PHD

Expression of archaeal proteases

Sugar, Frank John

Award date:
2001

Awarding institution:
University of Bath

Link to publication

Alternative formats
If you require this document in an alternative format, please contact: openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Expression of Archaeal Proteases

Submitted by Frank John Sugar
for the degree of Ph.D. of the
University of Bath, 2001

COPYRIGHT

Attention is drawn to the fact that copyright of this thesis rests with its author. This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author. This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

Frank J. Sugar
# Contents

Contents
List of Figures
Abbreviations
Acknowledgements
Abstract

<table>
<thead>
<tr>
<th>Chapter One: Introduction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Project Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 The Archaea</td>
<td>1</td>
</tr>
<tr>
<td>1.3 Extremophiles</td>
<td>4</td>
</tr>
<tr>
<td>1.3.1 Hyperthermophiles</td>
<td>4</td>
</tr>
<tr>
<td>1.3.2 Halophiles</td>
<td>7</td>
</tr>
<tr>
<td>1.3.3 Other Extremophiles</td>
<td>8</td>
</tr>
<tr>
<td>1.4 Proteases</td>
<td>9</td>
</tr>
<tr>
<td>1.4.1 Serine Endopeptidases</td>
<td>10</td>
</tr>
<tr>
<td>1.4.2 Other Endopeptidases</td>
<td>10</td>
</tr>
<tr>
<td>1.4.3 Exopeptidases</td>
<td>12</td>
</tr>
<tr>
<td>1.5 Thermophilic Proteases</td>
<td>12</td>
</tr>
<tr>
<td>1.5.1 Thermophilic Proteases from the Archaea</td>
<td>14</td>
</tr>
<tr>
<td>1.5.1.1 <em>Pyrococcus</em> Species</td>
<td>14</td>
</tr>
<tr>
<td>1.5.1.2 <em>Sulfolobus</em> Species</td>
<td>18</td>
</tr>
<tr>
<td>1.5.1.3 Other Thermophilic Archaeal Proteases</td>
<td>20</td>
</tr>
<tr>
<td>1.5.2 Thermophilic Proteases from the Bacteria</td>
<td>21</td>
</tr>
<tr>
<td>1.5.2.1 <em>Thermus</em> Proteases</td>
<td>21</td>
</tr>
<tr>
<td>1.5.2.2 <em>Bacillus</em> Proteases</td>
<td>23</td>
</tr>
<tr>
<td>1.5.2.3 Other Thermophilic Bacterial Proteases</td>
<td>24</td>
</tr>
</tbody>
</table>
1.6 Halophilic Proteases 25
1.7 Biotechnological Applications of Thermostable Proteases 26
1.8 Thermophilic Esterases 28
1.9 Expression of Extremophilic Genes 29
1.10 Aims of the Project 32

Chapter Two: Materials and Methods 34

2.1 General Laboratory Reagents 34
2.1.1 Cell Culture 34
2.1.2 Molecular Biology Reagents 34
2.1.3 General Laboratory Chemicals 35
2.1.4 Protein Related Materials 35
2.1.5 Laboratory Equipment 36

2.2 Project Specific Materials 36
2.2.1 Gene Cloning 36
2.2.2 Assay Materials 37

2.3 General Laboratory Methods 37
2.3.1 Cell Culture 37
2.3.1.1 Escherichia coli 37
2.3.1.2 Bacillus subtilis 38
2.3.1.3 Culture Storage 38

2.3.2 Preparation of Competent Cells 38
2.3.2.1 Escherichia coli 38
2.3.2.2 Bacillus subtilis 39

2.3.3 Transformation 39
2.3.3.1 Escherichia coli JM109 cells 39
2.3.3.2 Escherichia coli BL21(DE3) and BL21(DE3) pLysS cells 40
2.3.3.3 Bacillus subtilis 40

2.3.4 Agarose Gel Electrophoresis 40

2.3.5 DNA Purification 41
2.3.5.1 Gel Purification 41
2.3.5.2 Plasmid Preparations from

*Escherichia coli* 41
2.3.5.3 Plasmid Preparations from

*Bacillus subtilis* 41

2.3.6 Polymerase Chain Reaction (PCR) 42
2.3.7 Restriction Digestion 42
2.3.8 Ligation of DNA 43
2.3.9 Sequencing 43
2.3.10 Induction of Expression of pET and pTYB1 vectors 43
2.3.11 Protein Estimation 44
2.3.12 SDS PAGE 44
2.3.13 Preparation of Cell Extracts, Cell Debris and Culture Supernatants 45
2.3.14 Enzyme Assays 45

2.3.14.1 General Peptidase Assays 46
2.3.14.2 Specific Peptidase Assays 46
2.3.14.3 Carboxylesterase Assay 46

Chapter Three: Cloning and Expression of *hlyR4*

from *Halofex mediterranei* 47

3.1 Introduction 47
3.2 Methods 48
3.2.1 Amplification and Cloning of *hlyR4* 48
3.2.2 Expression of *hlyR4* in *E. coli* 50
3.3 Results 51
3.3.1 Amplification and Cloning of *hlyR4* 51
3.3.2 Expression of *hlyR4* in *E. coli* 52
3.4 Discussion 55
Chapter Four: Cloning and Expression of an Esterase  
from *Archaeoglobus fulgidus*  

4.1 Introduction 59  
4.2 Methods 61  
\[4.2.1\] Amplification and Cloning of AF1763 61  
\[4.2.2\] Expression of AF1763 in *E. coli* 62  
4.3 Results 62  
\[4.3.1\] Amplification and Cloning of AF1763 62  
\[4.3.2\] Expression of AF1763 in *E. coli* 64  
\[4.3.3\] Assaying for Activity of the Recombinant Enzyme 64  
4.4 Discussion 67

Chapter Five: Cloning and Expression of afpl  
from *Archaeoglobus fulgidus*  

5.1 Introduction 70  
5.2 Methods 72  
\[5.2.1\] Amplification and Cloning of afpl 72  
\[5.2.2\] Expression of afpl in *E. coli* 74  
\[5.2.3\] Chitin Column Chromatography 75  
5.3 Results 76  
\[5.3.1\] Amplification and Cloning of afpl 76  
\[5.3.2\] Expression of afpl from pET vectors in *E. coli* 78  
\[5.3.3\] Expression of afpl from pTYB1 vector in *E. coli* 79  
5.4 Discussion 82

Chapter Six: Cloning and Expression of an Extracellular Protease  
from *Aeropyrum pernix*  

6.1 Introduction 87  
6.2 Methods 90
List of Figures

Figure 1.1 The universal rRNA tree of life.
Figure 1.2 Phylogenetic tree of the Archaea.
Figure 1.3 Catalytic mechanism of serine proteases.
Figure 1.4 Other proteolytic species from *Pyrococcus furiosus*.
Figure 3.1 Sequence alignment of the signal sequences of bacteriorhodopsin and subtilisin with halolysin R4.
Figure 3.2 Separation of PCR products of pre-pro, pro and mature gene constructs.
Figure 3.3 *Ndel/Bam*HI restriction digestion of transformant plasmids.
Figure 3.4 SDS gels showing expression of pre-pro-protein, pro-protein and mature halolysin.
Figure 3.5 Table of activities of subtilisin against azocasein.
Figure 4.1 Agarose gel showing PCR products from AF1763 amplification.
Figure 4.2 Agarose gel showing digested plasmids from transformed cells.
Figure 4.3 Hanes plot for AF1763 against p-nitrophenyl butyrate.
Figure 4.4 The dependence of velocity on substrate concentration for AF1763 against p-nitrophenyl butyrate.
Figure 5.1 A diagram showing the basis of the IMPACT system.
Figure 5.2 Comparison of the C-terminal ends of the two afpl constructs with the native sequence.
Figure 5.3 Agarose gels showing PCR amplifications of afpl.
Figure 5.4 Agarose gels showing restriction digestion of transformant plasmids.
Figure 5.5 SDS gels showing expression of afpl from pET vectors.
Figure 5.6  SDS gels showing expression of afpl from pTYB1.
Figure 5.7  SDS gel showing renaturation of afpl.
Figure 5.8  SDS gel showing the protein obtained after stripping the chitin column.
Figure 6.1  Sequence alignment of subtilisin and APE0263 pre- and pro-regions.
Figure 6.2  Organisation of the subtilisin/APE0263 gene fusion in pBE3.
Figure 6.3  Genomic DNA extracted from Aeropyrum pernix cells.
Figure 6.4  Amplification of a segment of APE0263.
Figure 6.5  Amplification of the APE0263 gene still containing the internal BamHI site.
Figure 6.6  Amplification of APE0263 without the internal BamHI site.
Figure 6.7  Agarose gel to show digested plasmids.
Figure 6.8  Agarose gel showing digestion of pBE3 vector containing the insert.
Figure 6.9  Graph obtained from gel filtration showing protein and activity elution.
Figure 6.10 Calibration curve for the gel filtration column.
Figure 6.11 SDS gel showing separation of two of the most active fractions obtained after gel filtration.
Figure 6.12 Thermoactivity of APE0263.
Figure 6.13 Hanes plot for APE0263 against ala-ala-pro-phe-pNA.
Figure 6.14 The dependence of velocity on substrate concentration for APE0263 against ala-ala-pro-phe-pNA.
Figure 6.15 Comparison of $K_m$ values for ala-ala-pro-phe-pNA for various proteases.
Figure 6.16 pH profile of APE0263.
Figure 6.17 Thermal inactivation of gel filtration-purified APE0263 at 90 and 95 °C.
Figure 6.18 Thermal inactivation of concentrated APE0263 at 80, 90 and 99 °C.
Figure 6.19  Effect of various inhibitors on the activity of APE0263.
Figure 6.20  Effect of various cations on the activity of APE0263.
Figure 6.21  Hanes plot for APE0263 against ala-ala-pro-leu-pNA.
Figure 6.22  The dependence of velocity on substrate concentration for APE0263 against ala-ala-pro-leu-pNA.
# Abbreviations

<table>
<thead>
<tr>
<th>Alanine</th>
<th>Arginine</th>
<th>Asparagine</th>
<th>Aspartic Acid</th>
<th>Cysteine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>Arg</td>
<td>Asn</td>
<td>Asp</td>
<td>Cys</td>
</tr>
<tr>
<td>A</td>
<td>R</td>
<td>N</td>
<td>D</td>
<td>C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Glutamic Acid</th>
<th>Glutamine</th>
<th>Glycine</th>
<th>Histidine</th>
<th>Phenylalanine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu</td>
<td>Gln</td>
<td>Gly</td>
<td>His</td>
<td>Phe</td>
</tr>
<tr>
<td>E</td>
<td>Q</td>
<td>G</td>
<td>H</td>
<td>F</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Leucine</th>
<th>Lysine</th>
<th>Methionine</th>
<th>Tryptophan</th>
<th>Tyrosine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu</td>
<td>Lys</td>
<td>Met</td>
<td>Trp</td>
<td>Tyr</td>
</tr>
<tr>
<td>L</td>
<td>K</td>
<td>M</td>
<td>W</td>
<td>Y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Serine</th>
<th>Threonine</th>
<th>Serine</th>
<th>Phenylalanine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>Thr</td>
<td>Thr</td>
<td>Tyr</td>
</tr>
<tr>
<td>S</td>
<td>T</td>
<td>T</td>
<td>Y</td>
</tr>
</tbody>
</table>

- **A$_{410}$**: absorbance at 410 nm
- **APS**: ammonium persulphate
- **bp**: base pairs
- **BSA**: bovine serum albumin
- **CAPS**: 3-[Cyclohexylamino]-1-propanesulfonic acid
- **CE**: cell extract
- **CTAB**: hexadecyltrimethylamonium bromide
- **DMSO**: dimethylsulfoxide
- **DNA**: deoxyribonucleic acid
- **dNTP**: deoxynucleotide triphosphate
- **DTT**: dithiothreitol
- **EDTA**: (di sodium) ethylenediamine tetraacetate
- **EPPS**: N-[2-Hydroxyethyl]piperazine-N’-[3-propanesulfonic acid]
- **HEPES**: N-[2-Hydroxyethyl]piperazine-N’-[2-ethanesulfonic acid]
- **g**: relative centrifugal force
- **IPTG**: isopropyl-β-D-thiogalactoside
- **kbp**: kilobase pairs
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{\text{cat}}$</td>
<td>catalytic rate constant</td>
</tr>
<tr>
<td>$K_M$</td>
<td>Michaelis constant</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani (medium)</td>
</tr>
<tr>
<td>MES</td>
<td>2-[N-Morpholino]ethanesulfonic acid</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-[N-morpholino]propanesulphonic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>NA</td>
<td>nutrient agar</td>
</tr>
<tr>
<td>NB</td>
<td>nutrient broth</td>
</tr>
<tr>
<td>OD$_{600}$</td>
<td>optical density at 600 nm</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCMB</td>
<td>$p$-chloromercuribenzoate</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethane sulfonyl fluoride</td>
</tr>
<tr>
<td>pNA</td>
<td>para-nitroanilide</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SG</td>
<td>Schaefers (medium)</td>
</tr>
<tr>
<td>TAE</td>
<td>tris-acetate, EDTA (buffer)</td>
</tr>
<tr>
<td>TCA</td>
<td>trichloroacetic acid</td>
</tr>
<tr>
<td>TEMED</td>
<td>N, N', N'', tetramethylethylene diamine</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>U</td>
<td>units (of enzyme activity)</td>
</tr>
<tr>
<td>$V_e$</td>
<td>elution volume</td>
</tr>
<tr>
<td>$V_o$</td>
<td>void volume</td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>maximum velocity</td>
</tr>
<tr>
<td>(v/v)</td>
<td>volume to volume ratio</td>
</tr>
<tr>
<td>(w/v)</td>
<td>weight to volume ratio</td>
</tr>
</tbody>
</table>
Acknowledgements

Firstly, I would like to thank my supervisors, Mike and David for all their help, advice and encouragement throughout the project, and without whom it would not have been possible. Through their knowledge and teaching, I have learnt a lot during the project, which has hopefully turned me into a better scientist.

Secondly, I would like to thank all the post-docs who have helped me in the technical aspects of the project, specifically Carl for his endless patience and seemingly endless knowledge. Also, HC and Ursula for their help with halophiles and proteases, respectively. I would also like to thank our visiting post-docs, Ida and Eva for various helpful advice. A big thanks has got to go to Debs for the help with all the halophile work. The obligatory acknowledgement for helpful discussions goes to Douglas Clark from GeneSys Ltd. for advice about PCR and vectors, Francene Perler and her colleagues from NEB for help with the IMPACT system, Francis Arnold from CalTech for providing us with the pBE3 vector and enabling me to get some positive results, and Paul Whitely from next door for his help with all manner of questions. On the technical side, thanks to Paul Jones for doing the DNA sequencing.

Almost finally, a huge thank you to all the members, past and present, of Labs 1.33 and 1.28 for providing plenty of entertainment and putting up with the Molly and monkey impressions. From our lab, Debs, Mick, Pete, Krish, Carlos, Cat, Nina, Becky (also for putting up with me as a housemate for two years), Tree, Tina and Amie. From next door, JJ and Richie for being bonkers, Ben, Sue, Sharmy, Catherine, Hannah and Tracey for putting up with me talking at them. Thanks also to everyone else I have met whilst in Bath, specifically Simon for the use of his computer and my fellow year-mates for some great social occasions.

Finally, a big thank you goes to the BBSRC and Kymed GB for funding this project.
To Mum and Dad
Abstract

The gene encoding APE0263, an extracellular alkaline serine protease identified from the genome sequence of the aerobic hyperthermophilic Archeon *Aeropyrum pernix*, has been cloned and expressed in *Bacillus subtilis*. The gene was cloned into the *Escherichia coli* / *B. subtilis* shuttle vector, pBE3, as a fusion to the pre- and pro­regions of subtilisin E. When expressed in *B. subtilis*, APE0263 was secreted into the culture medium. A combination of ammonium sulphate precipitation and gel filtration were used to partially purify the recombinant enzyme.

APE0263 is active against the peptide substrate N-succinyl-ala-ala-pro-phe-p-nitroanilide with an optimum for activity at pH 7.5-8.5 and 80 °C and a $K_m$ of approximately 900 µM. It also exhibits activity against the peptide substrate N-succinyl-ala-ala-pro-leu-p-nitroanilide, but not N-succinyl-ala-ala-pro-p-nitroanilide or N-succinyl-gly-gly-phe-p-nitroanilide, suggesting it may require a proline residue at the P2 position. Its thermostability depends upon the protein concentration, with a dilute enzyme sample having half-lives of 9.6 and 8.2 min at 90 and 95 °C, respectively. A concentrated enzyme sample, however, exhibited a much higher thermostability, having half-lives of 180 and 26 min at 90 and 99 °C, respectively. APE0263 is a monomeric protein with a relative molecular mass of 35 k.

The intracellular protease, afpl, from another hyperthermophilic Archeon, *Archaeoglobus fulgidus*, was also cloned and expressed in two different pET vectors from *Escherichia coli*. However, despite obtaining high levels of recombinant protein, activity was not observed against a number of protein and peptide substrates. Sequencing of two genes from unsuccessful expression systems has revealed the presence of single mutations that are thought to inactivate the protein. It is thought that the native gene product is toxic to the host cells and so genes containing inactivating mutations are selected. The expression of afpl as a fusion in the IMPACT system was also attempted. However, the recombinant protein was expressed in an insoluble form. Despite being subsequently renatured, it failed to
cleave from an affinity column used to purify it, possibly due to the method of renaturation.

The pET system was successfully utilised to clone and express a thermophilic esterase, AF1763, also from *A. fulgidus* and an extracellular protease, *hlyR4*, from the halophilic Archeon, *Haloferax mediterranei*. Despite the presence of a number of mutations in the gene, the recombinant esterase was active against both *p*-nitrophenyl acetate and *p*-nitrophenyl butyrate, having $K_M$ values of approximately 4mM and 1.4 mM against the former and later, respectively. Recombinant *hlyR4* was expressed mainly in the form of insoluble inclusion bodies, but when soluble protein was obtained, the enzyme was found to be inactive against the protein substrates azocasein and azoalbumin.
1. Chapter One

Introduction

1.1 Project Background

Proteases have a large variety of applications in a range of biotechnological processes, mainly in the detergent and food industries; however, there is a continuing need for enzymes with novel specificities and high chemical and thermal stability. Enzymes from the hyperthermophilic and halophilic Archaea are generally much more tolerant to extremes of temperature and ionic strength, respectively, than their mesophilic counterparts and therefore offer great potential advantages.

The aims of this project are to produce extremophilic proteases in mesophilic hosts, to characterise them in terms of their substrate specificities and thermal properties, and to purify the recombinant proteins.

1.2 The Archaea

Classically, biologists grouped the living world into a basic dichotomy: that which was not plant was animal. Bacteria were considered plants until they were discovered to resemble plants even less than animals did. This initial division was therefore subsequently amended, until it was extended to divide the living world into five Kingdoms: Animalia, Plantae, Fungi, Protista and Monera. These Kingdoms or domains were based on phenotypic observations such as morphological, physiological and ecological characteristics. This view of life coexisted with an even more general dichotomy of structuring the living world into Prokaryotae and Eukaryotae, based cytologically on the presence or absence of a nucleus, which was
proposed in 1938 by Chatton and formalised in 1968 by Murray (reviewed by Woese et al., 1990).

However, an unsuspected phylogeny was revealed from work carried out by Carl Woese, based on molecular sequence analysis of 16S ribosomal RNAs. Molecular phylogeny revealed that there are not five evolutionary lines, represented by the five kingdoms of life, but three (Woese and Fox, 1977), two of which are exclusively prokaryotic, the third being the eukaryotic lineage. These three 'urkingdoms' or primary kingdoms were known as the eubacteria, the urkaryotes (corresponding approximately to the prokaryote and eukaryote division) and the archaeabacteria (Woese and Fox, 1977). This view was modified later, in a proposal that urkingdom be replaced by a new taxon known as Domain and that eubacteria, archaeabacteria and urkaryotes be changed to Bacteria, Archaea and Eukarya, respectively (Woese et al., 1990).

The universal phylogenetic tree of life shown below (figure 1.1) shows that evolution from the universal ancestor of all cells proceeded initially in two directions, the Bacteria and the Archaea/Eukarya lineages. Eukarya and Archaea later diverged to yield the three domains of life.

![Figure 1.1: The universal rRNA tree of life (adapted from Brown and Doolittle, 1997).](image-url)
It is seen that the archaeal domain is the closest domain to the universal origin of life, suggesting that it is the most primitive of the domains. This is supported by the fact that many Archaea inhabit environments that are thought to represent the conditions under which life originated. There are three kingdoms within the Archaea according to rRNA analysis, Crenarchaeota, Euryarchaeota and Korarchaeota, the first two consisting of recognised organisms and the third being a tentative assignment based on PCR analysis of rRNA sequences from environmental samples (Barns et al., 1996), see figure 1.2 below.

**Figure 1.2**: Phylogenetic tree of the Archaea (adapted from Barns et al., 1996).

The Crenarchaeota kingdom generally consists of hyperthermophilic and thermoacidophilic organisms, (e.g. Desulfiurococcus and Sulfolobus, respectively) whereas the kingdom Euryarchaeota is more diverse, comprising methanogens (e.g. Methanosarcina) and halophiles (e.g. Haloferax) as well as hyperthermophiles (e.g. Pyrococcus and Archaeoglobus). It is important to note that the characterisation of
archaeal diversity has been limited by the fact that only a small fraction (<0.1 %) of organisms from environmental samples can be cultured by standard methods (Hugenholtz and Pace, 1996) and so the recognised organisms may not represent the true composition of the Archaea. Methods to characterise diversity that are independent of culturing are becoming increasingly important and in fact rRNA sequence analysis of soil and marine microbial populations has revealed the presence of mesophilic (Bintrim et al., 1997) and psychrophilic Archaea (Preston et al., 1996; Stein et al., 1996). This has led to the awareness that Archaea do not just occupy a limited number of extreme environmental niches. Definition of an archaeal organism must not therefore be phenotypic, the more so as there are hyperthermophilic members of the Bacterial domain, e.g. *Aquifex aeolicus* and *Thermotoga maritima*.

The tripartite classification, however, has its opponents who challenge some of the assumptions applied (Mayr, 1990; Margulis and Guerrero, 1991) and an alternative tree of life has been proposed (e.g. Rivera et al., 1998).

### 1.3 Extremophiles

Organisms that thrive and indeed require environments previously considered inhospitable are known as extremophiles and are widely distributed in the ecosystem. They have various phenotypes reflecting adaptation to extremes of temperature (psychrophiles, thermophiles and hyperthermophiles), ionic strength (halophiles), pH (acidophiles and alkaliphiles) and/or pressure (barophiles). Only the hyperthermophiles and halophiles are directly related to this project and will be described in any detail.

#### 1.3.1 Hyperthermophiles

Whilst the term ‘thermophile’ is a general one for organisms growing at elevated temperatures, strictly it only describes those organisms having a growth optimum between 55 and 80 °C, with organisms having optima above 80 °C being described
as hyperthermophiles. Extremely thermophilic organisms have been described since
the early 1970s (Brock et al., 1972), but the first report of an organism capable of
growth above the normal boiling point of water was a decade later (Stetter, 1982).
Despite a report shortly afterwards (Baross and Deming, 1983) which described
growth of organisms at temperatures of at least 250 °C, and was subsequently
disputed (Trent et al., 1984; White, 1984), the upper known temperature for life has
not increased substantially since. Presently, the Archaeon Pyrolobus fumarii
represents the most extreme hyperthermophile known, growing optimally at 106 °C
and up to 113 °C (Blochl et al., 1997). This organism does not grow below 90 °C
(which is above the pasteurisation temperature of 85 °C) and a significant number of
cells from an exponentially growing culture survive autoclaving for one hour at
121°C. The authors speculated that optimising the culture conditions might further
extend this upper temperature, but admitted that due to the limiting hydrolytic
stability of cellular components above these temperatures (Bernhardt et al., 1984),
113 °C was probably nearing the absolute maximum growth temperature.

How hyperthermophilic proteins achieve their stability at such high
temperatures has been the subject of much investigation. Whilst the perhaps initially
surprising results showed that hyperthermophilic proteins not only shared the same
catalytic mechanisms but also have very similar structures and high sequence
homology to their mesophilic counterparts, further studies have elucidated a number
of factors important in conferring thermostability (e.g. Vieille and Zeikus, 1996;
Danson and Hough, 1998; Vetricani et al., 1998; Robb and Clark, 1999; Kumar et al.,
2000 and references within). Another perhaps unanticipated conclusion from this
eyarly research is that thermostability is not the result of a single factor. For example,
it is well known that hyperthermophiles have thermoprotectants that minimise
protein denaturation, including complex organic compounds (Adams, 1993), and
also potassium ions, which are present at concentrations of 0.8 M in Pyrococcus
woesei (Scholz et al., 1992) and at least 1.0 M in hyperthermophilic methanogens
(Hensel and König, 1988). However, the hyperthermophile Thermoproteus tenax
contains potassium ions at concentrations less than 100 mM (Hensel et al., 1987).
Thus, it is clear that protein thermostability is not due to a universal mechanism, but instead comes from a number of subtle interactions as a result of specific amino acid sequence and structural differences, briefly described below.

- **Amino acid composition**

Some amino acids undergo covalent modification at higher rates at the elevated temperatures that hyperthermophiles withstand and thermolabile amino acids that have a tendency to undergo oxidation (cysteine and methionine) and deamidation (asparagine and glutamine) are generally found at lower frequencies in hyperthermophilic proteins. Thermolabile residues in hyperthermophilic proteins are usually found within the interior of the protein where they are more stable (Hensel et al., 1992).

- **Increased packing density**

A trend commonly associated with protein hyperthermophilicity is an increase in compactness due to shorter surface loops (Russell et al., 1994). Hyperthermophilic proteins also tend to have a reduced number of internal cavities due to increased packing density (Anderson et al., 1993; Hurley et al., 1992; Britton et al., 1995). This is associated with an increase in core hydrophobicity (Haney et al., 1997).

- **Cooperative association and ionic networks**

In oligomeric proteins, complete thermal unfolding is thought to proceed via a two step process, whereby oligomer dissociation precedes the monomer unfolding (Robb and Clark, 1999). It therefore follows that strong intermolecular bonds such as ion pairs can inhibit this process. It appears that some hyperthermophilic proteins may have oligomeric structures, whereas the corresponding proteins are monomeric in mesophiles (Robb and Clark, 1999). For example phosphoribosyl anthranilate isomerase is found to be dimeric in *Thermotoga maritima* and monomeric in enteric bacteria (Hennig et al., 1997).

The importance of surface ion pairs in protein thermostability was recognised early on (Perutz and Raidt, 1975) and comparison of citrate synthase structures has
shown an increase in inter-subunit ion pairs in the \textit{Pyrococcus furiosus} enzyme, compared to its less thermostable counterparts (Russell \textit{et al.}, 1997). Glutamate dehydrogenase from \textit{Pyrococcus furiosus} was found to contain an extended network of ion pairs compared to the enzyme from the mesophile \textit{Clostridium symbiosum} (Yip \textit{et al.}, 1995) and the introduction of a network into the less thermostable enzyme from \textit{Thermococcus litoralis} has resulted in an increased thermostability (Vetriani \textit{et al.}, 1998).

- Rigidity

An increased rigidity of thermostable enzymes is required to preserve their active conformation under the more severe denaturing conditions they encounter (Danson and Hough, 1998). This change is however thought to decrease the catalytic efficiency of thermophilic enzymes due to it limiting their conformational flexibility.

1.3.2 Halophiles

The halophilic Archaea inhabit hypersaline environments where salt concentrations can approach saturation (up to 6 M NaCl). Such environments include marine salterns and hypersaline lakes such as the Dead Sea and the Great Salt Lake. These organisms are pigmented and their characteristic red blooms were noted in ancient times in the solar salterns used to mine salt from seawater. More recently, in the late nineteenth century, the first halophilic organisms were cultured from spoiled food products where salt had been used as a preservative.

Halophilic Archaea protect themselves from the high osmotic pressure of the environment by maintaining internal concentrations of salt, usually KCl, sufficient to establish an osmotic balance (Kushner, 1988). Consequently, the whole biochemistry of halophiles is adapted to function at high salt concentrations and in fact halophilic proteins can not only tolerate salt concentrations sufficient to denature and precipitate many mesophilic proteins, but actually require them for activity. This evolutionary strategy is in contrast to the halotolerant bacteria, which effect osmoregulation by accumulating low molecular weight organic solutes. Most
extreme halophiles grow optimally at NaCl concentrations around 4.5 M and virtually all can grow at saturation, albeit more slowly, with cell lysis occurring at concentrations of around 1 M and below (DasSarma, 1995).

NaCl and KCl are both salting-out agents, that is they tend to precipitate proteins by increasing the surface tension at the protein-water interface. Therefore, a halophilic protein must attract water back to its surface. Early observations of halophilic proteins showed that they tend to have an excess of acidic amino acids and a decreased number of hydrophobic amino acids (Reistad, 1970; Lanyi, 1974). A number of structures of halophilic proteins are available, [malate dehydrogenase from Haloarcula marismortui (Dym et al., 1995), ferredoxin from Haloarcula marismortui (Frolow et al., 1996), dihydrofolate reductase from Haloferax volcanii (Pieper et al., 1998) and tetrahydromethanopterin formyltransferase from the thermophilic halophile Methanopyrus kandlerin (Ermler et al., 1997)] and these have helped formulate the current hypothesis on halophilicity. These structures show that the surface of halophilic proteins is primarily composed of acidic amino acids, which are thought to attract water molecules to the surface of the protein in the form of hydrated salt ions.

1.3.3 Other Extremophiles

As mentioned above, organisms also exist that not only thrive above the mesophilic temperatures we consider normal, but also below them. Such organisms, termed psychrophiles, are defined as having a growth optimum below 20 °C, and have been isolated from marine and soil environments where the temperature is constantly low, e.g. Antarctica. A recent review has summarised our current knowledge on cold-activity of enzymes from psychrophilic organisms (Russell, 2000 and references within).

Barophiles (or piezophiles) are organisms that survive high pressure environments (up to 50 Mpa), which were traditionally seen as denaturing conditions. However, more recent research has shown pressure to enhance the thermostability of a significant number of thermostable proteins (Hei and Clark, 1994). Barophilic organisms have been isolated from the deep sea, which is
characteristically a low temperature environment, but in the vicinity of hydrothermal vents, hyperthermophilic barophiles exist. The molecular biology and physiology of barophiles has been reviewed elsewhere (Kato and Bartlett, 1997; Horikoshi, 1998).

Hyperthermophilic acidophiles (growth at pH <4), e.g. *Sulfolobus*, and halophilic alkaliphiles (growth at pH >9), e.g. *Natronococcus*, also exist within the Archaea, which although they survive those extreme pH values, have an internal environment close to neutrality (van de Vossenberg *et al.*, 1995). However, they must have extracellular enzymes adapted to these extreme conditions. Unfortunately, there have been no reports of structural data from these enzymes to date and so the basis of protein acidophily and alkaliphily is not understood.

1.4 Proteases

The Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) has set out recommendations for the classification of enzymes that act on peptide bonds, named peptidases (EC 3.4), (http://www.chem.qmw.ac.uk/iubmb/enzyme/EC34/). Peptidases are often difficult to define, due to their huge diversity of action and structure. They are currently classified via three criteria: type of reaction catalysed, chemical nature of the catalytic site and evolutionary relationship. It is recommended that the term "peptidase" or "peptide hydrolase" be used for any enzyme that hydrolys peptide bonds. Peptidases are further divided into enzymes that either cleave distal to the terminus of the peptide chain (endopeptidases; EC 3.4.21-25 and 3.4.99), or proximal to the terminus of the peptide chain (exopeptidases; EC 3.4.11-19).

It should be noted that the use of the term "peptidase" be synonymous with "protease", even though peptidase was previously restricted to the subclass exopeptidase. Also, the term "proteinase", previously taken to mean endopeptidase, has been replaced by endopeptidase for consistency.

The endopeptidases are further categorised into subclasses on the basis of their catalytic mechanisms.
1.4.1 Serine Endopeptidases

The serine endopeptidases (EC 3.4.21) comprise two main families, the chymotrypsin family and the subtilisin family. Both families share a common catalytic mechanism and active site geometry, but have differing overall structures. They all have a catalytic triad, comprising histidine, aspartic acid and serine groups, at their active site and are characterised by irreversible inactivation by, for example, phenylmethylsulfonyl fluoride (PMSF) (Turini \textit{et al.}, 1969).

Their catalytic mechanism (Fastrez and Fersht, 1973) is usually via a two step reaction, shown in figure 1.3 overleaf, commencing with an acylation step, resulting in the formation of a covalently linked enzyme-peptide intermediate between the substrate and the essential serine, and hydrolysis of the peptide fragment. A deacylation step follows in which water coordinates a nucleophilic attack on the enzyme-peptide intermediate, resulting in loss of the peptide.

1.4.2 Other Endopeptidases

Cysteine (or thiol) endopeptidases (EC 3.4.22) are characterised by the presence of a catalytic dyad of cysteine and histidine (Polgar, 1990; Storer and Menard, 1994) and their resulting susceptibility to sulfhydryl agents such as \textit{p}-chloromercuribenzoate (PCMB). The most studied member of this group is the plant protease, papain.

The aspartic endopeptidases (EC 3.4.23), also referred to as acidic endopeptidases, have an active-site aspartic acid, usually within the motif Asp-(Ser/Thr)-Gly and are inhibited by pepstatin. Most aspartic proteases belong to the pepsin family, although a second family exists, comprising viral proteases. Structural and kinetic studies (James \textit{et al.}, 1977; Pearl, 1987) have shown that aspartic acid peptidases depend on a pair of aspartic acid residues for catalytic activity, which is situated between the lobes of a bilobal structure. Each lobe contributes one of the essential residues.
Metalloendopeptidases (EC 3.4.24) require a divalent ion for activity, usually zinc (Holmes and Matthews, 1981), and are therefore inhibited by chelating agents such as diaminoethanetetra-acetic acid (EDTA). Most metallopeptidases contain the metal binding motif His-Glu-Xaa-Xaa-His.
Chapter One

The threonine endopeptidases (EC 3.4.25) contain a conserved N-terminal threonine residue, the hydroxyl group of which acts as the active-site nucleophile (Seemuller et al., 1995) and are inhibited by the peptide aldehyde inhibitor acetyl-leu-leu-norleucinal (calpain inhibitor 1). Proteasomes belong to this class of proteases, and interestingly mutation of the active-site threonine to a serine group in archaeal proteasomes does not abolish activity (Seemuller et al., 1995).

Endopeptidases of unknown catalytic mechanism are assigned to the subclass EC 3.4.99.

1.4.3 Exopeptidases

The exopeptidases are further categorised based on the specific reaction catalysed at the terminus of the peptide chain. Exopeptidases that liberate a single C-terminal amino acid are named carboxypeptidases (EC 3.4.16-18) and those that liberate a dipeptide are named peptidyl-dipeptidases (EC 3.4.15). Carboxypeptidases are assigned to the three groups based on their catalytic mechanism: the serine carboxypeptidases (EC 3.4.16), the metallopeptidases (EC 3.4.17), and the cysteine carboxypeptidases (EC 3.4.18). Exopeptidases acting at the N-terminus liberate a single amino acid [aminopeptidases (EC 3.4.11)] or a dipeptide or a tripeptide [dipeptidyl-peptidases and tripeptidyl-peptidases (EC 3.4.14)]. Other exopeptidases exist, those that are specific for dipeptides, named dipeptidases (EC 3.4.13) and those that remove terminal residues with peptide linkages other than those of α-carboxyl to α-amino groups, known as omega peptidases (EC 3.4.19).

1.5 Thermophilic Proteases

Proteases have long had great prominence in fundamental studies, with the earliest enzyme physiology investigations concentrating on their roles in nutrition and digestion. Gastric digestion has been studied since the 1700s, with the discovery of
pepsin by Wassman in 1839 (reviewed in Neidleman, 1989) and its crystallisation (Northrup, 1930) following later. As their regulatory and metabolic importance has become apparent, there has been an increasing interest in these areas as well. Proteases have also figured largely in early biotechnology applications, mainly in the detergent industry, with the first enzymatic detergent, “Burnus”, consisting of a crude pancreatic extract, appearing in 1913 (Rao et al., 1998).

This widespread interest in proteases has been more recently reflected in the growing attention of thermophilic and hyperthermophilic proteases. There are well known advantages of using thermophilic enzymes in industry (e.g. Sonnleiter and Fiechter, 1983; Daniel et al., 1985/6). Due to their high temperature environments, thermophilic and hyperthermophilic organisms contain enzymes that are highly temperature stable. It is also generally accepted that proteins that are resistant to heat are also resistant to chemical denaturation, as the interactions leading to enhanced thermal stability confer stability to other denaturing and chaotropic agents. There are three modes of activity loss by heat in proteases (Daniel et al., 1995), the first two being generic mechanisms, the last being specific to proteases. Firstly, heating results in denaturation, that is conformational unfolding of the protein. Secondly, irreversible covalent modifications can occur, for example deamidation of asparagine and glutamine residues (Daniel et al., 1996). Finally, in the case of proteases, autolysis becomes more crucial at higher temperatures. Denaturation is in theory reversible, but renaturation is often prevented by subsequent aggregation. Renaturation in proteases is also often prevented by autolysis, as unless the unfolding is rapid and complete, the partially unfolded proteins are more accessible to the remaining active enzyme molecules.

One of the earliest reports of a thermophilic protease was that of Heinen and Heinen (1972), describing an extracellular protease from a Bacillus species growing at 72 °C. More recently, a variety of reports of proteolytic species have been described from the hyperthermophiles.
1.5.1 Thermophilic Proteases from the Archaea

The first archaeal thermophilic protease to be described was archaelysin from *Desulfurococcus* st. Tok12S1 (Cowan *et al.*, 1987c). Archaelysin is an extracellular serine protease which was purified from the culture medium, with an apparent relative molecular mass of 52 k. It is extremely thermoactive, with activity being measured at 125 °C, although it denatures very rapidly above 115 °C, and has a half-life at 95 °C of 70-90 min. Neither EDTA nor calcium ions have any effect on activity, thus excluding it from being a metalloprotease.

Subsequently, numerous reports of proteases have appeared from a variety of archaeal species as well as the hyperthermophilic bacteria.

1.5.1.1 *Pyrococcus* Species

The most extensively studied organism of this genus, and in fact of the whole of the Archaea, is the hyperthermophile *Pyrococcus furiosus* DSM 3638 (Fiala and Stetter, 1986). This anaerobic, marine heterotroph has been found to produce a heterogeneity of proteases (Blumentals *et al.*, 1990; Connaris *et al.*, 1991; Snowden *et al.*, 1992). Blumentals *et al.* (1990) reported the presence of five intracellular proteases, two of which, named S66 and S106, were SDS-resistant and had extremely high thermostability, with an extract of the two showing a half-life of 33 h at 98 °C. In addition to being resistant to SDS, the two proteases were resistant to 8M urea, 80 mM dithiothreitol and 5 % β-mercaptoethanol. Inhibition experiments showed S66 to be a serine type protease without a requirement for divalent cations. Immunoblot analysis has indicated that S66 is related to a 200 k precursor that does not have proteolytic activity, and also that it is unrelated to both S106 and the other three proteases. The gene encoding S66 was later sequenced and expressed in *Escherichia coli* (Halio *et al.*, 1996) and found to encode an 18.8 k subunit of an oligomeric protein, termed pfpl, active as two species, the 66 k protein previously reported (Blumentals *et al.*, 1990) and an 86 k protein, as judged by SDS analysis. The recombinant form of the protein was significantly less thermostable than the native form purified by treatment with SDS, having a half-life of only 19 min at
95°C. The authors speculated that this was possibly due to the fact that it was necessary to express the gene with a fusion tag to reduce toxicity and that not only were a small fraction of the fusion proteins subsequently cleaved, but those that were still contained three N-terminal amino acids of the fusion protein. Subsequent purification of pfpl from *P. furiosus* cell extracts (Halio et al., 1997) under nondenaturing conditions showed pfpl to exist as two species of ~124 k and ~59 k, corresponding to hexameric and trimeric forms, respectively, re-emphasising the fact that migration of proteins in SDS-PAGE is not always a true indication of their size (Dunker and Rueckert, 1969).

Eggen et al. (1990) identified and characterised a serine protease, named pyrolysin from *P. furiosus*, which was reported to be associated with the cell envelope, although activity was observed in the cytoplasm. They noted five activity bands on substrate SDS-PAGE, including one at 65 k which may correspond to S66. Pyrolysin is extremely thermostable and thermoactive, with a half-life at 100 °C of 4h, a temperature optimum of 115 °C and activity over a broad pH range (6.5-10.5). Pyrolysin was subsequently purified and its gene cloned and sequenced (Voorhorst et al., 1996). Two active species were obtained, a high and low relative molecular mass form, with identical N-terminal sequences. Incubation of the high relative molecular mass form at 95 °C produced the low relative molecular mass form of 80k, probably as a consequence of autoproteolytic removal of a C-terminal part of pyrolysin, and the sequencing of the gene revealed a pre-pro-enzyme organisation. An homology-modelled three-dimensional structure has been produced and comparison of pyrolysin with mesophilic and thermophilic proteases has identified features that could be important in protein stability (Voorhorst et al., 1997).

In contrast to the five proteases identified by Blumentals et al. (1990) and Eggen et al. (1990), Connaris et al. (1991) reported the presence of 13 discrete proteases using substrate-containing SDS-PAGE, including one of 66 k which may correspond to S66. The authors stated that the substrate concentrations used by Blumentals et al. (1990) and Eggen et al. (1990) differed by as much as 50-fold from their experiments and attributed the disparity in results to this difference in sensitivity and resolution.

Snowden et al. (1992) investigated the effect of various nutritional conditions on the growth and intracellular proteolytic activity of *P. furiosus* and found that
peptides and proteins could support growth, but free amino acids alone were not sufficient, even when supplemented with additional carbon and energy sources. Pfpl was found to be induced under peptide-limiting conditions, such that it becomes one of the dominant proteins in the \textit{P. furiosus} cell extract, and to be repressed by the addition of an alternative carbon and energy source such as maltose.

A number of other proteolytically active species have been described from \textit{P. furiosus}. Their names and some properties are shown in figure 1.4 overleaf.

Other species of \textit{Pyrococcus} have also been shown to exhibit proteolytic activity. Dib \textit{et al.} (1998) reported the purification of an extracellular serine protease from the culture supernatant of \textit{P. abyssi} (Erauso \textit{et al.}, 1993) which has an optimum temperature for activity of 95 °C and pH 9. It can adopt several forms, displaying apparent relative molecular masses of 150, 105 and 60 k on SDS PAGE. \textit{Pyrococcus} sp. strain KOD1 was shown to produce at least three extracellular proteases (Morikawa \textit{et al.}, 1994), one of which (TT protease) was purified and shown to be a 44 k thiol protease with maximum activity at pH 7.0 and 110 °C. It is very thermostable, retaining all its activity after 2 h at 90 °C and having a half-life of 1 h at 100 °C.

In addition to the completed genome sequencing of \textit{P. furiosus} and \textit{P. abyssi}, \textit{P. horikoshii} (Gonzalez \textit{et al.}, 1998) has also had its genome sequenced and from this a number of protease-encoding genes have been identified (Ishikawa \textit{et al.}, 1997, 1998; Ando \textit{et al.}, 1999). The gene encoding a carboxypeptidase (CPP) has been cloned and expressed in \textit{E. coli} (Ishikawa \textit{et al.}, 1997) and found to exhibit maximum activity at over 90 °C and pH 7.5. Ishikawa \textit{et al.} (1998) cloned and expressed a gene encoding a 72.4 k acylamino acid-releasing enzyme with optimum activity over 90 °C and reported that heat treatment of the recombinant enzyme increased the specific activity 7-14 fold. A thermostable aminopeptidase has also been identified in \textit{P. horikoshii} and its gene has been cloned and expressed in \textit{E. coli} (Ando \textit{et al.}, 1999). The resulting protein cleaves the N-terminal amino acid from a variety of substrates with an optimum temperature for activity of over 90 °C. The recombinant enzyme was found to have a calcium ion bound to it, which was not essential for activity, but instead depended on the presence of a cobalt ion for activity.
<table>
<thead>
<tr>
<th>Protease</th>
<th>Properties</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyrrolidone carboxyl peptidase (pcp)</td>
<td>Pcp forms a tetramer composed of 22.9 k monomers and has optimum activity between pH 6 and 9, with a temperature optimum of around 90 °C. The gene encoding pcp has been cloned and expressed in E. coli.</td>
<td>Tsunasawa et al. (1998)</td>
</tr>
<tr>
<td>Proteasome</td>
<td>P. furiosus proteasome has been purified from cell extracts and has found to be a 640 k complex, consisting of alpha (25 k) and beta (22k) subunits with a temperature optimum of 95°C.</td>
<td>Bauer et al. (1997)</td>
</tr>
<tr>
<td>Methionine aminopeptidase</td>
<td>Methionine aminopeptidase catalyses the removal of N-terminal methionine residues from a growing peptide chain. It has a temperature optimum of 100 °C and a half-life of 4.5 h at 90°C. Its gene has been cloned and expressed in E. coli.</td>
<td>Tsunasawa et al. (1997); Tahirov et al. (1998)</td>
</tr>
<tr>
<td>Prolyl endopeptidase (PEPase)</td>
<td>PEPase is a 70 k serine protease with a temperature optimum of between 85 and 90 °C. Dilute samples are resistant to heating above 65°C, whilst concentrated samples are susceptible to autolysis. Its gene has been cloned and expressed in E. coli.</td>
<td>Harwood et al. (1997)</td>
</tr>
<tr>
<td>Proline dipeptidase, (prolidase)</td>
<td>Prolidase has been purified from cell extracts and has found to be an intracellular homodimer containing one cobalt atom per 39.4 k subunit. Its catalytic activity requires additional cobalt inferring a second metal ion-binding site. It has a narrow substrate specificity, cleaving only dipeptides with proline at the C-terminus and a nonpolar residue at the N-terminus. Optimal activity is at pH 7.0 and 100 °C. Its gene has been cloned and expressed in E. coli.</td>
<td>Ghosh et al. (1998)</td>
</tr>
<tr>
<td>Carboxypeptidase (PfuCP)</td>
<td>PfuCP has been purified and found to be a cobalt-activated metalloprotease consisting of two identical 59 k subunits. It has a temperature optimum of over 90 °C, a narrow pH optimum of 6.2-6.6 and a half-life of 40 min at 80 °C.</td>
<td>Cheng et al. (1999)</td>
</tr>
<tr>
<td>Aminopeptidase</td>
<td></td>
<td>Tsunasawa (1998)</td>
</tr>
</tbody>
</table>

**Figure 1.4:** Other proteolytic species from *Pyrococcus furiosus*. 
**1.5.1.2 Sulfolobus Species**

*Sulfolobus solfataricus* is a thermoacidophilic Archaeon that grows optimally at around pH 4.0 (De Rosa *et al.*, 1974). It contains six intracellular proteolytically active species (Fusi *et al.*, 1991), of which two are aminopeptidases, three are endoproteases and one is a carboxypeptidase.

Aminopeptidase I was reported to be a metalloprotease with a relative molecular mass of greater than 450 k and aminopeptidase II was also reported to be a metalloprotease, but of 170 k. However, a previous study by Hanner *et al.* (1990) reported the isolation and characterisation of a cobalt-dependent aminopeptidase that exists as a tetramer of approximately 320 k. The enzyme has a temperature optimum of 75 °C, which is surprisingly below the optimal growth temperature of the organism (85 °C), and a pH optimum of 6.5. Subsequently, Condo *et al.* (1998) reported the sequencing of a gene encoding a monomeric aminopeptidase of 90 k that is found associated with a chaperonin. The aminopeptidase has a preference for leucine residues and is dependent on zinc ions for activity, which is optimal at around 80 °C and pH 6, with 40% of its activity remaining at 95 °C. Neither of these reports substantiate the findings of Fusi *et al.* (1991) and in the absence of further information, the disparity remains.

Endoprotease I was described by Fusi *et al.* (1991) as a serine type, with a relative molecular mass of 115 k. This has been confirmed by Burlini *et al.* (1992) who purified it from cell extracts and found it to be a dimeric, serine-type protein (actual estimations of subunit and dimer sizes were 54 and 118 k, respectively), with a pH optimum of 7.5. It has high thermostability, having a half-life of 342 min at 92°C and is activated by manganese ions.

Endoproteases II and III were described by Fusi *et al.* (1991) as serine and cysteine proteases, respectively, with relative molecular masses of 32 and 27 k, respectively. With the exception of protease II, all of the enzymes retain most of their activity after heating at 90 °C for 15 min, with protease II being the most thermolabile, losing its activity rapidly above 60 °C, probably due to autolysis. Neither proteases II or III have been described further.

However, the carboxypeptidase, only described by Fusi *et al.* (1991) as being 160 k, has been purified (Colombo *et al.*, 1992) and found to be a zinc-dependent
Sulfolobus acidocaldarius (Brock et al., 1972) is another thermoacidophile that inhabits an even lower pH environment than S. solfataricus, having a growth optimum at pH 2.0. It produces an extracellular aspartic protease, named thermopsin, a monomeric protein with some unusual properties (Lin and Tang, 1990, 1995). It has an optimum for activity of 90 °C and pH 2, although activity has been detected up to pH 12 and it is stable at 80 °C for 48 h. The majority (>90 %) of the activity is tightly associated with the cell wall and is consequently more difficult to purify; therefore, thermopsin has been purified from the culture medium. The gene encoding thermopsin has been cloned and sequenced, and from this the amino acid sequence has been shown not only to possess no significant homology to other aspartic or cysteine proteases, but more interestingly, to lack the asp-thr-gly active site catalytic triad characteristic of aspartic proteases. Based on these findings, a further study compared its specificity and inhibition to another aspartic protease, pepsin (Fusek et al., 1990). Thermopsin is similar to pepsin with respect to its substrate specificity, having a preference for large, hydrophobic residues and a similar $K_M$. However, it is not totally inhibited, and even then only non-specifically, by the aspartic protease inhibitors diazoacetyl-DL-norleucine methyl ester (DAN) and 1,2-epoxy-3-(p-nitrophenoxy)propane (EPNP). The thiol-inhibitor mercuric chloride inhibited thermopsin, but via a mechanism unrelated to the thiol group. Thermopsin contains a single cysteine residue, which has been shown to be blocked by an unknown group that is removed during HCl hydrolysis, but not SDS denaturation of the enzyme. SDS PAGE analysis has indicated that thermopsin has a relative molecular mass of 51 k, whilst gel filtration has indicated a size of 46 k. Both of these values are greater than that predicted from the sequence (32.7 k), suggesting that thermopsin may be glycosylated at some of its 11 potential sites. The deduced peptide sequence highlighted a 41 amino acid region preceding the N-
terminus of the mature protein, of which the majority resembled a signal sequence, although the possibility of a short pro-region as well was hypothesised.

Thermopsin has been recombinantly expressed in *E. coli* as a fusion to the C-terminal of pepsinogen (Lin *et al.*, 1992), as expression of thermopsin alone yielded only very low levels of recombinant protein. The protein was expressed in the form of insoluble fusion bodies, which could be refolded after urea denaturation. Heat treatment of the fusion denatured the pepsinogen section, which was then accessible to thermopsin degradation, yielding active protease.

Finally, there has been a report (Vankley *et al.*, 1995) of an unclassified low relative molecular mass protease from *Sulfolobus shibatae* (Grogan *et al.*, 1990), only 9 k in size, with a pH optimum of 7.2, the activity of which is reported to change very little from 37-100 °C.

1.5.1.3 Other Thermophilic Archaeal Proteases

To date, there have only been two reports of proteases from the two thermophilic Archaea used in this project, both from the strictly aerobic hyperthermophile, *Aeropyrum pernix* K1 (Sako *et al.*, 1996). Sako *et al.* (1997) reported the purification and characterisation of an extracellular metalloprotease, designated aeropyrolysin. The enzyme has a relative molecular mass of 52 k as judged by SDS PAGE and an optimum activity at pH 6-8 and 110 °C in the presence of 1 mM calcium ions, and 100 °C in its absence. Aeropyrolysin is extremely thermostable, having a half-life of 1.2 h at 125 °C in the presence of 1 mM calcium ions. Crocker *et al.* (1999) described the purification and characterisation of an intracellular serine protease, named pemilase, with a relative molecular mass of 50 k as determined by SDS PAGE. Pemilase has an optimum for activity of 90 °C and pH 9.0, with a half-life of 12 min at 110 °C.

There have been no reports to date, however, describing proteolytic activity in the sulphate-reducing Archaeon, *Archaeoglobus fulgidus*.

The genes encoding the proteasome from *Thermoplasma acidophilum* have been expressed in *E. coli* in an assembled and active form which is not toxic to the
host cells (Zwickl et al., 1992), during which processing of the pro-region of the β-subunit occurs by an autocatalytic mechanism.

Other proteases of interest include STABLE and a hyperthermophilic, barophilic protease. STABLE (Mayr et al., 1996) is a surface layer protease from *Staphylothermus marinus* which, in its stalk-bound form, retains residual activity after a ten-minute incubation at 135 °C. Its gene has been sequenced using an entirely PCR-based approach, similar to that used for tetrabrachion (Peters et al., 1996). A barophilic protease (Michels and Clark, 1997) was partially purified from *Methanococcus jannaschii*, a hyperthermophilic deep-sea methanogen. The protease has a temperature optimum of 116 °C, and activity can be measured up to 130 °C. Furthermore, raising the pressure to 500 atm increased the reaction rate at 125 °C 3.4-fold and the thermostability 2.7-fold.

### 1.5.2 Thermophilic Proteases from the Bacteria

#### 1.5.2.1 *Thermus* proteases

The first protease to be characterised from *Thermus* was caldolysin (Cowan and Daniel, 1982), an extracellular serine protease from *Thermus aquaticus* st. T-351. Due to its sensitivity to EDTA and its apparent resistance to PMSF, caldolysin was originally thought to be a metalloprotease. However, it does not bind a catalytically active ion, but it does bind a total of six calcium atoms per molecule (Khoo et al., 1984) which it requires for thermostability, but not activity. For example, the half-life at 75 °C is greater than 193 h in the presence of 10 mM calcium, compared to 4.8 min in its absence (Cowan and Daniel, 1982). Interestingly, it has been found that lanthanide ions can not only replace calcium ions in caldolysin, but actually confer a greater thermostability, increasing its half-life at 95 °C from 1 h to more than 4 h (Khoo et al., 1984). Caldolysin has a relative molecular mass of 21 k and a pH optimum of 8.0, with activity being measured up to 100 °C.

The most extensively studied protease from *Thermus* species is Aqualysin I, another extracellular serine protease, from *Thermus aquaticus* st. YT-1. *T. aquaticus*
YT-1 secretes two serine proteases, aqualysin I and II (Matsuzawa et al., 1983), from the time the cells reach stationary phase until they cease growth. Aqualysin II is a neutral protease (pH optimum of 7.0) that is EDTA sensitive with a temperature optimum of around 95 °C; it has not been described further. Aqualysin I has been purified and found to be an alkaline serine protease with a relative molecular mass of 28.5 k, a pH optimum of around 10.0 and a temperature optimum of 80 °C (Matsuzawa et al., 1988). Its gene has been cloned into E. coli and sequenced (Kwon et al., 1988b) and the enzyme has been shown to contain two disulphide bonds (Kwon et al., 1988a). The gene has also been expressed in E. coli and from this the organisation of the protein has been deduced (Terada et al., 1990). Aqualysin I comprises four domains, an N-terminal signal sequence, an N-terminal pro-region, the mature protease and an additional C-terminal pro-region. This is in contrast to the proteases secreted by Bacillus, which typically contain an N-terminal signal sequence and an N-terminal pro-region preceding the mature domain (e.g. Ohta and Inouye, 1990) and also in contrast to the proteases secreted by Gram-negative bacteria such as Serratia marcescens (Yanagida et al., 1986) and Neisseria gonorrhoeae (Pohlner et al., 1987), which consist of an N-terminal signal sequence, the mature region and a C-terminal pro-region.

Attempts to clone a genomic fragment containing the entire gene for aqualysin and its 5' promoter region into E. coli were not successful; the authors speculated this was due to constitutive expression from the E. coli-like promoter being toxic to the cells (Terada et al., 1990). However, the gene was expressed from a controlled promoter (the tac promoter of pKK223-3), apparently without any lethal effects. Activity was not detected either in the culture medium or soluble fraction of the cells, although there was measurable activity in the membrane fraction which was shown to contain a precursor of aqualysin lacking its N-terminal pre- and pro-regions. Heat-treatment of the membrane fraction was shown to evoke proteolytic activity by self-cleavage of the C-terminal pro-region and active site mutagenesis showed a precursor lacking only the N-terminal signal sequence accumulating in the membrane fraction. Therefore, the signal sequence directs aqualysin through the inner membrane, but does not secrete it, and is concomitantly cleaved by a signal peptidase, and the cleavage of the N- and C-terminal pro-regions occurs sequentially due to the activity of aqualysin itself.
Chapter One

An aminopeptidase from *T. aquaticus* YT-1 has also been purified and characterised (Minagawa et al., 1988) and its gene has been cloned, sequenced and expressed in *E. coli* (Motoshima et al., 1990).

Cowan et al. (1987a) speculated on the possibility of using the properties of enzymes to discriminate between strains of *Thermus aquaticus* by comparing the extracellular serine proteases of four of its strains, but concluded that despite functional differences between them, they were probably not sufficient for taxonomic purposes.

Extracellular proteases have also been purified from *Thermus* strain ToK3 (Saravani et al., 1989), *T. caldophilus* strain GK24 (Taguchi et al., 1983) and *Thermus* sp. strain Rt41A (Freeman et al., 1993; Peek et al., 1992). The gene encoding the Rt41A protease purified by Peek et al. (1992) has also been cloned, sequenced and expressed in *E. coli* (Munro et al., 1995) as a fusion to glutathione-S-transferase.

The gene encoding the Lon protease from *Thermus thermophilus* HB8 has been cloned, sequenced and expressed in *E. coli* (Watanabe et al., 1999). Interestingly, the proteolytic activity of the recombinant enzyme was optimal at 70°C, which is lower than the growth optimum of the organism (85 °C), and in fact activity was almost completely lost at temperatures higher than 80 °C. The authors speculated that this may be due to the requirements of additional factors for the maintenance of the active structure, for example one of the substrates, ATP.

1.5.2.2 *Bacillus* proteases

Whilst not directly related to this project, a review of thermophilic proteases would not be complete without an overview of some proteases from *Bacillus*. The archetypal *Bacillus* proteases are the extracellular serine proteases, the subtilisins, specifically subtilisin Carlsberg (Smith et al., 1968) and subtilisin BPN’ (Markland and Smith, 1967) from *B. licheniformis* and *B. amyloliquefaciens*, respectively. However, neither of these proteins are especially thermostable, although thermolysin (Endo, 1962), a metalloprotease from *B. thermoproteolyticus*, is having a temperature optimum of 77 °C and a half-life of 1 h at 80 °C (Van den Burg et al.,
Thermolysin has been extensively studied (e.g. Izquierdo and Stein, 1990; Holland et al., 1995; O'Donohue and Beaumont, 1996; Inouye et al., 1998), its amino acid sequence has been determined (Titani et al., 1972) and a high resolution structure is available (Holmes and Matthews, 1982). It requires one zinc atom per molecule for activity and four calcium atoms are implicated in thermo- and autolytic-stability (Roche and Voordouw, 1978). Thermolysin has been heterologously expressed from *B. subtilis* (O'Donohue et al., 1994) and the recombinant form is indistinguishable from the wild type.

*B. stearothermophilus* st. MK232 has been reported to produce a metalloprotease (NprM) with a higher thermostability than thermolysin (Kubo et al., 1988), having a half-life of at least 25 min at 95 °C. The gene encoding NprM has been cloned and expressed in *B. subtilis* (Kubo and Imanaka, 1991).

The serine protease from *Bacillus* st. AK.1 is highly thermostable, having a half-life of 19 min at 90 °C in the presence of calcium (Peek et al., 1993) and its encoding gene has been cloned and expressed in *E. coli*.

### 1.5.2.3 Other Thermophilic Bacterial Proteases

The gene encoding an extracellular protease from *Thermoactinomyces* sp. E79 has been expressed in *E. coli* in a host-dependent manner (Lee et al., 1998). The enzyme was expressed intracellularly as an inactive pre-pro form that could be converted to the active form by heat treatment.

Another protease to have its gene expressed in *E. coli* in an active form is a serine protease from *Aquifex pyrophilus* (Choi et al., 1999). When the gene is expressed in its precursor form, it is translocated to the *E. coli* cell wall as a 43 k protein, indicating from its gene sequence that both the N- and C-terminal regions are processed. The gene encoding an aminopeptidase from *A. pyrophilus* has also been cloned and expressed in *E. coli* (Khan et al., 2000) in an active form.
1.6 Halophilic Proteases

Extracellular proteolytic enzymes are widely distributed in halophiles. These organisms rarely metabolise carbohydrates but instead have well-developed enzymes for metabolising amino acids and therefore utilise proteins from dead organisms in their environment. Protein hydrolysis activity can be detected in culture supernatants of halophiles and the earliest reports of proteolytic enzymes from halophilic organisms were primarily concerned with this phenomenon (Norberg and Hofsten, 1969).

The first reported purification and characterisation of a halophilic protease was that from *Halobacterium halobium* (Izotova et al., 1983). Enzymes from halophiles are difficult to purify as they lose their activity at low ionic strength and so procedures must be sought that are unaffected by high salt concentrations. The enzyme activity of the *H. halobium* protease is completely and irreversibly lost at NaCl concentrations of less than 2 M. Subsequently, a number of extracellular proteases from halophiles have been purified and characterised. A second protease from *H. halobium* was reported (Kim and Dordick, 1997) and extracellular proteases have been purified from *Natrialba asiatica* (Kamekura and Seno, 1990), *Halobacterium* sp. strain TuA4 (Schmitt et al., 1990), *Haloferax mediterranei* 1538 (Stepanov et al., 1992) and partially purified from an unidentified (but assumed to be a *Natronobacterium*) haloalkaliphile, strain A2 (Yu, 1991). A membrane-associated metalloprotease has also been reported from *H. halobium* (Fricke et al., 1993) and extracellular proteolytic activity has been characterised from *Natronococcus occultus* (Studdert et al., 1997).

The first protease from a halophile to have its gene sequenced was halolysin from *Natrialba asiatica* (Kamekura et al., 1992). The gene was expressed in another halophilic Archaeon, *Haloferax volcanii*. Only one more gene encoding an extracellular halophilic protease has been reported to date, *hlyR4* from *Haloferax mediterranei* (Kamekura et al., 1996).

With two halolysins to study, comparative and mutational analysis of the amino acid sequence is in progress to analyse the relative importance of certain residues to the proteins' halophilicity.
1.7 Biotechnological Applications of Thermostable Proteases

Proteases as a whole have a paramount position in commercial applications. They represent around 60% of the present $1 billion sales of industrial enzymes (Godfrey and West, 1996). The application of thermostable proteases has come about due to certain recognised advantages of using enzymes at elevated temperatures (Sonnleiter and Fiechter, 1983; Daniel et al., 1985/6; Daniel et al., 1995). For example, contamination of food processes is greatly reduced, and almost all pathogenic bacteria are killed, at temperatures above 70°C. Furthermore, diffusion rates and solubilities of chemicals are greater at higher temperatures, whilst viscosities are lowered. It is also accepted that thermostable enzymes are more tolerant to the denaturing effects of detergents and organic solvents and more resistant to proteolysis (e.g. Daniel et al., 1982; Owusu and Cowan, 1989), enabling these conditions to be employed.

An advantage specific to thermostable proteases is that as the temperature is increased, substrate proteins denature and become more susceptible to proteolysis. This can lead to temperature coefficients (Q_{10}) greater than 2, which means that thermostable proteases have higher specific activities than their mesophilic counterparts at their respective growth temperatures (Cowan et al., 1987b). This is important in the application of the hydrolysis of proteins, as opposed to peptides, and leads to an interesting consequence. For example, a Q_{10} value of 3 for an enzyme optimally active at 70°C, will lead to the enzyme having ten times less activity at 30°C compared to a Q_{10} value of 1.7 (Daniel et al., 1995), potentially eliminating the need for downstream removal of the protease. However, a stable protease could be detrimental to a product requiring long term storage.

The applications of proteases in general in the detergent, leather, food and pharmaceutical industries have been well documented (Kalisz, 1988; Outtrup and Boyce, 1990; Rao et al., 1998) and with a few exceptions, most of the uses of thermophilic proteases have been potential replacements of existing enzymes.

The major application to date of a thermophilic protease is in the synthesis of aspartame, a sweetening dipeptide, where the process utilises the organic solvent stability of thermolysin from *B. proteolyticus* (Isowa et al., 1979). Like all
enzymatic reactions, peptide hydrolysis is reversible. In the presence of water (aqueous solutions), the favoured reaction is hydrolysis. However, in low-water activity environments, for example organic solvents, peptide bond synthesis occurs. Aspartame is a dipeptide composed of the methyl ester of L-phenylalanine and L-aspartate, and thermolysin catalyses the condensation of the two components.

Another thermostable protease with a commercial application is a *Thermus* protease (marketed by BRL as Pretaq), used in the preparation of DNA, initially from blood and tissue samples, for PCR (McHale *et al.*, 1991), but subsequently as a substitute for proteinase K in genomic DNA purification (Borges and Bergquist, 1992).

Fujiwara *et al.* (1991) reported the use of a thermostable protease in the recovery of silver from used X-ray film, which improved the processing capacity of a bench-scale plant compared to the previously used mesophilic protease, due to increased temperatures being possible.

Thermostable proteases seem to fill the requirements for a suitable meat tenderising enzyme. Their thermal properties mean that they can be added to meat before the point of sale and not cause over tenderising during storage and only be active during a defined cooking period. Wilson *et al.* (1992) showed that application of a thermostable protease resulted in improved taste and good tenderisation.

Proteases have a major bulk use in laundry detergents and although improved stability is an obvious aspiration, current trends have been towards using lower temperatures, where psychrophilic enzymes flourish. A somewhat more appropriate application for thermophilic proteases could be in dishwasher formulations, allowing the use of less aggressive chemical and physical conditions.

Another process currently employing harsh cleaning conditions is the cleaning of ultrafiltration membranes used in whey processing. Coolbear *et al.* (1992) have proposed the use of thermophilic proteases to reduce the need for strong alkali and/or acid and have found a *Thermus* protease which, in the presence of an anionic detergent, was effective at cleaning, albeit at a slower rate.

Thermostable proteases will doubtless continue to attract interest from both an industrial and fundamental viewpoint and as efforts are undertaken to manipulate not only their thermostability, but also the other properties of them, for example solvent stability and substrate specificity, this can only advance in time.
1.8 Thermophilic Esterases

Esterases catalyse the hydrolysis of a wide range of ester bonds. Classification of esterases is based on substrate specificity and their sensitivity to inhibitors. Arylesterases (A-esterases) hydrolyse aromatic esters and are not inhibited by organophosphates. Carboxylesterases (B-esterases) have a wide substrate specificity and are inhibited by organophosphates. Finally, C-esterases hydrolyse choline esters and are inhibited by organophosphates and eserine.

It has been suggested (Hemila et al., 1994; Krejei et al., 1991) that esterases, lipases and cholinesterases belong to a large family of phylogenetically related proteins. Three subfamilies have so far been identified, the C group, group L and the hormone-sensitive lipase (HSL) subfamily. The C group includes several esterases, cholinesterases from vertebrates and invertebrates, lipases from fungi and some nonenzymic proteins. The group L includes lipases from vertebrates and bacteria, and lipoprotein lipases. The HSL subfamily was named after Hemila et al. (1994) reported the cloning and sequencing of a gene from Alicyclobacillus acidocaldarius sharing sequence homology with the HSL from human and rat.

To date there have been a number of reports of carboxylesterases (EC 3.1.1.1) from thermophilic species. Carboxylesterases have been reported from a number of strains of Bacillus stearothermophilus (Matsunaga et al., 1974; Owusu and Cowan, 1991; Wood et al., 1995). Matsunaga et al. (1974) reported the purification of an esterase from B. stearothermophilus NCA 214, with an optimum activity at pH 7.0 and 65 °C and an estimated relative molecular mass of 42-47 k. Owusu and Cowan (1991) reported the isolation of an esterase with an apparent relative molecular mass of 38-45 k from B. stearothermophilus G18A7. The enzyme is optimally active at 85°C and pH 9.5 and has very high thermostability, losing only 10 % of its activity after incubation at 105 °C for 150 min. An esterase from B. stearothermophilus Tok19A1 has been reported by Wood et al. (1995), having maximum activity at 60°C and pH 6.0 and an apparent relative molecular mass of 45-50 k. The enzyme is very thermostable, having a half-life of 170 h at 60 °C.

A carboxylesterase has been purified and characterised from the thermoacidophile B. acidocaldarius (Manco et al., 1994), with a relative molecular
Chapter One

mass of 34 k and an optimum for activity of 65 °C and pH 8.0. Its encoding gene has been cloned and sequenced, and overexpressed in *Escherichia coli* (Manco *et al.*, 1998) and its structural properties have been investigated (D’Auria *et al.*, 2000b).

The first report of an archaeal esterase is that from the thermoacidophilie *Sulfolobus acidocaldarius* (Sobeck and Gorisch, 1988, 1989). The enzyme is a tetramer, having a relative molecular mass of 128 k. It displays a broad pH optimum of between 7.5 and 8.5, and is very thermostable. It retains over 90% of its activity after incubation at 90 °C for 1 h, and has a half-life of 1 h at 100 °C.

Subsequently, an extracellular esterase from *S. shibatae* has been reported (Huddleston *et al.*, 1995). This esterase is active between pH 4.0 and 9.0, with an optimum for activity at pH 6.0 and 90 °C. It is very thermostable, having a half-life of 20 min at 120 °C. Comenc *et al.* (1998) screened 160 thermophilic and hyperthermophilic anaerobic archaea and bacteria for esterase activity and found 47 to be positive. Characterisation of the crude esterase activity from *Pyrococcus abyssi* was performed and the enzyme was found to have a optimum for activity between pH 5.8-7.8 and above 70 °C. Furthermore, it retained all of its activity after incubation at 90 °C for 8.5 h and exhibited a half-life of 13 min at 120 °C. A thermostable esterase from *P. furiosus* has been expressed in *E. coli* from a genomic library clone (Ikeda and Clark, 1998) that exhibits a temperature optimum of 100 °C and a half-life of 50 min at 126 °C.

There have been several reports describing an esterase from the thermophilic Archaeon, *Archaeoglobus fulgidus* (e.g. Manco *et al.*, 2000b), which is described in further detail in section 4.1.

Finally, there has been a report of the identification of esterase activities in seven *Thermus* strains (Berger *et al.*, 1995). *T. thermophilus* HB8 showed esterase activity in the culture supernatant.

1.9 Expression of Extremophilic Genes

Whilst commercial production of recombinant proteins generally employs the use of *Bacillus* species, the host of choice for expression of archaeal and bacterial genes for
purely research purposes is *E. coli*. This is due to its short generation time and the expanse of molecular and physiological information available for the organism.

However, its preferred low salt environment is in contrast to that of halophilic proteins and expression of halophilic genes in *E. coli* is known to be problematic. Initial attempts to express the gene encoding bacteriorhodopsin were less than satisfactory (Dunn *et al.*, 1987). Further work produced increased expression levels and active protein (Karnik *et al.*, 1987; Nassal *et al.*, 1987; Braiman *et al.*, 1987) but to date there have been no reports of a halophilic gene expressed in *E. coli* that has produced active and soluble protein. When soluble protein is obtained, as was the case for the expression of malate dehydrogenase (Cendrin *et al.*, 1993) and 3-Hydroxy-3-methylglutaryl-coenzyme A reductase (Bischoff and Rodwell, 1996), catalytic activity was not seen until the addition of 3 M NaCl or KCl. It has been found that soluble recombinant protein is sometimes achieved by lowering the growth temperature (Schein and Notebom, 1988). This has been successful in the expression in *E. coli* of soluble citrate synthase from *Hf. volcanii* (Connaris *et al.*, 1999). The recombinant enzyme, however, still required overnight incubation in 2M KCl for reactivation.

Recombinant proteins that are insoluble form inclusion bodies and require more substantial effects to reactivate. This involves solubilisation by denaturation and then renaturation, in the case of halophilic proteins in the presence of salt. One such example of a halophilic protein successfully renatured and reactivated is dihydrolipoamide dehydrogenase from *Hf. volcanii* (Connaris *et al.*, 1999), which was solubilised in 8 M urea and subsequently diluted into a buffer containing 2 M KCl, 10 μM FAD and 0.3 mM GSSG/3 mM GSH. The enzyme exhibited maximal activity after 3 days incubation at 4 °C. Two other halophilic proteins successfully reactivated are dihydrofolate reductase (Blecher *et al.*, 1993), which was dissolved in 6 M guanidine HCl and renatured in the presence of 1 M NaCl, and bacterio-opsin (Chen and Gouaux, 1996), which was dissolved in 8 M urea and reactivated in 200mM NaCl. On the other hand, despite numerous methodologies, the direct expression of seryl-tRNA synthetase failed to produce active protein from inclusion bodies, and this was only managed after expression as a fusion protein (Taupin *et al.*, 1997).
The development of halophilic expression systems has seen the emergence of halophilic genes expressed in halophiles. This has proven especially useful in the expression of genes that need to be secreted across the membrane, as *E. coli* may fail to recognise the signal sequence as such. Examples of extracellular proteins produced in this way include halolysin R4 (Kamekura *et al.*, 1996), α-amylase (Kobayashi *et al.*, 1994) and halocin H4 (Cheung *et al.*, 1997).

Expression of thermophilic genes in *E. coli* is generally less problematic, the major distinction being only the wide temperature difference in the organisms' respective growth optima, and early worries that thermophilic genes would not be successfully expressed in mesophilic hosts have been largely unfounded. Activity is generally observed when the recombinant protein is expressed in a soluble form. Multimeric proteins sometimes require heat treatment to trigger *in vitro* assembly, as has been shown to be the case for glutamate dehydrogenase from *Pyrococcus furiosus* (Diruggiero and Robb, 1995) and reverse gyrase from *Methanopyrus kandleri* (Krah *et al.*, 1997). A wide range of genes have successfully been expressed in an active form in *E. coli*, for example an esterase from *Archaeoglobus fulgidus* (Manco *et al.*, 2000b), an aldolase from *Sulfolobus solfataricus* (Buchanan *et al.*, 1999), inorganic pyrophosphatase from *Thermus thermophilus* (Satoh *et al.*, 1998), citrate synthase from *Pyrococcus furiosus* (Muir *et al.*, 1995) and elongation factor G from *Aquifex aeolicus* (Martemyanov *et al.*, 2000).

Expression of thermophilic genes in thermophilic hosts has only recently been described, although the integration and expression of the *leuB* gene from the mesophilic *Bacillus subtilis* has been achieved in the *Thermus thermophilus* chromosome (Akanuma *et al.*, 1998). There have been two recent reports of gene expression in thermophiles, both in the extremely acidophilic Archaeon *Sulfolobus solfataricus*. The first describes the expression of the *Bacillus stearothermophilis* alcohol dehydrogenase gene, *adh-hT* (Fiorentino *et al.*, 2000), utilising an *E. coli*/*S. solfataricus* shuttle vector (Cannio *et al.*, 1998), and the second describes the expression of the *Pyrococcus furiosus* β-glucosidase gene, *celB* (Stedman *et al.*, 2000). There has also been a recent report of the potential for expression in *Pyrococcus abyssi* (Lucas *et al.*, 2000) based on the extrachromosomal cryptic plasmid pGT5.
1.10 Aims of the Project

The above discussion has highlighted the fact that research into proteases is an expanding and significant field, as their relevance to both industrial and fundamental areas of science has become apparent. As the interest in extremophiles has also widened, this has brought with it not only a knowledge of their physiology, but a new source of industrial enzymes, able to survive and operate under conditions previously considered denaturing. Proteases, and specifically extracellular proteases, have been reported that exhibit activity under a variety of very extreme conditions, prompting speculation that whatever the application conditions, a suitable enzyme might be found. The purification of numerous proteases has been successfully carried out, providing a number of novel protocols. Although this procedure is a source of authentic protein, possibly free from N-terminal methionine residues or fusion tag artefacts, large-scale processes rely upon the expression of the encoding genes in an expression host.

The aims of this work are to produce such an expression system, possibly giving rise to activity under novel conditions, whilst at the same time understanding more about adaptation of the parent organism to its particular niche environment. It has been shown that heterologous expression of potentially lethal genes is possible in some cases. Varying levels of expression and processing have been achieved in some cases in a manner dependent upon the strain of the host organism. In other cases, active protein has been achieved following heat-activation of an inactive precursor, even so far as for the protease to cleave its own fusion. Consequently, the expression of an extremophilic protease in a mesophilic host is a valid objective, as an alternative to its purification from the host organism. Alternatively, given a successful secretion pathway, proteases have also been heterologously exported to the culture medium, thus avoiding the possibility of toxic activity within the cell. The cloning and expression of exopeptidases have shown that it is possibly easier to produce these enzymes recombinantly than endopeptidases, due to their limited specificity reducing their potential toxicity. However, exopeptidases tend to have specialised roles in industrial processes, rather than widespread applications.

The genome sequencing revolution has brought about a wealth of information and for the most part has eliminated the need to purify the native protein before
cloning the encoding gene. Traditional methods to clone and express a gene required the protein product from that gene to be initially purified and then part of its N-terminal sequence to be determined. A DNA probe would then be produced corresponding to that sequence which would be used to screen a genomic DNA library of the organism. This is no longer necessary, as PCR probes corresponding to sequences up and downstream of the gene of interest can be produced and the gene can be amplified in a single PCR step. Novel proteins are still being reported, however, including proteases that contain no known catalytic motifs. The aim of this project is to produce a novel recombinant protease capable of activity under extreme conditions.
2. Chapter Two

Materials and Methods

2.1 General Laboratory Reagents

2.1.1 Cell Culture

Tryptone, yeast extract, sodium chloride, ampicillin (sodium salt), carbenicillin (disodium salt), chloramphenicol, kanamycin monosulphate, manganese chloride, rubidium chloride, ferric ammonium citrate, L-aspartic acid and L-tryptophan were purchased from Sigma-Aldrich Company Ltd., Poole, UK. Bacto-agar, glycerol, magnesium sulphate, D-glucose, calcium chloride, di-sodium hydrogen orthophosphate, potassium dihydrogen orthophosphate, potassium acetate, tri-sodium citrate and di-potassium hydrogen orthophosphate were supplied by Fisher Scientific Ltd., Loughborough, UK. Nutrient broth and casamino acids were from Difco Laboratories Ltd., Surrey, UK. Ammonium chloride was purchased from BDH chemicals Ltd., Poole, UK.

2.1.2 Molecular Biology Reagents

Vent DNA polymerase, T4 DNA ligase and all restriction enzymes and buffers were supplied by New England Biolabs, Hitchin, UK. T4 DNA ligase supplied with the Rapid DNA Ligation Kit was purchased from Boehringer Mannheim UK (Diagnostics and Biochemicals) Ltd., Lewes, UK. BioTaq DNA polymerase and dNTPS were supplied by Bioline, London, UK. Pfu DNA polymerase was purchased from Promega Ltd., Southampton, UK. Cloned Pfu DNA polymerase was supplied by Stratagene, La Jolla, CA 92037 USA, and Taq/Pwo DNA polymerase
was from Hybaid Ltd., Middlesex, UK. 1 kbp DNA ladder and M13 sequencing primers were from Gibco BRL Life Technologies Ltd., Paisley, UK. SeaKem LE agarose was supplied by FMC BioProducts, Staffordshire, UK. Ethidium bromide, mineral oil, bromophenol blue and xylene cyanol FF were from Sigma-Aldrich Company Ltd., Poole, UK. Absolute ethanol was purchased from Hayman Ltd., Witham, UK. Sucrose was supplied by Fisher Scientific Ltd., Loughborough, UK., and IPTG was purchased from CN Biosciences UK, Nottingham, UK. QIAprep Spin Miniprep Kits, QIAGEN Plasmid Mini Kits and QIAEX II Gel Extraction Kits were from Qiagen Ltd., Crawley, UK. S.N.A.P. Miniprep Kits were purchased from Invitrogen BV, 9704 CH Groningen, Netherlands, and Geneclean III kits were from Bio 101, Inc., distributed by Anachem Ltd., Luton, UK. PCR primers were custom made by Perkin Elmer-Applied Biosystems (Cheshire, UK.) of ReadyPure quality and MWG-Biotech UK Ltd., (Milton Keynes, UK.) of HPSF quality. T4 promoter and terminator primers were supplied by Novagen, Inc, Madison, WI 53711, USA.

2.1.3 General Laboratory Chemicals

Lysozyme, MES, MOPS and Triton X-100 were supplied by Sigma-Aldrich Company Ltd., Poole, UK. Tris and sodium hydroxide were from BDH chemicals Ltd., Poole, UK. EDTA, butan-1-ol, potassium hydroxide and hydrochloric acid were from Fisher Scientific Ltd., Loughborough, UK.

2.1.4 Protein Related Materials

TEMED, SDS, Coomassie Blue R, Coomassie G-250, glutathione and L-cysteine and gel filtration standards were purchased from Sigma-Aldrich Company Ltd., Poole, UK. Methanol, glacial acetic acid, glycine and β-mercaptoethanol were supplied by BDH chemicals Ltd., Poole, UK. APS, protein molecular weight markers and poly-prep chromatography columns were from Bio-Rad Laboratories Ltd., Hertfordshire, UK. Phosphoric acid was from Fisher Scientific Ltd., Loughborough, UK. Chitin beads were supplied by New England Biolabs, Hitchin,
UK. DTT was purchased from Alexis Corporation (UK) Ltd., Bingham, UK. BSA was supplied by Pierce Chemical Company, Illinois, 61105, USA., and acrylamide was from National Diagnostics, Hull, UK. For sucrose, bromophenol blue and absolute ethanol, please refer to section 2.1.2.

2.1.5 Laboratory Equipment

Spectroscopy was carried out on Perkin-Elmer Lambda 11, Lambda Bio and Lambda 3B UV/Vis spectrophotometers (Perkin-Elmer Ltd., Buckinghamshire, UK.) using Perkin-Elmer Computerised Spectroscopy Software (PECSS) version 4.31. Centrifugation was carried out in Sorvall RC-5B and RC-5C Refrigerated Superspeed Centrifuges [Du Pont (UK) Ltd., Hertfordshire, UK.] using Sorvall SS-34 and GSA rotors. Cell disruption was carried out by ‘One Shot’ cell disrupter (Constant Systems Ltd., Warwickshire, UK.) and 150 W Ultrasonic Disintegrator Mk2 (MSE Scientific Instruments, Crawley, UK). PCR was carried out in a Perkin-Elmer Cetus DNA Thermal Cycler (Perkin-Elmer Ltd., Buckinghamshire, UK.).

2.2 Project Specific Materials

2.2.1 Gene Cloning

*Haloferax mediterranei* genomic DNA was kindly provided by Dr M. J. Bonete (University of Alacante, Spain). *Archaeoglobus fulgidus* VC-16 genomic DNA and cell paste were kindly provided by Prof. T. Lien (University of Bergen, Norway). *Aeropyrum pernix* K1 cell paste was kindly provided by Dr Neil Raven (CAMR, Salisbury). pUC18 was supplied by Invitrogen BV, 9704 CH Groningen, Netherlands. pET-3a, pET-11a, *Escherichia coli* competent BL21(DE3) [F, *ompT*, *hsdS*<sub>B</sub>, (r<sub>B</sub>, m<sub>B</sub>), *dcm*, *gal*, λ(DE3)] and competent BL21(DE3) pLysS cells [F, *ompT*, *hsdS*<sub>B</sub>, (r<sub>B</sub>, m<sub>B</sub>), *dcm*, *gal*, λ(DE3), pLysS, Cm<sup>†</sup>] were purchased from Novagen, Inc, Madison, WI 53711, USA. pTYB1 was supplied by New England
Biolabs, Hitchin, UK. pBE3 and *Bacillus subtilis* DB428 were kindly provided by Kentaro Miyazaki (c/o Prof. F. H. Arnold, California Institute of Technology, Pasadena, CA 91125, USA). *Escherichia coli* bacterial strain JM109 {*endA1, recA1, gryA96, thi, hsdR17 (rK−, mK+), relA1, supE44, Δ(lac-proAB), [F', traD36, proAB, lacFZΔM15]} and competent JM109 cells were purchased from Promega Ltd., Southampton, UK., or made in house.

### 2.2.2 Assay Materials

Azocasein, azoalbumin, gelatine, ala-ala-pro-phe-p-nitroanilide, ala-ala-phe-p-nitroanilide, *p*-nitrophenyl acetate, *p*-nitrophenyl butyrate, gly-gly-phe-p-nitroanilide, ala-ala-pro-leu-p-nitroanilide, *p*-nitroaniline, subtilisin Carlsberg (EC 3.4.21.62), PMSF, pepstatin A and PCMB were all purchased from Sigma-Aldrich Company Ltd., Poole, UK. TCA was supplied by BDH chemicals Ltd., Poole, UK.

### 2.3 General Laboratory Methods

Most of the methods used for molecular cloning were based on those of Sambrook *et al.* (1989).

#### 2.3.1 Cell Culture

Unless otherwise specified, all components were sterilised by autoclaving (121 °C for 15 min) before use. For solid media, 1.5 % (w/v) agar was added before autoclaving.

*2.3.1.1 Escherichia coli*

*E. coli* was routinely cultured at 37 °C in Luria-Bertani (LB) medium, consisting of 1% (w/v) tryptone, 0.5 % (w/v) yeast extract and 1 % (w/v) sodium chloride.
2.3.1.2 *Bacillus subtilis*

*B. subtilis* DB428 was routinely cultured at 30 °C in either Nutrient Broth (NB) medium, consisting of 0.3 % (w/v) beef extract powder and 0.5 % (w/v) peptone at 30 °C, or at 37 °C in Schaeffers (SG) medium, consisting of 1.6 % (w/v) nutrient broth, 0.05 % (w/v) MgSO₄, 0.2 % (w/v) KCl, 1 mM Ca(NO₃)₂, 100 μM MnCl₂, 1μM FeSO₄ and 0.1 % (w/v) glucose.

2.3.1.3 Culture Storage

For short-term storage, cultures were kept on plates at 4 °C and subcultured every two months. Permanent stocks were kept by the addition of 15 % (v/v) sterile glycerol to an overnight liquid culture, which was then aliquotted, snap frozen in liquid nitrogen and stored at −70 °C.

2.3.2 Preparation of Competent Cells

2.3.2.1 *Escherichia coli*

Competent *E. coli* JM109 cells were prepared according to Promega protocols and applications guide (3rd edition). A fresh, single colony was inoculated into 2.5 ml LB and incubated overnight at 37 °C. The following day, the entire culture was inoculated into 250 ml LB supplemented with 20 mM MgSO₄ and incubated with shaking to an OD₆₀₀ of 0.4–0.6. The cells were pelleted by centrifugation (1000 g, 3min) and gently resuspended in 100 ml ice-cold TFB1 [30 mM potassium acetate buffer, pH 5.8, 10 mM CaCl₂, 50 mM MnCl₂, 100 mM RbCl, 15 % (v/v) glycerol, filter-sterilised (0.2 μm)]. The resuspended cells were incubated on ice for 5 min and then pelleted by centrifugation as before. The cells were gently resuspended in 10 ml ice-cold TFB2 [10 mM MOPS buffer, pH 6.5, 75 mM CaCl₂, 10 mM RbCl, 15 % (v/v) glycerol, filter-sterilised (0.2 μm)]. The cells were then separated into 200 μl aliquots, snap-frozen in liquid nitrogen and stored at −70 °C.
2.3.2.2 *Bacillus subtilis*

Competent *B. subtilis* DB428 cells were made fresh for each transformation, based on the protocol of Kunst *et al.* (1995). A fresh, single colony was streaked onto a nutrient agar (NA) plate and incubated overnight at 30 °C. The following day, a small loopful of cells was inoculated into 10 ml MD medium (see below), supplemented with 50 μl of 20 % (w/v) casamino acids and incubated at 37 °C with vigorous shaking to an OD$_{600}$ of 1-1.5. 10 ml warm MD medium lacking casamino acids was added to the culture, which was then incubated at 37 °C with shaking for a further 1 h. The cells were now competent. MD medium consists of (per 100 ml):

- 10 ml 10x PC (see below)
- 5 ml 40 % (w/v) glucose
- 2.5 ml 0.2 % (w/v) L-tryptophan
- 0.5 ml filter-sterilised (0.2 μm) 0.22 % (w/v) ferric ammonium citrate
- 5 ml 5 % (w/v) L-aspartic acid (pH 7.0)
- 30 μl 1 M MgSO$_4$, pH 7.0, filter-sterilised (0.2 μm)
- 10x PC stock consists of (per 1 L): 107 g K$_2$HPO$_4$, 60 g KH$_2$PO$_4$, 8.5 g Na$_3$ citrate.2H$_2$O.

2.3.3 Transformation

2.3.3.1 *Escherichia coli* JM109 cells

200 μl of in-house prepared, or 50 μl of purchased, competent *E. coli* JM109 cells were removed from -70 °C storage and thawed on ice (~5 min). Once thawed, 1-5μl of a ligation reaction was added and gently mixed and the cells were incubated on ice for 30 min. The cells were then heat-shocked at 42 °C for 90 s before being immediately returned to ice for 5 min. The mixture was transferred to a sterile 15 ml centrifuge tube, made up to 1 ml with LB medium and incubated at 37 °C for 1 h with shaking. 100 μl of culture was then spread onto an LB agar plate supplemented with ampicillin (100 μg ml$^{-1}$) or the preferred carbenicillin (100 μg ml$^{-1}$) and incubated at 37 °C overnight. The remainder of the culture was centrifuged (13000g, 1 min), resuspended in 100 μl LB and treated as above.
2.3.3.2 *Escherichia coli* BL21(DE3) and BL21(DE3) pLysS Cells

20 µl competent *E. coli* BL21(DE3) or BL21(DE3) pLysS cells were removed from storage at -70 °C and thawed on ice (~5 min). Once thawed, 1 µl of a plasmid preparation was added and gently mixed and the cells were incubated on ice for 30 min. The cells were then heat-shocked at 42 °C for 40 s before being immediately returned to ice for 5 min. The mixture was transferred to a sterile 1.5 ml microcentrifuge tube containing 80 µl LB medium and incubated at 37 °C for 1 h with shaking. The culture was then spread onto an LB agar plate supplemented with 100 µg ml\(^{-1}\) ampicillin or the preferred carbenicillin [and 34 µg ml\(^{-1}\) chloramphenicol for BL21(DE3) pLysS cells] and incubated at 37 °C overnight.

2.3.3.3 *Bacillus subtilis*

800 µl freshly-prepared competent *B. subtilis* DB428 cells were added to a pre-warmed (37 °C) 15 ml centrifuge tube. 4 µl of a plasmid preparation was added and gently mixed and the cells were incubated at 37 °C with shaking for 20 min. 25 µl of 20 % (w/v) casamino acids was then added and the cells were incubated at 37 °C with shaking for 1.5 h. 100 µl of the culture was then spread on a NA plate supplemented with 5 µg ml\(^{-1}\) kanamycin and incubated at 30 °C overnight. The remainder of the culture was centrifuged (13000 g, 1 min), resuspended in 100 ml NB and treated as above.

2.3.4 Agarose Gel Electrophoresis

DNA solutions were examined by agarose gel electrophoresis, which allows the resolution of linear, duplex DNA molecules between the sizes of 200 bp and 50 kbp with various concentrations of agarose (Sambrook *et al.*, 1989). Gels were made by dissolving the appropriate amount of agarose in 1x TAE buffer (0.04 M Tris acetate, 0.001 M EDTA) by boiling, adding ethidium bromide to a final concentration of 500µg l\(^{-1}\) and pouring the solution into a prepared perspex gel mould. A well-forming comb was added and the gel allowed to cool and set. Once set, the gel was placed in an electrophoresis tank containing 1x TAE buffer, the comb was removed.
and DNA was loaded in an appropriate amount of 6x loading buffer [40 % (w/v) sucrose, 0.25 % (w/v) xylene cyanol, 0.25 % (w/v) bromophenol blue]. DNA samples were run against a 1 kb ladder (see Appendix seven) for size estimation. The gel was electrophoresed at a constant voltage of 65-85 V and the DNA was visualised using a UV transilluminator.

### 2.3.5 DNA Purification

#### 2.3.5.1 Gel Purification
DNA fragments were separated by electrophoresis, excised from the gel with a clean scalpel and purified via QIAEX II gel extraction or Geneclean III kits according to the manufacturers’ instructions.

#### 2.3.5.2 Plasmid Preparations from *Escherichia coli*
A single colony was inoculated into 10 ml LB supplemented with ampicillin (100 μg ml⁻¹) or the preferred carbenicillin (100 μg ml⁻¹) and incubated at 37 °C with shaking, overnight. The following day, 3 ml was centrifuged (13000 g, 1 min) and the supernatant was removed, leaving the pellet as dry as possible. Plasmid DNA was purified via QiaPrep spin or S.N.A.P miniprep kits according to the manufacturers’ instructions.

#### 2.3.5.3 Plasmid preparations from *Bacillus subtilis*
A single colony was inoculated into 5 ml NB supplemented with kanamycin (5 μg ml⁻¹) and incubated at 30 °C with shaking, overnight. The following day, 3 ml was centrifuged (13000 g, 1 min) and the supernatant was removed, leaving the pellet as dry as possible. Plasmid DNA was purified using Qiaprep spin miniprep kit according to the manufacturer’s instructions, with the addition that 2 mg ml⁻¹ lysozyme was added to P1 buffer and the cells were incubated at 37 °C for 5 min. The plasmid DNA was then transformed into *E. coli* JM109 cells and purified again (see section 2.3.5.2) to obtain DNA of sufficient quality for further applications.
2.3.6 Polymerase Chain Reaction (PCR)

PCR was routinely carried out in a 100 µl reaction volume containing approximately 100 ng of template DNA, 100 pmol of each primer, 25 nmol dNTPs, 2 mM Mg\(^{2+}\), 1U DNA polymerase, all overlaid with 100 µl mineral oil to prevent evaporation. The buffers used were those supplied with the polymerase as a 10x stock, with final 1x concentrations as follows: Vent DNA polymerase, 10 mM KCl, 20 mM Tris-HCl (pH 8.8 at 25 °C), 10 mM (NH\(_4\))\(_2\)SO\(_4\), 2 mM MgSO\(_4\), 0.1 % (v/v) Triton X-100; Pfu DNA polymerase, 10 mM KCl, 20 mM Tris-HCl (pH 8.8 at 25 °C), 10 mM (NH\(_4\))\(_2\)SO\(_4\), 2 mM MgSO\(_4\), 0.1 % (v/v) Triton X-100, 0.1 mg nuclease-free BSA; Cloned Pfu DNA polymerase, 10 mM KCl, 10 mM (NH\(_4\))\(_2\)SO\(_4\), 20 mM Tris-HCl (pH 8.75), 2 mM MgSO\(_4\), 0.1 % (v/v) Triton X-100, 100 µg ml\(^{-1}\) BSA; BioTaq DNA polymerase, 16 mM (NH\(_4\))\(_2\)SO\(_4\), 67 mM Tris-HCl (pH 8.8 at 25 °C), 0.01 % (v/v) Tween-20; Taq/Pwo DNA polymerase, 50 mM Tris-HCl (pH 9.1), 14 mM (NH\(_4\))\(_2\)SO\(_4\). Reactions generally consisted of a ‘hot-start’ at 96 °C for 5 min, prior to the addition of the polymerase, which was added whilst holding the reaction at 85 °C and then 30 cycles of 96 °C for 75 s (denaturation), 55 °C for 90 s (annealing) and 72°C for 2 min (extension), followed by a final incubation at 72 °C for 10 min. Single primer controls were run with each change of primer and polymerase.

2.3.7 Restriction digestion

Digestion of DNA was carried out using restriction endonucleases, following the manufacturer’s instructions. DNA was routinely incubated in the appropriate buffer (supplied as a 10x concentrate) with up to 20 U of enzyme(s) at 37 °C for up to 3 h. For reactions requiring two enzymes, the buffer was chosen that gave the maximal activity for both enzymes. The glycerol concentration was kept below 10 % (v/v) to avoid star activity (the ability to cleave sequences that are similar, but not identical to, their defined recognition sequence). If need be, the enzymes were inactivated by heating at 65 °C for 20 min. DNA fragments were purified either straight from the reaction mixture by Geneclean III or QIAEX II gel extraction kits, omitting the gel.
extraction steps, or separated by agarose gel electrophoresis and purified as before. Restriction digestion controls were carried out, comprising the double digestion and purification of a control plasmid, followed by attempted re-ligation and transformation. The presence of transformants indicated that double digestion had not occurred.

2.3.8 Ligation of DNA

DNA fragments with cohesive ends were ligated together using T4 DNA ligase, either in conjunction with a single buffer and incubated at 16 °C overnight, or as part of a Rapid DNA Ligation kit and incubated at room temperature for 30 min to 2 h. Ligation controls were carried out, comprising the single digestion and purification of a control plasmid, followed by attempted re-ligation and transformation. The presence of transformants indicated that ligation has been successful.

2.3.9 Sequencing

DNA sequencing was automated and performed in house using an ABI PRISM 377 DNA sequencer (Perkin-Elmer Applied Biosystems, Foster City CA, USA). The sequencing is based on the method of Sanger et al. (1977), where the insertion of dideoxy nucleotide (ddNTP) terminates the extension of the complementary strand. Each ddNTP is specifically labelled with one of four fluorescent dyes, so it was not necessary to perform four separate reactions.

2.3.10 Induction of expression of pET and pTYB1 vectors

Fresh, single colonies of E. coli BL21(DE3) and BL21(DE3) pLysS cells containing the appropriate vector were grown overnight in 10 ml LB medium containing the appropriate antibiotic(s) at 37 °C with shaking. The following day, 1-5 ml of culture was transferred to 50 ml fresh LB medium and grown at 25-37 °C with shaking to an
OD$_{600}$ of approximately 0.6, induced with either 0.4 mM [BL21(DE3) cells] or 1mM [BL21(DE3) pLysS cells] IPTG and grown for up to 9 h at 25-37 °C with shaking.

### 2.3.11 Protein Estimation

Protein estimation was based on the method of Bradford (1976), where 100 µl of protein sample was incubated with 900 µl Bradford solution [0.01 % (w/v) Coomassie brilliant blue G-250, 4.8 % (v/v) ethanol, 8.5 % (v/v) phosphoric acid] for 15 min at room temperature. Absorbance at 595 nm was compared against BSA standards of 0-20 µg.

### 2.3.12 SDS PAGE

SDS PAGE was performed according to the method of Laemmli (1970). Resolving gels, consisting of either 4 ml or 5 ml of 30 % (v/v) acrylamide [corresponding to either 10 % (v/v) or 12.5 % (v/v) final acrylamide, respectively], 3 ml resolving gel buffer [18.15 % (w/v) Tris, pH 8.8, 0.4 % (w/v) SDS], made up to 12 ml with distilled water, were polymerised with 50 µl of 10 % (w/v) ammonium persulphate and 12.5 µl TEMED. This solution was poured into a mini-gel apparatus, overlaid with water-saturated isobutanol and left to set (~20 min). When set, the butanol was washed off with distilled water and stacking gel solution, consisting of 0.9 ml of 30% (v/v) acrylamide, 2.4 ml stacking gel buffer [5.9 % (w/v) Tris, pH 6.8, 0.4 % (w/v) SDS], made up to 6.9 ml with distilled water and polymerised with 50 µl of 10% (w/v) ammonium persulphate and 10 µl TEMED, was poured over the resolving gel. A well-forming comb was added and the gel left to set (~20 min). When the gel had set, the gel apparatus was placed in an electrophoresis tank, which was filled with tank buffer [0.63 % (w/v) Tris, pH 8.3, 0.1 % (w/v) SDS, 0.4 % (w/v) glycine] and the comb was removed. Protein samples were prepared by adding an equal volume of 2x loading buffer [1.52 % (w/v) Tris HCl, pH 6.8, 4 % (w/v) SDS, 20 % (w/v) sucrose, 0.08 % (w/v) bromophenol blue, 10 % (v/v) β-mercaptoethanol] and
boiling for 5 min. Electrophoresis was carried out through the stacking gel at a constant current of 10 mA per gel and through the resolving gel at a constant current of 20 mA per gel. The electrophoresis was stopped when the dye front reached the bottom of the gel. Gels were stained with Coomassie staining solution [0.25 % (w/v) Coomassie blue R, 45.4 % (v/v) methanol, 45.4 % (v/v) water, 9.2 % (v/v) acetic acid] for 30 min and destained in destain solution [7.5 % (v/v) acetic acid, 5 % (v/v) methanol, 87.5 % (v/v) water] overnight.

2.3.13 Preparation of Cell Extracts, Cell Debris and Culture Supernatant

Cell cultures were centrifuged (2500 g, 10 min) and the culture supernatant was removed and filtered if required (0.65 μm). Cell pellets were generally resuspended in a one-tenth culture volume of appropriate buffer. Cells were then either passed through a cell disrupter at 25-30 kpsi until lysed, or sonicated as follows. Cells were treated with a one-tenth volume of 1% (v/v) Triton X-100, lysozyme (100 μg ml⁻¹), incubated at 37 °C for 30-60 min. The mixture was then incubated on ice for 30 min and then sonicated (15 s bursts) until lysed. The cell debris was then pelleted by centrifugation (8000 g, 20 min) and the cell extract removed.

2.3.14 Enzyme Assays

Enzyme assays were always run in parallel with a control containing substrate alone, and one containing untransformed *E. coli* BL21(DE3) cell extract or *B. subtilis* DB428 culture supernatant instead of the recombinant cell extract or culture supernatant, to ensure measured rates were solely due to the activity of the recombinant enzyme. Unless otherwise stated, results graphs are based on a single acquisition of data.
2.3.14.1 General Peptidase Assays

General peptidase assays used azocasein and azoalbumin at 0.1-1 % (w/v) concentrations. Routinely, protein samples made up to 1ml in buffer were added to 1ml of substrate and incubated for 30-60 min. TCA was then added [3 % (w/v) final concentration], the mixture was incubated on ice for 5 min, centrifuged (4000 g, 5min) and the absorbance of the supernatant was measured at 410 nm.

2.3.14.2 Specific Peptidase Assays

Specific peptidase assays generally used 80 μM (final concentration) ala-ala-pro-phe-p-nitroanilide or ala-ala-phe-p-nitroanilide in a 1 ml reaction volume. Formation of the p-nitroaniline product was followed continuously at 410 nm.

2.3.14.3 Carboxylesterase Assays

Carboxylesterase assays used p-nitrophenyl butyrate or p-nitrophenyl acetate at varying concentrations in a 1 ml reaction volume. Formation of the p-nitrophenoxide product was followed continuously at 410 nm.
3. Chapter Three

Cloning and Expression of hlyR4 from Haloferax mediterranei

3.1 Introduction

The gene encoding halolysin R4 (hlyR4), a halophilic extracellular serine protease from the halophilic Archaeon Haloferax mediterranei, has previously been expressed in another halophilic Archaeon, Haloferax volcanii, from an E. coli/Hf. volcanii shuttle vector pMDS30 (Kamekura et al., 1996). The authors reported that the isolation of the entire gene into Escherichia coli in a single step was impossible due to deletion events, a phenomenon also experienced by the group cloning the halolysin gene from a separate strain of Hf. mediterranei (V. M. Stepanov, cited by Kamekura et al., 1996). However, smaller pieces of Hf. mediterranei DNA could be cloned and used to construct the entire gene which was subsequently stably maintained in E. coli. Expression of the hly gene from Natrialba asiatica has also been attempted in the osmophilic yeast Zygosaccharomyces rouxii (Kamekura et al., 1992). However, the plasmid could not be stably maintained in this host and so no expression was observed.

Our group has experience in the expression of halophilic genes in E. coli (Connaris et al., 1999) and so it was decided to attempt the cloning and expression of the hlyR4 gene in E. coli. Using E. coli as an expression host has a number of advantages over Hf. volcanii, including its short generation time and simplified transformation protocol. The expression system chosen was the vector pET-3a, a T7 promoter-driven system. This is transformed into E. coli BL21(DE3), a strain deficient in both lon and ompT proteases (Studier and Moffatt, 1986; Rosenberg et al., 1987; Studier et al., 1990). The pET vectors are not based on E. coli promoters and so there should not be any transcription in the absence of the T7 RNA
Chapter Three

polymerase, which is under the control of the lacUV5 promoter, and induced by IPTG.

Halolysin has three distinct regions, firstly, the pre-region which directs the protein to the membrane for translocation, during which the pre-region is cleaved. Next is the pro-region which folds after translocation and directs the folding of the mature region. Cleavage of the pro-region by the mature region liberates the mature region, which is the active species secreted into the environment. It was decided to produce three constructs of the protein, the full length pre-pro-protein and two truncated versions, encoding only the pro-protein and mature regions. It was thought that E. coli may fail to recognise the signal sequence as such and so a mature protein would be the only way to produce a catalytically active protein. It was hoped that the low salt environment of the E. coli cytoplasm would inhibit the potentially toxic activity of the protease until the desired addition of salt. If this was not the case, then the construction of the pro-protein was hoped to fulfil the same role.

Expression of halolysin R4 from pMDS30 in Hf. volcanii is dependent on 40bp immediately downstream of the natural stop codon that are thought to contain a transcription termination signal (Kamekura et al., 1996) and, even though this expression was from a constitutive (rather than inducible) promoter and pET-3a has its own termination sequences, the 40 bp were still included in all the three constructs.

3.2 Methods

3.2.1 Amplification and Cloning of hlyR4

For the isolation of the hlyR4 gene, primers were designed based on the published sequence (Kamekura et al., 1996). In order to clone the gene into pET-3a, it was necessary to introduce Ndel and BamHI sites at the 5’ and 3’ ends of the gene, respectively, and so the gene was checked for internal sites. The Ndel sequence contains the ATG start codon and ensures the gene is inserted into the vector in the correct frame. As the cleavage site of the pre-region (signal sequence) has not been reported, it was necessary to identify this for amplification of the pro-protein. This
was achieved via sequence alignment of the \textit{hly}R4 primary sequence with signal sequences of \textit{Halobacterium halobium} bacteriorhodopsin and subtilisin from \textit{Bacillus amyloliquifaciens}, with consideration of the general format of signal sequences as stated by Watson (1984). Sequence alignment was performed by Clustalw 1.6 (Thompson \textit{et al.}, 1994), shown in figure 3.1 below, and a putative signal sequence was identified.

\begin{verbatim}
Bacteriorhodopsin        MLELLPTAVEGVSQAQ
Subtilisin               MRGKKVWIS---LLFALALIFTMAFGSTSSAQ
Halolysin               MAGTPNFDRRSFLRLAAAGLATGMAGVT^ATPGRSP
\end{verbatim}

\textbf{Figure 3.1: Sequence alignment of the signal sequences of bacteriorhodopsin and subtilisin with halolysin R4.}

Residues underlined indicate the three amino acids immediately after the cleavage site and the doubly underlined residue indicates the residue of halolysin immediately preceding the putative cleavage site.

Based on the alignment with the two signal sequences shown above, the putative cleavage site was taken to be after the serine residue doubly underlined. Furthermore, Watson (1984) concluded that the most striking feature of a signal sequence is the presence of residues with small, uncharged side chains at the C-terminal end. Of the characterised prokaryotic signal sequences, 98\% end with alanine, glycine, serine, cysteine or threonine (Watson, 1984). For amplification of the pre-pro-protein, the forward primer had the base sequence 5' TCA TGT TTG CAT ATG GCA GGC ACA CCG 3', the pro-protein forward primer had the base sequence 5' GCT GGT GTA CAT ATG GCG ACT CCC GGT 3' and the mature protein forward primer had the base sequence 5' ACG CAC GAA CAT ATG TAT ACG GCG AAC 3' (\textit{NdeI} sites underlined). The reverse primer used in all three cases had the base sequence 5' ACA GTC GTG TGA CAT ATG GCG ACT CCC GGT 3' (\textit{BamHI} site underlined). The presence of the \textit{NdeI} site in the pro-region and mature region constructs introduces the ATG start codon and therefore a methionine residue. Appendix one shows the gene sequence of halolysin, the positions of the primers and the deduced amino acid sequence of the three constructs.
PCR was performed as described in section 2.3.6 using Taq DNA polymerase and the products were separated on a 1 % agarose gel. The appropriate resulting fragments were excised from the gel with a clean scalpel and purified using the Geneclean protocol. The DNA was then digested with Ndel and BamHI at 37 °C for 3 h (see section 2.3.7) and purified from the reaction buffer and DNA fragments using the Geneclean protocol. The resulting DNA was ligated to pET-3a, which had been treated with the same enzymes, by incubation with DNA rapid ligase at room temperature for 30 min. The non-expression host *E. coli* JM109 cells (strain lacking the T7 RNA polymerase) were transformed to ampicillin resistance with the ligation reaction as described in section 2.3.3.1. The following day ampicillin resistant colonies were picked and grown in LB medium to enable plasmid purification as described in section 2.3.5.2. The eluted plasmids were screened by treatment with Ndel and BamHI at 37 °C for 3 h and separation on a 1 % agarose gel. Plasmids showing the correctly-sized vector and insert bands were selected for expression studies.

### 3.2.2 Expression of hlyR4 in *E. coli*

Once the presence of the vector was confirmed by ampicillin-resistant growth of transformants and the presence of a correctly-sized insert in the vector by restriction analysis, the vector was transferred to the expression host, *E. coli* BL21(DE3), as described in section 2.3.3.2. The following day, single colonies were subcultured onto a fresh LB ampicillin plate thereby providing a 'master plate' to ensure a homogeneous population was used in subsequent experiments. Growth of transformants and induction of expression were performed as described in section 2.3.10. The cultures were harvested by centrifugation (2500 g), the supernatant was separated and the cell extracts (soluble fraction) and insoluble fractions were prepared by resuspending the cell pellets in 5 ml of 10 mM Tris-HCl, pH 7.6, containing 4.3 M NaCl and disrupting the cells by sonication as described in section 2.3.13. The cell debris was pelleted by centrifugation (8000 g), the cell extract (supernatant) was removed, and the cell debris was resuspended in 5 ml of 10 mM Tris-HCl, pH 7.6, containing 4.3 M NaCl. The cell extract was assayed for
proteolytic activity against azocasein, azoalbumin and N-succinyl-ala-ala-pro-phe-p-nitroanilide at 37 °C as described in sections 2.3.14.1 and 2.3.14.2 using 10 mM Tris-HCl, pH 7.6, containing 2 or 4.3 M NaCl. The assays were verified using subtilisin as a positive control.

3.3 Results

3.3.1 Amplification and Cloning of hlyR4

Following PCR amplification, bands were obtained of the expected size for all three constructs (1640 bp for the pre-pro-protein, 1556 bp for the pro-protein and 1295 bp for the mature halolysin) that were distinct from the single primer controls. Figure 3.2 below shows samples of the PCR products run on 1 % agarose gels. The arrows indicate the bands of interest.

![Figure 3.2: Separation of PCR products of pre-pro, pro and mature gene constructs.](image)

a) Lanes 1 & 9, 1 kbp marker (arrow indicates 517 bp); lanes 2 & 3, pre-pro-protein PCR products; lanes 4 & 5, mature protein PCR products; lane 6, pre-pro-protein forward primer control; lane 7, mature protein forward primer control; lane 8, reverse primer control.

b) Lane 1, 1 kbp marker (arrow indicates 1636 bp); lane 2, pro-protein PCR product; lane 3, forward primer control.
Following restriction digestion, ligation into pET-3a and transformation into \textit{E. coli} JM109 cells, numerous colonies were picked and screened for the presence of an appropriately-sized insert. Such restriction patterns were observed for each of the constructs and the following figure shows the digests run on 1\% agarose gels. The arrows indicate the genes cut out of the plasmid.

![Figure 3.3: Ndel/BamHI restriction digestion of transformant plasmids.](image)

a) Lane 1, 1 kbp marker (arrow indicates 1018 bp); lanes 2-5, transformants a-d.
b) Lane 1, 1 kbp marker (arrow indicates 517 bp); lanes 2-4, transformants e-g.

Both the above agarose gels show the expected characteristic band pattern consisting of the linear plasmid and the excised gene for each construct. Transformants a (pre-pro-protein; gel a, lane 2), d (mature protein; gel a, lane 5) and f (pro-protein; gel b, lane 3) were chosen for expression studies.

3.3.2 Expression of \textit{hlyR4} in \textit{E. coli}

The pET-3a vectors containing the appropriate construct were transformed successfully into the expression host, \textit{E. coli} BL21(DE3). Expression of the three gene constructs was measured via SDS PAGE analysis of whole cell pellet samples as well as the soluble and insoluble fractions. The following figure shows the
separation of the proteins by SDS PAGE and expression levels of the three constructs. The arrows indicate the putative bands of interest.

It is clear from figure 3.4 that the levels of expression from pET-3a were very high, with the recombinant proteins constituting the major bands on the SDS gels. As judged by SDS PAGE, the relative molecular masses of the putative pre-pro-protein, pro-protein and mature protein are 75, 57 and 64 k, respectively. These values are higher than expected from the translated gene sequences, namely 53.5, 50.6 and 41.3 k, respectively. The pre-pro-protein construct of halolysin was not secreted across the E. coli membrane, as it is clear that the protein is localised in the cytoplasm. All three constructs were expressed to measurable levels in a soluble
form, although a significant proportion of both the pre-pro-protein and the pro-protein was also expressed as insoluble inclusion bodies. The mature construct was also expressed as inclusion bodies on a number of separate occasions, which seemed to be independent of culture volume and the method of cell lysis. Also, insoluble protein was seen after growth and expression at 25 °C. Attempts were made to renature the insoluble inclusion bodies of all three constructs. This was attempted via denaturation with 8 M urea and 50 mM DTT, followed by dilution to a final concentration of 10-20 µg ml\(^{-1}\) into renaturation buffers. The buffers contained 3mM reduced glutathione, 0.3 mM oxidised glutathione and either 2 or 4.3 M NaCl. The renaturation mixtures were incubated at 4 °C overnight and assayed for activity against ala-ala-pro-phe-pNA. However, the protein either visibly precipitated immediately after dilution, or was inactive following the overnight incubation.

Following cell lysis, soluble fractions containing the three constructs were assayed for proteolytic activity against azocasein, azoalbumin and ala-ala-pro-phe-pNA; however, no reproducible activity was measured. For example, the control reactions varied under similar conditions and the absorbances measured for the recombinant enzymes varied independently of the length of the assay or the amount of enzyme added. The enzyme used to verify the assay (subtilisin) gave reproducible activity and was even active in the presence of the concentrations of salt used for the recombinant proteins. The following table shows the results from one such experiment, performed as described in section 2.3.14.1. (0-4 U of subtilisin was incubated at 37 °C for 1 h in the presence of 1% azocasein in 10 mM Tris-HCl buffer, pH 7.6. In addition, 0.4 U of subtilisin was incubated at 37 °C for 1 h in the presence of 1% azocasein in 10 mM Tris-HCl buffer, pH 7.6, containing 2 M NaCl).

<table>
<thead>
<tr>
<th>Amount of Subtilisin (U)</th>
<th>(A_{410}) (no salt)</th>
<th>(A_{410}) (2 M NaCl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>0.04</td>
<td>0.663</td>
<td>-</td>
</tr>
<tr>
<td>0.20</td>
<td>1.082</td>
<td>-</td>
</tr>
<tr>
<td>0.40</td>
<td>1.303</td>
<td>0.668</td>
</tr>
<tr>
<td>1.00</td>
<td>1.632</td>
<td>-</td>
</tr>
<tr>
<td>4.00</td>
<td>1.917</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 3.5: Table showing activities of subtilisin against azocasein.
3.4 Discussion

This work has been successful in expressing *hlyR4* in *E. coli*. All three of the desired constructs, the pre-pro-protein, the pro-protein and the mature protein were expressed to high levels. As judged by SDS PAGE analysis of cell fractions and comparison with a non-transformed control, it is clear that the major proteins produced upon induction were the halolysin constructs. The apparently anomalous migration on SDS PAGE is most likely due to the high surface acidic charge characteristic of halophilic proteins. The high negative charge repels the SDS detergent, causing a slower gel migration and resulting in an increased estimation of size. Anomalous migration of some proteins on SDS PAGE was noted early on (Dunker and Rueckert, 1969) and this has been reported for other halophilic proteins (e.g. Bischoff and Rodwell, 1996; Izotova *et al.*, 1983).

However, the results of this work have re-emphasised some of the recognised problems of expressing a halophilic gene in *E. coli*. The internal environment of *E. coli* is a low-salt environment, in contrast to that of the extreme halophiles, and the accumulation of expressed halophilic proteins as insoluble inclusion bodies is often the result (e.g. Taupin *et al.*, 1997; Madern *et al.*, 1995; Blecher *et al.*, 1993). However, once soluble expression, or renaturation of the inclusion bodies, is achieved a recombinant halophilic protein generally only requires the addition of physiological levels of salt to become catalytically active. This was not achieved in this work. The design of the pre-pro- and pro-proteins was such, however, that activity was not anticipated, as the pre-pro- and pro-regions are expected to prevent the production of a catalytically active protease. As discussed in section 3.1, the pre-region acts as a signal sequence to direct translocation across the inner membrane, and the pro-region acts as a chaperone to promote folding of the protein. Both of these regions prevent the functional activity of the enzyme, which is only seen after cleavage of the pro-region. The halolysin signal sequence is obviously not recognised by *E. coli*, as the protein remains in the cytoplasm, and so the host’s signal peptidase can not act to cleave this region, inhibiting the action of the pro-region in the further folding and processing of this construct. An obvious solution to this problem is to use a signal sequence from an *E. coli* protein that is secreted. Unfortunately, *E. coli* does not normally secrete many proteins, and the exploitation
of the secretion pathways through both the inner and outer membranes is not particularly efficient (Makrides, 1996). A second approach is to use a signal sequence in tandem with agents that induce non-specific permeability of the outer membrane, resulting in the release of protein into the growth medium. For example, the fusion of the signal sequence from \textit{ompA} to streptokinase has effected its secretion from \textit{E. coli} (Ko \textit{et al.}, 1995). The addition of glycine and Triton X-100 to the culture medium has been shown to increase the levels of a tumour necrosis factor alpha fusion protein secretion 170-fold (Yang \textit{et al.}, 1998).

It was hoped that the pro-region might enable the folding of the pro-protein construct; however, this also appears not to have occurred, based on the absence of activity. The reason for this is possibly due to the fact that the pro-region itself may not have folded into its active conformation in the \textit{E. coli} cytoplasm and therefore was not able to act as a chaperone to assist in the folding of the mature region.

The absence of activity of the mature halolysin construct is unexpected. Whilst the expression of many protease genes can give rise to problems involved with toxicity, a halophilic protease is almost in an ideal situation, whereby the expression of an active protein is inhibited by the lower concentrations of salt and reconstituted following cell lysis, by the addition of salt. Even following soluble expression of the mature region of halolysin, activity was not observed after the addition of varying concentrations of salt.

A major factor in the lack of activity could be the sequence of the recombinant protein. PCR amplification of the genes was performed with \textit{Taq} polymerase, an enzyme without proof-reading activity, and hence a lower fidelity. Sequencing of the three constructs was not performed and so the presence of inactivating mutations can not be ruled out. Another factor that can have an effect on the structure and stability of a recombinant protein is the presence or absence of the N-terminal methionine (Uversky \textit{et al.}, 1999). In \textit{E. coli}, the translation of all proteins is initiated with an \textit{N}-formylmethionine residue, which is subsequently processed by a deformylase enzyme (Adams, 1968) and cleaved by methionine aminopeptidase. However, around 40\% of the cytosolic proteins present in \textit{E. coli} retain their N-terminal methionine residue, as the action of methionine aminopeptidase has been found to depend on the penultimate amino acid (Hirel \textit{et al.}, 1989). The catalytic efficiency of methionine aminopeptidase decreases with
increasing side-chain length of the penultimate amino acid, with glycine, alanine, proline, serine, threonine, valine and cysteine the only amino acids effecting significant cleavage.

The penultimate amino acid in the mature region of halolysin is tyrosine (Kamekura et al., 1996), and therefore it has a very high probability of retaining its N-terminal methionine after expression. Chaudhuri et al. (1999) reported that recombinant α-lactalbumin expressed in *E. coli* possessed a remarkable decrease in conformational stability and the destabilising effect of the N-terminal methionine has also been shown for hen egg white lysozyme (Mine et al., 1997) expressed in *E. coli*. It should be noted, however, that there are no obvious rules relating to the structural consequences of the modification, as Uversky et al. (1999) reported that the presence of the N-terminal methionine has a stabilising effect on a permuted variant of S6 ribosomal protein expressed in *E. coli*.

Following the lack of activity in the soluble fractions of the recombinant proteins, it may seem unwise to dwell on the insoluble fractions. However, protein folding is known to be a complex process, which proceeds via the formation of intermediate protein structures, and so it is possible that the soluble mature halolysin construct has misfolded into an inactive conformation. Denaturation and subsequent refolding of either fraction could direct the renaturation into a native state, although inclusion bodies offer a simple purification step. Strategies for protein refolding are therefore still desirable. Although a generic protocol for this has not been developed, successful strategies have been reported which entail growing *E. coli* under osmotic stress (Blackwell and Horgan, 1991), using size-exclusion chromatography (Batas and Chaudhuri, 1996) and hydrostatic pressure (Foguel et al., 1999). A novel procedure has also been reported for the renaturation of trypsin, using an engineered trypsin inhibitor (Nohara et al., 1999). Time constraints prohibited extensive investigation into the renaturation of the halolysin inclusion bodies, but the scheme used failed to recover a soluble protein. The inclusion bodies were solubilised with 8M urea and 50 mM DTT, followed by dilution into a renaturation buffer containing 3 mM reduced glutathione, 0.3 mM oxidised glutathione and either 2 or 4.3 M NaCl. Typically, the protein immediately precipitated upon dilution. However, overnight dialysis of the denatured inclusion bodies, to slowly remove the high concentrations of urea, would perhaps be a more successful alternative to the dilution step. The
growth and induction of cultures at a lower temperature, for example 25 °C, has been shown to direct expression of a soluble product (Schein and Notebom, 1988). This has been successful within our group in the expression in *E. coli* of citrate synthase from *Haloferax volcanii* (Connaris *et al.*, 1999). Lowering of the growth temperature in this work, however, has not succeeded in producing soluble recombinant halolysin.

Large-scale expression of halophilic genes in *E. coli* overcomes the problems of growing halophiles, such as the need for high salt media which can cause corrosion in stainless steel fermenters. However, it has been suggested that expression in *Haloferax volcanii* could be a generic approach for all halophilic enzymes (Jolley *et al.*, 1996) following the construction of shuttle vectors for this organism (Lam and Doolittle, 1989; Nieuwlandt and Daniels, 1990; Holmes *et al.*, 1994). It appears then that expression of halolysin in a halophilic host is, despite the technical difficulties involved in for example transformation, a more appropriate procedure than the expression system utilised in this work.

At this point, it was decided not to attempt the expression of halolysin in *H. volcanii*, as has already reported (Kamekura *et al.*, 1996), but instead to concentrate on another class of extremophiles as a source of proteolytic enzymes, the hyperthermophiles. As already discussed (see section 1.8) genes from hyperthermophilic organisms are generally less problematic to express in *E. coli* than those from halophiles and the sequencing of their genomes has identified numerous potential proteolytic enzymes.
4. Chapter Four

Cloning and Expression of an Esterase from Archaeoglobus fulgidus

4.1 Introduction

Archaeoglobus fulgidus VC-16 is a hyperthermophilic, sulphate-reducing Archaeon (Stetter et al., 1987). It is strictly anaerobic and has a growth optimum of 83 °C, with a doubling time of 4 h, although it does grow between 60 and 95 °C. It can grow chemoorganotrophically on, amongst other substrates, casamino acids, peptone, casein and yeast extract with sulphate as the electron acceptor (Stetter, 1988). The genome of A. fulgidus has been sequenced (Klenk et al., 1997) and found to contain 2436 open reading frames (ORFs) within 2.2 Mbp, over half of which have no assigned role.

To date, there have been no reports describing proteolytic activity from A. fulgidus, although since this project was initiated there have been reports of an esterase gene (AFEST), that has been cloned and expressed in Escherichia coli (Manco et al., 2000b). The gene product was expressed in an active and soluble form and found to encode a 35.5 k carboxylesterase with an optimum for activity at around 80 °C and pH 7.1. The enzyme is very thermostable, having half-lives of 28 and 26 min at 90 and 95°C, respectively. Subsequent work has investigated its conformational dynamics at high temperatures (D’Auria et al., 2000a), a homology model has been built (Manco et al., 2000a) and preliminary crystallographic studies have been reported (Liu et al., 2000).

One disparity in the published information has arisen though. The original paper (Manco et al., 2000b) reported the amplification of ORF AF1716 (A. fulgidus ORF number 1716), which according to The Institute for Genomic Research (TIGR; http://www.tigr.org) is a gene of 933 bp, encoding for a carboxylesterase of 35.5 k.
The subsequent papers, however, (D’Auria et al., 2000a; Manco et al., 2000a; Liu et al., 2000) describe the ORF as AF1763, which according to TIGR is a gene of 1422bp, encoding a putative lipase of 52.8 k, although the authors reported that no lipase activity was detected in the expressed protein using trioleoylglycerol as substrate (Manco et al., 2000b).

The ORF amplified in this work is also AF1763, which was chosen at the time to demonstrate the ability of the chosen expression system to produce an active hydrolase. Proteases have been shown to demonstrate esterase activity (Holmquist and Vallee, 1976), although it is not necessarily expected that an esterase will have proteolytic activity. However, the hypothesised in vivo action of an unknown protein may well depend upon only one characterised activity in vitro. It should be noted that AF1763 has comparable similarity to both an esterase from Burkholderia gladioli and a lipase from Bacillus licheniformis and that the assigned gene function in a genome annotation is not always correct (see section 6.1), a fact borne out by the demonstration of Manco et al. (2000b) that AF1763 is in fact an esterase. Even single base changes can potentially give either a truncated or elongated gene, or removal of an identifying motif. A protein whose main function is ester hydrolysis might be expected to show less toxic effects when produced recombinantly than one which is peptide hydrolysis and the main aim of this chapter is to demonstrate the capability of the expression system to produce an active hydrolytic enzyme.

The vector used in this work is the pET vector pET-11a, which is similar to pET-3a (see section 3.1), with additional repression both at the level of the host strain and also the T7 promoter itself. pET-11a contains the T7lac promoter, which contains a lac operator sequence immediately downstream of the promoter region. Binding of the lac repressor at this site reduces transcription by T7 RNA polymerase. pET-11a is transformed into E. coli BL21(DE3) pLysS cells, which are identical to BL21(DE3) cells, but they also contain pLysS plasmids encoding T7 lysozyme, which is a natural inhibitor of T7 RNA polymerase. Thus, transcription as a whole is reduced and therefore so is uninduced transcription.
4.2 Methods

4.2.1 Amplification and Cloning of AF1763

For the amplification of AF1763, primers were designed based on the published sequence (TIGR; http://www.tigr.org). In order to clone the gene into pET-11a, it was necessary to introduce \textit{NdeI} and \textit{BamHI} recognition sites into the 5' and 3' ends of the gene, respectively, and so the sequence was checked for internal sites.

For amplification of AF1763, the forward primer had the base sequence 5' TGA AGG TAA TTC CAT ATG AGA GGG CTT GCC 3' (\textit{NdeI} site underlined) and the reverse primer had the base sequence 5' TTC TTT CCT TCT GGA TCC CTT AGC AAA AAA G 3' (\textit{BamHI} site underlined); see Appendix two.

PCR was performed as described in section 2.3.6 using \textit{Pfu} DNA polymerase and the product was separated on a 1 % agarose gel. The appropriate resulting fragment was excised from the gel with a clean scalpel and purified using the Geneclean protocol. The DNA was then digested with \textit{NdeI} and \textit{BamHI} at 37 °C for 4 h (see section 2.3.7) and purified from the reaction buffer and DNA fragments using the Geneclean protocol. The resulting DNA was ligated to pET-11a, which had been treated with the same enzymes, by incubation with DNA rapid ligase at room temperature for 2 h. The non-expression host \textit{E. coli} JM109 cells (strain lacking the T7 RNA polymerase) were transformed to ampicillin resistance with the ligation reaction as described in section 2.3.3.1. The following day, ampicillin resistant colonies were picked and grown in LB medium to enable plasmid purification as described in section 2.3.5.2. The eluted plasmids were screened by treatment with \textit{NdeI} and \textit{BamHI} at 37 °C for 3 h and separation on a 1 % agarose gel. Plasmids showing the correctly-sized vector and insert bands were selected for expression studies. Sequencing of the gene was performed as described in section 2.3.9, from pET-11a prepared from \textit{E. coli} JM109 using T7 promoter and termination primers.
4.2.2 Expression of AF1763 in *E. coli*

Once the presence of the vector was confirmed by ampicillin resistant growth of transformants and the presence of a correctly-sized insert in the vector by restriction analysis, the vector was transferred to the expression host, *E. coli* BL21(DE3) pLysS as described in section 2.3.3.2. The following day, single colonies were subcultured onto a fresh LB ampicillin plate thereby providing a 'master plate' to ensure a homogeneous population was used in subsequent experiments. Growth of transformants and induction of expression was performed as described in section 2.3.10. The cultures were harvested by centrifugation (2500 g), the supernatant was discarded and the cell extract (soluble fraction) and cell debris (insoluble fraction) were prepared by resuspending the cell pellets in 5 ml of 50 mM EPPS buffer, pH 8.9, 2 mM EDTA, and lysing the cells by cell disruption as described in section 2.3.13. The cell debris was pelleted by centrifugation (8000 g), the cell extract (supernatant) was removed, and the cell debris was resuspended in 2 ml of 50 mM EPPS buffer, pH 8.0, 2 mM EDTA. The cell extract was assayed for proteolytic activity against 0.1 % (w/v) azocasein and azoalbumin, and 80μM ala-ala-pro-phe-pNA and ala-ala-phe-pNA, at 70 °C, as described in sections 2.3.14.1 and 2.3.14.2, respectively, using the above buffer. It was also assayed for esterase activity against varying concentrations of *p*-nitrophenyl acetate and *p*-nitrophenyl butyrate at 70 °C, as described in section 2.3.14.3. Enzyme kinetics were analysed using Enzpack for Windows v1.4 software (BIOSOFT).

4.3 Results

4.3.1 Amplification and Cloning of AF1763

Following PCR amplification, a band was obtained of the expected size for the AF1763 gene (1478 bp) that was distinct from the single primer controls. Figure 4.1 shows the PCR product run on a 1 % agarose gel. The arrow indicates the band of interest.
Chapter Four

Figure 4.1: Agarose gel showing PCR product from AF1763 amplification.
Lane 1, AF1763 PCR product; lane 2, forward primer control; lane 3, reverse primer control; lane 4, 1 kbp marker (arrow indicates 1636 bp).

Following restriction digestion, ligation into pET-11a and transformation into E. coli JM109 cells, numerous colonies were picked and screened via restriction digestion for the presence of an appropriately-sized insert. Such restriction patterns were observed for AF1763 and figure 4.2 below shows the digests run on a 1 % agarose gel. The arrow indicates the gene cut out of the plasmid.

Figure 4.2: Agarose gel showing digested plasmids from transformed cells.
Lane 1, 1 kb marker (arrow indicates 1018 bp); lanes 2 & 3, AF1763 transformants a & b.

The gel shows the expected characteristic band pattern consisting of the linear plasmid and the excised gene for AF1763. Transformant ‘a’ was chosen for expression studies.
4.3.2 Expression of AF1763 in E. coli

The pET-11a vector containing AF1763 was transformed successfully into the expression host, E. coli BL21(DE3). Expression of the gene was measured via SDS PAGE analysis of whole cell pellet samples as well as the soluble and insoluble fractions. The expression of the gene, however, was too low to be measured via SDS PAGE, and no band was observed that was distinct from an untransformed sample of BL21(DE3) cells. Expression was attempted on more than one occasion with the same result.

4.3.3 Assaying for Activity of the Recombinant Enzyme

AF1763 did not exhibit activity against either of the peptide substrates, or the protein substrates, even when up to 400 μg protein was added to the reaction. However, AF1763 was active against both p-nitrophenyl butyrate and p-nitrophenyl acetate. Typical assays were carried out in duplicate at 70 °C, in 50 mM EPPS buffer, pH 8.0, 2mM EDTA, and contained 85 μg of protein. One unit of activity for this assay is defined as that required to produce 1 nmol of product per minute. The specific activity is defined as the number of units per mg of protein. Taking into account the temperature coefficient of the buffer’s pKₐ (-0.015), the reaction pH at 70 °C will be 7.3. The extinction coefficient (ε) of the resulting nitrophenolate anion at this pH is found to be 10800 1 M⁻¹ cm⁻¹, calculated according to the following equation (John, 1992):

\[ ε_{pH} = 18000 \times \left[ \frac{10^{(pH-7.15)}}{10^{(pH-7.15)} + 1} \right] \] (where pH=7.3 in this case)

Initial rates were measured and found to vary proportionally to the amount of enzyme added, one of the indications of a true enzymatic reaction. The effect of substrate concentration on specific activity was investigated, and values of \( K_M \) and \( V_{max} \) were determined by the direct linear plot of Cornish-Bowden and Eisenthal (1974). Calculated values for \( K_M \) and \( V_{max} \) from the direct linear plot are 3.99 ±1.29
mM and 704 ±185 U mg\(^{-1}\), respectively. To determine whether the data are a good fit to the Michaelis-Menten equation, a linear transformation is performed on it, and the data are plotted in a form that is more easily interpreted. The Michaelis-Menten equation can be rearranged from

\[
V = \frac{V_{\text{max}} \cdot S}{K_M + S}
\]

to give

\[
\frac{s}{v} = \frac{1}{V_{\text{max}}} \cdot \frac{V_{\text{max}}}{S} + \frac{K_M}{V_{\text{max}}}
\]

where a graph of s/v against s is linear. Taking the estimates of \(K_M\) and \(V_{\text{max}}\) from the direct linear plot, a trendline can be drawn, with a gradient of \(1/V_{\text{max}}\) and a y-intercept of \(K_M/V_{\text{max}}\). How closely the experimental data fit the line is an indication of how closely the data fit Michaelis-Menten kinetics. The following figure shows the data plotted as s/v against s.

---

**Figure 4.3:** Hanes plot for AF1763 against \(p\)-nitrophenyl butyrate. S, mM; v, U/mg. Assays were carried out in duplicate and the error bars represent the range of the data about the mean. The trendline is a transformation of the Michaelis-Menten equation and indicates where perfect data are expected to lie.

---

It is clear from this graph that the data do not fit the line very closely and hence do not fit Michaelis-Menten kinetics very closely. The following figure shows the data plotted as specific activity against substrate concentration. The line fitted through the experimental points is that generated from the Michaelis-Menten equation using \(K_M\) and \(V_{\text{max}}\) values determined by the direct linear plot.
Figure 4.4: The dependence of velocity on substrate concentration for AF1763 against $p$-nitrophenyl butyrate. The upper and lower lines indicate the extent of the errors of $K_M$ and $V_{\text{max}}$.

It can be seen from figure 4.4 that the substrate concentrations used in the assays are not only far from saturating, they are also mainly below $K_M$.

The main aim of this chapter was to express a recombinant hydrolase to ensure the expression system chosen was capable of doing so. Therefore, extensive characterisation of the resulting esterase was not carried out, but AF1763 also exhibited activity against $p$-nitrophenyl acetate, and estimates were obtained for $K_M$ of $1.43 \pm 0.42$ mM and $V_{\text{max}}$ of $1390 \pm 400$ U mg$^{-1}$. Due to the limited solubility of the substrate in aqueous solution, the substrate concentrations used in the assay were significantly below $K_M$, which is reflected in the large errors in the values determined for $K_M$ and $V_{\text{max}}$. In addition, initial thermostability studies indicated that the esterase retains approximately 75% of its activity against $p$-nitrophenyl butyrate after incubation at 70 °C for 30 min.

Sequencing of the gene has revealed the presence of at least five nucleotide mutations that each result in a mutation in the amino acid sequence. The resulting mutations are aspartate to tyrosine at position 61, serine to glycine at position 80, serine to isoleucine at position 87, lysine to aspartate at position 101 and serine to asparagine and position 102.
4.4 Discussion

The results presented in this work have demonstrated that the expression system chosen is capable of the expression of an active, recombinant hydrolase, in this case an esterase. The levels of expression of the esterase as judged by SDS PAGE, however, were such that they were indistinguishable from the non-expression control. The reasons for this are unknown without further work, but possible explanations are discussed below. There have been recent reviews discussing the optimisation of protein expression in *E. coli* (e.g. Hannig and Makrides, 1998; Makrides, 1996; Weickert *et al.*, 1996; Hockney, 1994), which have summarised a number of properties known to influence the levels of expression of a recombinant gene.

One of the problems encountered with expression of recombinant proteins in *E. coli* is proteolysis, that is the selective removal of abnormal and incorrectly folded proteins. This problem can be minimised by the secretion of the recombinant protein to the periplasm or the extracellular medium, although as discussed in section 3.4, *E. coli* is generally not considered a particularly good host for secretion. The alternative is cytosolic expression using protease-deficient host strains such as BL21(DE3), although protein degradation is still recognised as a problem in heterologous expression.

As discussed in section 3.4, the presence or absence of the N-terminal methionine can have an effect on the stability and structure of a recombinant protein. The penultimate amino acid determines the efficiency of removal of the N-terminal methionine, which is arginine in AF1763. According to Hirel *et al.* (1989), arginine residues inhibit the removal of the N-terminal methionine.

One other determining factor in the levels of expression is secondary structure in the mRNA transcript, which can block the translation initiation and Shine-Dalgarno regions (Looman *et al.*, 1986; Tessier *et al.*, 1984). Without a detailed analysis and a structural model for the mRNA of the gene expressed, however, it is not possible to discuss the extent of structure formation of the mRNA and therefore its potential significance.

Another factor known to influence expression levels is codon usage within the target gene. Genes show a non-random usage of synonymous codons, which
usually relates to the abundance of their cognate tRNAs (Ikemura, 1985). By implication, heterologous genes containing codons that are rarely used by *E. coli* may therefore be expressed inefficiently. For example, the presence of the arginine-encoding codons AGA and AGG has been found to be the limiting factor in the expression of several mammalian genes in *E. coli*, as the organism uses those codons so rarely (Brinkmann *et al.*, 1989). The authors reported that levels of expression increased when the gene was co-expressed with the *argU* (*dnaY*) gene encoding tRNA\(^\text{Arg}^\text{AGG/AGA}\). Furthermore, six archaeal genes have been expressed in *E. coli* to levels 5 to 20-fold higher when expressed in the presence of a plasmid expressing the genes encoding tRNA\(^\text{Arg}^\text{AGA/AGG}\) and tRNA\(^\text{Ile}^\text{AUA}\) (Kim *et al.*, 1998). It has also been reported that translational inhibition is much stronger when rare codons are located near the 5' end of the mRNA (Goldman *et al.*, 1995). AF1763 contains the rare arginine-encoding codons AGA and AGG 25 times and the isoleucine-encoding codon AUA 10 times and this may have been one of the causes of the low levels of expression of AF1763.

Even in the absence of observed expression, however, cell extracts of *E. coli* expressing AF1763 still exhibited measurable activity against both \(p\)-nitrophenyl butyrate and \(p\)-nitrophenyl acetate. The \(K_M\) for \(p\)-nitrophenyl butyrate was 3.99 ± 1.29mM, which is considerably higher than the value of 21 μM reported by Manco *et al.* (2000b). A comparison of the \(K_M\) values for \(p\)-nitrophenyl acetate is not possible as it has not been reported previously, although the authors did report that with the exception of \(p\)-nitrophenyl octanoate, the \(K_M\) values decreased with an increase in the acyl chain length (Manco *et al.*, 2000b), implying a high \(K_M\). Estimations of \(K_M\) can vary between samples of cell extract and purified protein if there is a species present that also binds the substrate, but this is a considerable difference and so there must be other factors influencing the measurement. A comparison with the remaining data described by Manco *et al.* (2000b) is not possible, as this work was finalised before their report. The data presented do not exactly fit the Michaelis-Menten curve, as judged by the non-linear behaviour when \(s/v\) is plotted against \(s\). One possible reason for this is that there is substrate inhibition. However, the most likely reason is inaccurate measurements. This can be a consequence of using substrate concentrations that are far below \(K_M\), as the gradient of the curve is such that a small error in the substrate concentration results in a large error in measured velocity.
The presence of mutations in the gene sequence is an important factor. Manco et al. (2000b) sequenced AF1763 and found it to be identical to that reported in the genome sequence (Klenk et al., 1997). The mutations in this work are therefore most likely to be genuine and not a mistake in the genome sequence. The comparatively high $K_M$ and low $k_{cat}$ values might be due to disruption in the region around the active site, although based on the homology modelling reported by Manco et al. (2000a), serine 160 is the nucleophile, with aspartate 255 and histidine 285 constituting the other residues of the catalytic triad, and the mutations found in the AF1763 gene in this work do not disrupt the catalytic triad. Complete sequencing of the gene, however, was not carried out and also without a homology model of this protein, the effects of the mutations can not accurately be established. The inability to produce a ‘native’ gene is cause for concern, as the PCR was performed using a high-fidelity polymerase and low magnesium concentrations to minimise errors. Gene toxicity is known to be a problem when expressing proteases (see section 5.4 for a discussion), although it is possible that an esterase exhibits a lethal effect on cells, even more so as the native esterase could have proteolytic activity. The result of this is that only genes containing activity-reducing mutations are selected. Even though the growth and induction of the transformants was carried out at 37 °C, which is far below the temperature optimum of AF1763, it has been reported that even at 10 °C it shows around 10 % of its maximal activity (Manco et al., 2000b). So it would not be unreasonable to expect AF1763 to have a significant activity at the induction temperatures.

One method of producing recombinant proteases is to express them as fusion proteins, which are generally used as an aid to purification, but can also inhibit the folding of the target protein and therefore the toxic activity of the protease. As explained above, toxicity was possibly the cause of selection of mutations in the esterase gene expressed in this work. One potential problem with expressing genes as fusions, however, is the need for cleavage to remove the fusion partner and the presence of additional amino acids after the fusion protein has been cleaved. Therefore, direct expression is still a desired route, and both expression systems are the basis for work reported in chapters five and six.
As already mentioned in the introduction (section 1.5.1.1), a multimeric protease from *Pyrococcus furiosus*, named pfpl (*Pyrococcus furiosus* protease I), has been characterised (Blumentals *et al.*, 1990; Halio *et al.*, 1997) and its encoding gene has been cloned and expressed in *Escherichia coli* (Halio *et al.*, 1996). It has been expressed in *E. coli* with a poly-histidine tag, which was subsequently not completely cleaved, and this was thought to be the cause of the recombinant protein having a much lower thermostability than the native enzyme.

The sequencing of the *Archaeoglobus fulgidus* genome (Klenk *et al.*, 1997) has revealed the presence of an open reading frame (AF1281) with high homology to pfpl (56 % identity, 76 % similarity), referred to subsequently as afpl (*Archaeoglobus fulgidus* protease I).

The aims of this chapter are to clone and express afpl in an active form that is as close to that of the native enzyme as possible. This involves the use of differing expression systems, firstly the pET vectors pET-3a and pET-11a used previously (see sections 3.1 and 4.1) and also the IMPACT system.

The pET-3a vector was chosen as it is a conventional, high-level expression vector, and pET-11a was used in addition as it has more stringent control over basal levels of expression, which is desirable as protease genes are potentially toxic to the host cells. An alternative method of expressing a toxic gene is to express it as a fusion protein. This has been successful not only in the case of pfpl, but also for example thermopsin from *Sulfolobus acidocaldarius* (Lin *et al.*, 1992).
A system was chosen named the IMPACT system, which utilises a vector containing the gene for an intein. Inteins in protein coding genes were first discovered in a DNA polymerase from *Thermococcus litoralis* (Perler et al., 1992). They are precursor proteins which self-cleave, yielding an active enzyme. Inteins have been utilised in a novel purification system that enables a recombinant protein to be purified in a single step (Chong et al., 1997, 1998). The vector used in this study (pTYB1) has an intein fused to the 5' end of a chitin binding domain. The gene of interest is cloned into the 5' end of the fusion and expressed; the fusion protein is then bound to a chitin column and self-cleavage is induced, releasing the recombinant protein (figure 5.1).

The IMPACT plasmids, like the pET system, are based on the T7 promoter which is induced by IPTG, and are transformed into *E. coli* BL21(DE3) cells. For cloning, the vector pTYB1 utilises a Sapl restriction site at the 3' end of the insert. This enzyme has the following recognition sequence, with / indicating the cleavage sites:

\[
\begin{align*}
5' & \text{GCTCTTCn/} & 3' \\
3' & \text{CGAGAAGnnnn/} & 5'
\end{align*}
\]

(where n denotes any nucleotide)

As the INTEIN and chitin binding domain fusion is at the 3' end of the gene, the PCR product must not contain a stop codon, which means that the Sapl recognition site will be the terminal sequence of that product. This means that the sequence contained within the Sapl recognition site will be translated in the protein. Thus, the protein translated will contain an amino acid translated from nGC (serine, arginine, glycine or cysteine from AGC, CGC, GCG

---

**Figure 5.1**: A diagram showing the basis of the IMPACT system.
or TGC, respectively) and two C-terminal serine residues (translated from the sequence TCT TCn). It was decided at this point to produce two constructs of afpI, the first containing the entire native protein sequence and the three additional amino acids from the incorporated Sapl site, and the second construct whereby the additional three amino acids replace two of the native C-terminal amino acids. In the first case, it was decided to have an additional glycine residue with the two serine residues, as it is a smaller amino acid than arginine or cysteine and not as hydrophobic as serine. One of the C-terminal residues in the native enzyme is already arginine and so in the second construct, this was taken as the residue encoded by the recognition sites nGC sequence, followed by two serine residues. The following figure summarises the constructs to be made:

\[
\begin{align*}
\text{Construct one} & : & \text{ELIKILKRSS} \\
\text{Construct two} & : & \text{ELIKILKRYGSS} \\
\text{Native sequence} & : & \text{ELIKILKRY}
\end{align*}
\]

Figure 5.2: Comparison of the C-terminal ends of the two afpI constructs with the native sequence.

As the overhanging sequence of the Sapl recognition site can be any amino acid, it must be chosen such that it is complementary to that in the vector, and not to the target gene sequence.

5.2 Methods

5.2.1 Amplification and Cloning of afpI

For the amplification of afpI, primers were designed based on the published sequence (TIGR; http://www.tigr.org). In order to clone the gene into pET-3a and pET-11a, it was necessary to introduce NdeI and BamHI recognition sites into the 5' and 3' ends of the gene, respectively; to clone the gene into pTYB1, it was necessary to introduce
NdeI and SapI recognition sites into the 5' and 3' ends of the gene, respectively, so the sequences were checked for internal sites.

The forward primer in all four cases had the base sequence 5' TTT TGA GCT GTT CAT ATG AGG GTG CTG ATT 3' (NdeI site underlined) and the reverse primer for the pET vectors had the base sequence 5' ATG CTG AGC CAC GGA TCC TTA AGG GTA GGG 3' (BamHI site underlined). The reverse primers for pTYB1 had the base sequence 5' GCA TTG GTG TTT GCA GGA AGA GCG CTT AAG GAT TTT GCA GGA AGA GCC GTA TCG CTT AAG GAT TTT G 3' (construct one) and 5' AGG GGG GCA TTG GCA GGA AGA GCC GTA TCG CTT AAG GAT TTT G 3' (construct two), Sapl sites underlined; see Appendix three.

PCR was performed as described in section 2.3.6 using Cloned Pfu and Vent DNA polymerases and the products were separated on a 1 % agarose gel. The appropriate resulting fragments were excised from the gel with a clean scalpel and purified using the Geneclean protocol. The DNA was then digested with either NdeI and BamHI at 37 °C for 3-6 h (see section 2.3.7) for ligation into the pET vectors, or NdeI and SapI at 37°C for 4 h for ligation into pTYB1, and purified from the reaction buffer and DNA fragments using the Geneclean protocol. The resulting DNA was ligated to pET-3a, pET-11a or pTYB1, which had been treated with the corresponding enzymes, by incubation with DNA rapid ligase at room temperature for 30 min to 1 h. The non-expression host E. coli JM109 cells (strain lacking the T7 RNA polymerase) were transformed to ampicillin resistance with the ligation reaction as described in section 2.3.3.1. The following day ampicillin resistant colonies were picked and grown in LB medium to enable plasmid purification as described in section 2.3.5.2. The eluted pET plasmids were screened by treatment with NdeI and BamHI at 37 °C for 3 h and the eluted pTYB1 plasmids were screened by treatment with NdeI and KpnI (as the SapI site is not recreated after cloning), and separation on a 1 % agarose gel. Plasmids showing the correctly-sized vector and insert bands were selected for expression studies.

Sequencing of the genes was performed as described in section 2.3.9 from the expression vectors using T7 promoter and termination primers for the pET vectors and T7 promoter and INTEIN reverse primers for pTYB1.
5.2.2 Expression of afpl in *E. coli*

Once the presence of the vector was confirmed by ampicillin resistant growth of transformants and the presence of a correctly-sized insert in the vectors by restriction analysis, the vectors were transferred to the expression hosts, *E. coli* BL21(DE3) for pET-3a and pTYB1, and BL21(DE3) pLysS for pET-11a, as described in section 2.3.3.2. The following day, single colonies were subcultured onto a fresh LB ampicillin plate thereby providing a ‘master plate’ to ensure a homogeneous population was used in subsequent experiments. Growth of transformants and induction of expression was performed as described in section 2.3.10. Growth and expression of the cells transformed with the pTYB1 vector was also attempted at 15 °C overnight and with 50 μM IPTG to try to induce expression of a soluble product. The cultures were harvested by centrifugation (2500 g), the supernatant was discarded and the cell extract (soluble fraction) and cell debris (insoluble fraction) were prepared by resuspending the cell pellets in 5 ml of 50 mM sodium phosphate, HEPES or MES buffer, pH 7.5-7.7, and lysing the cells by cell disruption or sonication as described in section 2.3.13. The cell debris was pelleted by centrifugation (8000 g), the cell extract (supernatant) was removed, and the cell debris was resuspended in 2 ml of 50 mM sodium phosphate, HEPES or MES buffer, pH 7.5-7.7. The cell extracts from the pET vectors were assayed for proteolytic activity against 0.1 % (w/v) azocasein and azaalbumin, and 80 μM ala-ala-pro-phe-pNA and ala-ala-phe-pNA, at 70 °C, as described in sections 2.3.14.1 and 2.3.14.2, respectively, using the above buffers. *Archaeoglobus fulgidus* cell extract was also assayed for activity against these substrates. Cell extract was prepared by resuspending 1 g of frozen cell paste in 4 ml of 20 mM sodium phosphate buffer, pH 7.6, lysing them by cell disruption as described in section 2.3.13 and pelleting the cell debris (8000 g, 20 min).

A sample of the cell extract from pET-3a was treated with TCA prior to visualising on SDS PAGE. This was to ensure the complete dissociation of the monomers, which did not occur with pfpl after heating to 95 °C in the presence of 2 % (w/v) SDS (Halio *et al.*, 1996). To a sample of cell extract, a half volume of ice cold 15 % (w/v) TCA was added. The mixture was incubated on ice for 20 min and
the precipitate centrifuged (13000 g, 20 min) at 4 °C. The supernatant was discarded and the pellet was washed with 50 µl of ice cold acetone and the precipitate centrifuged (13000 g, 2 min) at 4 °C. The supernatant was discarded and the pellet was left to air dry for 5 min. The pellet was then resuspended in SDS loading buffer and visualised by SDS PAGE.

The recombinant enzyme obtained from pET-3a was also assayed for activity using gelatine substrate PAGE and substrate overlay gels. Substrate PAGE was performed as SDS PAGE (see section 2.3.12), with the addition that 0.5 % (w/v) gelatin was copolymerised with the polyacrylamide. For the substrate overlay gel, SDS PAGE was performed as before (section 2.3.12). The gel was then washed in 50 mM sodium phosphate buffer, pH 7.5, 2.5 % (v/v) Triton X-100 for 1 h at room temperature to remove the SDS and a second polyacrylamide gel containing 0.5 % (w/v) gelatin was placed on top. The gels were immersed in 50 mM sodium phosphate buffer, pH 7.5, weighted down and incubated at 75 °C overnight. The gelatin containing gel was stained and destained as before.

The cell debris from the pTYB1 vector expression was washed twice with 5 ml of 20 mM HEPES buffer, pH 8.0, containing 500 mM NaCl, 1 mM EDTA and resuspended in the above buffer supplemented with 50 mM DTT and 10 % (w/v) SDS. The cell debris was then solubilised by incubating it overnight at room temperature. The following day, it was incubated at 37 °C for 30 min to dissolve any precipitated SDS and then added slowly to a renaturation buffer (50 mM EPPS buffer, pH 8.0, 1 mM EDTA, 0.3 mM oxidised glutathione, 3 mM reduced glutathione) to a final concentration of 10 µg ml⁻¹ and incubated at 4 °C overnight. The following day, the renatured protein was loaded onto a chitin column as described below.

5.2.3 Chitin Column Chromatography

Poly-prep chromatography columns (0.8 x 4 cm) were used for the affinity purification. The chitin beads were supplied in 40 % (v/v) ethanol. Routinely, 4 ml of chitin bead slurry was used to prepare a column with a final bed volume of 2 ml.
The column was washed with 3 bed volumes of water and then equilibrated with 10 bed volumes of 20 mM HEPES buffer, pH 8.0, containing 500 mM NaCl, 1 mM EDTA. Up to 2 mg of enzyme sample was loaded onto the column, which was then washed with 10 bed volumes of 20 mM HEPES buffer, pH 8.0, containing 500 mM NaCl, 1 mM EDTA. The cleavage reaction was induced by loading 3 bed volumes of 20 mM HEPES buffer, pH 8.0, containing 500 mM NaCl, 1 mM EDTA plus 50 mM DTT, cysteine or β-mercaptoethanol and incubating overnight at 4, 16 or 23 °C. The target protein was eluted by adding 3 bed volumes of 20 mM HEPES buffer, pH 8.0, containing 500 mM NaCl, 1 mM EDTA and assayed for activity against azocasein, azoalbumin and ala-ala-phe-pNA. The column was regenerated by washing with 3 bed volumes of 1 % (w/v) SDS (to remove the chitin/INTEIN precursor), followed by 5 bed volumes of water. If the column was needed immediately, it was equilibrated with 10 bed volumes of 20 mM HEPES buffer, pH 8.0, containing 500 mM NaCl, 1 mM EDTA, or if not it was washed with 3 bed volumes of 20 % (v/v) ethanol and stored at 4 °C.

5.3 Results

5.3.1 Amplification and Cloning of afpI

Following PCR amplification, bands were obtained of the expected size for all four of the constructs (568 bp for the pET constructs, 536 bp for the pTYB1 construct number one and 542 bp for the pTYB1 construct two) that were distinct from the single primer controls. Figure 5.3 overleaf shows samples of the PCR products run on 1 % agarose gels. The arrows indicate the bands of interest.

Following restriction digestion, ligation into pET-3a, pET-11a and pTYB1, and transformation into *E. coli* JM109 cells, numerous colonies were picked and screened for the presence of an appropriately-sized insert. The efficiency of transformation, however, was noted to be approximately 100x lower than that of the control plasmid. The appropriate restriction patterns were observed for each of the
constructs and figure 5.4 below shows the digests run on 1% agarose gels. The arrows indicate the genes cut out of the plasmid.

![Figure 5.3: Agarose gels showing PCR amplifications of afpl.](image)

**Figure 5.3:** Agarose gels showing PCR amplifications of afpl.

a) Lane 1, 1 kbp marker (arrow indicates 1636 bp); lane 2, PCR amplification for pET vectors; lane 3, forward primer control; lane 4, reverse primer control.

b) Lane 1 & 7, 1 kbp marker (arrow indicates 1018 bp); lane 2, PCR amplification of construct one for pTYB1 vector; lane 3, PCR amplification of construct two for pTYB1 vector; lane 4, forward primer control; lane 5, construct one reverse primer control; lane 6, construct two reverse primer control.

![Figure 5.4: Agarose gels showing restriction digestion of transformant plasmids.](image)

**Figure 5.4:** Agarose gels showing restriction digestion of transformant plasmids.

- a) Lane 1, 1 kbp marker (arrow indicates 517 bp); lanes 2-5, transformants a-d from pET-3a.
- b) Lanes 1, 1 kbp marker; lanes 2 & 3, transformants e & f from pET-11a.
- c) Lanes 1 & 8, 1 kbp marker; lanes 2-7, transformants g-l from pTYB1.

All three of the above agarose gels show the expected characteristic band pattern consisting of the linear plasmid and the excised gene for each construct. Transformants d (afpl in pET-3a), e (afpl in pET-11a), h (construct one in pTYB1) and i (construct two in pTYB1) were chosen for expression studies.
5.3.2 Expression of afpl from pET vectors in *E. coli*

The pET vectors containing the appropriate construct were transformed successfully into the expression hosts, *E. coli* BL21(DE3) and BL21(DE3) pLysS. Expression of the genes was measured via SDS PAGE analysis of whole cell pellet samples as well as the soluble and insoluble fractions. The following figure shows the separation of the proteins by SDS PAGE and expression levels of the gene. The arrows indicate the putative bands of interest.

As can be seen from figure 5.5, significant expression was achieved from both pET-3a and pET-11a. The levels of expression were lower from pET-11a; this is probably due to the additional repression of transcription that this vector has, over pET-3a, even though induction was for longer from pET-11a. The additional repression is clear from a comparison of the uninduced cell pellet from the pET-11a expression (gel c, lane 2) with that from the pET-3a expression (gel a, lane 3). The
cell pellet from the pET-3a expression contains bands corresponding to putative afpl present in both the uninduced and induced samples, whereas the cell pellet from the pET-11a expression only contains visible bands in the induced samples.

After SDS PAGE, afpl appears in two forms, one corresponding to approximately 80 k and the other corresponding to approximately 22 k. After TCA precipitation, only the 22 k band remains (gel b, lane 4).

Despite high levels of recombinant protein present in the cell extracts, no activity was observed against azocasein, azoalbumin, ala-ala-pro-phe-pNA, or ala-ala-phe-pNA using cell extracts from pET-3a or pET-11a. Furthermore, activity was not observed on substrate PAGE or substrate overlay gels using pET-3a cell extracts.

*Archaeoglobus fulgidus* cell extract, however, was found to be active against both of the peptide substrates, ala-ala-pro-phe-pNA and ala-ala-phe-pNA and also azocasein, but not azoalbumin. When 500 μg of cell extract was incubated at 80 °C, the reaction against ala-ala-pro-phe-pNA went to completion after 2 min and the reaction against ala-ala-phe-pNA went to completion after 1.5 min. When 1 mg of CE was incubated at 70 °C for 1 h in the presence of 0.1 % (w/v) azocasein, the A410 value obtained was 0.48.

Sequencing of the two genes cloned into pET-3a and pET-11a revealed the presence of a single mutation in each. The gene cloned into pET-3a has a base mutation of adenine (A) to thymine (T), resulting in an altered codon of GAG to GTG and an amino acid mutation of glutamic acid to valine at position 84 (E84V). The gene cloned into pET-11a has a base mutation of thymine (T) to cytosine (C), resulting in an altered codon of GTG to GCG and an amino acid mutation of valine to alanine at position 41 (V41A).

### 5.3.3 Expression of afpl from pTYB1 vector in *E. coli*

Expression of the two gene constructs was measured via SDS PAGE analysis of whole cell pellet samples as well as the soluble and insoluble fractions. Figure 5.6 below shows the separation of the proteins by SDS PAGE and expression levels of the genes. The arrows indicate the putative bands of interest.
Figure 5.6: SDS gels showing expression of afpl from pTYB1.

a) Lane 1, broad range marker; lane 2, construct two, uninduced cell pellet; lane 3, construct two, induced cell pellet; lane 4, construct one, uninduced cell pellet; lane 5, construct one, induced cell pellet.

b) Lane 1, broad range marker; lane 2, construct two, cell extract; lane 3, construct two, cell debris; lane 4, construct one, cell extract; lane 5, construct one, cell debris.

As the SDS gels in figure 5.6 show, both constructs of afpl were expressed from pTYB1 in an insoluble form and so attempts were made to resolubilise them. This was carried out as described in section 5.2.2 with SDS, and also urea. When the urea-denatured samples were added to the renaturation buffer, the protein precipitated immediately. However, the SDS-denatured sample did not show any obvious precipitation when it was diluted into the renaturation buffer and so, after removing any invisible precipitated protein by centrifugation (8000 g, 20 min), a sample of the renaturation buffer was visualised on a SDS gel, shown in figure 5.7 below, and the remainder was loaded onto a chitin column.

Figure 5.7: SDS gel showing renaturation of afpl.

Lane 1, broad range marker; lane 2, construct one; lane 3, construct two.
Following overnight incubation of the column under various conditions designed to induce cleavage of the intein and release of the target protein, the elute was examined by SDS PAGE and assayed for activity. Neither a band of the expected size, nor activity against azocasein, azoalbumin or ala-ala-phe-pNA, was observed.

To check that the protein had in fact undergone cleavage from the column, the column was stripped with SDS and the elute was visualised by SDS PAGE. The following figure shows the resulting SDS gel and the arrow indicates the putative fusion protein.

![SDS gel showing the protein obtained after stripping the chitin column.](image)

Lane 1, broad range marker; lane 2, construct two; lane 3, construct one.

It is clear from this that the afpl protein has not been cleaved from the precursor, as the 74 k fusion protein is still present on the column. Afpl is therefore not going to be present in the eluate of the column after the overnight incubation. This was case after numerous attempts of cleavage under varying conditions. The two lower-sized bands are probably due to non-specific binding to the chitin column, as they do not represent the 55 k intein-chitin binding domain fusion.

Sequencing of the two constructs revealed the presence of mutations in the coding sequences of the genes. Construct one contains six base mutations, three of which do not result in a change to the amino acid sequence (silent mutations). Of the remaining three, the first mutation is one from cytosine to adenine, resulting in an altered codon of CTC to ATC, producing an amino acid mutation of leucine to isoleucine at position 37 (L37I). The second mutation is again cytosine to adenine, resulting in an altered codon of CCT to CAT, producing an amino acid mutation of
proline to histidine at position 73 (P73H). The final mutation is guanine to adenine, resulting in an altered codon of ATG to ATA, producing an amino acid mutation of methionine to isoleucine at position 150 (M150I). Construct two contains one base mutation, guanine to thymine, producing an amino acid mutation of alanine to serine at position 130 (A130S).

5.4 Discussion

This work has been successful in achieving both the direct expression of afpl from the pET vectors and also the expression of afpl as a fusion in the IMPACT system. A number of problems have arisen, however. The first of these is that the final goal was not achieved, namely, the expression of an active protease in E. coli. However, the reasons behind this should enable an expression scheme that does produce active protein to be created. Considering first the expression from the pET vectors, it is clear that a protein has been expressed, and it displays similar properties on SDS PAGE to pfpl. For example, even though afpl would be expected to have a lower thermo- and chemical stability than pfpl, coming from an organism with a growth temperature optimum of around 15 °C lower, it still displays resistance to SDS in that its multimeric form is not completely dissociated during SDS PAGE. TCA was required to dissociate the multimer completely, although even then it may be that complete unfolding has not occurred, as the monomer displays an apparent relative molecular mass of 22 k, which is higher than the 19 k predicted from the translated gene sequence. From SDS PAGE, it is apparent that the multimeric form of afpl is a tetramer. There is no evidence from this work to support the presence of either a hexameric or trimeric species, analogous to the active forms of pfpl.

Sequencing of the cloned gene has revealed that the expressed products are indeed afpl, and not artefacts, although the recombinant proteins do contain mutations. This is the significant factor in this work, as it appears that these mutations have been selected during the transformation process. It is highly unlikely that two genes containing mutations were selected by chance out of the many 'native' sequences, in which case all that is required is more colonies to be picked after
transformation. This, therefore, is believed not to be the case as PCR conditions were chosen to reduce the possibility of errors. Firstly, a high fidelity polymerase was chosen and secondly, the magnesium concentration in the reaction buffer was kept low. Cloned *Pfu* polymerase has been reported to have an error rate of $1.3 \times 10^{-6}$ (Cline *et al.*, 1996), which is defined as the number of mutations per nucleotide. Vent polymerase is reported to have an error rate of $57 \times 10^{-6}$ (Mattila *et al.*, 1991). Neither of these polymerases, therefore, would be expected to produce the number of mutations seen in this work under normal circumstances, especially as the gene to be amplified is quite short.

Proteases are known to exhibit toxic effects on the host cells expressing them (e.g. Choi *et al.*, 1999; Lee *et al.*, 1998; Saul *et al.*, 1996; Terada *et al.*, 1990). Hailo *et al.* (1996) reported selection of vectors having deletions in the *pfpl* gene when they tried expressing it as a nonfusion protein. Therefore, tight control is required over basal transcription to ensure expression only occurs after induction. The pET system requires the vector to be initially transformed into a host (JM109), which does not contain T7 RNA polymerase, to ensure the plasmid can be stably maintained. Even once the vector is transformed into the expression host [BL21(DE3) or BL21(DE3) pLysS], it is claimed that the regulation is such that there is virtually no transcription of the target gene in the absence of a source of T7 RNA polymerase (Mierendorf *et al.*, 1994). It appears, however, that during the transformation process of the pET vectors containing *afpl*, a significant level of transcription and translation has occurred such that any host cells containing a copy of a ‘native’ *afpl* gene have died from the toxic effects of a heterologous protease, and so the only host cells to survive are those containing an inactivating mutation within the *afpl* gene. The mutations are thought to have arisen during PCR, as even the highest fidelity enzymes available still have a measurable error rate, albeit very low. This is probably the reason behind the apparent low transformation efficiency of the pET vectors containing the *afpl* gene.

As the genes were sequenced from the pET vectors prepared from *E. coli* JM109 cells, it is clear that the mutations have been selected before the vectors have been transformed into the expression host and so some degree of transcription must have occurred in this host. As already mentioned, JM109 cells do not contain a copy of T7 RNA polymerase and so in theory, no transcription should occur from the T7
promoter. However, any growing cell must contain a source of RNA polymerase, and therefore it appears that the JM109 host RNA polymerase has sufficient activity on the T7 promoter to allow a low, but still toxic, level of transcription. Therefore, even though the pET-11a vector has exhibited a higher level of control over basal transcription in the expression host cells, this expression system has still produced an inactive enzyme. As a side point, the expression from pET-3a appears to have been poorly controlled, as the level of afpI is similar in the uninduced and induced samples. This is thought to have been due to the fact that one of the components of the medium used to grow the cells, yeast extract, contains a homologue to the inducing agent, IPTG (Duncan Clark, GeneSys Ltd.; personal communication). Therefore, tighter control over basal expression would be expected in medium not containing yeast extract.

Strategies to produce an active protease using this system must involve either preventing transcription in the non-expression host, or reducing the toxic activity of the recombinant protease once it is expressed. Stopping basal transcription in the JM109 cells would appear not to be possible, as the transcription is due to the action of the host’s own RNA polymerase. One way, however, of reducing the activity of the enzyme would be to lower the temperature in all stages of the transformation and expression processes. However, afpI appears to be toxic at 37 °C, which is 46 °C below the optimum growth temperature of *Archaeoglobus fulgidus*. This corresponds to well below 10 % of its maximum activity, and therefore reducing the temperature another 10-20 °C would not be expected to produce a significant difference. Transformation of *E. coli* JM109 with pET-11a containing the afpI gene, and growth on a minimal medium not containing yeast extract at 4 °C, was attempted but, as expected, the cells failed to grow under these conditions (results not shown).

If the level of basal transcription in the expression host could be inhibited until induction, given a high enough cell density, a significant level of recombinant protein might be obtained before the cells lysed. This would, however, involve preventing the toxic levels of transcription in the non-expression host first. One solution to that problem is to attempt transformation straight into the expression host.

It appears then, that the expression system as it stands is not a suitable one for the production of recombinant proteases. Halio *et al.* (1996) expressed pfpl in *E. coli* from another pET vector, pET-15b, which contains a poly-histidine fusion tag.
This route was not chosen for this work initially as the recombinant protein contained at least three histidine residues that were not cleaved, producing a less active and less thermostable recombinant protein than the native. However, this method has at least produced an active recombinant protease.

The IMPACT system, utilising a fusion protein, also failed to produce an active protease. The main factor was the failure of the intein precursor to undergo self-cleavage, releasing the target protein. Whilst it is not possible to assess the reason for this definitively, it is most likely due to SDS binding to the protein and preventing the required protein splicing and cleavage reaction. SDS was used as the denaturing agent when the recombinant fusion protein was expressed as insoluble inclusion bodies, as the samples denatured with urea immediately precipitated when diluted into a renaturation buffer. The chemical denaturant, guanidine hydrochloride was not used; however, this may have been more successful in the attempts to renature the inclusion bodies.

Triton X-100 was used during substrate PAGE and substrate overlay gel assays at a concentration of 2.5 % (v/v), to remove the SDS in the gels. This was not attempted in the column chromatography and may have proven successful in removing the bound SDS on the recombinant protein, although concentrations above 0.5 % (v/v) will inhibit binding to the chitin column (Inca Ghosh, NEB; personal communication).

A further problem encountered in this work was the presence of mutations in the recombinant protein. The target gene is the only portion of the entire fusion that arises from PCR, as the intein-chitin binding domain is vector-encoded, and therefore it is the only part where mutations would be expected. However, as the sequencing reactions only covered the target gene, it is unknown whether the remainder of the fusion contained any mutations, which could explain the inhibition of the cleavage reaction. The reasons behind the high rate of errors in the two constructs produced are unknown. One possibility is that the gene was, like in the pET vectors, toxic to the host cells and so only the vectors with inactivating mutations in the genes were selected. This would be unlikely, as the presence of the intein fusion protein should prevent folding of the recombinant protein into an active conformation; furthermore, pfpl, and therefore by implication afpl, is not active as a monomer, instead requiring assembly into a multimeric form for activity. The
conditions for PCR were chosen, as for the pET vectors, to reduce such errors. If the mutations were due to toxic activity, it therefore follows that had the cleavage reaction occurred, the enzyme would have been inactive. If the mutations are due to another factor, had the protein undergone self-cleavage, it is still unknown whether these mutations would have had a deleterious effect on activity.

There is an alternative method for expressing genes that is particularly applicable to toxic genes, namely to secrete the protein into the culture medium. The gene must still initially be cloned and transformed into the expression host though and so it must be ensured that the resulting protein does not exhibit any toxic activity whilst it is in the cytoplasm. One way to achieve this is to clone the gene initially in the reverse orientation in a non-expression vector, so that even if any transcription occurs, it will not result in a proteolytically-active product. Once the gene has been sequenced, and a ‘native’ sequence cloned, the gene is then cloned in the correct orientation in an expression vector. This idea is the basis for work reported in the next chapter.
6. Chapter Six

Cloning and Expression of an Extracellular Protease from Aeropyrum pernix

6.1 Introduction

Aeropyrum pernix K1 is the first strictly aerobic, hyperthermophilic Archaeon described to date (Sako et al., 1996). It grows optimally at 90-95 °C, with a doubling time of 200 min. It is heterotrophic, utilising such proteinaceous compounds as tryptone and yeast extract, which supply most of the carbon, energy and nitrogen demands. As such, it would be expected to produce extracellular proteases, and at this time, one such enzyme termed aeropyrolysin has been purified and characterised (Sako et al., 1997), see section 1.5.1.3. To date, there has only been one further report describing proteolytic activity in A. pernix (Croocker et al., 1999), that of the intracellular protease, pemilase (see section 1.5.1.3). The genome of A. pernix K1 has been sequenced (Kawarabayasi et al., 1999) and from this an ORF has been identified (APE0263), encoding an extracellular protease.

Initially, an alternative protease from Archaeoglobus fulgidus was considered as a suitable target protein (AF1653) for expression. However, sequence alignment against halolysin R4 from Haloferax mediterranei, aerolysin from Pyrobaculum aerophilum and a putative subtilisin-like extracellular protease from A. pernix, indicated the ORF identified was in fact too small to be a protease (see Appendix 4). Furthermore, sequence alignment against subtilisin E from Bacillus subtilis has suggested that the active site serine residue is absent from AF1653 (see Appendix 5). The ORF encoding the subtilisin-like extracellular protease from A. pernix does contain the serine protease motif (see Appendix 4) and was therefore chosen for this study. As APE0263 is a putative protease and has not been previously reported, the sequences of the pre- and pro-regions are unknown at this stage. The pre- and pro-
regions were identified by sequence alignment, using Clustalw, of subtilisin E from B. subtilis against APE0263. Figure 6.1 below shows the resulting alignment.

The residues at the C-terminal of the pre-region are highly homologous, making identification of the pre-region in APE0263 simple, and even though the residues at the C-terminal of the pro-region are not so highly conserved, the resulting residues at the N-terminal of the mature protein are more so, clearly identifying the pro-region. As discussed in section 3.2.1, Watson (1984) found that 98 % of prokaryotic signal sequences examined contained either an alanine, glycine, serine, cysteine or threonine residue at the C-terminal end. This is found to be the case for both subtilisin E and APE0263, based on the above alignment.

The previous work has suffered from problems with the selection of non-lethal mutations arising from the toxicity of the native protease. This is due to the fact that the expression system used was leaky, i.e. there was some degree of uninduced transcription and translation. One way to guarantee no expression of a proteolytically active species is to clone the gene in the reverse orientation initially. The gene can then be sequenced to ensure it is free from errors and then cloned in the correct orientation into the expression vector.

Furthermore, an expression system was chosen that would enable the recombinant protein to be secreted to the extracellular medium, primarily to avoid
the toxicity problems associated with afpl, but also to facilitate purification. As discussed in section 3.4, *Escherichia coli* is generally not considered a particularly good host for secretion purposes. However, *Bacillus* species have long been used industrially for the secretion of proteins and a large number of both homologous and heterologous proteins have been produced in *Bacillus* (reviewed by e.g. Simonen and Palva, 1993).

There were two vectors chosen for this work, the general cloning vector pUC18 into which the target gene would be cloned in the reverse orientation for sequencing purposes, and the expression vector pBE3, an *E. coli/B. subtilis* shuttle vector containing the gene for subtilisin E (Zhao et al., 1997). The mature region of the subtilisin-encoding gene is removed and the mature region of the target gene is inserted, producing a fusion at the 3' end of the subtilisin pre-pro-region. Figure 6.2 below shows the organisation of the fusion.

![Diagram](image)

**Figure 6.2:** Organisation of the subtilisin/APE0263 gene fusion in pBE3.

Consequently, during expression, the target gene is exported through the cell membrane to the culture medium. The vector is transformed into *B. subtilis* DB428 cells, a strain deficient in the extracellular protease subtilisin. The C-terminal 11 amino acids of the subtilisin pro-region are replaced by those from the APE0263 pro-region, to ensure that if cleavage of the pro-region by autolysis is sequence-specific, APE0263 will have its own recognition sequence.
6.2 Methods

6.2.1 Preparation of *A. pernix* genomic DNA

The method used was a modification of that reported by Yamano *et al.* (1999). 0.5 g of frozen *A. pernix* cell paste was thawed and resuspended in 9.5 ml lysis buffer [50 mM Tris-HCl buffer, pH 8.0, containing 150 mM NaCl, 100 mM EDTA, 2 % (w/v) SDS, 0.1 mg ml\(^{-1}\) proteinase K] and incubated at 65 °C for 30 min. The lysate was extracted (with equal volumes), once with phenol, once with phenyl/chloroform (1:1) and then once with chloroform/isoamyl alcohol (24:1). A 0.8 volume of ice-cold isopropanol was added and the DNA was collected with a glass loop, washed with 70 % (v/v) ethanol and resuspended in 2 ml TE buffer (10 mM Tris-HCl buffer, pH 8.0, 1 mM EDTA). The DNA solution was incubated with 0.1 mg ml\(^{-1}\) RNase A at 37 °C for 3.5 h, and then with 0.1 mg ml\(^{-1}\) proteinase K and 0.5 % (w/v) SDS at 50°C for 90 min. Next, 5 M NaCl and 10 % (w/v) hexadecyltrimethylammonium bromide (CTAB) in 0.7 % (w/v) NaCl were added to give final concentrations of NaCl and CTAB of 0.7 M and 1 % (w/v), respectively. After incubation at 65 °C for 20 min, the DNA was extracted once with chloroform/isoamyl alcohol (24:1), followed by four extractions with phenol/chloroform (1:1). A 0.8 volume of ice-cold isopropanol was added and the DNA was collected with a glass loop, washed with 70 % (v/v) ethanol and resuspended in 1 ml TE buffer.

The purity of the DNA was calculated by measuring the absorbance at 260 and 280 nm, a value of 1.8 for the \(A_{260}:A_{280}\) ratio indicating pure DNA. The quantity of DNA was estimated by measuring the absorbance at 260 nm, with an \(A_{260}\) of 1.0 approximating to 50 µg ml\(^{-1}\) double-stranded DNA. The quality of DNA was examined by running a 15 µl sample of the DNA on a 0.8 % (w/v) agarose gel.

6.2.2 Amplification and Cloning of APE0263

For the amplification of APE0263, primers were designed based on the published sequence (Biotechnology Center, National Institute of Technology and Evaluation,
In order to clone the gene into both pUC18 and pBE3, it was necessary to introduce \textit{NdeI} and \textit{BamHI} recognition sites into the 5' and 3' ends of the gene, respectively, and so the sequence was checked for internal sites. APE0263 was found to contain an internal \textit{BamHI} site and so an initial PCR was required to alter the recognition site, but not the amino acid sequence. The sequence GGG GAT CCA (\textit{BamHI} site underlined) encoding Gly-Asp-Pro was mutated to GGA GAC CCA (mutated bases underlined), still encoding Gly-Asp-Pro. The basis of the mutagenesis is similar to that used in insertional mutagenesis (Kammann \textit{et al.}, 1989; Good and Nazar, 1992) whereby the desired mutation is incorporated into one PCR primer and a portion of the target sequence is amplified using a second, flanking primer. The product from that reaction is then used as a ‘megaprimer’ in a second PCR, with another flanking primer from the opposite end of the target sequence.

For the mutagenic amplification, the forward primer had the base sequence 5' GTC GTC GCT GGA GAC CCA GAC GAT 3' and the reverse primer had the base sequence 5' GGA GAA GTA GTT GGA TCC AGG TAG GCC 3' (\textit{BamHI} site underlined). For the second round PCR, the forward primer had the base sequence 5' TAC CTG GTG TTG CAT ATG TTT CCG AGG ACG 3' (\textit{NdeI} site underlined) and the product from the first round PCR was used as the reverse primer; see Appendix six.

The first round (mutagenic) PCR was performed as described in section 2.3.6 using Vent DNA polymerase and \textit{A. pernix} genomic DNA as template, and the products were separated on a 1 % agarose gel. The appropriate resulting fragment was excised from the gel with a clean scalpel and purified using the Geneclean protocol. To provide a template for the second round PCR, a separate PCR was performed as above using the two flanking primers (those containing \textit{NdeI} and \textit{BamHI} restriction sites) to amplify the APE0263 gene containing the unwanted \textit{BamHI} site. This would reduce the possibility of obtaining non-specific products from genomic DNA in the second round PCR. The appropriate resulting fragment was purified as above. The second round PCR was performed as above using Vent polymerase, with the exceptions that the annealing temperature was reduced to 40 °C and the annealing time was increased to 15 min, to give the ‘mega-primer’ sufficient opportunity to anneal. In addition, the amount of template was reduced as the
product and template will be identical in size, but only the product, lacking the internal BamHI site is required. The appropriate resulting fragment was purified as above.

The DNA was then digested with NdeI and BamHI at 37 °C for 4 h (see section 2.3.7) and purified from the reaction buffer and DNA fragments using the Geneclean protocol. The resulting DNA was ligated to pUC18, which had been treated with the same enzymes, by incubation with DNA rapid ligase for 2 h at room temperature. The non-expression host E. coli JM109 cells were transformed to ampicillin resistance with the ligation reaction as described in section 2.3.3.1. The following day ampicillin resistant colonies were picked and grown in LB medium to enable plasmid purification as described in section 2.3.5.2. The eluted plasmids were screened by treatment with NdeI and BamHI at 37 °C for 3 h and separation on a 1% agarose gel. Plasmids showing the correctly-sized vector and insert bands were sequenced to check for a ‘native’ gene. Sequencing of the genes was performed as described in section 2.3.9 using M13 reverse primers, with the addition that 5% (v/v) DMSO was added to the reaction. The M13 forward sequencing primer could not be used in this instance, as the binding site had been removed in the NdeI/BamHI digestion.

Once the presence of a gene with an authentic sequence was confirmed, it was cut from the pUC18 vector and ligated in the correct orientation into pBE3, which had been treated with NdeI and BamHI to remove the mature region of the subtilisin gene, by incubation with DNA rapid ligase for 2 h at room temperature. B. subtilis DB428 cells were transformed to kanamycin resistance as described in section 2.3.3.3. The following day kanamycin resistant colonies were picked and streaked onto a ‘master plate’ and also grown in NB medium to enable plasmid purification as described in section 2.3.5.3. Plasmids showing the correctly-sized vector and insert bands were selected for expression studies.

6.2.3 Expression of APE0263 in B. subtilis

Once the presence of the vector was confirmed by kanamycin resistant growth of transformants in B. subtilis and the presence of a correctly-sized insert in the vector
by restriction analysis, the master plate corresponding to the verified colony was used for expression studies.

A single colony was inoculated into 10 ml SG medium (see section 2.3.1.2) supplemented with 5 μg ml\(^{-1}\) kanamycin and incubated at 37 °C with shaking, overnight. The following day, 1-10 ml of the culture was used to inoculate 50 ml-1 l fresh SG medium supplemented with 5 μg ml\(^{-1}\) kanamycin, which was incubated at 37 °C with shaking for 24-36 h. The cells were harvested by centrifugation (5000 g, 20 min) and the supernatant was removed and filtered (0.65 μm). Ammonium sulphate was then added to the supernatant to a final concentration of 100 % (w/v), which was constantly stirred for a further 1 h at room temperature. The precipitated protein was centrifuged (5000 g, 30 min) and the supernatant discarded. The pellet was resuspended in a minimum volume of 50 mM EPPS buffer, pH 8.0 and then subjected to gel filtration as described below. Protein samples were visualised by SDS PAGE as described in section 2.3.12. Following inhibition studies, later SDS PAGE samples were prepared containing 1 mM EDTA in the standard loading buffer.

### 6.2.4 Assaying for Activity of APE0263

Activity assays were routinely carried out at 80 °C as described in section 2.3.14.2 against the peptide substrate ala-ala-pro-phe-pNA, in 50 mM EPPS buffer, pH 8.0. Absorbance measurements of the product, p-nitroaniline, were carried out at varying pH values to determine the extinction coefficient under the standard assay conditions. One unit of activity for this assay is defined as that required to produce 1μmol of product per minute. The specific activity is defined as the number of units per mg of protein. Enzyme kinetics were analysed using Enzpack for Windows v1.4 software (BIOSOFT).
6.2.5 Gel Filtration

Gel filtration was carried out using an HR 16/60 Superdex 200 column pre-equilibrated with 50 mM EPPS buffer, pH 8.0, containing 100 mM NaCl. Samples were centrifuged (13000 g, 10 min) and the supernatant filtered (0.22 μm) and then loaded onto the column at a flow rate of 1 ml min⁻¹. 1.5 ml fractions were collected, assayed for activity as described above and the active fractions were pooled. This pooled sample was used in all the following experiments. The column was calibrated using horse heart cytochrome c (12.4 k), bovine carbonic anhydrase (29 k), BSA (66 k), yeast alcohol dehydrogenase (150 k), and blue dextran (2000 k) standards.

6.2.6 Temperature Profile

Assays to determine the apparent temperature optimum were carried out in 50 mM EPPS buffer, against ala-ala-pro-phe-pNA as described in section 2.3.14.2, with the pH adjusted to account for the pKₐ shift, giving an assay pH of 8.0. 2 μg of enzyme sample was assayed at 20-90°C.

6.2.7 pH Profile

Assays to determine the pH optimum used 2-10 μg enzyme sample against ala-ala-pro-phe-pNA as described in section 2.3.14.2 and were carried out at 80 °C in 50mM MES (pH 5.5-6.5), MOPS (pH 6.5-7.5), EPPS (pH 7.5-8.5), Glycine (pH 8.5-10.0) or CAPS (pH 10.0-10.5) buffers, taking care to use each buffer within its respective pKₐ range. When changing from one buffer to the next, an overlapping pH value was used, to ensure any difference in activity was solely due to the pH change and not any ionic effects by the buffer.
6.2.8 Thermal Inactivation Studies

Thermostability studies were performed by preheating 1.65 ml of 50 mM EPPS buffer, adjusted to pH 8.0 at the desired temperature, in a sealed tube and then adding 550 µl of enzyme sample (110 µg) to the heated buffer and mixing. At specific time points, a 120 µl sample was removed and immediately cooled on ice. The control, or time-zero point, was performed by duplicating the dilution of the enzyme, but heating neither the buffer nor the enzyme and removing a sample immediately. 40 µl enzyme samples (8 µg) were then assayed against ala-ala-pro-phe-pNA at 80 °C as described above.

6.2.9 Substrate Specificity and Effect of Inhibitors

The alternative substrates, ala-ala-phe-pNA, ala-ala-pro-leu-pNA and gly-gly-phe-pNA, were also examined for susceptibility to the enzyme. Assays were carried out at 80 °C in 50 mM EPPS buffer, pH 8.0, using 10 µg of enzyme sample. The effect of the inhibitors PMSF, PCMB and pepstatin was tested by incubating them with 40 µg samples of the enzyme at room temperature for at least 15 min and then assaying 4 µg of enzyme sample against ala-ala-pro-phe-pNA at 80 °C as described above. Furthermore, the effect of divalent cations after the addition of EDTA was investigated by preincubating 100 µg of enzyme samples with 1 mM EDTA for at least 15 min, and then adding 10 mM ZnSO₄, MnCl₂, CaCl₂ or MgSO₄ and assaying 10 µg enzyme samples against ala-ala-pro-phe-pNA as above. The effect of the divalent cations was also investigated without the preincubation with EDTA by incubating them with 40 µg samples of the enzyme at room temperature for at least 15 min and then assaying 4 µg of enzyme sample against ala-ala-pro-phe-pNA at 80 °C as described above.
6.3 Results

6.3.1 Preparation of *A. pernix* genomic DNA

Figure 6.3 below shows a sample of the extracted genomic DNA from *A. pernix* run on a 0.8 % (w/v) agarose gel.

It is apparent from the gel that high molecular weight DNA has been extracted from the *A. pernix* cells, although it looks to have been significantly sheared in the process. The amount of shearing seen would perhaps be reduced, however, by resuspending the cell paste in a buffer lacking the SDS. It is probable that during the resuspension of the cells, the genomic DNA released from the cells that had already been lysed would be very susceptible to shearing. Once the cells had been completely resuspended, the SDS could be added. There is no obvious contamination by RNA. The $A_{260}:A_{280}$ ratio of the DNA sample is 1.8, indicating pure DNA and the $A_{260}$ value is 0.985, indicating a concentration of approximately 50 ng μl⁻¹ double stranded DNA.

![Figure 6.3: Genomic DNA extracted from *Aeropyrum pernix* cells.](image)

Lane 1, 1 kb marker; lane 2, genomic DNA

6.3.2 Amplification and Cloning of APE0263

Following PCR amplification, using the mutagenic primer to remove the internal *Bam*HI site, a product was obtained of the expected size (599 bp) that was distinct
from the single primer controls. The following figure shows the first round PCR products run on a 1 % agarose gel and the arrow indicates the band of interest.

**Figure 6.4:** Amplification of a segment of APE0263.

Lane 1, 1 kbp marker (arrow indicates 1018 bp); lane 2, segment of APE0263; lane 3, forward primer control; lane 4, reverse primer control.

Following PCR amplification of the entire gene from genomic DNA, a product was obtained of the expected size (1108 bp) that was distinct from the single primer controls. Figure 6.5 below shows the product run on a 1 % agarose gel and the arrow indicates the band of interest.

**Figure 6.5:** Amplification of the APE0263 gene still containing the internal *Bam*HI site.

Lane 1, 1 kbp marker (arrow indicates 1018 bp); lane 2, APE0263 amplification; lane 3, forward primer control; lane 4, reverse primer control.
Using the entire gene from the above figure as template, the gene fragment obtained in figure 6.4 as a reverse primer and a forward, flanking primer, a band was obtained in the expected region. The following figure shows the products separated on a 1 % agarose gel and the arrow indicates the band of interest.

Figure 6.6: Amplification of APE0263 without the internal BamHI site.
Lane 1, 1 kb marker (arrow indicates 1636 bp); lane 2, APE0263 amplification.

Following restriction digestion and ligation into pUC18, and transformation into E. coli JM109 cells, numerous colonies were picked and screened for the presence of an appropriately-sized insert. The appropriate restriction pattern was observed for APE0263 and figure 6.7 below shows the digests run on 1 % agarose gels. The arrow indicates the gene cut out of the plasmid.

Figure 6.7: Agarose gel to show digested plasmids.
Lane 1, 1 kb marker (arrow indicates 1018 bp); lanes 2-5, transformants a-d.
Sequencing of transformant d revealed the gene to be identical to the published APE0263 sequence (with the exception of the mutated BamHl site) and so it was chosen for expression studies. Following restriction digestion to remove the gene from pUC18 and ligation into pBE3, the vector was then transformed into B. subtilis DB428 cells. Numerous colonies were picked and screened for the presence of an appropriately-sized insert. The appropriate restriction pattern was observed for APE0263 and figure 6.8 below shows the digest run on a 1 % agarose gel. The arrow indicates the gene cut out of the plasmid.

![Figure 6.8: Agarose gel showing digestion of pBE3 vector containing the insert.](image)

Lane 1, 1 kb marker (arrow indicates 1636 bp); lane 2, digested vector.

The gene had already been sequenced in pUC18 to check the sequence was correct and it was chosen for expression studies without further verification as the only amplification steps in the meantime had been in vivo and hence assumed to be very stringent.

6.3.3 Expression of APE0263 in B. subtilis

Following growth of B. subtilis with pBE3 containing the subtilisin/APE0263 gene fusion, samples of the culture supernatant separated by SDS PAGE did not appear to contain a band of the expected sizes (45.7, 43.0 or 35.3 k for the pre-pro-protein, pro-protein and mature protein, respectively). The expression of proteolytic activity was attributed to APE0263, as the transformation of pBE3 vector still containing the original subtilisin gene should not have been possible. During the digestion of the
vector to prepare it for ligation of the PCR product, the subtilisin gene was cut out and the resulting vector was separated from any undigested vector by agarose gel electrophoresis. Furthermore, host *B. subtilis* cells were grown in parallel to an equivalent optical density that had not been transformed with pBE3 and the culture supernatant was assayed for activity. It was discovered that the untransformed *B. subtilis* culture supernatant had little or no activity against ala-ala-pro-phe-pNA at the elevated temperatures used.

### 6.3.4 Assaying for Activity of APE0263

Measurements of the absorbance of *p*-nitroanilide (the product of proteolytic cleavage) gave an estimation of its extinction coefficient of 8700 M\(^{-1}\) cm\(^{-1}\) at pH 8.0. Following expression of APE0263 in a 1 l culture, the supernatant was assayed for activity against ala-ala-pro-phe-pNA at 80 °C and found to contain a total activity of 105 U. Specific activities could not be determined and therefore a purification table is not available, due to the fact that estimation of protein concentrations were not obtained until after gel filtration. Ammonium sulphate precipitation was used as a means of concentrating the protein sample and not as a crude purification step, as in theory, all the protein present in the sample should precipitate in 100 % (w/v) saturation. Following ammonium sulphate precipitation, the resulting pellet contained a total activity of 98 U, giving a 93 % recovery.

### 6.3.5 Gel Filtration

Following ammonium sulphate precipitation, the pellet was resuspended into a total volume of approximately 9 ml of 50 mM EPPS buffer, pH 8.0. This was loaded in three aliquots onto the gel filtration column, and the following figure shows a typical elution profile.
The A_{280} trace has gone off the scale due to the very high protein concentration present in the enzyme sample after ammonium sulphate precipitation. This is partly due to the fact that a high volume sample (1 litre) was concentrated, but also due to the fact that the culture medium contains proteins and peptides that will still be present in the sample. From the three runs, the 26 most active fractions were pooled [fractions 30-39 (elution volume 70.5-85.5 ml), 31-39 (elution volume 72-85.5) and 28-34 (elution volume 67.5-78 ml) from runs 1-3, respectively], giving a final enzyme sample volume of approximately 39 ml. The pooled fractions were assayed for activity against ala-ala-pro-phe-pNA, and found to contain a total activity of 46 U. This indicates a total recovery from the culture supernatant of 44 %. The following table summaries the recovery of APE0263 after the two stages of purification.

The following figure shows the calibration curve obtained for the gel filtration column.
Ve, elution volume; Vo, void volume.
The line fitted through the points is a linear regression ($R^2 = 0.9812$).

From the gel filtration, the active species in the sample is estimated to have a relative molecular mass of $33.9 \pm 16.7$ k. Figure 6.11 below shows SDS PAGE analysis of the most active fractions from two separate runs. The arrows indicate the putative bands of interest.

It appears from the above figure that the dominant expression product is the pro-protein, which has an estimated relative molecular mass of 45 k from the gene.
sequence, and 43 k from SDS PAGE. The minor band below that of the pro-protein is likely to be the mature protein, which has an estimated relative molecular mass of 35 k from the gene sequence and 33 k from SDS PAGE.

### 6.3.6 Temperature Profile and Kinetic Parameters

Figure 6.12 below shows the activity of APE0263 against ala-ala-pro-phe-pNA at varying temperatures. It can be seen from this that APE0263 has an apparent temperature optimum of around 80 °C, with the activity decreasing quite rapidly above this temperature.

![Thermoactivity of APE0263](image)

**Figure 6.12: Thermoactivity of APE0263.**

Assays above 30 °C were carried out in triplicate and assays at 20 and 30 °C were carried out in duplicate. Error bars represent the range of data about the mean and the line fitted through the points is a polynomial regression.

The effect of the concentration of the ala-ala-pro-phe-pNA substrate on specific activity was investigated, and values of $K_M$ and $V_{\text{max}}$ were determined by the direct linear plot of Cornish-Bowden and Eisenthal (1974). Calculated values for $K_M$ and $V_{\text{max}}$ from the direct linear plot are $0.912 \pm 0.024$ mM and $70.7 \pm 0.87$ U mg$^{-1}$, respectively. As discussed in section 4.3.3, it is not readily obvious how closely the data fit the Michaelis-Menten equation from the Michaelis-Menten curve and it is
therefore necessary to plot the data in a linear fashion. The following figure shows
the data plotted as \( s/v \) against \( s \).

![Hanes plot for APE0263 against ala-ala-pro-phe-pNA.](image)

**Figure 6.13:** Hanes plot for APE0263 against ala-ala-pro-phe-pNA. \( s \), mM; v, U/mg.

It appears from this graph that the data do in fact fit the Michaelis-Menten
equation. The following figure shows the data plotted as specific activity against
substrate concentration. The line fitted through the experimental points is that
generated from the Michaelis-Menten equation using \( K_M \) and \( V_{\text{max}} \) values determined
by the direct linear plot.

![The dependence of velocity on substrate concentration for APE0263 against ala-ala-pro-phe-pNA.](image)

**Figure 6.14:** The dependence of velocity on substrate concentration for APE0263 against ala-ala-pro-phe-pNA.
It is clear from the above figure that some of the substrate concentrations used were above $K_M$, which is reflected in the relatively low errors of the estimates of $K_M$ and $V_{\text{max}}$. The following table shows the $K_M$ values for ala-ala-pro-phe-pNA for a number of other proteases.

<table>
<thead>
<tr>
<th>Protease</th>
<th>Organism</th>
<th>$K_M$ (mM)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>APE0263</td>
<td><em>Aeropyrum pernix</em></td>
<td>0.912</td>
<td>This work</td>
</tr>
<tr>
<td>Subtilisin E</td>
<td><em>Bacillus subtilis</em></td>
<td>0.56</td>
<td>Chen and Arnold (1993)</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>1.9</td>
<td>Takagi et al. (1990)</td>
</tr>
<tr>
<td>Subtilisin Savinase</td>
<td><em>Bacillus lentus</em></td>
<td>0.385</td>
<td>Zhao and Arnold (1999)</td>
</tr>
<tr>
<td>Chymotrypsin</td>
<td>Bovine</td>
<td>0.95</td>
<td>Egmond et al. (1994)</td>
</tr>
<tr>
<td>Halophilic</td>
<td><em>Halobacterium halobium</em></td>
<td>0.11</td>
<td>Ryu et al. (1994)</td>
</tr>
<tr>
<td>Rt41A protease</td>
<td><em>Thermus sp. strain Rt41A</em></td>
<td>2.5</td>
<td>Peek et al. (1992)</td>
</tr>
<tr>
<td>Thermophilic</td>
<td><em>Sulfolobus solfataricus</em></td>
<td>0.85</td>
<td>Burlini et al. (1992)</td>
</tr>
<tr>
<td>Thermitase</td>
<td><em>Thermoactinomyces vulgaris</em></td>
<td>0.0146</td>
<td></td>
</tr>
<tr>
<td>Aqualysin I</td>
<td><em>Thermus aquaticus</em></td>
<td>1.2</td>
<td>Tanaka et al. (1998)</td>
</tr>
</tbody>
</table>

*Figure 6.15: Comparison of $K_M$ values for ala-ala-pro-phe-pNA for various proteases.*

6.3.7 pH Profile

Figure 6.16 below shows the pH profile obtained for APE0263.

*Figure 6.16: pH profile of APE0263.*

Assays were carried out in triplicate in a variety of buffers and the error bars represent the range of data about the mean (\(< 10\%\)).
As can be seen from the above figure, APE0263 has a pH optimum for activity of 7.5-8.5. Furthermore, it exhibits activity over quite a range of pH values, although the activity does tail off rapidly below pH 6.5 and above pH 9.0. The substrate was also particularly unstable in CAPS buffer, suggesting that buffer to be a very unsuitable assay and storage buffer.

### 6.3.8 Thermal Inactivation Studies

The following figure shows the thermostability of APE0263 at 90 and 95 °C, using the gel filtration-purified enzyme sample.

![Figure 6.17: Thermal inactivation of gel filtration-purified APE0263 at 90 and 95 °C.](image)

Assays were carried out in duplicate and the data varied by less than 10 % about the mean. The line fitted through the data is a linear regression ($R^2$ values > 0.985).

It can be seen from the graph that APE0263 does not have an extremely high thermostability. Estimated half-lives are 9.6 and 8.2 min at 90 and 95 °C, respectively. This may be affected in a significant manner by the protein
concentration of the enzyme samples used. A separate thermostability experiment performed using concentrated culture supernatant that had not been purified by gel filtration exhibited a much higher thermal stability. The results are shown in the following figure.

![Graph showing thermal inactivation of APE0263 at different temperatures](image)

Figure 6.18: Thermal inactivation of concentrated APE0263 at 80, 90 and 99 °C. The line fitted through the data is a linear regression.

Estimates of half-lives from this graph are 54 h, 180 min and 26 min at 80, 90 and 99 °C, respectively. It appears, therefore, that the thermostability of APE0263 depends in quite a significant manner on the protein concentration of the sample used, although whether this is due to the concentration of APE0263 itself or just the protein concentration per se, was not determined.

### 6.3.9 Effect of Inhibitors and Substrate Specificity

The following table below shows the relative activities of APE0263 against ala-ala-pro-phe-pNA in the presence of various inhibitors.
Chapter Six

Inhibitor Relative Activity (%)

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Relative Activity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>100</td>
</tr>
<tr>
<td>1 mM EDTA</td>
<td>0</td>
</tr>
<tr>
<td>0.1 mM PMSF</td>
<td>37.0</td>
</tr>
<tr>
<td>0.2 mM PMSF</td>
<td>13.0</td>
</tr>
<tr>
<td>0.5 mM PMSF</td>
<td>3.4</td>
</tr>
<tr>
<td>10 μM PCMB</td>
<td>92.9</td>
</tr>
<tr>
<td>100 μM PCMB</td>
<td>88.2</td>
</tr>
<tr>
<td>1 mM PCMB</td>
<td>99.7</td>
</tr>
<tr>
<td>10 mM PCMB</td>
<td>89.9</td>
</tr>
<tr>
<td>100 μM pepstatin A</td>
<td>85.6</td>
</tr>
</tbody>
</table>

Figure 6.19: Effect of various inhibitors on the activity of APE0263.

It can be seen that EDTA and PMSF have a dramatic effect on the activity of APE0262, whilst the cysteine protease inhibitor, PCMB, and the aspartic protease inhibitor, pepstatin A, have little effect. It seems, therefore, as if APE0263 is a serine protease, although its activity appears to depend on the presence of divalent cations. The following figure shows the effect of divalent cations on, firstly the inhibition by EDTA, and secondly on the absolute activity of APE0263.

<table>
<thead>
<tr>
<th>Inhibitor and/or Cation</th>
<th>Relative activity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>100</td>
</tr>
<tr>
<td>1 mM EDTA</td>
<td>0</td>
</tr>
<tr>
<td>1 mM EDTA + 10 mM ZnSO₄</td>
<td>7.3</td>
</tr>
<tr>
<td>1 mM EDTA + 10 mM MnCl₂</td>
<td>30.3</td>
</tr>
<tr>
<td>1 mM EDTA + 10 mM CaCl₂</td>
<td>33.2</td>
</tr>
<tr>
<td>1 mM EDTA + 10 mM MgSO₄</td>
<td>44.6</td>
</tr>
<tr>
<td>10 mM ZnSO₄</td>
<td>33.0</td>
</tr>
<tr>
<td>10 mM MnCl₂</td>
<td>70.0</td>
</tr>
<tr>
<td>10 mM CaCl₂</td>
<td>88.2</td>
</tr>
<tr>
<td>10 mM MgSO₄</td>
<td>77.9</td>
</tr>
</tbody>
</table>

Figure 6.20: Effect of various cations on the activity of APE0263.

APE0263 was not active against the alternative substrates ala-ala-phe-pNA (tested at 80 μM final concentration) and gly-gly-phe-pNA (tested at up to 1 mM final concentration). It was, however, active against ala-ala-pro-leu-pNA. The effect
of substrate concentration on specific activity was investigated, and values of $K_M$ and $V_{max}$ were determined by the direct linear plot of Cornish-Bowden and Eisenthal (1974). Calculated values for $K_M$ and $V_{max}$ from the direct linear plot are $2.27 \pm 0.17\text{mM}$ and $55.0 \pm 3.4 \text{ U mg}^{-1}$, respectively. It would appear then, that the specific activity of APE0263 is lower against ala-ala-pro-leu-pNA than against ala-ala-pro-phe-pNA. This has been reported for both subtilisin Carlsberg and a cold-active serine alkaline protease also (Kulakova et al., 1999). The Rt41A protease from *Thermus* sp. strain Rt41A has also been reported to exhibit a higher $K_M$ against ala-ala-pro-leu-pNA than against ala-ala-pro-phe-pNA (Peek et al., 1992). The following figure shows the data obtained for APE0263 against the ala-ala-pro-leu-pNA substrate plotted as $s/v$ against $s$.

![Figure 6.21: Hanes plot for APE0263 against ala-ala-pro-leu-pNA. s, mM; v, U/mg.](image)

From the above figure, it appears that these data obey the Michaelis-Menten equation. The following figure shows the data plotted as specific activity against substrate concentration. The line fitted through the experimental points is that generated from the Michaelis-Menten equation using $K_M$ and $V_{max}$ values determined by the direct linear plot.
6.4 Discussion

This chapter has succeeded in its aims to clone and express an active protease from *Aeropyrum pernix* in *Bacillus subtilis*. The expression system set up has also succeeded in secreting the recombinant precursor to the extracellular medium. The fact that the protein was secreted, and also its size as estimated by SDS PAGE, indicates that the pre-region was successfully processed. However, after visualising the products from gel filtration, it appears the major product is the, presumably inactive, pro-protein, with the active mature protein comprising a smaller proportion. N-terminal sequencing of the two protein species was not attempted, although this would determine exactly what forms the bands represent. Cleavage of the pro-region is known to be an autocatalytic event in *Bacillus* (e.g. Shinde and Inouye, 1995; Jain *et al.*, 1998), which is why the cleavage site and the 11 upstream amino acids were taken from the *A. pernix* pro-region and not the subtilisin E pro-region. As the pro-region acts to facilitate folding of the mature region, it might have been wise to use the complete *A. pernix* pro-region to ensure the correct folding of the mature region. By implication, this would ensure total cleavage of the pro-region by the mature region. However, it has been demonstrated in Gram-negative bacteria that the pro-regions positioned at the C-terminal end of the protein play a critical role in protein
translocation (Gray et al., 1986; Pohlner et al., 1987) and the pro-regions of Bacillus proteins have been suggested to be important in secretion (Honjo et al., 1985). It was therefore deemed important to have the subtilisin pro-region as well, because if secretion was inhibited, the pre-region would not have been cleaved and the protein would not be active at all. Investigations into increasing the proportion of the active protein produced would be profitable, although time constraints prohibited this to be carried out. Inspection of the sequence of the pro-region revealed the absence of a convenient site for cleavage by an exogenous protease, e.g. factor Xa, thrombin etc., which would be the most simple protocol available.

Ammonium sulphate precipitation to concentrate the protein, followed by gel filtration proved to be a successful procedure. The final yield of active protein recovered from the culture supernatant was 44%. SDS PAGE analysis of the most active fractions showed the recombinant protein to be the dominant protein in the sample, with few contaminating species present. The lower-sized bands present in the pooled fractions are probably due to autolysis, as contaminating species of a lower molecular mass would be eluted in the latter volumes from the gel filtration column. Estimation of the relative molecular mass of the active species of APE0263 by gel filtration produced a value of 33.9 ± 16.7 k, which is only slightly lower than that calculated from the gene sequence of the mature protein (35.5k).

From the thermoactivity graph, it can be seen that APE0263 has an apparent temperature optimum of around 80 °C. This is perhaps surprising, given that the growth optimum of A. pernix is 90-95 °C. However, there have been reports of thermophilic proteases having an optimum temperature for activity below the optimum growth temperature of their respective host organisms. Watanabe et al. (1999) reported that recombinant lon protease from Thermus thermophilus HB8 has an optimum temperature for activity of 70 °C, with T. thermophilus growing optimally at 85 °C. This phenomenon is not limited to recombinant proteins, however, as Hanner et al. (1990) have reported the purification and characterisation of an aminopeptidase from Sulfolobus solfataricus with an optimum temperature for activity below the growth optimum of S. solfataricus.

The contrasting thermostability results raises an interesting point. The enzyme from the gel filtration-prepared sample has a half-life of 9.6 min at 90 °C,
Chapter Six

whereas the enzyme from a sample of concentrated culture supernatant has a half-life of approximately 180 min at the same temperature. It appears from these results, therefore, that the thermostability of APE0263 depends in no small part upon the protein concentration of the sample tested. Thermostability as defined by the resistance to thermal unfolding is known to be due to a number of intrinsic factors of a particular protein, which would be constant for the same protein. However, it has been shown (McEvily and Harrison, 1986) that the protein concentration can affect the thermostability of an enzyme. At concentrations of 5 μM and above, porcine citrate synthase shows no loss of activity after a 6 h incubation at 35 °C, whereas incubating concentrations of 5 nM and below at the same temperature results in almost complete loss of activity after 3 h. Furthermore, BSA is routinely added to some restriction enzymes to increase their stability, as well as to absorb impurities in the reaction mixture. The thermostability of proteases is somewhat more complex than other proteins, however, as loss of activity can also occur through autolysis. At the elevated temperatures at which the thermostability experiments were carried out, the protease would be active against other proteins present in the sample. In a concentrated sample containing other proteins, for example, a concentrated culture supernatant, a lot of the proteolytic activity would be directed against those other proteins. However, in a sample mainly containing the protease in question, any proteolytic activity will be directed against other molecules of the same enzyme and so the loss of activity due to autolysis would be attributed to denaturation.

APE0263 has been described in the A. pernix genome annotation as an alkaline, serine protease and the results of this work would seem to support this. It has been shown that APE0263 has a pH optimum of around 7.5-8.5, although it does have significant activity over the pH range 6.5-9.5. Furthermore, it is sensitive to the effects of PMSF. Its sensitivity to EDTA might suggest APE0263 to be a metalloprotease; however, to date there have been no reports of a serine protease that also requires a catalytically active ion, but there are known to be serine proteases that require ions for thermostability. Caldolysin, for example, has been shown to bind six calcium atoms that are not required for activity (Khoo et al., 1984). Experiments were not carried out to determine the effect of ions on thermostability in this study, only activity. The presence of ions, or indeed salt, in the concentrated enzyme sample may also contribute to the increased thermostability of APE0263, compared
to the gel filtration-prepared sample. Thermostability is also affected by medium components that act as thermoprotectants. The growth medium, SG medium, contains both calcium and manganese, which will still be present in the concentrated enzyme sample at the same concentration as the culture medium. However, those ions will have been excluded during gel filtration. The addition of divalent cations to the enzyme was shown to be able to relieve some of the inhibition by EDTA, re-emphasising the potential role of cations in maintaining a catalytically active conformation. Magnesium, manganese and calcium were the most efficient cations tested in relieving the inhibition. The addition of divalent cations to an uninhibited enzyme, however, only succeeded in reducing the activity, with magnesium, manganese and calcium having less effect than zinc.

The initial substrate specificity experiments carried out suggest that the presence of a proline residue at the P2 position is essential for cleavage. The presence of either the hydrophobic residues leucine or phenylalanine at the P1 position confers susceptibility to APE0263, although the presence of alternative residues was not examined. The presence of a bulky, hydrophobic residue at the P1 position appears not to be sufficient, based on the resistance of both ala-ala-phe-pNA and gly-gly-phe-pNA to cleavage by APE0263. These results are in contrast to those reported for other enzymes. Aqualysin has been shown to be active against ala-ala-phe-pNA and gly-gly-phe-pNA, in addition to ala-ala-pro-phe-pNA (Tanaka et al., 1998). Furthermore, it is active against the tripeptide series phe-X-ala, where X is leucine, valine or alanine (Tanaka et al., 1998). The subtilisins, Sendai, Carlsberg and BPN do not depend upon a proline residue at the P2 position either, exhibiting activity against a leu-leu-val-tyr peptide substrate with \( K_M \) values in the micromolar range (Yamagata et al., 1995). It has been shown that the \( K_M \) values of APE0263 against ala-ala-pro-phe-pNA are comparable to those previously reported for a variety of other proteases. The \( K_M \) value for chymotrypsin is considerably lower than the values for other proteases; this is perhaps expected as the substrate was synthesised specifically for that enzyme.

A goal in all expression studies is to increase the levels of the recombinant protein obtained. For this work, that requires the induction of the subtilisin promoter to either higher levels of expression, or at earlier stage of the growth cycle. It is established that the expression of subtilisin from \textit{B. subtilis} II68 is closely associated
with the production of endospores, in response to nutrient deprivation (Stahl and Ferrari, 1984). However, it is not fully understood whether the role of subtilisin is for development and/or assembly of the spore coat protein, or simply as a scavenging enzyme for nutrients. If the latter case is true, then growing *B. subtilis* under low nutrient conditions might increase the expression of the subtilisin promoter, and therefore the recombinant protein. An alternative method might be to grow the cells at a higher temperature. It has been shown (Kubo and Imanaka, 1991) that increasing the growth temperature brings about a concomitant increase in the expression of a neutral protease gene from *B. stearothermophilus*, using *B. subtilis* as a host strain. When the cultivation temperature was below 39 °C, enzyme expression occurred in the stationary phase of growth. However, if the temperature was above 40 °C, expression was initiated in the exponential phase. Selection of the host strain is sometimes quite a significant factor in determining the levels of expression. *B. subtilis* is known to secrete several proteases in addition to subtilisin (Simonen and Palva, 1993 and refs within), which significantly affect the production of heterologous proteins from it. An alternative approach is to use a species that naturally secretes lower levels of proteolytically active species. For example, *B. megaterium* has the advantage over other *Bacillus* species that it does not produce any alkaline proteases (Vary, 1994). Furthermore, *B. brevis* HPD31 does not secrete any detectable protease activity (Takagi *et al.*, 1989), which would increase the half-life of a heterologous protein, making it potentially a more suitable host.

The methodology produced in this study would seem to be applicable for most proteins as a generic approach to expressing both toxic and non-toxic genes. Intracellular proteins, which lack a pro-region, could also be produced in this way, with the introduction of an exogenous protease cleavage site at the C-terminus of the subtilisin pro-region to facilitate downstream processing. Alternatively, the target gene could be fused directly to the subtilisin pre-region, avoiding the need for a pro-region. The low expression levels are a barrier to wide-spread applications, although industrial applications would use a set of previously well-defined alternative conditions, such as strains and culture conditions, to optimise the levels of expression and secretion.
This project has succeeded in its original aims, to clone and express an extremophilic protease in a mesophilic host. This was achieved using a *Bacillus subtilis* expression system consisting of a fusion of the target protein, the extracellular protease APE0263 from *Aeropyrum pernix*, with the pre- and pro-regions of subtilisin E. The subtilisin pre-region effected secretion of the fusion protein to the extracellular medium, during which the pre-region was removed by processing. Complete processing of the pro-region was not seen and the reasons behind this are unknown. However, there was significant cleavage of the pro-region and production of active enzyme. The pro- and mature constructs were not purified from each other and so the possibility of the pro-region having a low level of activity can not be ruled out.

The enzyme was partially purified by a combination of ammonium sulphate precipitation and gel filtration. The ammonium sulphate step, however, was not employed solely as a purification procedure, but as a means to concentrate the protein in the culture supernatant. Based on the elution profile, there appears to be high molecular mass species present in the pooled fractions from gel filtration. However, lower-sized bands are seen on SDS PAGE and these are thought to be products of autolysis. This could indeed be verified by N-terminal sequencing of the individual bands.

The suitability of APE0263 to an industrial application is so far unknown. However, its thermostability in concentrated protein solutions would make it an ideal candidate for high temperature processes. Also, its optimum activity under alkaline pH conditions would make it suitable for these processes occurring at higher pH values. However, its substrate specificity has not been fully investigated, limiting its application thus far.
Cytoplasmic expression of an active protease was also attempted using an *Escherichia coli* expression system. However, despite obtaining high levels of expression of an intracellular protease, afpl from *Archaeoglobus fulgidus*, no activity was observed against a number of protein and peptide substrates. This is thought to have been due to toxic activity of the recombinant protein. Subsequent sequencing of the cloned genes from the expression host revealed the presence of discrete mutations and suggested that inactive mutants were selected due to toxicity of the native, active recombinant protein. It has been reported by others that expression of an active protease in *E. coli* is possible, usually via a fusion protein expression system that does not always produce a ‘native’ enzyme after *in vitro* processing. It would appear from this, that expression is better attempted via a secretion system, to avoid the toxicity problems seen in this work. *E. coli* is generally regarded as an unsuitable host for secretion experiments, due to its secretion pathways being thus far not fully understood, and so expression of secreted protein is usually attempted in a Gram-positive host, such as *B. subtilis*. The expression system produced for APE0263 is not necessarily limited to the expression of extracellular enzymes. Intracellular enzymes do not have a pro-region, and so the processing of the subtilisin pre-region following translocation would produce a ‘native’ enzyme, even lacking the N-terminal methionine. Even intracellular proteases could perhaps be expressed via this system, utilising the subtilisin pro-region which may induce the correct folding of a foreign protein. The introduction of a protease cleavage site may be required to effect subsequent cleavage of the pro-region, as the action of the target protease may not be sufficient. Alternatively, the protease could be fused directly to the subtilisin pre-region, which would produce a native protease in the extracellular medium, but the target protease may require the presence of a pro-region for correct folding.

The expression of an extracellular halophilic protease was also attempted, halolysin R4 from *Haloferax mediterranei*, in a number of differing forms. Although high levels of expression were obtained, the recombinant proteins were inactive and also in the majority of cases, insoluble. Attempts to renature the insoluble inclusion bodies were unsuccessful. It appears then, not only from these results, but those from others, that the expression of a halophilic protein in a mesophilic host is very
Chapter Seven

problematic. Therefore, it is perhaps best attempted in a halophilic expression system, using one of the many shuttle vectors now available for *Halofex volcanii*.

There remains much work to be carried out on this project. Time constraints did not permit the detailed characterisation of the recombinant enzyme from *Aeropyrum pernix*. Further substrate specificity investigations would be interesting, to determine whether a proline residue at the P2 position is essential for susceptibility to the enzyme. Investigating the activity of the enzyme against intact proteins to discover if it cleaves after the sequence pro-(phe/leu) in macromolecular substrates would also be worthwhile. Also, knowing whether APE0263 does in fact have a preference for any residue at the P1 position, or whether it is limited to hydrophobic residues, would increase its potential applications. Further investigations into the thermostability of APE0263 would also be useful. For example, the addition of BSA to the enzyme sample would determine if the increased stability seen in the concentrated sample was due to the increase in enzyme concentration, or simply the decrease in autolysis due to an increase in total protein concentration. As mentioned above, the reasons behind the incomplete processing of the pro-region are unknown. Investigations into these reasons would be profitable, which may take the introduction of an exogenous protease recognition site to resolve the question. Whilst the complete processing of the recombinant enzyme would indeed be desirable, an extra *in vitro* processing step may be cost prohibitive for industrial purposes. A comparison of the effect on activity of the complete removal of the pro-region with that of the intact pro-region would determine if the pro-region does in fact exhibit low levels of activity. An alternative idea is to utilise the APE0263 pro-region in its entirety, in conjunction with the subtilisin pre-region. As discussed in section 6.4, this could potentially inhibit secretion of the precursor, as the pro-region has been shown in some cases to be involved in translocation. However, if the protein was secreted and the pre-region processed correctly, it would subsequently be expected that the pro-region would be fully cleaved, after assisting in the correct folding of the mature region. As also mentioned in section 6.4, efforts to increase the levels of expression of APE0263 would be useful. This would involve the induction of the subtilisin promoter at an earlier stage of the growth cycle. One way to achieve this is perhaps via growth on nutrient-limiting medium. Alternatively, any protocol that simply increases the levels of expression at any stage
of the growth cycle would be helpful. The use of a different *Bacillus* host that secretes lower levels of proteases may provide a greater yield of recombinant protein, by increasing its half-life. Two such species, *B. brevis* or *B. megaterium*, could be used in place of *B. subtilis*, assuming pBE3 replicates in those hosts.

Should afpl be expressed in an active form, it would be interesting to see how this was obtained. Its homologue from *Pyrococcus furiosus*, pfpl, was expressed in a soluble and active form in *E. coli* from pET15-b. As discussed in section 5.4, the subsequent processing of the poly-histidine fusion tag was not fully accomplished. Not all of the recombinant protein molecules were cleaved, and those that were still contained three histidine residues. This is considered to be the reason behind the recombinant enzyme having a much lower thermostability than the native enzyme. As such, this route was not chosen for afpl although, in hindsight, it may be one of the limited options available for producing afpl in an active form in *E. coli*. Furthermore, if the crystal structure of either pfpl or afpl is solved, it would very intriguing to observe the active conformation, as pfpl does not contain any known protease motif. Also, it would be good to observe the structural effect of the two separate mutations selected during the expression of afpl from the pET system.

The work involved in the cloning and expression of hlyR4 was terminated in favour of a protease that might produce results in a shorter space of time. There did remain, however, a lot of work to be carried out on the refolding of the inclusion bodies produced. The sequencing of the genes would be the first step in any continuation of the project, to check for the presence of any mutations that would explain the lack of activity when soluble protein was obtained.

Heterologous expression of thermophilic proteases remains a major technological problem and at present there are no generic, reliable routes for production of these enzymes for further biochemical characterisation or for investigation of their possible commercial applications.
Appendix 1:
Nucleotide and Amino Acid sequence of hlyR4

<table>
<thead>
<tr>
<th>50</th>
<th>30</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCTAAATTTCAGTCAAGCTTACACATAATATCGTAGCTGAGTGATGTTTGCCCATG</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>GCAGGCACTCACCACCTTGACAGGCAAAGCTTACCTTCTACGACCTGCCGAGGCGAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>90</td>
<td>110</td>
</tr>
<tr>
<td>ACCGGCATGGGCTGTAACCTAGTCCGACTCCCGGTCCGCTCACCAGGAAAGCAGTGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>150</td>
<td>170</td>
</tr>
<tr>
<td>GAATACCTGCCTGGGCTACGACGTGACTCCCGGTCCGACAGGAAAGGTGCAGTGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>190</td>
<td>210</td>
<td>230</td>
</tr>
<tr>
<td>GATGCCGAAATGGTCAAGACGAGACCCTACGACCAGCTCAGGATACGAGGCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>270</td>
<td>290</td>
</tr>
<tr>
<td>CCATCCCTCTAAGGCCACCGGAAAGCGCGTGAAAACCTTCTACGAGGCAAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>310</td>
<td>330</td>
<td>350</td>
</tr>
<tr>
<td>GATGAGGTAGTACGACGTCCGACTCCCGGTCCGACAGGAAAGGTGCAGTGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>370</td>
<td>390</td>
<td>410</td>
</tr>
<tr>
<td>CCAGAAGACGACGAGGCGACAGTCCGACTCCCGGTCCGACAGGAAAGGTGCAGTGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>430</td>
<td>450</td>
<td>470</td>
</tr>
<tr>
<td>ACGCTGGAAGTATCCCTCAGTGCAGTACCGACCCACAGGGAAGACGACGACGAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>490</td>
<td>510</td>
<td>530</td>
</tr>
<tr>
<td>GACCGACTTCCGACGCTTGGAAGTTCCGACTCCCGGTCCGACAGGAAAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>570</td>
<td>590</td>
</tr>
<tr>
<td>GACCCGATATTCCGACGAGTCCGACTCCCGGTCCGACAGGAAAGG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[\text{agtpnfdrrs flrlaaaagl tmagvts ATPGRSPGPKKD vpgnaihnetlsyaavkr fpskapkqarenfisaitkr devkyaekehthelaytyand pkysqyapqovningsawdt tlgssskvkiavvqdkgvkydhpdlssqfgsnkgrdffvdndg dpypdldldesyhghthvagia}\]
Preregion is shown in lower case, proregion in normal size and mature region is shown in bold. Start and stop codons are doubly underlined. The sequence on which the primers were based is singularly underlined.
Appendix Two

Appendix 2:
Nucleotide and Amino Acid sequence of AF1763

-10 10 30
TTATATAATATTTTGAGGTAATTCCATATGAGAGGCTTGCCGTGCTGGTGTTGCTG
MRGLAVLLL

50 70 90
GTTTTTGGCGTGACAGGGCGACGCAGGAAGACTTTAGACCGTAGTGTGTGGTGGCAAGGT
VFAVQVAAEDEFRPVVFVGH

110 130 150
CTTGCGGTCTTCGTCAGTCAGGAGCACAGGGAAATCGGCTGAAATCAGCCAGATAT
LAGSAGQFESQGMRFPAANGY

170 190 210
CCGCGAGATATGGATCAAAACCTTCGAGATACGACGATTTCGTGGGCTTTGGTTGGA
PAEYVKTFEYDTSWALVVE

230 250 270
ACTGACATGCCTCTAGGAGAATTTGACGATGACTTCCAGGACGAAATCGGCTGAAATCAGCCAGATAT
TDMLFSGLESFGLNISQII

290 310 330
GACCCCTGAACGGTTCGAAAGACTTTGCAAGACAGGACGCCAGGGAGATTGCAGATT
DPETLDRKILSKSRRLIDET

350 370 390
TTTCAGCAGGCCTAGGATAGCAGGCGTTCGGTCTGCCTGGAAAGCCGAGCGGTAAGCTG
FSRLDRVIDEAALAESGADV

410 430 450
GATCTGGTGGCGACCTCAATGCGGACCTGGTGGATCAGATCAGTTGCTGCTCGCTCTCG
DLVGHSMGFVFLVRYVNSSP

470 490 510
GAGAGGCTGCGAAAGGTTGAGACGCTCGCTGCGCTGCGCTGGGGGTGATGCTG
ERAALKVAHILLDDGVWGVDA

530 550 570
CCCGAGGAGGATTCCGAGCCTGGGCTGAAATCCGAAAGCTTGCTGCCGCTCTGGCG
PEGIPTLAVFGNPALKALG

590 610 630
CTTCGAGGAGAAGGTTTCAACGCAACCAGCTTTACCTCAACAACATGAGCAGC
LPREEKVYNNATNVYFNNMTTH

650 670 690
TCGGATTACATTTAAATTTTTACTTTTTGCTAAGCGAATCAGAAGGAAAGAACTCAATC
S D Y I * F F T F L L S E S E G K N S I

Start and stop codons are doubly underlined. The sequence on which the primers were based is singularly underlined.
Appendix Three

Appendix 3:
Nucleotide and Amino Acid sequence of afpI

-40  -20  10
AAGATGGGAAAAGGTAAATATTAGTTTTGGAGCTGGTTATCATGAGGGTGCTGATTCTG
 M R V L I L

20  40  60
GCTGAGAATGAAATTGAGGACCTCGAGCTTTTCTACCCCGCTACAGCCTCCGGGAAAGAG
A E N E F E D E L E L F Y P L Y R L R E E

80  100  120
GGTTAGAGGTGAAGGGAGTGCATGCTCTCTGAGAATGAGGTGAGGATCGGGGGGAAGAG
G L E V K V A S S S L E V R V G K K G Y

140  160  180
CAGGTGAGGCCGGATTTAAGGGGCGAGCACTGAGTAAACGGGAAGGCTGGTGAATAGTC
Q V R P D L T Y E D V K V E D Y A G L V

200  220  240
ATCCCTGCGGCAAAGTCCCGTCAGAATAGTTAAGGGAAGGCTGGTTGAATAGTC
I P G G K S P E R V R I N E R A V E I V

260  280  300
AAGGACTTTTGGTAGCTTAAAGGGAGCAGGACCTGAGTCTGGTGAGGATTACGCCGGGCTGG
K D F L E L G K P V A A I C H G P Q L L

320  340  360
ATATCGGCGATGGCGTTAAGGGAAGAAGAATGAGCTGGTGAATAGTC
I S A M A V K G R R M T S W I G I R D D

380  400  420
TGATGATCAAGCCGGAGCCCTTTTCTACGAAGAAGACAGCCGGGTTGTGGTCTACGGAG
L I A A G A L Y E D R P V V V D G N V I

440  460  480
ACCTCCCGTAATGCCCCGAGCCTTTCTGAGGGAGCTCCATCAAAATCTCTTAAG
T S R M P D D L P Y F C G E L I K I L K

500  520  540
CGATACCTATATAAACACAATGCCCCCCTACCCCTAAGGGGGCCTGCTAGCATGGAT
R Y * Y K H Q C P P T L K G P W L S M D

Start and stop codons are doubly underlined. The sequence on which the primers were based is singularly underlined (dotted underline is the sequence for the reverse primer for pET vectors).
Amino acids highlighted in grey and black indicate residues that are similar and identical, respectively. Active site aspartic acid, histidine and serine residues are numbered 166, 206 and 385, respectively. Based on this alignment, AF1653 is too short to be a serine protease and does not contain the serine numbered 385.
Appendix 5:
Alignment of AF1653 with Subtilisin Carlsberg

Amino acids highlighted in **grey** and **black** indicate residues that are similar and identical, respectively. Active site aspartic acid, histidine and serine residues are numbered 137, 170 and 327, respectively. Based on this alignment, AF1653 does not contain the serine numbered 327.
Appendix 6:
Nucleotide and Amino Acid sequence of APE0263

0                          -20
AATCGTCGTTTCCTGATTACAAATATGTATTTTTATTGTATCTGGGTAATAGGGTG

20                          40
GTGGTCGCTGTTAGTAACTGCTATACAGGTTGGGAAGTACAGTAGTCGCGCTATTTGCTA

m v a v v t g v i q v g t k i a a i a i

60
m v a v v t g v i q v g t k i a a i a i

80                          100
GCGCTGATCTTCATTTCTGCTTTATACGGGATCAGCGGCGGCTGGCTA

a L I F I L P L F P V Y T G S A A G A S

120
GCGCTGATCTTCATTTCTGCTTTATACGGGATCAGCGGCGGCTGGCTA

T V V I A K I N P E E F N P K A V E A L

140                         160
ACGGTTGTGATAGCTATAGGATAATACCTGAGGAGGTTAACCCTAAAGGGCGGAGGCTGCT

m v a v v t g v i q v g t k i a a i a i

180
GCGCTGATCTTCATTTCTGCTTTATACGGGATCAGCGGCGGCTGGCTA

T V V I A K I N P E E F N P K A V E A L

200                        220
CAGGCGAGGTATATATATGCTGAGTATAGCCTGAGGAGGTTAACCCTAAAGGGCGGAGGCTGCT

Q G K V I Y V A D L A P V A I I S I P G

240
CAGGCGAGGTATATATATGCTGAGTATAGCCTGAGGAGGTTAACCCTAAAGGGCGGAGGCTGCT

Q G K V I Y V A D L A P V A I I S I P G

260                        280
AAGGCTGTAGGCTTCATTCAGCTGCTGTTATCCGTAGCCTGAGGCTGCTGCT

K A V G L L S K L P G V V S V S E D G V

300
AAGGCTGTAGGCTTCATTCAGCTGCTGTTATCCGTAGCCTGAGGCTGCTGCT

K A V G L L S K L P G V V S V S E D G V

320                        340
GTCAGCGGCGTGTGTGTCGCTGTTATCTGAGGATCAGCGGCGGCTGGCTA

V Q A M A K P W A G G G N K S Q P A E

360
GTCAGCGGCGTGTGTGTCGCTGTTATCTGAGGATCAGCGGCGGCTGGCTA

V Q A M A K P W A G G G N K S Q P A E

380                        400
GTCAGCGGCGTGTGTGTCGCTGTTATCTGAGGATCAGCGGCGGCTGGCTA

V L P W G V D Y I D A E L V W P D G V T

420
GTCAGCGGCGTGTGTGTCGCTGTTATCTGAGGATCAGCGGCGGCTGGCTA

V L P W G V D Y I D A E L V W P D G V T

440                        460
GGCTGGTTGAGCTTATACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGAC

G W V D V N G D G D G E I E V A V I D T

480
GGCTGGTTGAGCTTATACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGAC

G W V D V N G D G D G E I E V A V I D T

500                        520
GGCTGGTTGAGCTTATACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGAC


540
GGCTGGTTGAGCTTATACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGAC


560                        580
AACGCGAGGAGAATCCCACTGCCAGCGGAATAGAAACCGCCAGGCTACACAAGGTAACCGG

N G R I S S N Y Q D R N G H G T H V T G

600
AACGCGAGGAGAATCCCACTGCCAGCGGAATAGAAACCGCCAGGCTACACAAGGTAACCGG

N G R I S S N Y Q D R N G H G T H V T G
Preregion is shown in lower case, proregion in normal size and mature region is shown in bold. Start and stop codons are doubly underlined. The sequence on which the primers were based is singularly underlined.
0.5 μg of DNA was separated on a 1 % (w/v) agarose gel and visualised using a UV transilluminator after staining with ethidium bromide.


References


References


References


References


References


References


References


References


References


References


References


References


