primary assumption of this work was that the environmental DNA cloned from geothermal sediments was predominantly prokaryotic, it is possible, especially at the lower biotope temperature range (50 to 60°C), that eukaryotic DNA was also cloned. Functional expression of a eukaryotic gene in *E. coli* TOP10/pCR-XL-TOPO is very unlikely because the eukaryotic ORF would have to have been cloned in the proper orientation and reading frame with respect to the vector-encoded *lac* promoter and ribosome-binding site. Typically, eukaryotic post-transcriptional processing, such as transcript splicing, and co/post-translational modification such as trafficking and glycosylation, cannot be achieved in *E. coli*. Heterologous expression of eukaryotic metagenomic DNA must therefore require a eukaryotic cloning system. Microbial eukaryotic expression systems, such as *Saccharomyces cerivisiae*, *Pichia pastoris* and *Aspergillis* sp., are likely alternatives for cloning and expressing environmental DNA derived from eukaryotes.

An alternative way to achieve heterologous expression of cloned environmental nucleic acids derived from eukaryotes, is through cDNA-cloning. Future work may involve isolating mRNA from uncultivated microorganisms present within high-temperature biotopes and preparing cDNA for cloning into *E. coli* or yeast expression systems.

7.4.2: Library size

Given an average of 1 kb per gene, the number of genes that are represented by an environmental DNA library can be estimated by multiplying the number of unique recombinant clones by the average insert size (kb) (Table 7.3). ICE16 and ICE22 catalogued approximately 170000 genes which is in concordance with other environmental DNA libraries reported in the literature. The number of genes represented by various unamplified environmental expression libraries ranges from

97000, for a soil metagenome BAC library (Rondon *et al.*. 2000), to 2.6 million for soil DNA cloned into pBluescript (Henne *et al.*, 1999).

7.4.3: Detection of gene products encoded by cloned metagenomic DNA

Expression libraries may also be evaluated by the frequency at which unique activities or phenotypes are detected (Table 7.3). The dark phenotypes identified for two clones from ICE16 (Section 5.4.2) were detected at a frequency of 1 for every 90000 genes cloned. Thermostable alkaline phosphatase activity was detected in ICE22 at a frequency of 1 in 160000 genes cloned. Of the environmental DNA expression libraries reported in the literature, the highest frequency of detection was achieved with a BAC library (1 in 8000, Rondon *et al.*, 2000). The lowest frequency of detection occurred with marine environmental DNA cloned into λZAPII (1 in 38 million genes; Cottrell *et al.*, 1999). The frequency of detecting expressed proteins may potentially be increased by increasing the number of different assays performed. Future work for ICE16 and ICE22 may include implementing additional expression assays in order to increase the likelihood of detecting heterologously expressed activities.

What are some of the reasons for not detecting more enzyme activities during expression screening? Perhaps the most straightforward explanation would be that the genes encoding the target proteins were not cloned. The Clarke and Carbon (1976) formula (Equation 5.1) can be used to estimate whether ICE16 and ICE22 libraries possesses a target gene and is repeated here as Equation 7.1.

$$N = \frac{\ln(1-p)}{\ln\left(1-\frac{x}{y}\right)} \quad \text{(Equation 7.1)}$$

where, N= the number clones to screen in order to have a certain probability (p) of finding a single-copy gene in a DNA population
x = the size of the DNA fragment
y = genome size

Assuming that the DNA extracted from the geothermal sediments was prokaryotic with a genome size of 2 X 10^6 bp and that all genomes present within the sediments were equally represented in the libraries, then, given an insert size of 5kb, 1839 clones would have to have been be screened in order to have a 99% likelihood (p=0.99) of finding the target gene. In other words, the expected frequency of this target gene is 1 clone in 1839.

Environmental source	*E. coli* cloning vector	No. of clones screened [a]	Insert size (kb)	No. of genes represented (X 10^6) [b]	Activities [c]	Frequency [d]	Reference
Geothermal sediment (Ice16)	pCR-XL-TOPO	37000 (U)	~5	0.18	2 (5)	1/90000	This work
Geothermal sediment (Ice22)	pCR-XL-TOPO	33000 (U)	~5	0.16	1 (5)	1/160000	This work
Coastal seawater	λZapII	750000 (A)	~5	3.8	1 (2)	1/38000000	Cottrell *et al.* (1999)
Coastal seawater	pBluescript	230000 (A)	~5	1.2	13 (1)	1/92000	Cottrell *et al.* (1999)
Estuarine water	λZapII	75000 (A)	~5	0.30	9 (1)	1/33000	Cottrell *et al.* (1999)
Meadow soil	pBluescript	340000 (U)	~6.5	2.2	28 (1)	1/79000	Henne *et al.* (1999)
Soil from sugarbeet field	pBluescript	190000 (U)	~6.5	1.2	1 (1)	1/1200000	Henne *et al.* (1999)
Soil from Neime river valley	pBluescript	400000 (U)	~6.5	2.6	7 (1)	1/370000	Henne *et al.* (1999)
Soil	BAC	3600 (U)	27	0.097	12 (11)	1/8000	Rondon *et al.* (2000)
Soil	BAC	25000 (U)	44.5	1.1	29 (1)	1/38000	Rondon *et al.* (2000)

Table 7.3: Summary of some environmental DNA expression libraries. **a:** the number of clones screened per assay; U is unamplified library, A is amplified library. **b:** given an average of 1kb per gene; number of clones multiplied by insert size (kb). **c:** the number of unique activities or phenotypes detected per assay. The number of different assays performed are given in parentheses. **d:** the number of unique activities divided by the total number of genes represented.

Thus ICE16 and ICE22, respectively, should contain ~20 and ~18 unique recombinants that possess the target sequence. Whether these target genes are expressed and detected depends on several factors.

If heterologous expression of the target gene relies on transcription being initiated from the vector-encoded *lac* promoter, then the foreign gene must be inserted in the proper orientation with respect to the promoter. pCR-XL-TOPO does not encode a *lac* repressor protein (*Lac*I^q). Because 'leaky' initiation from the *lac* promoter is sufficient for transcription of appropriate sequences, the addition of IPTG for induction was not required (TOPO-XL-PCR Cloning Kit instruction manual, Invitrogen).

Whether encoded by vector or heterologous sequences, transcriptional control sequences are subject to trans-acting effector molecules (e.g. repressors), the presence of which, may down regulate initiation of transcription. If such trans-acting repressors were present during growth of recombinant *E. coli* TOP 10/pCR-XL-TOPO strains, then heterologous transcription may have been down regulated to a point were target gene product was not produced and, as a consequence, not detected.

If translation of the foreign transcript relies on the pCR-XL-TOPO-encoded ribosome-binding site, the target gene must be fused to the vector's *lac*Zα protein-coding region in the proper reading frame. Because this is such a specific requirement, it is thought that such an event would be unlikely.

If heterologous expression is to rely on transcriptional as well as translational control sequences encoded by the foreign gene, then the *E. coli* TOP10 host transcriptional and translational machinery must be able to recognise these sequences. For example, the foreign promoter must contain consensus –35 and –10 DNA elements that can be recognised by the host RNA polymerase holoenzyme (TTGACA and TATAAT, respectively for $E\sigma^{70}$ promoters) (Gross *et al.*, 1992). The foreign ribosome-binding site must also be recognised by the *E. coli* ribosome. Ideally this should consist of a SD consensus sequence (AGGAGG) located ~6-12 nucleotides upstream of an AUG start codon.

The heterologous transcript must also be free of secondary structures that may cause premature transcriptional termination or translational attenuation. The cloned transcript must also lack rare codons that may reduce the rate of translation.

If transcription and translation of the foreign gene is successfully achieved, then the protein must fold properly as well as be co- and post-translationally processed. The failure of the recombinant *E. coli* host to achieve this may result in an unfolded or inactive target protein that cannot be detected during expression screening. Furthermore, the target protein must neither be toxic to the *E. coli* TOP10 host nor be degraded by the host's proteolytic enzymes.

Finally, if a foreign gene is successfully expressed and the encoded protein is fully functional, the target activity may still be missed simply as a result of the assay method. Substrate specificity, cofactors, metal ions, pH, temperature and the presence of inhibitors are all factors that may contribute to the inability of an assay to detect a target activity.

Future work could concentrate on developing assays in which multiple substrates or other constituents are tested simultaneously. Screening ICE16 and ICE22 with different substrates, tested under various assay conditions, may increase the likelihood of detecting target gene products. Indeed, the prototype BAC library (Table 7.3) (Rondan *et al.*, 2000) which had the lowest number of genes cloned, actually had the highest frequency of detection presumably because more activities were screened than any other library.

7.5: Sequence and Expression Studies

As described in Chapter 6, molecular and expression analyses were performed on recombinant *E. coli* clones recovered from ICE16 and ICE22 environmental libraries. Two clones, *E. coli* TOP10/p5ICE16 and *E. coli* TOP10/p6ICE16, were identified in ICE16 library during amylase detection on starch agar plates (Section 5.4.2). Although these clones did not possess amylase activity, they did display an atypical dark phenotype after a heat inactivation step and upon addition of iodine solution. A third clone, *E. coli* TOP10/pPhos22, was detected in ICE22 library during screening for thermostable phosphatase activity (Section 5.7.1).

7.5.1: Sequence studies

In an effort to identify the gene(s) conferring the observed phenotypes, the DNA inserts of plasmids p5ICE16, p6ICE16 and pPhos22 were sequenced and compared to the sequences in the NCBI databases (Section 2.10). The physical maps of the p5ICE16-,

p6ICE16- and pPhos22-DNA inserts are shown in Figure 7.1. The coding sequences identified for the cloned DNA fragments are summarised in Table 7.4.

There are several lines of evidence indicating that the isolated environmental DNA fragments were derived from thermophilic prokaryotes.

i. In all cases, the putative genes identified in this work possessed highest homologies to bacterial and/or archaeal sequences. Except for p5*orf2* and tRNA-Gly, high-scoring homologues were from thermophilic or extremophilic species.

ii. The homologues identified for p5*orf3* (*YitJ* methionine biosynthetic gene), p6*orf1* (site-specific recombinase gene), pPhos*orf1* and pPhos*orf2*-3 (transposase genes) are each specific to prokaryotes.

iii. A putative operon encoding *hem* genes was identified in p6ICE16. Operons or similar gene clusters are features specific to prokaryotes.

iv. In bacteria, the mRNA element that directs the initiation of translation is the ribosome binding site. This includes an initiation codon (AUG, GUG, UUG, AUU or AUA) and a Shine-Dalgarno sequence (AGGAGGU) appropriately spaced (9±3 nt) upstream to the translational start (Stormo, 1986). Archaeal mRNAs possess similar elements for translation initiation (Dennis, 1997). All ORFs identified in the cloned DNA-inserts of p5ICE16, p6ICE16 and pPhos22 were preceded by putative ribosome binding sites indicating that the environmental DNA fragments were prokaryotic.

v. Appropriately placed bacterial-like but not archaeal-like promoter elements were identified in the cloned DNA fragments. It is therefore likely that the isolated clones were derived form bacterial sources.

7.5.2: Expression studies

7.5.2.1: p5ICE16 and p6ICE16

p5*orf2*, encoded by p5ICE16, was the likely candidate for conferring the dark phenotype to *E. coli* TOP10. It possessed a moderate SD sequence 13nt upstream to an ATG start codon and was located downstream of a bacterial-like promoter sequence. p5*orf1* and p5*orf2* (p5*met*X) were not considered likely candidates because the former was severely truncated at the cloning sight and the latter was positioned in the wrong orientation with respect to the vector-encoded *lac* promoter. Because p5*met*X appeared to lack its own promoter, heterologous expression was not likely.

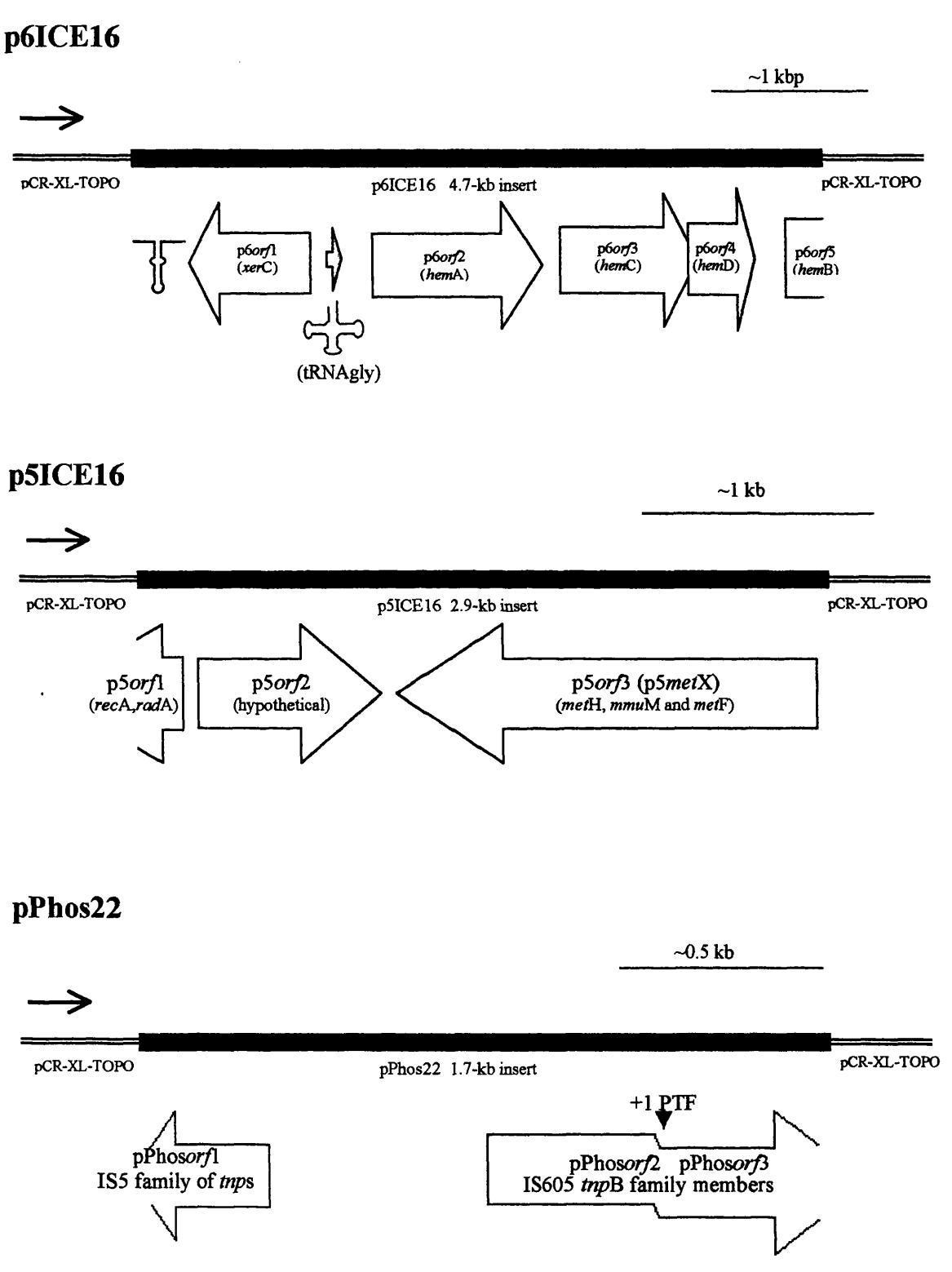


Figure 7.1: Physical maps of DNA inserts (black bars) from p5ICE16 (Section 6.3), p6ICE16 (Section 6.4) and pPhos22 (Section 6.5). Block arrows show the relative localisation of each gene and the orientation of coding sequences found. Truncated ORFs are indicated by incomplete arrows. The genes having homologous sequences to the coding sequences identified are shown in parentheses. The black arrow indicates the direction of transcription from the vector-encoded *lac* promoter. Vector sequences are indicated by double lines. Maps are at different scales. For detailed descriptions see the respective sections.

DNA Insert/ Coding sequence	Size of product	Putative gene	Organism	Identity of gene product
p5ICE16/				
p5*orf*1	trunc. 37 aa	RecA Recombinase	*Aquifex aeolicus*	55% over 37 aa
p5*orf*2	251 aa	Hypothetical	*Bacillus subtilis*	24% over 215 aa
p5*orf*3	612 aa	Methionine biosynthesis	*Bacillus subtilis*	36% over 612 aa
			Thermotoga maritima	N 27% over 401 aa
				C 32% over 278 aa
p6ICE16/				
p6*orf*1	312 aa	site-specific recombinase	*Haemophilus influenzae*	27% over 288 aa
			Methanobacterium thermoautotrophicum	41% over 159 aa
p6*orf*2	377 aa	Glutamyl-tRNA Reductase	*Archaeoglobus fulgidus*	32% over 303 aa
			Deinococcus radiodurans	37% over 263 aa
p6*orf*3	299 aa	Porphobilinogen deaminase	*Aquifex aeolicus*	38% over 214 aa
p6*orf*4	143 aa	Uroporphyrinogen synthase	*Deinococcus radiodurans*	40% over 143 aa
p6orf5	trunc. 76 aa	Porphobilinogen synthase	*Thiobacillus ferrooxidans*	50% over 40 aa
tRNA	76 nt	tRNA-Gly	*Stigmatella aurantiaca*	96% over 76 nt
pPhos22/				
pPhos*orf*1	trunc. 114 aa	IS5-like transposase	*Sphingomonas* Sp. LB126	45% over 107 aa
			Streptococcus thermophilus	44% over 101 aa
pPhos*orf*2-3	trunc. 273 aa	IS605-like transposase	*Helicobacter pylori*	34% over 271 aa
			Thermotoga maritima	32% over 262 aa

Table 7.4: Protein- and tRNA- encoding genes identified in environmental DNA sequences. Trunc. indicates that the coding sequence was truncated at the cloning site. N and C refer to p5*Orf*3 N-terminal and C-terminal sequences, respectively.

p5*orf2* encoded a hypothetical protein, the function of which could not be inferred (Section 6.3.3). The activity of the p5*orf2* gene product, therefore, could not be investigated directly.

For p6ICE16, the putative *hem* operon was considered the likely candidate for conferring the dark phenotype to *E. coli* TOP10. The *hem* genes each possessed a ribosome binding site. Bacterial-like promoter elements were also identified upstream of the *hem* gene cluster (Section 6.4.3).

BLAST searches using the deduced protein sequences encoded by p6*orfs*2-5 revealed homologies to enzymes involved in the early steps of tetrapyrrole biosynthesis (Section 6.4). The accumulation of tetrapyrrole pigments was thought to have been responsible for the dark phenotype observed with *E. coli* TOP10/p6ICE16; however, tetrapyrroles, which fluoresce when exposed to long-wave UV light, were not detected in *E. coli* TOP10/p6ICE16. This suggested that the putative *hem* genes were not being expressed. This possibility was further supported by the finding that plasmid p6ICE16 did not complement *E. coli hem*A mutant strain SASX41B (Section 6.4.9).

In an attempt to identify heterologously expressed proteins, SDS-PAGE was performed on *E. coli* TOP10/p5ICE16 and *E. coli* TOP10/p6ICE16 crude cell extracts (Sections 6.3.5 & 6.4.9); however, no protein band could be attributed to the heterologous expression of the cloned DNA in *E. coli* TOP10 cells. Despite the fact that cloned Ice16-DNA was shown to confer the dark phenotype to *E. coli* host cells (Section 5.4.2), no gene(s) or gene product(s) encoded by cloned sequences from either p5ICE16 or p6ICE16 could be identified as the determinant(s) responsible for the observed phenotype.

The only feature that p5ICE16 and p6ICE16 apparently had in common was that the plasmids conferred a dark phenotype to their *E. coli* hosts. With the aim of identifying any sequences that may be common to both, the nucleotide and deduced amino acid sequences of the p5ICE16-DNA insert were compared to those of the p6ICE16-DNA insert (Section 2.10). No significant homology at either the nucleotide level or at the amino acid level was identified for p5ICE16- and p6ICE16-DNA inserts. A common determinant responsible for the dark phenotype could not be identified for these clones.

The way forward with p5ICE16 and p6ICE16 is to subclone the candidate genes and re-test for the dark phenotype. This may be achieved by digesting the cloned p5ICE16- and p6ICE16-DNA fragments with various restriction enzymes. The digestion products from each clone may then be ligated into pCR-XL-TOPO or any other suitable vector. Alternatively, individual candidate ORFs may be amplified using PCR primers designed to anneal to flanking DNA sequences. The PCR-amplified products may then be ligated to an appropriate PCR expression vector. After transforming *E. coli* cells with the subcloned recombinant plasmids, the back phenotype can then screened in the same manner as that used for the original clones.

7.5.2.2: pPhos22

pPhos*orf*2-3, whose gene product is thought to arise from a +1 translational frameshift between pPhos*orf*2 and pPhos*orf*3, was the likely candidate for conferring the thermostable alkaline phosphatase activity to *E. coli* TOP10 cells harbouring pPhos22. pPhos*orf*2-3, which encoded a putative IS605-like transposase, was preceded by a putative ribosome binding site and two bacterial-like promoters (Section 6.5.5). No IS605 transposase has been isolated and characterised to date and the active site of IS605/TnpB-like proteins has yet to be assigned. Although it was not possible to reconcile sequence similarities to protein function for pPhos*Orf*2-3, it was thought that the phosphoryl transferase activity of the putative transposase was what had been detected in ICE22. To investigate whether any of the peptides encoded by pPhos22 were heterologously expressed in *E. coli* TOP10 cells, SDS-PAGE was performed using *E. coli* TOP10/pPhos22 crude cell extracts. Protein bands thought to correspond to alternatively translated products encoded by the IS605/tnpB-like gene were identified (Section 6.5.6). In an effort to characterise the phosphatase activity conferred upon *E. coli* TOP10 cells, a thermostability study was performed using *E. coli* TOP10/pPhos22 (Sections 2.11 & 6.5.7). During the course of the thermostability study, the phosphatase activity was lost from *E. coli* TOP10/pPhos22. The originally isolated clone, in the master microtitre plate was reconfirmed as being phosphatase positive, however, activity was lost each time the recombinant was propagated. The source clone was eventually depleted.

Plasmid DNA, possessing the same restriction pattern as that from the original clone, was recovered from the *E. coli* strains that had lost the heterologous phosphatase activity. This suggests that a cause other than plasmid instability was responsible for the

loss of the heterologously expressed phosphatase. Sequencing the DNA inserts of these inactive clones will reveal whether a mutation has occurred and may help determine why thermostable phosphatase activity was lost from these recombinant *E. coli* TOP10 strains.

Future work may also include subcloning the putative ORF for expression as described above for p5ICE16 and p6ICE16 (Section 7.5.2.1). If pPhos*Orf*2-3 was deleteriously mutated, then heterologous expression of the subcloned sequences can not be achieved. If this is the case, then the only way to retrieve the positive clone is by re-screening ICE22 for phosphatase activity as described for the original clone.

7.6: Future Investigations

While identifying the molecular determinants thought to be responsible for the observed phenotypes for p5ICE16, p6ICE16 and pPhos22, several putative genes encoded by the cloned DNA inserts were identified. These sequences are listed in Table 7.4 and are described in great detail in Chapter 6. Although several of these coding sequences were not considered to be candidate genes for heterologous expression as cloned, studying these putative genes in detail provided a glimpse into the uncultivated metagenome. Indeed, as a supplement to expression cloning, shotgun sequencing may provide another route to accessing novel target activities. Upon identifying an interesting coding sequence, the candidate ORF may be subcloned and expressed and the gene product functionally characterised. It must be emphasised however that, although this approach may provide access to the uncultivated metagenome, it is not culture-independent. This is because identification of putative genes relies on the sequence information of previously cultivated organisms. Nevertheless, this route provides an opportunity to study the uncultivated metagenome.

Future work with ICE16 and ICE22 libraries may also involve metagenome walking. By using the cloned DNA inserts as probes, contiguous DNA fragments present within the cloned environmental DNA may be identified. This approach may also identify the complete sequences of those genes that were truncated during the cloning procedures. Alternatively, specific coding sequences identified within the DNA inserts may be amplified by PCR and used as probes for screening libraries. This would provide another route to accessing homologous genes encoded by other uncultivated organisms.

7.7: Conclusions

i. With respect to DNA yield, purity and fragment size, both the mortar-and-pestle-plus-SDS method and the bead-beating method were suitable for extracting DNA from geothermal sediments. Further work is required to assess the diversity of the DNA obtained by these two methods.

ii. Compared to other *E. coli*-based cloning systems investigated in this work, the TOPO-TA cloning procedure, using pCR-XL-TOPO as vector, proved superior at generating environmental DNA libraries. Subsequently, the TOPO-TA cloning procedure was used to construct ICE16 and ICE22 environmental libraries using DNA derived from two different geothermal sediments.

iii. With respect to cloning efficiency, the number of 1-kb genes represented and the frequency at which heterologous activity was detected, ICE16 and ICE22 environmental libraries were in concordance with other environmental expression libraries reported in the literature. This indicates that ICE16 andICE22 are suitable for expression cloning.

iv. The generic applicability of ICE16 and ICE22 was demonstrated by screening for a number of different enzyme activities using various assay methods. Increasing the number of different screening assays may increase the frequency of detecting heterologous activity.

v. One transformant possessing phosphatase activity at 60°C (Phos22) and two transformants showing atypical phenotypes on starch agar plates at 50°C (5ICE16, 6ICE16) were recovered from ICE22 and ICE16 DNA libraries, respectively. These results demonstrated that novel activities can be accessed from the metagenome in a culture-independent manner.

vi. Further work is required to functionally characterise the heterologously expressed gene products detected in ICE22 and ICE16 libraries.

The work described in this study has demonstrated that sequences encoded by DNA, that has been extracted from geothermal sediments and cloned into pCR-XL-TOPO, can be expressed and detected in *E. coli*. As more environmental DNA libraries are constructed and functionally screened in the coming years, culture-independent methodologies, such as that described here, will become established as routine approaches for accessing and exploiting the natural biodiversity .

References

Abremski, K.E., Hoess, R.H. (1992) Evidence for a second conserved arginine residue in the integrase family of recombination proteins. Protein Eng. 5, 87-91.

Adams, M.W.W. (1999) The biochemical diversity of life near and above 100°C in marine environments. J. Appl. Microbiol. 85, 108S-117S.

Alfredsson, G.A., Kristjansson, J.K., Hjorleifsdottir, S., Stetter, K.O. (1988) *Rhodothermus marinus* gen. nov., sp. nov., a thermophilic halophilic bacterium from submarine hot springs in Iceland. J. Gen. Microbiol. 134, 299-306.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local sequence alignment search tool. J. Mol. Biol. 215, 403-410.

Amann, R. I., Ludwig, W., Schleifer, K. H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol. Rev. 59, 143–169.

Aragno, M. (1992) The aerobic, chemolithoautotrophic, thermophilic bacteria. In Thermophilic Bacteria. (Kristjansson, J.K. ed.) CRC Press, Boca Raton, Florida. pp.77-103.

Aravalli, R.N., Garrett, R.A. (1997) Shuttle vectors for hyperthermophilic archaea. Extremophiles 1, 183-191.

Arber, W., Dussoix, D. (1962) Host specificity of DNA produced by *Escherichia coli.* 1. Host controlled modification of bacteriophage λ. J. Mol. Biol. 5, 18-36.

Arnold, R.J., Polevoda, B., Reilly, J.P.Shermani, F. (1999) The Action of N-terminal Acetyltransferases on Yeast Ribosomal Proteins. J. Biol. Chem. 274, 37035-37040.

Atkins, J., Weiss, R.B., Gesteland, R.F. (1990) Ribosome gymnastics: degree of difficulty 9.5, style 10.0. Cell 62, 413-423.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidmen J.G., Smith J.A., Struhl, K. (1989) Current Protocols in Molecular Biology. Wiley, New York.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidmen J.G., Smith J.A., Struhl, K. (1995) Short Protocols in Molecular Biology, Third Edition. John Wiley & Sons, Inc. U.S.A.

Avissar, Y.J., Beale, S.I., (1989) Identification of the enzymatic basis for delta-aminolevulinic acid auxotrophy in a *hem*A mutant of *Escherichia coli.* J. Bacteriol. 171, 2919-2924.

Bachmann, B.J. (1990) Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. 54, 130-197.

Bakken, L.R., Lindahl, V. (1995) Recovery of bacterial cells from soil. In Nucleic Acids in the Environment: Methods and Applications. (Trevors, J.T., van Elsas, J.D. eds.) Springer, New York. pp. 153-177.

Balbas, P., Bolivar, F. (1990) Design and construction of expression plasmid vectors in *Escherichia coli*. Methods Enzymol. 185, 14-37.

Banerjee, R.V., Frasca, V., Ballou, D.P., Matthews, R.G. (1990a) Participation of cob(I) alamin in the reaction catalyzed by methionine synthase from *Escherichia coli*: a steady-state and rapid reaction kinetic analysis. Biochemistry 29, 11101-11109.

Banerjee, R.V., Harder, S.R., Ragsdale, S.W., Matthews, R.G. (1990b) Mechanism of reductive activation of cobalamin-dependent methionine synthase: an electron paramagnetic resonance spectroelectrochemical study. Biochemistry 29, 1129-1135.

Banerjee, R.V., Johnston, N.L., Sobeski, J.K., Datta, P., Matthews, R.G. (1989) Cloning and sequence analysis of the *Escherichia coli metH* gene encoding cobalamin-dependent methionine synthase and isolation of a tryptic fragment containing the cobalamin-binding domain. J. Biol. Chem. 264, 13888-13895.

Barns, S.M., Delwiche, C.F., Palmer, J.D., Pace, N.R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc. Natl. Acad. Sci. USA 93, 9188-9193.

Barns, S.M., Fundyga. R.E. Jeffries, M.W., Pace, N.R. (1994) Remarkable archaeal diversity in a Yellowstone National Park hot spring environment. Proc. Natl. Acad. Sci. USA 91, 1609-1613.

Baross, J. A., Holden, J.F. (1996) Overview of hyperthermophiles and their heat-shock proteins. Adv. Protein Chem. 48, 1-34.

Bell, S. D., Jackson, S.P. (1998) Transcription and translation in *Archaea*; a mosaic of eukaryal and bacterial features. Trends Microbiol. 6, 222-228.

Bernard, P., Gabant, P., Bahassi. E.M., Couturier, M. (1994) Positive selection vectors using the F plasmid ccdB killer gene. Gene 148, 71-74.

Blakely, G.W., May, G., McCulloch, R., Arciszewska, L.K., Burke, M., Lovett, S.T., Sherratt, D.J. (1993) Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. Cell 75, 351-361.

Blakely, G.W., Sherratt, D.J. (1994) Interactions of the site-specific recombinases *Xer*C and *Xer*D with the recombination site *dif*. Nucleic Acids Res. 22, 5613-5620.

Blakely,G.W., Sherratt, D.J. (1996) *Cis* and *trans* in site-specific recombination. Mol. Microbiol., 20, 234–237.

Blanchin-Roland, S., Masson, J.M. (1989) Protein secretion controlled by a synthetic gene in *Escherichia coli*. Protein Eng. 2, 473-480.

Blattner, F.R., Burland, V., Plunkett, G. III., Sofia, H.J., Daniels, D.L. (1993) Analysis of the *Escherichia coli* genome. IV. DNA sequence of the region from 89.2 to 92.8 minutes. Nucleic Acids Res. 21, 5408-5417.

Blochl E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H.W., Stetter, K.O. (1997) *Pyrolobus fumarii* gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113°C. Extremophiles 1, 14-21.

Bourgoin, F., Guedon, G., Gintz, B., Decaris, B. (1998) Characterization of a novel insertion sequence, IS1194, in *Streptococcus thermophilus*. Plasmid 40, 44-49.

Britschgi, T.B., Giovannoni, S.J. (1991) Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequenceing Appl. Enivron. Microbiol. 57, 1707-1713.

Brock T.D., Brock, K.M., Belly, R.T., Weiss, R.L. (1972) *Solfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch. Microbiol. 84, 54-68.

Brock, T.D., Freeze, H. (1969) *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. J. Bacteriol. 98, 289-297.

Brock, T.D., Madigan, M.T. (1991) Biology of Microorganisms, Sixth Edition. Prentice-Hall International, London.

Brunschwig, E., Darzins, A. (1992) A two-component T7 system for the overproduction of genes in *Pseudomonas aeruginosa*. Gene 111,35-41.

Cao, Y., Hallet, B., Sherratt, D. J., Hayes, F. (1997) Structure-function correlations in the *Xer*D site-specific recombinase revealed by pentapeptide scanning mutagenesis. J. Mol. Biol. 274, 39-53.

Cao, Y., Hayes, F. (1999) A newly identified, essential catalytic residue in a critical secondary structure element in the integrase family of site-specific recombinases is conserved in a similar element in eucaryotic type IB topoisomerases. J. Mol. Biol. 289, 517-527.

Chandler, D.P., Schreckhise, R.W., Smith, J.L., Bolton, H. (1997) Electroelution to remove humic compounds from soil DNA and RNA extracts. J. Microbiol. Methods 28, 11-19.

Chandler, M., Fayet, O. (1993) Translational frameshifting in the control of transposition in bacteria. Mol. Microbiol. 7, 497-503.

Chen, W., Russell, C.S., Murooka, Y., Cosloy, S.D. (1994) 5-Aminolevulinic acid synthesis in *Escherichia coli* requires expression of *hem*A. J. Bacteriol. 176, 2743-2746.

Chen, L.H., Liu, M.L., Hwang, H.Y., Chen, L.S., Korenberg, J., Shane, B. (1997) Human methionine synthase: cDNA cloning, gene localization, and expression. J. Biol. Chem. 272, 3628-3634.

Chevaldonne, P., Desbruyeres, D., Childress, J.J. (1992) Some like it hot and some even hotter. Nature 359, 593-594.

Christiensen T., Woeldike, H., Boel, E., Mortensen, S.B., Hjortschoejk, Thim, L., Hansen, M.T. (1988) High level expression of recombinant genes in *Aspergillus oryzae*. Biotechnology 6, 1419-1422.

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

Clarke, L., Carbon, J. (1976) A colony bank containing synthetic Col El hybrid plasmids representative of the entire E. coli genome. Biotechnology 24, 179-187.

Claus, D., Berkeley, R.C.W. (1986) Genus *Bacillus* Cohn 1872, 174^{AL} , In Bergey's Manual of Systematic Bacteriology (Sneath, P.H.A., Mair, N.S., Sharpe, M.E., Holt, J.G., eds) Vol. 2, Williams & Wilkins, Baltimore. p 1105.

Clegg, C.D., Ritz, K., Griffiths, B.S. (1997) Direct extraction of microbial community DNA from humified upland soils. Lett. Appl. Microbiol. 25, 30-33.

Colloms, S.D., Sykora, P., Szatmari, G., Sherratt, D.J. (1990) Recombination at *Col*E1 *cer* requires the *Escherichia coli xer*C gene product, a member of the lambda integrase family of site-specific recombinases. J. Bacteriol. 172, 6973-6980.

Cottrell, M.T., Moore, J.A., Kirchman, D.L. (1999) Chitinases from uncultured marine microorganisms. Appl. Environ. Microbiol. 65, 2553-2557.

Cue, D., Greggory, S.B., Beckler, S., Reeve, J.N., Konisky, J. (1985) Structure and sequence divergence of two archaebacterial genes. Proc. Natl. Acad. Sci. USA 82, 4201-4211.

Cullen, D.W., Hirsch, PR. (1998) Simple and rapid method for direct extraction of microbial DNA from soil for PCR. Soil. Biol. Biochem. 30, 983-993.

Daniel, R.M. (1996) The upper limits of enzyme stability. Enzyme Microb. Technol. 19, 74-79.

Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olson, G.J., Swanson, R.V. (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature 392, 353-358.

Demerec, M., Adelberg, A.E., Clark A.J., Hartman, P.E. (1966) A proposal for a uniform nomenclature in bacterial genetics. Genetics 54, 61-76.

Dennis, P.P. (1997) Ancient ciphers: translation in *Archaea*. Cell 89, 1007-1010.

Drummond, J.T., Huang, S., Blumenthal, R.M., Matthews, R.G. (1993) Assignment of enzymatic function to specific protein regions of cobalamine-dependent methionine synthase from *Escherichia coli*. Biochemistry 32, 9290-9295.

Edwards, C. (1990) Thermophiles. In Microbiology of Extreme Environments. (Edwards, C. ed.) Open University Press, Oxford. pp 1-32.

Eggen, R.I.L., Geerling, A.C.M., Boshoven, A.B.P., Devos, W.M. (1991a) Cloning, sequence analysis and functional expression of the acetyl coenzyme-A synthase gene from *Methanothrix soehgenii* in *Escherichia coli*. J. Bacteriol. 173, 6383-6389.

Eggen, R.I.L., Geerling, A.C.M., Jetten, M.S.M., Devos, W.M. (1991b) Cloning, expression and sequence analysis of the genes for carbon monoxide dehydrogenase of *Methanothrix soehgenii*. J. Biol. Chem. 266, 6883-6887.

Elledge, S.J., Mulligan, J.T., Ramer, S.W., Spottswood, M., Davis, R.W. (1991) λYES: A multifunctional cDNA expression vector for the isolation of genes by complementation of yeast and *Escherichia coli* mutations. Proc. Natl. Acad. Sci. USA 88, 1731-1735.

Emory, S.A., Bouvet, P., Belasco, J.G. (1992) A 5' terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. Genes Devel. 6, 135-148.

Enfors, S-.O. (1992) Control of *in vivo* proteolysis in the production of recombinant proteins. Trends Biotechnol. 10, 310-315.

Escoubas, J. M., Lane, D., Chandler, M. (1994) Is the IS1 transposase, *Ins*AB, the only IS1-encoded protein required for efficient transposition? J. Bacteriol. 176, 5864-5867.

Escoubas, J.M., Prere, M.F., Fayet, O., Salvignol, I., Galas, D., Zerbib, D., Chandler, M. (1991) Translational control of transposition activity of the bacterial insertion sequence IS1. EMBO J. 10, 705-712.

Evans, B R., Chen, J.W., Parsons, R.L., Bauer, T.K., Teplow, D.B., Jayaram, M. (1990) Identification of the active site tyrosine of *Flp* recombinase: possible relevance of its location to the mechanism of recombination. J. Biol. Chem. 265, 18504-18510.

Faegri, A., Torsvik, V.I., Goksöyr, J. (1977) Bacterial and fungal activities in soil: separation of bacteria by a rapid fractionation centrifugation technique. Soil Biol. Biochem. 9,105-112.

Farabaugh, P.J. (1996) Programmed translational frameshifting. Microbiol. Rev. 60, 103-134.

Fayet, O., Ramond, P., Polard, P., Prere, M.F., Chandler, M. (1990) Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? Mol. Microbiol. 4, 1771-1777.

Fiala, G., Stetter, K.O., Jannasch, H.W., Langworthy, T.A., Madon, J. (1986) *Staphylothermus marinus* sp. nov. represents a novel genus of extremely thermophilic submarine heterotrophic archaebacteria growing up to 98°C. Syst. Appl. Micrbiol. 8, 106-113.

Fiala, G., Stetter, K.O. (1986) *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. Arch. Microbiol. 145, 56-61.

Fleer, R., Yeh, P., Amellal, N., Maury, I., Fournier, A., Baccheta, F., Baduel, P., Jung, G., L'Hote, H., Becquart, J. (1991) Stable multicopy vectors for high-level secretion of recombinant human serum albumin by *Kluyveromyces* yeasts. Biotechnology 9, 968-975.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., Fitzhugh, W., Fields, C.A., Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., Venter, J.C. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269, 496-512.

Flemming, C.A., Leung, K., Lee, H Trevors, J.T., Greer, C (1994) Survival of *lux-lac* marked biosurfactant-producing *Pseudomonas aeruginosa* UC2L in soil monitored by non-selective plating and polymerase chain reaction. Appl. Environ. Microbiol. 60, 1606-1613.

Fogarty, W.M. (1983) Microbial Enzymes and Biotechnology. Applied Science Publishers, Belfast.

Fontana, A. (1991) Analysis and modulation of protein stability. Curr. Opin. Biotechnol. 24, 551-560.

Frasca, V., Banerjee, R.V., Dunham, W.R., Sands, R.H., Matthews, R.G. (1988) Cobalamin-dependent methionine synthase from *Escherichia coli* B: electron paramagnetic resonance spectra of the inactive form and the active methylated form of the enzyme. Biochemistry 27, 8458-8465.

Fujii, K., Galivan, J.H, Huennekens, F.M. (1977) Activation of methionine synthase: further characterization of the flavoprotein system. Arch. Biochem. Biophys. 178, 662-670.

Fujino, E., Fujino, T., Karita, S., Sakka, K., Ohmiya, K. (1995) Cloning and sequencing of some genes responsible for porphyrin biosynthesis from the anaerobic bacterium *Clostridium josui*. J. Bacteriol. 177, 5169-5175.

Gerhardt, P., Murry, R.G.E., Costilow, R.N., Nester, E.W., Wood, W.A., Krieg, N.R., Phillips, G.B. (1981) Manual of Methods for General Bacteriology. ASM Press, Washington, DC.

Gesteland, R.F., Atkins, J.F. (1996) Recoding: dynamic reprogramming of translation. Annu. Rev. Biochem. 65, 741-768.

Giovannoni, S.J., Britschgi, T., Moyer, C.L. Field, K.G. (1990) Genetic diversity in Sargasso sea bacterioplankton, Nature (London) 345, 60-63.

Gish, W., States, D.J. (1993) Identification of protein regions by database similarity search. Nat. Genet. 3, 266-272.

Glover, D.M., Hames, B.D. (1995) DNA Cloning 2: Expression Systems, A Practical Approach. Oxford University Press, Oxford.

Goeddel, D.V. (1990) Methods in Enzymology, Volume 185: Gene Expression Technology. Academic Press, Inc., London.

Gottesman, S. (1990) Minimizing proteolysis in *Escherichia coli*: Genetic solutions. Methods Enzymol. 185, 119-129.

Goulding, C.W., Matthews, R.G. (1997) Cobalamin-dependent methionine synthase from *Escherichia coli*: involvement of zinc in homocysteine activation. Biochemistry 36, 15749-15757.

Goulding, C.W., Postigo, D., Matthews, R.G. (1997) Cobalamin-dependent methionine synthase is a modular protein with distinct regions for binding homocysteine, methyltetrahydrofolate, cobalamin, and adenosylmethionine. Biochemistry 36, 8082-8091

Goyette, P., Sumner, J.S., Milos, R., Duncan, A.M., Rosenblatt, D.S., Matthews, R.G., Rozen, R. (1994) Human methylenetetrahydrofolate reductase: isolation of cDNA, mapping and mutation identification. Nature Genet. 7, 195-200.

Gralla, J.D. (1990) Promoter recognition and mRNA initiation by *Escherichia coli* $E\sigma 70$. Methods Enzymol. 185, 37-60.

Grandchamp, B., De Verneuil, H., Beaumont, C., Chretien, S., Walter, O. Nordmann, Y. (1987) Tissue-specific expression of porphobilinogen deaminase. Two isoenzymes from a single gene. Eur. J. Biochem. 162, 105-110.

Grindley, N.D., Leschziner, A.E. (1995) DNA transposition: from a black box to a color monitor. Cell 83, 1063-66.

Gross C.A., Lonetto, M., Losick, R. (1992) Bacterial transcription factors. In Transcriptional regulation. (McKnight, S.L., Yamamoto, K.R. eds.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. pp. 129-178.

Grundy, F.J., Henkin , T.M. (1998) The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. Mol. Microbiol. 30, 737-749.

Guenther, B.D., Sheppard, C.A., Tran, P., Rozen, R., Matthews, R.G., Ludwig, M.L. (1999) The structure and properties of methylenetetrahydrofolate reductase from *Escherichia coli* suggest how folate ameliorates human hyperhomocysteinemia. Nat. Struct. Biol. 6, 359-365.

Haggoud, A., Reysset, G., Azeddoug, H., Sebald, M. (1994) Nucleotide sequence analysis of two 5-nitroimidazole resistance determinants from *Bacteroides* strains and of a new insertion sequence upstream of the two genes. Antimicrob. Agents Chemother. 38, 1047-1051.

Hanahan, D., Jessee, J., Bloom, F. R. (1995) Techniques for transformation of *E. coli*. In DNA Cloning 1: Core Techniques, A Practical Approach. (Glover, D.M., Hames, B.D. eds) Oxford University Press, Oxford. pp. 1-36.

Hannig, G., Makrides, S. C. (1998) Strategies for optimizing heterologous protein expression in *Escherichia coli*. Trends Biotechnol. 16, 54-60.

Hansson, M., Rutberg, L., Schroder, I., Hederstedt, L. (1991) The *Bacillus subtilis hem*AXCDBL gene cluster, which encodes enzymes of the biosynthetic pathway from glutamate to uroporphyrinogen III. J. Bacteriol. 173, 2590-2599.

Harwood, C.R., Cutting, S.M. (1990) Molecular Biology Methods for *Bacillus*. John Wiley, Chichester

Hatch, F.T., Larrabee, A.R., Cathou, R.E., Buchanan, J.M. (1961) Enzymatic synthesis of the methyl group of methionine. I. Identification of the enzymes and cofactors involved in the system isolated from *Escherichia coli*. J. Biol. Chem. 236, 1095-1101.

Henne, A., Daniel, R., Schmitz, R.A., Gottschalk, G. (1999). Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Appl. Environ. Microbiol. 65, 3901–3907.

Henne, A., Schmitz, R.A., Bömeke, M., Gottschalk, G., Daniel, R. (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. Appl. Environ. Microbiol. 66, 3113-3116.

Holben, W.E., Jansson, J.K., Chelm, B.K., Tiedje, J.M. (1988) DNA probe method for the detection of specific microorganisms in the soil community. Appl. Environ. Microbiol. 54, 703-711.

Holmes, M.L., Dyall-Smith, M.L. (1990) A plasmid vector with a selectable marker for halophilic archaebacteria. J. Bacteriol. 172, 756-761.

Hopwood, D.A., Bibb, M.J., Chater, K.F.,Kieser, T. (1987) Plasmid and phage vecors for gene cloning and analysis in *Streptomyces*. Methods enzymol. 153, 116-166.

Hu, S.-T., Lee, L.-C., Lei, G.-S. (1996) Detection of an IS2-encoded 46-kilodalton protein capable of binding terminal repeats of IS. Bacteriol. 178, 15652-15659.

Huber, G., Spinnler, C., Gambacorta, A., Stetter, K.O. (1989) *Metallosphaera sedula* gen. and sp. nov., represents a new genus of Aerobic, metal-metabolizing, thermoacidophilic archaebacteria. System. Appl. Microbiol. 12, 38-47.

Huber, R., Kurr, T.A., Jannasch, H.W., Stetter, K.O. (1989) A novel group of abyssal methanogenic archaebacteria (Methanpyrus) growing at 110°C. Nature 342, 833-834.

Huber, R., Langworthy, T.A., Konig, H.T., Thomm, M., Woese, C.R., Sleytr, V.B., Stetter, K.O. (1986) *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C.. Arch. Microbiol. 144, 324-333.

Huber, R., Wilharm, T., Huber, D., Trincone, A., Burggraf, S., Konig, H., Rachel, R., Rockinger, I., Fricke, H., Stetter, K.O. (1992) *Aquifex pyrophilus* gen. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. Syst. Appl. Microbiol. 12, 32-37.

Hugenholtz, P., Pitulle, C., Hershberger, K.L., Pace, N.R. (1998) Novel division level bacterial diversity in a Yellowstone hot spring. J. Bacteriol. 180, 366–376.

Ikeda, M., Clark, D.S, (1998) Molecular cloning of extremely thermostable esterase gene from hyperthermophilic archaeon *Pyrococcus furiosus* in *Escherichia coli*. Biotechnol. Bioeng. 57,624-629.

Ikemi, M., Murakami, K., Hashimoto, M., Murooka, Y. (1992) Cloning and characterization of genes involved in the biosynthesis of delta-aminolevulinic acid in *Escherichia coli*. Gene 121, 127-132.

Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2, 13-34.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F., Boyer, H.W. (1977) Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science, 198, 1056-1063.

Jaenicke, R., Schurig, H., Beaucamp, N., Ostendorp, R. (1996) Structure and stability of hyperstable proteins: Glycolytic enzymes from hyperthermophilic bacterium *Thermotoga maritima* Advan. Prot. Chem. 48, 181-269.

Johansson, P., Hederstedt, L. (1999) Organization of genes for tetrapyrrole biosynthesis in Gram-positive bacteria. Microbiology 145, 529-538.

Jones, W.J., Leigh, J.A., Mayer, F, Woese, C.R., Wolfe, R.S. (1983) *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. Arch. Microbiol. 136, 254-261.

Jordan, P.M., Seehra, J.S. (1980) ^{14}C NMR as a probe for the study of enzyme catalysed reactions. Mechanism of action of 5-aminolevulinic acid dehydratase. FEBS Lett. 114, 283-286.

Jordan, P.M., Mgbeje, B.I.A., Thomas, S.D., Alwan, A.F. (1988) Nucleotide sequence for the *hem*D gene of *Escherichia coli* encoding uroporphyrinogen III synthase and initial evidence for a *hem* operon. Biochem. J. 249,613-616.

Kafala, B., Sasarman, A. (1997) Isolation of the *Staphylococcus aureus hem*CDBL gene cluster coding for early steps in heme biosynthesis. Gene 199, 231-239.

Kaiser, K., Murray, N. E., Whittaker, P. A. (1995) Construction of representative genomic DNA libraries using phage lambda replacement vectors. in DNA Cloning 1: Core Techniques, A Practical Approach. (Glover, D.M., Hames, B.D. eds) Oxford University Press, Oxford. pp.37-84.

Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M., Tabata,S. (1995) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. DNA Res. 2, 153-166.

Kannangara, C.G., Andersen, R.V., Pontoppidan, B., Willows, R., von Wettstein, D. (1994) Enzymic and mechanistic studies on the conversion of glutamate to 5-aminolaevulinate. In The Biosynthesis of the Tetrapyrrole Pigments (Chadwick, D.J., Ackrill, K. eds.) John Wiley and Sons, Chichester. pp. 3-25.

Katzman, M., Mack, J.P., Skalka, A.M., Leis, J. (1991) A covalent complex between retroviral integrase and nicked substrate DNA. Proc. Natl. Acad. Sci. USA 88, 4695-4699.

Khan, E., Mack, J.P., Katz, R.A., Kulkosky, J., Skalka, A.M. (1991) Retroviral integrase domains: DNA binding and the recognition of LTR sequences. Nucleic Acids Res. 19, 851-60 (Erratum. (1991) Nucleic Acids Res. 19,1358).

Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Utterback, T., Cotton, M.D., Spriggs, T., Artiach, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O., Woese, C.R., Venter, J.C. (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. Nature 390, 364-370.

Knaebel, D.B., Crawford, R.L. (1995) Extraction and purification of microbial DNA from petroleum-contaminated soils and detection of low numbers of toluene, octane and pesticide degraders by multiplex polymerase chain reaction and Southern analysis. Mol. Ecol. 4, 579-591.

Kobayashi, K., Kato, M., Miura, Y., Kettoku, M., Komeda, T., Iwamatsu, A. (1996) Gene cloning and expression of new trehalose-producing enzymes from the hyperthermophilic archaeum *Sulfolobus solfataricus* KM1. Biosci. Biotechnol. Biochem. 60, 1882-1885.

Kowalczykowski, S.C. (1991) Biochemical and biological function of *Escherichia coli* RecA protein: behavior of mutant RecA proteins. Biochimie 73, 289-304.

Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D., Rehrauer, W.M. (1994) Biochemistry of homologous recombination in *Escherichia coli*. Microbiol. Rev. 58, 401-465.

Kristjansson, J.K., Stetter, K.O. (1992) Thermophilic bacteria. In Thermophilic Bacteria. (Kristjansson, J.K. ed.) CRC Press, Boca Raton, Florida. pp. 1-18.

Krsek, M., Wellington, E.M.H. (1999) Comparison of different methods for the isolation and purification of total community DNA from soil. J. Microbial. Methods 39, 1-16.

Kuhlemeier, C.J., van Arkel, G.A. (1987) Host-vector systems for gene cloning in cyanobacteria. Methods enzymol. 153, 199-215.

Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P., Skalka, A.M. (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. 12, 2331-2338.

Ladenstein R., Antranikian, G. (1998) Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. Adv. Biochem. Eng. Biotechnol. 61, 37-85.

Lander, M., Pitt, A.R., Alefounder, P.R., Bardy, D., Abell, C., Battersby, A. (1991) Studies on the mechanism of hydroxymethylbilane synthase concerning the role of arginine residues in substrate binding. Biochem. J. 275, 447-452.

Leclerc, D., Campeau, E., Goyette, P., Adjalla, C.E., Christensen, B., Ross, M., Eydoux, P., Rosenblatt, D.S., Rozen, R., Gravel, R.A. (1996) Human methionine synthase: cDNA cloning and identification of mutations in patients of the *cblG* complementation group of folate/cobalamin disorders. Mol. Genet. 5, 1867-1874.

Leff, L.G., Dana, J.R., McArthur, J.V., Shimkets, L.J. (1995) Comparisons of methods of DNA extraction from stream sediments. Appl. Environ. Microbiol. 61, 1141-1143.

Li, J.M., Brathwaite, O., Cosloy S.D., Russell, C.S. (1989) 5-Aminolevulinic acid synthesis in *Escherichia coli*. J. Bacteriol. 171, 2547-2552.

Li, Y.N., Gulati, S., Baker, P.J., Brody, L.C., Banerjee, R., Kruger, W.D. (1996) Cloning, mapping and RNA analysis of the human methionine synthase gene. Hum. Mol. Genet. 5, 1851-1858.

Liesack, W., Stackebrandt, E. (1992) Occurrence of novel groups of the domain Bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment. J. Bacteriol. 174, 5072–5078.

Lim, S.H., Witty, M., Wallace-Cook, A.D., Ilag, L.I., Smith, A.G. (1994) Porphobilinogen deaminase is encoded by a single gene in *Arabidopsis thaliana* and is targeted to the chloroplasts. Plant Mol. Biol. 26, 863-872.

Louie, G.V. (1993) Porphobilinogen deaminase and its structural similarity to the bi-domain binding proteins. Curr. Opin. Struct. Biol. 3, 401-408.

Lowe, T.M., Eddy, S.R. (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequences. Nucleic Acids Res. 25, 955-964.

Luthi, K., Moser, M., Ryser, J., Weber, H. (1990) Evidence for a role of translational frameshifting in the expression of transposition activity of the bacterial insertion element IS1. Gene 88, 15-20.

Machida, C., Machida, Y. (1989) Regulation of IS1 transposition by the *ins*A gene product. J. Mol. Biol. 208, 567-574.

Machida, Y., Machida, C., Ohtsubo, E. (1984) Insertion element IS1 encodes two structural genes required for its transposition. J. Mol. Biol. 177, 229-245.

Madigan, M.T. (1986) *Chlorobium tepidum* sp. nov., a thermophilic photosynthetic bacterium of the family Chromatiaceae. Int. J. Syst. Bacteriol. 36, 222-227.

Mahillon, J, Chandler, M. (1998) Insertion sequences. Microbiol. Mol. Biol. Rev. 62, 725-774.

Marmur, J. (1963) A procedure for the isolation of deoxyribonucleic acid from microorganisms. Methods Enzym. VI, 726-738.

Matthews, B.W., Nicholson, H., Becktel, W.J. (1987) Enhanced protein stability from site-directed mutations that decrease the entropy of unfolding. Proc. Natl. Acad. Sci. USA 84, 6663-6667.

Matthews, R.G., Sheppard, C., Goulding, C. (1998) Methylenetetrahydrofolate reductase and methionine synthase: biochemistry and molecular biology. Eur. J. Pediatr. 157, S54–S59.

Meeks, J.C., Castenholz, R.W., (1971) Growth and photosynthesis in an extreme thermophile, *Synechococcus lividus* (Cyanophyta). Arch. Mikrobiol. 78, 25-41.

Meile, L., Reeve, J.N. (1985) Potential shuttle vectors based on the methanogen plasmid pME2001. Biotechnology 3, 69-72.

Miller, D.N., Bryant, J.E., Madsen, E.L. Ghiorse, W.C. (1999) Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. Appl. Environ. Microbiol. 65, 4715-4724.

Miller, A.D., Hart, G.J., Packman, L.C., Battersby, A.R. (1988) Evidence that the pyrromethane cofactor of hydroxymethylbilane synthase (porphobilinogen deaminase) is bound to the protein through the sulphur atom of cysteine-242. Biochem. J. 254, 915-918.

More, M.I., Herrick, J.B., Silva, M.C., Ghoirse, W.C., Madsen, E.L. (1994) Qauntitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. Appl. Environ. Microbiol. 60, 1572-1580.

Murai, N., Kamata, H., Nagashima, Y., Yagisawa, H., Hirata, H. (1995) A novel insertion sequence (IS)-like element of the thermophilic bacterium PS3 promotes expression of the alanine carrier protein-encoding gene. Gene 163, 103-107.

Myrold, D., Martin, K.J., Ritchie, N.J. (1995) Gel purification of soil DNA extracts. In Molecular Microbial Ecology Manual. (Akkermans, A.D.L., van Elsas, J.D., deBruijn, F.J. eds.) Kluwer Academic Publishers, The Netherlands. pp. 1.3.5/1-1.3.5/9.

Nash, H.A. (1996) Site-specific recombination: integration, excision, resolution, and inversion of defined DNA segments. In *Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology (Neidhardt, F.C., Curtiss, R., III, Ingraham, J.L., Lin, E.C. C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Shaechter, M., Umbarger, H.E. eds) Second Edition, ASM Press, Washington, DC. pp. 2363-2376.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C., Fraser, C.M. (1999) Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature 399, 323-329.

Nesme, X., Picard, C., Simonet, P. (1995) Specific DNA sequences for detection of soil bacteria. In Nucleic Acids in the Environment: Methods and Applications (Trevors, J. T., van Elsas J.D., eds.) Springer, New York. pp. 111-139.

Neuhierl, B., Thanbichler, M., Lottspeich, F., Bock, A. (1999) A Family of S-Methylmethionine-dependent Thiol/Selenol Methyltransferases: role in selenium tolerance and evolutionary relation. J. Biol. Chem. 274, 5407–5414.

Ogawa, T., Yu, X., Shinohara, A., Egelman, E.H. (1993) Similarity of the yeast RAD51 filament to the bacterial *Rec*A filament. Science 259, 1896–1899.

Ogram, A., Sayler, G.S., Barkay, T. (1987) Th extraction and purification of microbial DNA from sediments. J. Microbial. Methods 7, 57-66.

Ohnuma, S., Suzuki, M., Nishino, T., (1994) Archaebacterial ether-linked lipid biosynthetic gene; expression cloning, sequencing and characterization of geranylgeranyl-diphosphate synthase. J. Biol. Chem. 269,14792-14797.

Old, I.G., Margarita, D., Glass, R.E., Saint Girons, I. (1990) Nucleotide sequence of the *met*H gene of *Escherichia coli* K-12 and comparison with that of *Salmonella typhimurium* LT2. Gene 87, 15-21.

Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. Science 276, 734-740.

Pan, G.H., Sadowski, P.D. (1992) Ligation activity of FLP recombinase: the strand ligation activity of a site-specific recombinase using an activated DNA substrate. J. Biol. Chem. 267, 12397-12399.

Pargellis, C.A., Nunes-Duby, S.E., Moitoso de Vargas, L., Landy, A. (1988) Suicide recombination substrates yield covalent lambda integrase-DNA complexes and lead to identification of the active site tyrosine. J. Biol. Chem. 263, 7678-7685.

Parsons, R.L., Prasad, P.V., Harshey, R.M., Jayaram, M. (1988) Step-arrest mutants of FLP recombinase: implications for the catalytic mechanism of DNA recombination. Mol. Cell. Biol. 8, 3303-3310.

Patenge, N.,Haase, A., Bolhuis, H., Oesterhelt, D.(2000) The gene for a halophilic beta-galactosidase (*bga*H) of *Haloferax alicantei* as a reporter gene for promoter analyses in *Halobacterium salinarum.* Mol. Microbiol. 36, 105-113.

Paul, J.H., Pichard, S.L. (1995) Extraction of DNA and RNA from aquatic environments. In Nucleic Acids in the Environment: Methods and Applications. (Trevors, J.T., van Elsas, J.D. eds.) Springer, New York. pp. 153-177.

Peariso, K., Goulding, C.W., Huang, S., Matthews, R.G., Penner-Hahn, J.E. (1998) Characterization of the zinc binding site in methionine synthase enzymes of *Escherichia coli*: The role of zinc in methylation of homocysteine. J. Am. Chem. Soc. 120, 8410-8416.

Petricek, M., Rutberg, L. Schroder. I., Hederstedt, L. (1990) Cloning and characterization of the *hem*A region of the *Bacillus subtillis* chromosome. J. Bacteriol. 172, 2250-2258.

Picard, C., Ponsonnet, C., Paget, E., Nesme, X., Simonet, P. (1992) Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain-reaction. Appl. Environ. Microbiol. 58,2717-2722.

Pierson, B.K., Castenholz, R.W. (1974) A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus auranticus* gen. and sp. nov. Arch. Microbiol., 100, 5-24.

Pley, U., Schipka, J., Gambacorta, A., Jannasch, H.W., Fricke, H, Rachel, R., Stetter, K.O. (1991) *Pyrodictium abyssi* sp. nov. represents a novel heterotrophic marine archaeal hyperthermophile growing at 110°C. Syst. Appl. Microbiol. 14, 245-253.

Plunkett, G. III., Burland, V., Daniels, D.L., Blattner, F.R. (1993) Analysis of the *Escherichia coli* genome. III. DNA sequence of the region from 87.2 to 89.2 minutes. Nucleic Acids Res. 21, 3391-3398.

Polard, P., Prere, M. F., Chandler, M., Fayet, O. (1991) Programmed translational frameshifting and initiation at an AUU codon in gene expression of bacterial insertion sequence IS911. J. Mol. Biol. 222, 465-477.

Polard, P., Prere, M. F., Fayet, O., Chandler, M. (1992) Transposase-induced excision and circularization of the bacterial insertion sequence IS911. EMBO J. 11,5079-5090.

Polevoda, B., Norbeck, J., Takakura, H., Blomberg, A., Sherman, F. (1999) Identification and specificities of N-terminal acetyltransferases from *Saccharomyces cerevisiae*. EMBO J. 18, 6155-6168.

Pollitt, S., Zalkin, H. (1983) Role of primary structure and disulfide bond formation in beta-lactamase secretion. J. Bacteriol. 153, 27-32.

Raich, N., Romeo, P.H., Dubart, A., Beaupain, D., Cohen-Solal, M., Goossens, M. (1986) Molecular cloning and complete primary sequence of human erythrocyte porphobilinogen deaminase. Nucleic Acids Res. 14, 5955-5968.

Reese, M.G., Harris, N.L., Eeckman, F.H. (1996) Large scale sequencing specific networks for promoter and splice site recognition. in Biocomputing: Proceedings of the 1996 Pacific Symposium. (Hunter, L., Klein, T.E., eds.) World Scientific Publishing Co. Singapore.

Rezsohazy, R., Hallet, B., Delcour, J., Mahillon, J. (1993) The IS4 family of insertion sequences: evidence for a conserved transposase motif. Mol. Microbiol. 9, 1283–1295.

Riccio, M.L., Rossolini, G.M., Lombardi, G., Chiesurin, A., Satta, G. (1997) Expression cloning of different bacterial phosphatase-encoding genes by histochemical screening of genomic libraries onto an indicator medium containing phenolphthalein diphosphate and methyl green. J. Appl. Microbiol. 82, 177-185.

Rice, P., Craigie, R., Davies, D.R. (1996) Retroviral integrases and their cousins. Curr. Opin. Struct. Biol. 6, 76–83.

Richard, G.M. (1974) Modifications of the diphenylamine reaction giving increased sensitivity and simplicity in the estimation of DNA. Anal. Biochem. 57, 369-376.

Roberts, D.L., Zhao, S., Doukov, T., Ragsdale, S.W. (1994) The reductive acetyl coenzyme A pathway: sequence and heterologous expression of active methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase from *Clostridium thermoaceticum*. J. Bacteriol. 176, 6127-6130.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J., Goodman, R.M. (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. 66, 2541-2547.

Romanos, M.A., Clare, J.J., Beesley, K.M., Rayment, F.B., Ballantine, S.P., Makoff, A.J., Dougan, G., Fairweather, N.F., Charles, I.G., (1991) Recombinant *Bordetella pertussis* pertactin (P69) from the yeast *Pichia pastoris*: high-level production and immunological properties. Vaccine 9, 901-906.

Saano, A., Lindstrom, K. (1995) Small scale extraction with spun column cleanup. In Molecular Microbial Ecology Manual. (Akkermans, A.D.L., van Elsas, J.D., deBruijn, F.J. eds.) Kluwer Academic Publishers, The Netherlands. pp. 1.3.4/1-1.3.4/6.

Saiki, T., Kobayashi, Y., Kawagae, K., Beppu T (1985) *Dyctioglomus thermophilum* gen. nov., sp. nov., a chemoorganotrophic anaerobic, thermophilic bacterium. Int. J. Syst. Bacteriol. 35, 253-259.

Saint-Girons, I., Duchange, N., Zakin, M.M., Park, I., Margarita, D., Ferrara, P., Cohen, G.N. (1983) Nucleotide sequence of *met*F, the *E. coli* structural gene for 5-10 methylene tetrahydrofolate reductase and of its control region. Nucleic Acids Res. 11, 6723-6732.

Saint-Girons, I., Parsot, C., Zakin, M.M., Barzu, O., Cohen, G.N. (1988) Methionine biosynthesis in Enterobacteriaceae: biochemical, regulatory, and evolutionary aspects. Crit. Rev. Biochem. 23, S1–S42.

Sako, Y., Nomura, N., Uchida, A., Ishida, Y., Morii, H., Koga, Y., Hoaki, T., Maruyama, T. (1996) *Aeropyrum pernix* gen. nov., sp. nov., a novel aerobic hyperthermophilic archaeon growing at temperatures up to 100 degrees C. Int. J. Syst. Bacteriol. 46, 1070-1077.

Sambrook, J., Fritsch, E. F, Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schein, C.H. (1991) Optimizing protein folding to the native state in bacteria. Curr. Opinion Biotechnol. 2, 746-50.

Schein, C.H., Noteborn, M.H.M. (1988) Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperatures. Biotechnology 6, 291-294.

Scordis, P., Flower, D.R., Attwood, T.K. (1999) FingerPRINTScan: intelligent searching of the PRINTS motif database. Bioinformatics 15, 799-806.

Segerer, A., Neuner, A., Kristjansson, J.K., Stetter, K.O. (1986) *Acidianus infernus* gen. nov., sp. nov., and *Acidianus brierleyi* comb. nov.: Facultatively aerobic, extremely acidophilic thermophilic sulfur-metabolizing archaebacteria. Int. J. Syst. Bacteriol. 36, 559-564.

Segerer, A.H., Trincone, A., Gahrtz, M., Stetter, K.O. (1991) *Stygiolobus azoricus* gen. and sp. nov. represents a novel genus of anaerobic, extremely thermoacidophiic archaea of the order *Sulfolobales*. Int. J. Syst. Bacteriol. 41, 495-501.

Seitz, E.M., Brockman, J.P., Sandler, S.J., Clark, A.J., Kowalczykowski, S.C. (1998) *Rad*A protein is an archaeal *Rec*A protein homolog that catalyzes DNA strand exchange. Genes Dev. 12, 1248-1253.

Sekine, Y., Eisaki, N., Ohtsubo, E. (1994) Translational control in production of transposase and in transposition of insertion sequence IS3. J. Mol. Biol. 235, 1406-1420.

Sekine, Y., Ohtsubo, E. (1989) Frameshifting is required for production of the transposase encoded by insertion sequence 1. Proc. Natl. Acad. Sci. USA 86, 4609–4613.

Selenska, S., Klingmuller, W. (1991a) DNA recovery and direct detection of Tn5 sequences from soil. Lett. Appl. Microbiol. 13, 21-24.

Selenska, S., Klingmuller, W. (1991b) Direct detection of *nif* gene sequences of *Enterobacter agglomerans* in soil. FEMS Microbiol. Lett. 80, 243-256.

Seow, K.-T., Meurer, G., Gerlitz, M. Wendt-Pienkowski, E., Hutchinson, C.R., Davies, J. (1997) A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. J. Bacteriol. 179, 7360–7368.

Sharp, .PM., Bulmer, M. (1988) Selective differences among translation termination codons. Gene 63, 141-145.

Sharp, R.J., Riley, P.W., White, D. (1990) Heterotrophic thermophilic *Bacilli.* In Thermophilic Bacteria. (Kristjansson, J.K. ed.) CRC Press, Boca Raton, Florida. pp. 19-50.

Sheppard, C.A., Trimmer, E.E.; Matthews, R.G. (1999) Purification and properties of NADH-dependent 5,10-methylenetetrahydrofolate reductase from *Escherichia coli.* J. Bacteriol. 181, 718–725.

Short, J. (1997) Recombinant approaches for accessing biodiversity. Nat. Biotechnol. 15, 1322-1223.

Shuman, S. (1994) Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. J. Biol. Chem. 269, 32678-32684.

Smalla, K., Cresswell, N., Mendonca-Hagler, L.C., Wolters, A., van Elsas, J.D. (1993) Rapid DNA exraction protocol from soil for polymerase chain reaction-amplification. J. Appl. Bacteriol. 74, 78-85.

Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H.-M., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicare, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrovski, S., Church, G.M., Daniels, C.J., Mao, J.I., Rice, P., Nolling, J., Reeve, J.N. (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* delta H: functional analysis and comparative genomics. J. Bacteriol. 179, 7135-7155.

Soppa, J. (1999) Normalized nucleotide frequencies allow the definition of archaeal promoter elements for different archaeal groups and reveal base-specific TFB contacts upstream of the TATA box. Mol. Microbiol. 31, 1589-1601.

Stader, J.A, Silhavy, T.J. (1990) Engineering *Escherichia coli* to secrete heterologous gene products. Methods Enzymol. 185, 166-187.

Steffan, R.J., Goksoyr, J., Bej, A.K., Atlas, R.M. (1988) Recovery of DNA from soils and sediments. Appl. Environ. Microbiol. 54, 2908-2915.

Stetter, K.O. (1986) Diversity of extremely thermophilic archaebacteria. In Thermophiles: General, Molecular, and Applied Microbiology. (Brock, T.D. ed.) John Wiley & Sons, New York, pp 40-74.

Stetter, K.O. (1988) *Archaeoglobus fulgidus* gen. nov. sp. nov.: a new taxon of extremely thermophilic archaebacteria. Syst. Appl. Microbiol. 10, 172-173.

Stetter, K.O. (1998) Hyperthermophiles: isolation, classification and properties. In Extremophiles: Microbial Life in Extreme Environments. (Horikoshi, K., Grant, W.D. eds.) Wiley-Liss, New York. pp. 1-24.

Stormo, G.D. (1986) Translation initiation. In Maximizing Gene Expression. (Reznikoff, W., Gold, L. eds.) Butterworth, Stoneham, Massachusetts. pp. 195-224.

Stormo, G.D., Schneider, T.D., Gold, L.M. (1982) Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. 10, 2971-2996.

Subramanya, H.S., Arciszewska, L.K., Baker, R.A., Bird, L.E., Sherratt, D.J., Wigley, D.P. (1997) Crystal structure of the site-specific recombinase, *Xer*D. EMBO J. 16, 5178-5187.

Takagi, H., Morinaga, Y., Tsuchiya, M., Ikemura, H., Inouye, M. (1988) Control of folding of proteins secreted by a high expression secretion vector, pIN-111-ompA: 16-fold increase in production of active subtilisin E in *Escherichia coli*. Biotechnology 6, 948-950.

Tanaka, R., Yoshida, K., Nakayashiki, T., Masuda, T., Tsuji, H., Inokuchi, H., Tanaka, A. (1996) Differential expression of two *hem*A mRNAs encoding glutamyl-tRNA reductase proteins in greening cucumber seedlings. Plant Physiol. 110, 1223-1230.

Tebbe, C. C., Vahjen, W. (1993) Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and yeast. Appl. Environ. Microbiol. 59, 2657-2665.

Thanbichler, M., Neuhierl, B., Bock, A. (1999) *S*-Methylmethionine metabolism in *Escherichia coli*. J. Bacteriol. 181, 662-665.

Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680.

Torsvik, V. (1980) Isolation of bacterial DNA from soil. Soil Biol. Biochem. 12, 15-21.

Torsvik, V., Daae F. L., Goksoyr, J. (1995) Extraction, purification and analysis of DNA from soil bacteria. In Nucleic Acids in the Environment: Methods and Applications. (Trevors, J.T., van Elsas, J.D. eds.) Springer, New York. pp. 29-48.

Torsvik, V., Goksoyr, J., Daae F. L. (1990) High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. 56, 782–787.

Trevors, J.T., Lee, H., Cook, S. (1992) Direct extraction of DNA from soil. Microbial Releases 1, 111-115.

Tsai, Y-L., Olson, B.H. (1991) Rapid method for direct extraction of DNA from soil and sediments. Appl. Environ. Microbiol. 57, 1070-1074.

Tsai, Y-L., Olson, B.H. (1992a) Detection of low numbers of bacterial cells in soils and sediments by polymerase chain reaction. Appl. Environ, Microbiol. 58, 754-757.

Tsai, Y-L., Olson, B.H. (1992b) Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. Appl. Environ, Microbiol. 58, 2292-2295.

Tuan, R.S. (1997) Recombinant Gene Expression Protocols, Humana Press, Totowa, New Jersey.

Tumbula, D.L., Whitman, W.B. (1999) Genetics of Methanococcus: possibilities for functional genomics in Archaea. Mol. Microbiol. 33, 1-7.

Van Elsas, J.D., Smalla, K. (1995) Extraction of microbial community DNA from soils. In Molecular Microbial Ecology Manual. (Akkermans, A.D.L., van Elsas, J.D., deBruijn, F.J. eds.) Kluwer Academic Publishers, The Netherlands. pp. 1.3.3/1-1.3.3/11.

Vieille, C., Zeikus, G. (1996) Thermozymes: identifying molecular determinants of protein structural and functional stability. Trens. Biotechnol. 14, 183-189.

Vieira, J., Messing, J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19, 259-268.

Vogele, K., Schwartz, E., Welz, C., Rak, B. (1991) High level ribosomal frameshifting directs synthesis of IS150 gene products. Nucleic Acids Res. 19, 4377-4385.

Volkl, P., Huber, R., Drobner, E., Rachel, R., Burggraf, S., Trincone, A., Stetter, K.O. (1993) *Pyrobaculum aerophilum* sp. nov., a novel nitrate-reducing hyperthermophilic archaeum. Appl. Environ. Microbiol. 59, 2918-2926.

Voorhorst, W.G.B., Rik, I.L., Luesink, E.J., Devos, W.M. (1995) Characterization of the *celB* gene coding for Beta-galactosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site-directed mutation in *Escherichia coli*. J. Bacteriol. 177, 7105-7111.

Wagner, L.A., Weiss, R.B., Driscoll, R., Dunn, D.S., Gesteland, R.F. (1990) Transcriptional slippage occurs during elongation at runs of adenine or thymine in *Escherichia coli*. Nucleic Acids Res. 18, 3529–3535.

Ward, D.M., Bateson, M.M., Weller, R., Ruff-Roberts, A. (1992) Ribosomal analysis of microorganisms as they occur in nature. Adv. Microbial. Ecol. 12, 219-286.

Ward, D.M., Weller, R., Bateson, M.M. (1990) 16SrRNA sequences reveal numerous uncultured inhabitants in a natural community. Nature 345, 63-65.

Weinstein, J.D., Beale, S.I. (1983) Separate physiological roles and subcellular compartments for two tetrapyrrole biosynthetic pathways in *Euglena gracilis*. J. Biol. Chem. 258, 6799-6807.

Weller, R., Weller, J.W., Ward, D.M. (1991) 16SrRNA sequences of uncultivated hot spring cyanobacterial mat inhabitants retrieved as randomly primed cDNA. Appl. Environ, Microbiol. 57, 1146-1151.

White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C., Fraser, C.M. (1999) Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science 286, 1571-1577.

Wiegel, J., Braun, M., Gottschalk, G. (1981) *Clostridium thermoautotrophicum* species novum, a thermophile producing acetate from molecular hydrogen and carbon dioxide. Curr. Microbiol. 5, 255-260.

Wierenga, R.K., Terpstra, P., Hol, W.G.J. (1986) Prediction of the occurrence of the ADP-binding β-α-β fold in proteins using an amino acid sequence fingerprint. J. Mol. Biol. 187, 101-107.

Williams, D.C., Van Frank, R.M., Muth, W.L., Burnett, J.P. (1982) Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. Science 215, 687-688.

Woese, C.R., Fox, G.E.(1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. U S A. 74, 5088-5090.

Yanisch–Perron, C., Viera, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. Gene 33, 103–119.

Zhilina, T.N., Zavarzin, G.A. (1987) *Methanohalobium evestigatum* gen. nov. sp. nov. an extremely halophilic methane-producing archaebacterium. Dokl.Aka. Nauk. 293, 464-468.

Zhou, J., Bruns, M.A., Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. Appl. Environ. Microbiol. 62, 316-322.

Zillig, W., Stetter, K.O. Prangishvilli, C., Schafer, W., Wunderl, S., Janecovic, D., Holz, I., Palm, P. (1982) *Desulfurococcaceae*, the second family of the extremely thermophilic anaerobic, sulfur-respiring *Thermoproteales.* Zbl. Bakt. Hyg, I. Abt. Orig. C3, 302-317.

Zillig, W., Holz, I., Janecovic, D., Schafer, W., Reiter, W.D. (1983) The archaebacterium *Thermococcus cele* represents a novel genus with the thermophilic branch of the archaebacteria. Syst. Appl. Microbiol. 4, 88-94.