Nitrile degrading enzymes from extreme environments

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

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

2002
Abstract

In a study prior to this work, screening experiments for thermophilic nitrile degrading organisms led to the identification of a moderately thermophilic *Bacillus* sp. (strain RAPc8). The organism was shown to constitutively express a nitrile hydratase (NHase), which was purified and extensively characterised.

This work reports the cloning of a gene cluster containing the NHase genes of *Bacillus* sp. RAPc8 into *Escherichia coli*. Sequencing of a 5.9kb section of cloned DNA revealed the presence of eight complete open reading frames, encoding (in order), amidase (belonging to the nitrilase related aliphatic amidase family), nitrile hydratase β and α subunits (of the cobalt containing class), a putative 122 amino acid protein of unknown function, designated P14K, a homologue of the 2Fe-2S class of ferredoxins and three putative proteins with distinct homology to the cobalt uptake proteins cbiM, cbiN and cbiQ of the *S. typhimurium* LT2 cobalamin biosynthesis pathway. The NHase operon shares an exceptionally high sequence identity with that of *Bacillus* sp. BR449. The deduced NHase α subunit N-terminal sequence bore no similarity the empirically determined sequence of the previous study. However, in this study, N-terminal sequencing of the recombinant protein confirmed the sequence predicted by gene translation.

The amidase and nitrile hydratase genes of *Bacillus* sp. RAPc8 were subcloned and actively expressed in *E. coli*. The recombinant NHase was partially purified and found to exhibit catalytic behaviour very similar to the native protein. Substrate specificity trials showed that the recombinant enzyme, as with the native enzyme, has broad aliphatic substrate specificity, but no detectable activity on aromatic nitriles.

A rational design approach, based on the observation that the enzyme was irreversibly inhibited by benzonitrile, was applied in attempts to broaden the substrate specificity of the enzyme to include aromatic nitriles. A crude homology model, of the enzyme...
structure was built using the crystal structure of the NHase of *Rhodococcus* sp. R312 as a template. Aromatic residues in the active site, identified from the model were replaced with non-polar or charged residues using site directed mutagenesis. Of the six mutants investigated, none demonstrated any catalytic activity on the aromatic nitriles tested. However, two mutants, βY67E and βW76G, did display an apparent decrease benzonitrile inhibition (rise in $K_i$ of ~6.5mM and 29.25mM respectively). A further two mutants (β36L and βY67A) also showed a marked increase in specific activity in crude cell extracts compared with that of the wild type enzyme.

A PCR based method for the identification of NHase genes in environmental soil and sediment samples was developed and employed to verify the presence of such genes in samples taken from New Zealand hot pools. Six unique sequences corresponding to the partial sequence of the α subunit gene were identified. At the DNA level, these sequences showed significant homology with the *Bacillus* NHase sequences (between 83% and 93% sequence identity with the *Bacillus* sp. RAPc8 gene). DNA isolated from NHase positive sediment samples was used to create a substantial multi-genomic library for the purpose of screening for full-length NHase genes. However, attempts to identify functional nitrile hydratase genes through complementation screening on succinate-nitrile minimal media were unsuccessful.
Acknowledgements

I would like to acknowledge the Biotechnology and Biological Sciences Research Council for provision of the funding for my research at UCL.

I thank my supervisor, Professor Don Cowan for his support, his faith in my abilities and for providing me with the opportunity to undertake this research.

I gratefully acknowledge the invaluable help and advice of Professor Peter Piper, Dr. John Ward, Dr. Haitham Hussain, Dr. Stefan Millson and Dr. Mehdi Mollapour in various aspects of microbiology and molecular biology: without their help, this work would have taken a good deal more blood, sweat and tears than it already has!

Many thanks to all past and present members of the Ground Floor Labs for the requisite distractions and for helping me to see the light side of the dark days.

Paul, Emma, Rekha, Latha, George, Sam, Steve and Hannah I remain deeply indebted to you for your friendship, for your belief in me, for lending me your strength when mine was waning and generally for helping to keep my life rolling as smoothly as possible down a path that, over the past eight years has had it's fair share of daunting twists. I'll not forget you.

Finally, to my family I owe a similar debt for their love and continued support since year dot - I won't forget them either!

This thesis is dedicated to the memory of my grandmother, Evelyn Cameron, without whom…
Table of contents

**Abstract** ........................................................................................................... 2

**Acknowledgements** ......................................................................................... 4

**Table of Contents** ............................................................................................. 5

**List of Figures** .................................................................................................. 7

**List of Tables** .................................................................................................... 10

**Abbreviations** ................................................................................................... 11

**Chapter 1: Introduction** .................................................................................. 13

1.1 Nitriles .............................................................................................................. 13
1.2 Nitrilase ............................................................................................................ 16
1.3 Nitrile Hydratase ............................................................................................ 20
1.4 Amidase ............................................................................................................ 37
1.5 Applications of NHases ................................................................................... 40
1.6 Thermophiles .................................................................................................. 42
1.7 Thermostable enzymes ................................................................................... 43
1.8 Bacillus sp. RAPc8 .......................................................................................... 46
1.9 Bacillus sp. RAPc8 NHase .............................................................................. 46
1.10 Aims ................................................................................................................ 48

**Chapter 2: Materials and Methods** ................................................................. 49

2.1 Chemicals and reagents .................................................................................. 49
2.2 Enzymes ........................................................................................................... 49
2.3 Media ................................................................................................................ 49
2.4 Buffers and solutions ....................................................................................... 52
2.5 Bacterial strains ............................................................................................... 54
2.6 Plasmids .......................................................................................................... 55
2.7 Analytical Procedures ...................................................................................... 56
2.8 Growth and maintenance of cultures ............................................................... 59
2.9 Preparation of competent E. coli cells ............................................................. 59
2.10 Transformation of competent E. coli cells ...................................................... 60
2.11 DNA Purification ............................................................................................ 61
2.12 Polymerase chain reaction (PCR) ................................................................. 63
2.13 Site-directed mutagenesis ............................................................................. 63
2.14 Restriction endonuclease digestion of DNA ................................................... 64
2.15 DNA cloning .................................................................................................. 64
2.16 DNA hybridisation ......................................................................................... 67
2.17 Protein expression .......................................................................................... 70
2.18 Cell free extract preparation .......................................................................... 71
2.19 Protein purification ......................................................................................... 71
2.20 Growth and expression curves ....................................................................... 72
2.21 NHase characterisation ................................................................................ 72
## Contents

2.22 Sequencing protocols .................................................................74  
2.23 Computational techniques ........................................................74

**CHAPTER 3: CLONING THE NITRILE HYDRATASE OPERON FROM *BACILLUS SP.* RAPc8** ..........................................................76  
3.1 Introduction .................................................................................76  
3.2 Verification of the identity of Bacillus sp. RAPc8 .........................77  
3.3 Cloning Strategies .......................................................................78  
3.4 Nitrile hydratase operon structure ...................................................99  
3.5 Discussion .....................................................................................110

**CHAPTER 4: EXPRESSION AND CHARACTERISATION** .........118  
4.1 Introduction ..................................................................................118  
4.2 Expression strategy .......................................................................119  
4.3 Subcloning of NHase and amidase genes ........................................120  
4.4 Expression studies .......................................................................125  
4.5 Purification .....................................................................................130  
4.6 Characterisation of recombinant Nitrile hydratase ......................136  
4.7 Discussion .....................................................................................146

**CHAPTER 5: ENZYME ENGINEERING** .................................153  
5.1 Introduction ..................................................................................153  
5.2 Homology modelling ....................................................................153  
5.3 Site directed mutagenesis ...............................................................163  
5.4 Expression of mutants ...................................................................170  
5.5 Thermostability ............................................................................172  
5.6 Benzonitrile Inhibition .................................................................173  
5.7 Discussion .....................................................................................177

**CHAPTER 6: THE SEARCH FOR NOVEL NITRILE-METABOLISING ENZYMES** ........................................184  
6.1 Introduction ..................................................................................184  
6.2 DNA Extraction ...........................................................................185  
6.3 PCR analysis of WHA30 environmental DNA sample ....................187  
6.4 Library construction ......................................................................191  
6.5 Library screening .........................................................................194  
6.6 Discussion .....................................................................................195

**CHAPTER 7: GENERAL DISCUSSION** ....................................201

**REFERENCES** .............................................................................205

**APPENDICES** ............................................................................217  
Appendix I Standard curves ...............................................................217  
Appendix II Blast searches .................................................................218
List of figures

Fig. 1.1: Metabolism of nitriles via hydrolysis. 15
Fig. 1.2: Nitrilase mode of action. 17
Fig. 1.3: Stereoview of the αβ heterodimers of Rhodococcus sp. R312 NHase. 22
Fig. 1.4: Superimposition of Co-type NHase and Fe-type NHase. 25
Fig. 1.5: Three putative reaction mechanisms of NHase. 27
Fig. 1.6: Alignment of NHase β subunit N-terminal sequences. 32
Fig. 1.7: Alignment of NHase α subunit N-terminal sequences. 32
Fig. 1.8: Arrangement of NHase gene clusters 35
Fig. 1.9: Mechanism of the amidase acyl transfer 39
Fig. 1.10: Hydration of adiponitrile. 41
Fig. 1.11: RAPc8 NHase β subunit N-terminal sequence BLAST search. 47
Fig. 3.1: Organisation of the NHase operon of R. erythropolis. 80
Fig. 3.2: PCR amplification of R. erythropolis JCM6823 NHase α subunit gene. 80
Fig. 3.3: Hybridisation with Rhodococcus NHase α subunit gene. 81
Fig. 3.4: Alignment of the α subunit sequences of 10 recently cloned NHases. 82
Fig. 3.5: Degenerate oligonucleotides design. 83
Fig. 3.6: Hybridisation with DIG end-labelled NHF1 probe. 85
Fig. 3.7: Hybridisation with DIG end-labelled NHR2 probe. 85
Fig. 3.8: Hybridisation with DIG end-labelled NHF2 probe. 86
Fig. 3.9: Schematics outlining the arrangement of RAPc8 NHase PCR primers. 88
Fig. 3.10: PCR amplification using primers NHFl + NHR1 and NHF2 + NHR2. 91
Fig. 3.11: Annealing temperature optimisation of PCR reaction. 91
Fig. 3.12: Translated BLAST searches with the inserts of pNHF1 and pNHF2. 93
Fig. 3.13: Sequences of the inserts from pNHF1 and pNHF2. 94
Fig. 3.14: Hybridisation with DIG labelled PCR products NHF1 and NHF2. 95
Fig. 3.15: Organisation of NHase genes within the insert of pNHB4. 97
Fig. 3.16: Organisation and sequences of primers AMD3, AMD2 and C8B2. 98
Fig. 3.17: Amplification of the amidase and partial NHase β subunit genes. 99
Fig. 3.18: Alignment of NHase α subunit peptide sequences. 101
<table>
<thead>
<tr>
<th>Fig. 3.19:</th>
<th>Alignment of NHase β subunit peptide sequences.</th>
<th>102</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 3.20:</td>
<td>Alignment of the translation product of ORF4 with homologous proteins.</td>
<td>103</td>
</tr>
<tr>
<td>Fig. 3.21:</td>
<td>Hairpin loop structure found just downstream of the \textit{P14K} gene.</td>
<td>104</td>
</tr>
<tr>
<td>Fig. 3.22:</td>
<td>Layout of the \textit{Bacillus} sp. RAPc8 gene cluster containing the NHase genes.</td>
<td>106</td>
</tr>
<tr>
<td>Fig. 3.23:</td>
<td>Nucleotide sequence of the RAPc8 NHase gene cluster.</td>
<td>107</td>
</tr>
<tr>
<td>Fig. 3.24:</td>
<td>The order of the \textit{cobI} operons of \textit{B. megaterium} and \textit{S. typhimurium}.</td>
<td>114</td>
</tr>
<tr>
<td>Fig. 3.25:</td>
<td>Comparison of hydrophobicity plots of \textit{S. typhimurium} cbiQ and cbiN with RAPc8 homologues.</td>
<td>115</td>
</tr>
<tr>
<td>Fig. 4.1:</td>
<td>Primers used for cloning the RAPc8 NHase genes for expression studies.</td>
<td>120</td>
</tr>
<tr>
<td>Fig. 4.2:</td>
<td>PCR amplification of genes of the \textit{Bacillus} sp. RAPc8 NHase operon.</td>
<td>121</td>
</tr>
<tr>
<td>Fig. 4.3:</td>
<td>Cloning and subcloning of the RAPc8 amidase gene.</td>
<td>122</td>
</tr>
<tr>
<td>Fig. 4.4:</td>
<td>Cloning and subcloning of the RAPc8 NHase genes.</td>
<td>123</td>
</tr>
<tr>
<td>Fig. 4.5:</td>
<td>Expression profile of recombinant amidase</td>
<td>125</td>
</tr>
<tr>
<td>Fig. 4.6:</td>
<td>SDS-PAGE analysis amidase expressed in \textit{E. coli} (BL21) \textit{DE3} pLysS.</td>
<td>125</td>
</tr>
<tr>
<td>Fig. 4.7:</td>
<td>Expression profile of recombinant NHase.</td>
<td>127</td>
</tr>
<tr>
<td>Fig. 4.8:</td>
<td>SDS-PAGE analysis NHase expressed in \textit{E. coli} (BL21) \textit{DE3} pLysS.</td>
<td>128</td>
</tr>
<tr>
<td>Fig. 4.9:</td>
<td>Heat purification of RAPc8 amidase.</td>
<td>130</td>
</tr>
<tr>
<td>Fig. 4.10:</td>
<td>SDS-PAGE showing amidase purification.</td>
<td>131</td>
</tr>
<tr>
<td>Fig. 4.11:</td>
<td>Heat purification of RAPc8 NHase.</td>
<td>133</td>
</tr>
<tr>
<td>Fig. 4.12:</td>
<td>SDS-PAGE of fractions from NHase ion exchange chromatography.</td>
<td>134</td>
</tr>
<tr>
<td>Fig. 4.13:</td>
<td>Comparison of results of Edman sequencing of native and recombinant NHase with the deduced NHase α and β subunit N-terminal sequences.</td>
<td>135</td>
</tr>
<tr>
<td>Fig. 4.14:</td>
<td>Thermostability profile for the purified NHase.</td>
<td>137</td>
</tr>
<tr>
<td>Fig. 4.15:</td>
<td>Thermostability of recombinant NHase.</td>
<td>137</td>
</tr>
<tr>
<td>Fig. 4.16:</td>
<td>Kinetics of acetonitrile hydrolysis.</td>
<td>142</td>
</tr>
<tr>
<td>Fig. 4.17:</td>
<td>Kinetics of acrylonitrile hydrolysis.</td>
<td>143</td>
</tr>
<tr>
<td>Fig. 4.18:</td>
<td>Benzonitrile inhibition of acetonitrile hydrolysis.</td>
<td>144</td>
</tr>
<tr>
<td>Fig. 4.19:</td>
<td>Diagram of the arrangement of the electron orbitals of pyridine.</td>
<td>149</td>
</tr>
<tr>
<td>Fig. 5.1:</td>
<td>Amino acid alignments of the α and β subunits of bacterial NHases.</td>
<td>155</td>
</tr>
<tr>
<td>Fig. 5.2:</td>
<td>PHD secondary structure prediction for RAPc8 NHase sequences.</td>
<td>158</td>
</tr>
<tr>
<td>Fig. 5.3:</td>
<td>Space filled model of the active site region of RAPc8 NHase β subunit.</td>
<td>159</td>
</tr>
</tbody>
</table>
Contents

List of figures

Fig. 5.4: Space filled model of the active site region of RAPc8 NHase α subunit. 159
Fig. 5.5: Space filled model of the RAPc8 NHase holoenzyme. 160
Fig. 5.6: Space filled model of the RAPc8 NHase holoenzyme. 160
Fig. 5.7: Space filled model of an αβ dimer of RAPc8 NHase. 161
Fig. 5.8: Schematic of the Quik-change mutagenesis mutagenesis procedure. 163
Fig. 5.9: Restriction map of the RAPc8 NHase genes. 165
Fig. 5.10: Confirmation of the mutation of αW124 to G by treatment with Xmal. 166
Fig. 5.11: Confirmation of the mutation of βF36 to L by treatment with HaeIII. 166
Fig. 5.12: Confirmation of the mutation of βF52 to G by treatment with HaeIII. 167
Fig. 5.13: Confirmation of the mutation of βY67 to A by treatment with FokI. 167
Fig. 5.14: Confirmation of the mutation of βY67 to E by treatment with FokI. 168
Fig. 5.15: Confirmation of the mutation of βW76 to G by treatment with NcoI. 168
Fig. 5.16: SDS PAGE gel of cell free extracts of the expressed mutant proteins. 170
Fig. 5.17: Thermostability profile of wild type and mutated NHases. 171
Fig. 5.18: Benzonitrile inhibition of mutant βF36L. 173
Fig. 5.19: Benzonitrile inhibition of mutant βF52G. 173
Fig. 5.20: Benzonitrile inhibition of mutant βY67A. 174
Fig. 5.21: Benzonitrile inhibition of mutant βY67E. 174
Fig. 5.22: Benzonitrile inhibition of mutant βYW76G. 175
Fig. 5.23: Benzonitrile inhibition of mutant αW124G. 175
Fig. 6.1: Typical result of environmental DNA extraction from sample WHA30. 185
Fig. 6.2: Amplification of NHase fragments using primers NHFl and NHR1. 186
Fig. 6.3: Alignments of the DNA and deduced amino acid sequences of the six environmental PCR amplicons. 189
Fig. 6.4: Overview of environmental library construction using pCR-XL-TOPO. 191
Fig. 6.5: EcoRI-digested plasmids recovered en masse from the library. 192
Fig. 6.6: Primers used in amplification of NHase genes from soil in relation to the NHase operons of R. erythropolis and R. rhodochrous J1 (H-NHase). 196
List of tables

Table 1.1: Molecular and functional characteristics of nitrilases. 19
Table 1.2: Thermal stability of NHases. 30
Table 1.3: Structural and functional properties of NHases. 36
Table 2.1: Buffers and solutions. 52
Table 2.2: Bacterial strains. 54
Table 2.3: Plasmids. 55
Table 2.4: SDS PAGE resolving gel constituents. 57
Table 2.5: SDS PAGE stacking gel constituents. 57
Table 2.6: Typical PCR reaction constituents. 63
Table 3.1: Sequence identity and similarity of RAPc8 NHase with other NHases. 100
Table 3.2: Summary of the properties of the genes in the RAPc8 NHase locus. 105
Table 3.4: Comparison of the MW of RAPc8 amidase and NHase with the native proteins. 110
Table 4.1: Effect of cobalt addition to cell cultures on NHase activity. 127
Table 4.2: Heat purification of RAPc8 Amidase 131
Table 4.3: Purification of RAPc8 NHase. 134
Table 4.4: Comparison of NHase thermostability. 138
Table 4.5: Comparison of NHase substrate specificity 140
Table 4.6: Comparison of the kinetic data of the native and recombinant RAPc8 NHase on acetonitrile and acrylonitrile substrates. 141
Table 5.1: Details of the mutagenic primers used. 164
Table 5.2: Specific activities of each mutant of the recombinant RAPc8 NHase. 170
Table 5.3: Estimated benzonitrile $K_i$ for each of the RAPc8 NHase mutants. 172
Table 6.1: Identity and similarity of the DNA and translation products of the six amplicons with the DNA and amino acid sequences of \textit{Bacillus} sp. BR449 and \textit{Bacillus} sp. RAPc8. 188
Table 6.2: Comparison of environmental expression libraries. 197
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>angstrom</td>
</tr>
<tr>
<td>abs</td>
<td>absorbance (-log Transmitance)</td>
</tr>
<tr>
<td>Amp</td>
<td>ampicillin</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DIG</td>
<td>digoxigenin</td>
</tr>
<tr>
<td>g</td>
<td>gram</td>
</tr>
<tr>
<td>GC</td>
<td>gas chromatography</td>
</tr>
<tr>
<td>h</td>
<td>hour</td>
</tr>
<tr>
<td>k</td>
<td>kilo (10^3)</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>K&lt;sub&gt;M&lt;/sub&gt;</td>
<td>Michaelis Menten constant</td>
</tr>
<tr>
<td>L</td>
<td>liter</td>
</tr>
<tr>
<td>m</td>
<td>milli (10^-3)</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>Mw</td>
<td>relative molecular weight</td>
</tr>
<tr>
<td>NB</td>
<td>nutrient broth</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>T&lt;sub&gt;opt&lt;/sub&gt;</td>
<td>temperature optimum</td>
</tr>
<tr>
<td>U</td>
<td>enzyme activity: μmoles product formed/minute</td>
</tr>
<tr>
<td>v/v</td>
<td>volume per volume</td>
</tr>
<tr>
<td>V&lt;sub&gt;max&lt;/sub&gt;</td>
<td>rate of enzyme-catalysed reaction at infinit substrate concentration</td>
</tr>
<tr>
<td>w/v</td>
<td>weight per volume</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4chloro-3-indolyl-β-D-galactoside</td>
</tr>
<tr>
<td>β-gal</td>
<td>β-galactosidase</td>
</tr>
<tr>
<td>μ</td>
<td>micro (10^-6)</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenosine triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxyctosine triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>deoxyguanosine triphosphate</td>
</tr>
<tr>
<td>dTTP</td>
<td>deoxythymidine triphosphate</td>
</tr>
<tr>
<td>dUTP</td>
<td>deoxyuridine triphosphate</td>
</tr>
<tr>
<td>DIG-dUTP</td>
<td>DIG-conjugated dUTP</td>
</tr>
</tbody>
</table>
### The Amino Acid Codes

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

### The Nucleotide Codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Nitriles

1.1.1 Origins of nitriles

The nitriles, also known as organic cyanides are defined by the general formula R-C≡N. The presence throughout the natural environment of compounds containing this cyano group has been well documented, with reports of their synthesis or metabolism by members of the plant, animal and bacterial Kingdoms (Legras et al., 1990).

Many organisms rely on the toxicity of organocyanide compounds as a defence mechanism against herbivores, predators and parasites (Jones, 1979). For example, some insects have the ability to spray cyanide and mandelonitrile (Duffey, 1981), whilst certain strains of *Streptomyces* species produce toyamycin, an antimicrobial nitrile (Kikuchi, 1955). However, the greatest group of nitrilated compounds, the cyanogenic glycosides, are mainly produced by plants, often in seeds, roots and the epidermal layers of leaves (Legras et al., 1990). As the plant is masticated, the cyanogenic glycosides decompose and give rise to cyanide. Nitrile based compounds, such as 3-indoleacetonitrile also serve as growth hormones during germination (Mascharak, 2002).

Nitriles are also produced (on the megaton scale) by man, to be used as base chemicals for a wide range of syntheses, these include acrylonitrile and the polymers derived from it in the manufacture of plastics and acrylic fibres, and adiponitrile, used as a precursor in nylon-66 production (Wyatt and Linton, 1988). Nitriles are also used in the production of herbicides, pharmaceuticals and amino acids (Jallegeas et al., 1980, Nagasawa and Yamada, 1989). The industrial effluents of these syntheses, as well as the herbicides containing the -C≡N moiety, such as 3,5-dibromo-4-hydroxybenzonitrile (Bromoxynil) and 3,5-di-iodo-4-hydroxybenzonitrile (Ioxynil) are the chief sources of anthropogenic nitriles in the environment (Jallegeas et al, 1980, Mascharak, 2002). Acrylonitrile and the nitrile containing herbicides are released to such an extent that they are now considered environmental threats. Furthermore the dilute, aqueous nature
of nitrile-containing industrial effluents precludes disposal by incineration; thus they are usually disposed of either at sea or below the water table (Mascharak, 2002, Wyatt and Knowles, 1996, Wyatt and Linton, 1988).

Despite the commercial interest in these chemicals, the presence of sensitive functionalities on some of the nitrile compounds prevents their chemical modification – due to the harsh conditions required (e.g. heating with acid or alkali). For this reason, the industry has turned to microbial biotransformations for the conversion of nitriles into commodity chemicals. There is also widespread interest in their use in bioremediation of industrial effluents (Nagasawa and Yamada, 1989).

1.1.2 Enzymatic degradation of nitriles

Five possible mechanisms have been described for the detoxification of cyanide and nitriles (Legras, 1990):

1. Hydrogenation (of a limited number of nitriles) by nitrogenase containing nitrogen-fixing bacteria.
2. Oxidation by *Pseudomonas fluorescens* involving dioxygenase or monooxygenase followed by a cyanase.
3. Decomposition to cyanide followed by conversion to β-cyanoalanine by *Enterobacter* sp. and *Chromobacterium violaceum*.
4. Decomposition to cyanide followed by conversion to thiocyanate by a rhodanese.
5. Hydration/hydrolysis of the -C≡N function by nitrilase, or hydration by nitrile hydratase followed by an amidase catalysed hydrolysis.

The fifth mechanism is the most widely characterised, and perhaps the most useful reaction to microorganisms since the nitrile is metabolised to provide a source of carbon and/or nitrogen for growth (Nagasawa and Yamada, 1989). The nitrilase (Nase; EC 3.5.5.1) pathway catalyses the direct conversion of a nitrile to its corresponding carboxylic acid (Fig. 1.1 A). The nitrile hydratase (NHase; EC 4.2.1.84) pathway catalyses the conversion of a nitrile to its corresponding amide, which is then converted to a carboxylic acid by an amidase (Fig. 1.1B).
Fig. 1.1: Metabolism of nitriles via hydrolysis: A – nitrilase pathway, B – nitrile hydratase pathway.

A

\[
\begin{align*}
\text{Nitrilase} & \\
\text{R-CN + 2H}_2\text{O} & \rightarrow \text{R-COOH + NH}_3
\end{align*}
\]

B

\[
\begin{align*}
\text{Nitrile hydratase} & \\
\text{R-CN + H}_2\text{O} & \rightarrow \text{R-CONH}_2 \\
\text{Amidase} & \\
\text{R-CONH}_2 + \text{H}_2\text{O} & \rightarrow \text{R-COOH + NH}_3
\end{align*}
\]

The two pathways exist in both the Eucaryotic and Prokaryotic domains, but, to the authors knowledge there are currently no reports of nitrile-degrading activity in the Archaea. Although the majority of the enzymes described are bacterial, there are reports of nitrilase activity in three of the 21 plant families (Graminae, Cruciferae and Musaceae) (Thimann and Mahadevan, 1964), as well as several fungal species (Fusarium solani, Aspergillus niger and Penicillium chrysogenum) (Harper, 1977). However, it is impossible to assess the true diversity of these enzymes without more extensive screening data (Cowan et al., 1998).
1.2 Nitrilase

Nitrilases catalyse the direct conversion of nitriles to their corresponding acids. They form a branch of the superfamily (consisting of 13 branches) of carbon-nitrogen hydrolysing thiol enzymes, to which they give their name: the Nitrilase superfamily (Pace and Brenner, 2001). Members of the nitrilase branch of the superfamily are broadly distributed throughout the living world and can be found in plants, animals (C. elegans), fungi and many types of bacteria.

1.2.1 Composition

All Nases characterised to date are composed of a single subunit. Subunit sizes vary considerably, from 23kDa to 76kDa, with the average size being 40-45kDa (Table 1.1). Whilst several organisms are reported to produce functional Nase monomers (Comamonas testosteroni (Levy-schil et al., 1995) and Arthrobacter sp. J-1 (Bandyopadhyay et al., 1986)), the majority are homomultimers. Aggregation of non-specific numbers of subunits has also been reported (Goldlust and Bohak, 1989). This phenomenon appears to be due to a substrate activation process, whereby the presence of substrate induces a conformational change, which leads to aggregation and a resultant increase in activity, known as cooperative activation (Stevenson et al., 1992, Yamamoto et al., 1996).

Generally, the sequence similarities between Nases are low (Cowan et al., 1998, Kobayashi et al., 1992a). However, three residues: glutamate, lysine and cysteine are strictly conserved throughout the nitrilase superfamily, the cysteine is involved in a CWE motif that is conserved throughout the nitrilase branch (Pace and Brenner, 2001).

Mutagenesis studies on the Nases of Rhodococcus rhodochrous J1; R. rhodochrous K22 and Alcaligenes faecalis JM3 have demonstrated that C165 (R. rhodochrous J1 numbering) is essential for catalytic activity (Kobayashi et al., 1993, Kobayashi et al., 1992a, Kobayashi et al., 1992c).
1.2.2 Reaction mechanism

The centrally conserved cysteine residue of the nitrilase family is believed to act as the active site nucleophile (Novo et al., 1995). This hypothesis is further supported by several studies that demonstrate that mutation of this residue yields completely inactive enzymes (Kobayashi et al., 1993, Kobayashi et al., 1992a, Kobayashi et al., 1992c) and by a recent report of the generation of nitrilase activity in the cysteine protease papain by a Gln19Glu mutation (Dufour et al., 1995).

The reaction mechanism proposed for the Nase enzymes involves a nucleophilic attack on the nitrile carbon atom by the sulphhydril group of this cysteine with concomitant protonation of the nitrogen to form a thioimidate. This complex is then hydrolysed to an acylenzyme complex - generating ammonia, followed by a second hydrolysis – liberating the acid and regenerated enzyme (Fig.1.2) (Mahadevan and Thimann, 1964, Pace and Brenner, 2001, Stevenson et al., 1992)

Fig. 1.2: Nitrilase mode of action.

Hydrolysis of benzonitrile by *F. solani* Nase and phenylacetonitrile by *Rhodococcus* ATCC 39484 Nase results in small percentages of the corresponding amide in the product (Stevenson et al., 1992). It has been suggested that since the rate of breakdown of the covalent intermediate is the rate limiting step, that a tetrahedral intermediate is formed, which for certain substrates can break down to amide rather than the usual acid product (Stevenson et al., 1992).
1.2.3 Specificity

Early reports generally described the Nases as aromatic specific (Harper, 1977, McBride et al., 1986). However, over the past 12 years, a number of enzymes with a wide range of activities have been reported (Table 1.1), including the hydrolysis of aliphatic mono- and dinitriles (Cramp et al., 1997b, Kobayashi et al., 1990). Many of the Nases exhibit enantioselectivity and their ability to produce optically pure products is of particular commercial interest. The Nase from Acinetobacter sp. AK226 has been shown to convert racemic 2-(4'-isobutylphenyl) propionitrile to (S)-(+) 2-(4'-isobutylphenyl) propionic acid (S-(+)-ibuprofen) with a 180 fold higher activity on the S-(+) substrate than the R-(−)-substrate (Yamamoto et al., 1990). Similar enantioselectivity has been demonstrated by R. rhodochrous NCIMB 11216 Nase for the (+) enantiomer of racemic (+/−) 2-methylhexanitrile (Gradley et al., 1994).
<table>
<thead>
<tr>
<th>Microorganism</th>
<th>Native and (subunit) MW (kDa)</th>
<th>pH&lt;sub&gt;opt&lt;/sub&gt;</th>
<th>T&lt;sub&gt;opt&lt;/sub&gt; (°C)</th>
<th>Substrate specificity</th>
<th>Inducer</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Acinetobacter</em> sp. AK226</td>
<td>580 (43)</td>
<td>8</td>
<td>50</td>
<td>*Aliphatic, aromatic, heterocyclic</td>
<td>constitutive</td>
<td>(Yamamoto et al., 1990)</td>
</tr>
<tr>
<td><em>Alcaligenes faecalis</em> JM3</td>
<td>275 (44)</td>
<td>7.5</td>
<td>45</td>
<td>Arylacetonitriles</td>
<td>isovaleronitrile</td>
<td>(Kobayashi et al., 1993)</td>
</tr>
<tr>
<td><em>A. faecalis</em> ATCC 8750</td>
<td>460 (32)</td>
<td>7.5</td>
<td>40-45</td>
<td>*Substituted aliphatic nitriles</td>
<td>n-butyronitrile</td>
<td>(Yamamoto et al., 1992)</td>
</tr>
<tr>
<td><em>Arthrobacter</em> sp. J-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nase A</td>
<td>30</td>
<td>8.5</td>
<td>40</td>
<td>Aromatic, heterocyclic</td>
<td>benzonitrile</td>
<td>(Bandyopadhyay, 1986)</td>
</tr>
<tr>
<td>Nase B</td>
<td>23</td>
<td>7.5</td>
<td>30</td>
<td>Aromatic, heterocyclic</td>
<td>benzonitrile</td>
<td></td>
</tr>
<tr>
<td><em>B. pallidus</em> DAC521</td>
<td>165 (33)</td>
<td>6-9</td>
<td>65</td>
<td>Aliphatic, aromatic, heterocyclic</td>
<td>benzonitrile</td>
<td>(Cowan et al., 1998)</td>
</tr>
<tr>
<td><em>F. oxysporium</em></td>
<td>550 (37)</td>
<td>6-11</td>
<td>40</td>
<td>Aromatic, heterocyclic</td>
<td>Aromatic, heterocyclic</td>
<td>(Goldlust and Bohak, 1989)</td>
</tr>
<tr>
<td><em>F. solani</em></td>
<td>620 (76)</td>
<td>7.8-9.1</td>
<td></td>
<td>Aromatic, heterocyclic</td>
<td>benzonitrile</td>
<td>(Harper, 1977)</td>
</tr>
<tr>
<td><em>Klebsiella ozaenae</em></td>
<td>72 (37)</td>
<td>-</td>
<td>-</td>
<td>Bromoxynil, related aromatic</td>
<td>Bromoxynil</td>
<td>(McBride et al., 1986)</td>
</tr>
<tr>
<td><em>Nocardia</em> NCIMB 11215</td>
<td>560 (46)</td>
<td>7.0-7.5</td>
<td></td>
<td>Aromatic, heterocyclic</td>
<td>Benzonitrile</td>
<td>(Harper, 1985)</td>
</tr>
<tr>
<td><em>Rhodococcus</em> ATCC 39484</td>
<td>560 (40)</td>
<td>7.5</td>
<td>40</td>
<td>*Aromatic, heterocyclic</td>
<td>isovaleronitrile</td>
<td>(Stevenson, 1992)</td>
</tr>
<tr>
<td><em>R. rhodochrous</em> K22</td>
<td>650 (41)</td>
<td>5.5</td>
<td>50</td>
<td>Aliphatic</td>
<td>crotonitrile</td>
<td>(Kobayashi et al., 1992c)</td>
</tr>
<tr>
<td><em>R. rhodochrous</em> J1</td>
<td>78 (41)</td>
<td>7.6</td>
<td>45</td>
<td>*Aromatic, heterocyclic</td>
<td>isovaleronitrile</td>
<td>(Kobayashi et al., 1992a, Kobayashi et al., 1989)</td>
</tr>
</tbody>
</table>

Table 1.1: Molecular and functional characteristics of nitrilases.

¹Thermophilic organism  *Enantioselective  ºRegioselective
1.3 Nitrile Hydratase

Since the initial isolation of an NHase from *Nocardia rhodochrous* LL-100-21 (Asano *et al.*, 1980), a wide variety of microorganisms producing NHases have been described and their enzymes characterised (Table 1.3). The *Rhodococcus* genera dominate the known NHase producers, although recently a number of NHases from thermophilic *Bacillus* sp. have been described (Cramp and Cowan, 1999, Padmakumar and Oriel, 1999, Pereira *et al.*, 1998, Takashima *et al.*, 1998).

1.3.1 Size

Typically, NHases are heteromultimers composed of two distinct subunits, α and β. The subunits have similar molecular weights ranging from 22 to 28kDa; by convention the β subunit is the larger. It is commonly held that whilst the minimal functional unit is an αβ dimer, NHases are usually heterotetrameric in structure - (αβ)₂. *R. rhodochrous* J1 produces two NHases: L-NHase, which is tetrameric and H-NHase, which forms multimers of up to 11 αβ dimers (Nagasawa *et al.*, 1991). The molecular weight of *Rhodococcus* sp. N-771 in solution appears to be in equilibrium between dimer and tetramer, in the concentration range from $10^{-2} - 1\text{mg/ml}$, where the tetramer predominates (Nakasako *et al.*, 1999), similarly in the three crystal structures reported so far, the asymmetric unit consists of two dimers (Huang *et al.*, 1997, Miyanaga *et al.*, 2001, Nagashima *et al.*, 1998). As NHase concentration in the cell is often high, these results suggest that the natural state *in vivo* is the heterotetramer.

1.3.2 Cofactors

Early studies of NHases indicated that pyrroloquinoline quinone (PQQ) was a cofactor for *Rhodococcus* sp. R312 (Nagasawa and Yamada, 1987). However, this hypothesis was definitively disproved with the resolution of the enzyme's crystal structure (Huang *et al.*, 1997). The crystal structures described to date (Huang *et al.*, 1997, Miyanaga *et al.*, 2001, Nagashima *et al.*, 1998), together with numerous spectroscopic studies, have shown that NHases contain either a low-spin non-heme Fe (III) ion per αβ dimer, or a non-corronoid Co (III) ion per αβ dimer. These differences form the basis of the subdivision into two groups: (1) the Fe-type NHases and (2) the Co-type NHases. The association of these cofactors with the active centre of the enzyme is discussed in the following section.
<table>
<thead>
<tr>
<th>Microorganism</th>
<th>Native MW (kDa)</th>
<th>Subunit MW (kDa)</th>
<th>Cofactor</th>
<th>$T_{out}$ (°C)</th>
<th>$pH_{out}$</th>
<th>Substrate preference</th>
<th>Production</th>
<th>Coding of structural genes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. tumefaciens</td>
<td>69</td>
<td>$\alpha27 \beta27$</td>
<td>-</td>
<td>40</td>
<td>7.0</td>
<td>Aromatic, cyclic enantioselective</td>
<td>Inducible</td>
<td>-</td>
<td>(Bauer et al., 1994)</td>
</tr>
<tr>
<td>Bacillus sp. BR449*</td>
<td>-</td>
<td>$\alpha25 \beta27$</td>
<td>-</td>
<td>55</td>
<td>7.5</td>
<td>-</td>
<td>Constitutive</td>
<td>$\beta-\alpha$</td>
<td>(Kim and Oriel, 2000)</td>
</tr>
<tr>
<td>Bacillus sp. RAPc8*</td>
<td>110</td>
<td>$\alpha28 \beta29$</td>
<td>-</td>
<td>60</td>
<td>7.0</td>
<td>Aliphatic, cyclic aliphatic</td>
<td>Constitutive</td>
<td>-</td>
<td>(Pereira, 1998)</td>
</tr>
<tr>
<td>B. pallidus DAC 521*</td>
<td>110</td>
<td>$\alpha27 \beta29$</td>
<td>Co$^{\text{III}}$</td>
<td>50</td>
<td>7.0-7.5</td>
<td>Aliphatic</td>
<td>Constitutive</td>
<td>-</td>
<td>(Cramp, 1997)</td>
</tr>
<tr>
<td>B. smithii*</td>
<td>130</td>
<td>$\alpha26 \beta29$</td>
<td>Co$^{\text{III}}$</td>
<td>40</td>
<td>7.2</td>
<td>Aliphatic</td>
<td>Constitutive</td>
<td>$\beta-\alpha$</td>
<td>(Takashima et al., 1998)</td>
</tr>
<tr>
<td>P. chlororaphis</td>
<td>100</td>
<td>$\alpha22 \beta25$</td>
<td>Fe$^{\text{III}}$</td>
<td>20</td>
<td>6.0-7.5</td>
<td>Aliphatic</td>
<td>Inducible</td>
<td>$\alpha-\beta$</td>
<td>(Nagasawa et al., 1987)</td>
</tr>
<tr>
<td>P. putida</td>
<td>54, 95</td>
<td>$\alpha23 \beta24$</td>
<td>Co$^{\text{III}}$</td>
<td>30</td>
<td>7.2</td>
<td>Aliphatic, dinitriles enantioselective</td>
<td>-</td>
<td>$\alpha-\beta$</td>
<td>(Payne et al., 1997)</td>
</tr>
<tr>
<td>Ps. thermophila *</td>
<td>-</td>
<td>$\alpha29 \beta32$</td>
<td>Co$^{\text{III}}$</td>
<td>55</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$\beta-\alpha$</td>
<td>(Yamaki et al., 1997)</td>
</tr>
<tr>
<td>R. rhodochrous J1 H-NHase</td>
<td>530</td>
<td>$\alpha23 \beta26$</td>
<td>Co$^{\text{III}}$</td>
<td>35-40</td>
<td>6.5-6.8</td>
<td>Aliphatic, aromatic</td>
<td>Inducible</td>
<td>$\beta-\alpha$</td>
<td>(Komeda et al., 1996a)</td>
</tr>
<tr>
<td>R. rhodochrous J1 L-NHase</td>
<td>96</td>
<td>$\alpha23 \beta25$</td>
<td>Co$^{\text{III}}$</td>
<td>40</td>
<td>8.8</td>
<td>Aliphatic, aromatic</td>
<td>Inducible</td>
<td>$\beta-\alpha$</td>
<td>(Komeda et al., 1996b)</td>
</tr>
<tr>
<td>R. erythropolis</td>
<td>-</td>
<td>$\alpha23 \beta24$</td>
<td>Fe$^{\text{III}}$</td>
<td>-</td>
<td>-</td>
<td>Aliphatic, aromatic</td>
<td>Constitutive</td>
<td>$\alpha-\beta$</td>
<td>(Duran et al., 1993)</td>
</tr>
<tr>
<td>Rhodococcus sp. R312</td>
<td>95</td>
<td>$\alpha23 \beta24$</td>
<td>Fe$^{\text{III}}$</td>
<td>35</td>
<td>7-8.5</td>
<td>Aliphatic, cyclic</td>
<td>Constitutive</td>
<td>$\alpha-\beta$</td>
<td>(Nagasawa et al., 1986)</td>
</tr>
<tr>
<td>Rhodococcus sp. N-774</td>
<td>90</td>
<td>$\alpha23 \beta24$</td>
<td>Fe$^{\text{III}}$</td>
<td>35</td>
<td>7.0-8.5</td>
<td>Aliphatic</td>
<td>Constitutive</td>
<td>$\alpha-\beta$</td>
<td>(Hashimoto et al., 1994)</td>
</tr>
</tbody>
</table>

**Table 1.3:** Structural and functional properties of NHases. *Denotes the thermophilic organisms - Denotes undetermined.
1.3.3 Photoactivation

The NHases from *Rhodococcus* sp. R312 (Duran *et al.*, 1992), *Rhodococcus* sp. N-771 (Nagamune *et al.*, 1990), *Rhodococcus* sp. N-774 (Nakajima *et al.*, 1987) (which share identical amino acid sequences), *R. erythropolis* (Cummings *et al.*, 1995), and *C. testosteroni* NII (Bonnet *et al.*, 1997) have all been shown to display a unique photoreactivity; both *in vivo* and *in vitro*. The enzymes are inactive in the dark but activated upon light irradiation. Incubation in the dark inactivates the active form. Spectroscopic studies of the enzyme of *Rhodococcus* sp. N-771 indicated that the photoreactive site was an iron complex located on the alpha subunit (Tsujimura *et al.*, 1996). The inactivation was found to be due to a nitric oxide (NO) bound directly to the non-heme iron (Tsujimura *et al.*, 1997), which dissociates upon light irradiation causing a local conformational change of the iron centre α subunit residues and subsequent activation of the enzyme (Nagashima *et al.*, 1998). This was the first example of NO regulation of enzyme activity in bacteria, though at present its physiological significance is not known.

1.3.4 Structure

The structural characterisation of NHase has, to date, been chiefly focused on the iron-NHases, particularly that of *Rhodococcus* sp. R312 NHase (Brennan *et al.*, 1996b, Jin *et al.*, 1993, Nelson *et al.*, 1991, Scarrow *et al.*, 1996). Initial studies involving a variety of spectroscopic techniques indicated that the active site contained a six-coordinate, low spin Fe^{3+} ion with an oxygen, two sulphur and three nitrogen ligands. The nitrogen ligands were concluded to be the imidazole nitrogens of histidine residues and the oxygen was thought to be an hydroxide ion (Jin *et al.*, 1993, Nelson *et al.*, 1991).

These data have been re-evaluated following the determination of the crystal structures of photoactivated *Rhodococcus* sp. R312 NHase (Huang *et al.*, 1997) and nitrosylated NHase from *Rhodococcus* sp. N-771 (Nagashima *et al.*, 1998) to 2.3Å and 1.7Å respectively. The two structures are very similar, indicating that the conformation is conserved between active and inactive states (Nagashima *et al.*, 1998). In each case, the crystallographic asymmetric unit is made up of two heterodimers (αβ)_2. The alpha subunit consists of a long N-terminal arm containing two α-helices, which interact with the β-subunit helix domain, and a globular C-terminal domain made up of a novel αββα fold. The β subunit also consists of a long N-terminal loop that wraps around the α
subunit, a helical domain, and a C-terminal domain that is folded into a β roll. The two subunits form a tight dimer (Fig. 1.3), stabilised by the N-terminal loops of each subunit that wrap around the opposite subunit; these interactions suggest that dimer formation is a dynamic rather than simple docking process (Nakasako et al., 1999). The electrostatic properties of the interfacial surfaces of each subunit are quite different; Nakasako et al. (1999) proposed that up to 50 hydration water molecules which are densely distributed over the negatively charged β subunit surface may aid in dimer stabilisation by moderating this charge and hence decreasing the free energy in heterodimer association.

![Fig. 1.3: Stereoview of the αβ heterodimers of Rhodococcus sp. R312 NHase. The location of the iron centre is indicated by a red sphere, the α subunit is in blue, the β subunit in yellow. (Taken from Huang et al., 1997)](image)

The iron centre is located in a central cavity formed by the subunit interface. All protein ligands to the iron are provided by the α subunit. In accordance to the earlier spectroscopic data (Brennan et al., 1996b, Jin et al., 1993, Nelson et al., 1991, Scarrow et al., 1996), the ligands were found on five vertices of an octahedron. However they did not comply with the 2S, 3N and O prediction (Jin et al., 1993, Nelson et al., 1991) and no histidine residues were found within the cavity. In fact, the ligands came from three cysteine thiolates (αC110, 113 and 115) and two main-chain nitrogen atoms (αS114 and αC115). Huang et al. (1997) reported that the sixth position was unoccupied, whilst Nagashima et al. (1998) reported a nitric oxide ligand at this position.
position, illustrating the difference between the active and inactive states. The residues Cys113 and Cys115 of *Rhodococcus* sp. N-771 are post-translationally oxidised to cysteine-sulphinic (-SO\(_2\)H) and sulphenic acids (-SOH) respectively in the NO inactivated enzyme whilst just the sulphinic group is found in the light-activated enzyme (Kobayashi and Shimizu, 1998). Despite having an identical amino acid sequence and photoreactive behaviour, the crystal structure of *Rhodococcus* sp. R312 NHase does not display such modifications. This difference was attributed to the lower resolution achieved in this study rather than structural differences with species N-771 (Nagashima et al., 1998). The NO molecule is thought to be stabilised by the oxygen atoms of the three cysteines, which protrude into the iron centre termed by Nagashima et al. (1998) as a “claw setting”. Light irradiation is thought to break the Fe-N (NO) bond and induce a structural change in the claw setting. \(\alpha\)Cys113 and \(\alpha\)Cys115 form hydrogen bonds with \(\beta\)Arg56 and \(\beta\)Arg141, which are conserved through all known NHases (Huang et al., 1997, Nagashima et al., 1998). Mutation of these residues results in loss of activity, suggesting that they are involved in the stability of the claw setting and that this setting is important for NHase activity (Piersma et al., 2000). A channel from the bulk solvent to the active centre is formed by the subunit interface, involving five residues from the \(\alpha\) subunit and four from the \(\beta\) subunit. A cluster of 10 aromatic residues is found around this region and is believed to stabilise the channel (Nagashima et al., 1998). The entrance of the inactive N-771 enzyme, 10 Å from the centre and only 4 Å wide, is too narrow for substrates to enter. It is suggested therefore that dynamic structural changes, as are observed during light-activation, are required to open and close the channel during the catalytic cycle.

Until recently, little structural information was available concerning cobalt NHases. EPR and EXAFS spectroscopy on the H-NHase of *R. rhodochrous* J1 indicated that that the ligand environment was very similar to the iron NHases (Brennan et al., 1996a). The enzyme was found to contain a non-corrin Co\(^{3+}\) ion with a mixed S and N octahedral ligand field. However, these studies ruled out the possibility of NO coordination to the cobalt ion. The -V-C-(T/S)-L-C-S-C- sequence that makes up the cofactor liganding part of the active site is highly conserved amongst all NHases, the single difference being that cobalt NHases have threonine in the third position whereas the iron NHases have serine. The difference in cofactors may be ascribed to the difference in sequence (Kobayashi and Shimizu, 1998). The crystal structure of
Pseudonocardia thermophila NHase - a thermostable Co-type NHase was published very recently (Miyanaga et al., 2001). Resolved to 1.8 Å, the structure exhibits a very high similarity with the Fe-type NHase. The structure around the active centre is almost identical (Fig. 1.4), and the cofactor ligands αC111 and αC113 are post translationally modified to form a similar claw setting to the Fe-type NHase.

The major deviation from the Fe-type structures involves the residues between β95 and β138. The central portion of this region corresponds in alignments (Figs. 3.19 and 5.1) to an insertion in the Co-type NHases (cf. the Fe-type) of between 12 and 21 residues (20 residues in the Ps. thermophila enzyme). The Ps. thermophila NHase structure demonstrates that this region comprises an α helix (residues β111-β125), which interacts with a second α helix on the external surface of the α subunit. It has been proposed that this interaction contributes to the thermostability of the Ps. thermophila enzyme (Miyanaga et al., 2001). However, it should be noted that similar sequences are also found in the Co-type NHases of Rhodococcus species. (Mayaux et al., 1991) and both R. rhodochrous J1 NHases (Kobayashi et al., 1991), none of which are particularly thermostable enzymes (Section 1.3.7). It is also notable that an active site residue βW72 of the Ps. thermophila enzyme (conserved in the Co-type NHases) is orientated in a different direction to βY76 of Rhodococcus sp. N-771 (conserved in Fe-type NHases) (Nagashima et al., 1998). This difference in orientation leads to a larger space in the substrate-binding site, which may account for the general differences in substrate specificities (Section 1.3.6) of the two groups (Miyanaga et al., 2001).

Recent studies with the NHase of Rhodococcus sp. N-771 revealed that the enzyme showed activity when the iron cofactor was substituted with cobalt ions, though oxidation of both the cobalt to Co$^{3+}$ and αCys113 to the cysteine-sulphinic acid was essential for this activity (Nojiri et al., 2000). This study supports the hypothesis that Co and Fe-type NHases share very similar structures and, presumably, reaction mechanisms.
Fig. 1.4: Superimposition of the Co-type NHase and the Fe-type NHase from *Rhodococcus* sp. N-771. The α (green) and β (blue) subunits of Co-type NHase, and α (purple) and β (orange) subunits of Fe-type are shown in different colors. The magenta sphere is the cobalt atom in the active center, and the cyan one is the iron atom. A: Active sites of the Co-type and the Fe-type NHase. The side chains of residues of the Co-type NHase are represented by thick sticks and those of the Fe-type NHase by thin ones. B: Ribbon diagram of the Co-type and Fe-type NHases. Helices interacting between the α and β subunits (α36-α49, and β111-β125) are enriched by red dotted lines. (Taken from Miyanaga *et al.* 2001).
1.3.5 Reaction Mechanism
A number of mechanisms have been proposed for the hydrolysis of nitriles by NHases (Huang et al., 1997, Kobayashi, 1997, Nagasawa et al., 1986). Based on the effects of addition of substrate or competitive inhibitors on the EPR and Resonance Raman spectra, it is thought that the metal ion is involved as a Lewis-acid in catalysis (Huang et al., 1997).

Three reaction models involving the metal ion as a Lewis acid have been proposed (Huang et al., 1997):

1. The nitrile substrate binds directly to the metal ion, thereby activating it for hydration. The nitrile carbon atom is then subjected to nucleophilic attack by a water molecule held in place by a proximal residue (Tyr 72 has been indicated) (Fig. 1.5a).

2. The nitrile substrate approaches a metal bound hydroxide ion, which acts as a nucleophile on the nitrile carbon atom (Fig. 1.5b).

3. The final mechanism does not involve an enzyme bound intermediate, thus avoiding ligand exchange. A metal-bound hydroxide activates a water molecule, which then acts as the nucleophile, attacking the nitrile carbon atom (Fig. 1.5c).

Early hypotheses suggested that one of the conserved cysteine thiolates acted as a nucleophile, followed by hydrolysis of the resulting thioimidate (Kopf et al., 1996). These have largely been discarded following the observation that those residues are coordinated to the metal ion (Huang et al., 1997, Miyanaga et al., 2001, Nagashima et al., 1998). Due to the slow rates of ligand exchange for trivalent metals, in particular for Co"", it has been argued that mechanism 3 is the most likely since this does not involve ligand exchange (Huang et al., 1997). Furthermore, the observation that the rates of hydration of certain nitriles are very similar between some Co-type and Fe-type NHases, whilst the exchange kinetics of low spin Fe"" and Co"" are very different has also been interpreted as an indication that mechanism 3 is used (although there is currently no empirical evidence for this) (Mascharak, 2002). However, as reported in the crystal structure analysis of Rhodococcus sp. R312 NHase (Huang et al., 1997), the
substrate analogue iodoacetonitrile is directly associated with the metal ion – replacing the solvent derived hydroxide ligand (Doan et al., 1999). Furthermore, Co-chelate complexes have shown to bind the nitrile reagent, leading to the suggestion that direct association model is the correct mechanism (Nakasako et al., 1999).

Fig. 1.5: Three putative reaction mechanisms of NHase.

A

B

C
1.3.6 Specificity

The general trend of substrate specificity amongst the NHases characterised to date seems to be almost exclusively for aliphatic nitriles. However, some enzymes with activity on aromatic nitriles have been described (Table 1.3). It has been claimed that the iron-containing NHases are specific for aliphatic nitriles whereas the cobalt-containing NHases are specific for aromatic nitriles (Kobayashi and Shimizu, 1998, Miyanaga et al., 2001), although this is not strictly true. The Fe-type NHase of R. erythropolis showed broad substrate specificity, and was capable of hydrating aliphatic, aromatic and heterocyclic nitrile compounds (Duran et al., 1993) whilst the Co-type enzymes of B. smithii, B. pallidus DAC521, Bacillus sp. RAPc8 and P. putida demonstrate no activity on aromatic nitriles (Cramp and Cowan, 1999, Payne et al., 1997, Pereira et al., 1998, Takashima et al., 1998).

It has been suggested that the lack of aromatic specificity is due to the narrow entrance channel from the bulk solvent, which enables only small, aliphatic molecules to enter the catalytic site (Nakasako et al., 1999). A second hypothesis involving the spatial arrangement of the catalytic cavity suggests that a larger cavity caused by the displacement by a leucine residue and a phenylalanine residue of βW72 of Ps. thermophila NHase (in comparison to the corresponding βY76 of Rhodococcus sp. N-771) may be responsible for the differences in substrate preferences between Co-type and Fe-type NHases.

Several NHases have been described with very broad substrate ranges. Bacillus sp. RAPc8 demonstrated activity only on aliphatic nitriles but catalysed the hydration of branched, linear, cyclic-aliphatic and dinitriles. The highest affinity recorded was for acetonitrile, with a $K_M$ value of 8.8mM (Pereira, 1998). Both the L-NHase and H-NHase of R. rhodochrous J1 have broad substrate ranges, with similar $K_M$ values for the aliphatic nitriles (e.g., for acrylonitrile, $K_M$ values of 2.67mM and 1.89mM and specific activities of 828U/mg and 1760U/mg for L-NHase and H-NHase respectively). However, the affinity and activity of L-NHase for aromatic nitriles was significantly higher than H-NHase (Wieser et al., 1998).

There are also reports of NHases with enantioselectivity. Agrobacterium tumefaciens d3 NHase displays activity on aromatic and heterocyclic nitriles and produces
(S)-amides enantioselectively (>90% enantiomeric excess) from a number of racemic phenylpropionitriles (Bauer et al., 1994). *P. putida* NHase also produces (S)-amides, for example racemic 2-(4-chlorophenyl)-3-methylbutyronitrile was converted to the (S)-amide at >90% enantiomeric excess (Fallon et al., 1997).

### 1.3.7 Stability

Most NHases are particularly thermolabile although it has been demonstrated that *n*-butyric acid acts as a potent stabilising agent (Nagasawa et al., 1987). There appears to be a relation between stability and cofactor, with Co-type NHases being generally more stable than Fe-type (Tables 1.2 and 1.3). In the past five years, at least five NHases from moderate thermophiles have been characterised, four of which are thermophilic *Bacillus* isolates, the fifth a *Ps. thermophila* strain. With the exception of *B. pallidus* DAC521 (for which the cofactor is not known), all are Co-type enzymes.

Due to the different approaches of the various researchers in describing thermal stability, comparison between the enzymes is not straightforward. Nevertheless, Table 1.2 outlines the stability of some of the better-characterised enzymes.
Table 1.2: Thermal stability of NHases as defined by remaining activity following incubation for a predetermined time period at a specific temperature.

<table>
<thead>
<tr>
<th>Microorganism</th>
<th>Temperature (°C)</th>
<th>Incubation time</th>
<th>% Activity remaining&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Bacillus</em> BR449&lt;sup&gt;*&lt;/sup&gt;</td>
<td>60</td>
<td>2hrs</td>
<td>100</td>
<td>(Padmakumar and Oriel, 1999)</td>
</tr>
<tr>
<td><em>Bacillus</em> RAPc8&lt;sup&gt;*&lt;/sup&gt;</td>
<td>50</td>
<td>2.5hrs</td>
<td>50</td>
<td>(Pereira, 1998)</td>
</tr>
<tr>
<td><em>B. pallidus</em> DAC521&lt;sup&gt;*&lt;/sup&gt;</td>
<td>50</td>
<td>51min</td>
<td>50</td>
<td>(Cramp, 1997)</td>
</tr>
<tr>
<td><em>Bacillus smithii</em>&lt;sup&gt;*&lt;/sup&gt;</td>
<td>55</td>
<td>1.5hrs</td>
<td>50</td>
<td>(Takashima et al., 1998)</td>
</tr>
<tr>
<td><em>Ps. thermophila</em>&lt;sup&gt;a&lt;/sup&gt;</td>
<td>50</td>
<td>2hrs</td>
<td>100</td>
<td>(Yamaki et al., 1997)</td>
</tr>
<tr>
<td><em>P. chlororaphis</em>&lt;sup&gt;a&lt;/sup&gt;</td>
<td>20</td>
<td>10min</td>
<td>100</td>
<td>(Nagasawa et al., 1987)</td>
</tr>
<tr>
<td><em>P. putida</em></td>
<td>50</td>
<td>10min</td>
<td>0</td>
<td>(Fallon et al., 1997)</td>
</tr>
<tr>
<td><em>Rhodococcus</em> sp. N-774</td>
<td>30</td>
<td>30min</td>
<td>100</td>
<td>(Nagasawa and Yamada, 1995)</td>
</tr>
<tr>
<td><em>R. rhodochrous</em> J1 H-NHase&lt;sup&gt;a&lt;/sup&gt;</td>
<td>50</td>
<td>30min</td>
<td>100</td>
<td>(Nagasawa et al., 1991)</td>
</tr>
<tr>
<td><em>R. rhodochrous</em> J1 L-NHase</td>
<td>30</td>
<td>30min</td>
<td>100</td>
<td>(Wieser et al., 1998)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Denotes the thermophilic organisms
<sup>1</sup>Where 100% is given as remaining activity, authors reported that activity levels dropped dramatically above those temperatures.
<sup>b</sup>Stabilised in 34mM n-butyric acid
<sup>c</sup>Stabilised in 22mM n-butyric acid
<sup>d</sup>Stabilised in 44mM n-butyric acid

1.3.8 Gene structure

Over the past 13 years, the genes for a large number of the nitrile-modifying enzymes described so far have been cloned using a variety of techniques. The classic approach has been to determine N-terminal and internal amino acid sequences of the purified proteins, from which oligonucleotides were designed either for use as labelled probes (e.g. cloning of *Rhodococcus* sp. N-774 NHase (Ikehata et al., 1989)) or to produce PCR products for use as probes (e.g. cloning of *Ps. thermophila* NHase genes (Yamaki et al., 1997)). Other cloning strategies have involved the use of the N-774 NHase genes as probes in Southern hybridisations in order to isolate the NHase genes of
R. rhodochrous J1 (Kobayashi et al., 1991) and R. erythropolis (Duran et al., 1993), whilst PCR primers designed against the DNA sequence of the B. smithii NHase genes (Takashima et al., 1995) have been used in PCR for the isolation of the corresponding genes of Bacillus BR449 (Kim and Oriel, 2000).

The α and β subunit genes are of similar size and range from 609bp to 660bp for the α subunit and from 636bp to 706bp for the β subunit. The genes are found in both α-β and β-α orders, usually separated by a short sequence of between 16bp and 29bp (Duran et al., 1993, Kim and Oriel, 2000). With the exception of the R. rhodochrous J1 genes, all NHases characterised to date are found just downstream of an amidase gene. No amidase sequences were found close to the H-NHase genes of R. rhodochrous J1 although an amidase gene was found 1.9kb downstream of the α subunit gene of the L-NHase gene (Kobayashi et al., 1992b). The close location of the NHase and amidase genes in nitrile metabolising organisms supports the hypothesis that these enzymes are involved in a two-step reaction in the nitrile degradation pathway. No transcription terminators have been found between the amidase, α or β genes although stem-loop structures that may act as transcription terminators have been identified immediately downstream of the β subunit genes of P. chlororaphis (Nishiyama et al., 1991) and Rhodococcus sp. R312 (Bigey et al., 1999).

The translated peptide sequences of the NHases reveal a remarkably high degree of sequence conservation despite the fact that many originate from distantly related organisms. Generally the α subunit is more conserved than the β, containing 43 strictly conserved amino acids, whilst the β subunit contains 28 (Payne et al., 1997). A number of conserved sections in the amino acid sequences reflect regions of structural and/or functional significance. Of particular note is the sequence homology of the β subunit N-terminal sequences (Fig. 1.6), which may be explained by the observation from crystallographic data that this region is involved in the stabilisation of the subunit interface (Huang et al., 1997, Miyanaga et al., 2001, Nagashima et al., 1998). The region of the α subunit containing the cysteine ligands is also highly conserved, containing 17 strictly conserved residues conserved residues in a sequence composed of 28 residues. The region contains the characteristic cofactor-binding motif VC(T/S)LCSC(Y/T) – a stretch of eight residues – six of which are strictly conserved and two positions strictly conserved as threonine and tyrosine or serine and threonine in
the Co-type and Fe-type NHases respectively (Fig. 1.7) (Payne et al., 1997). It has been suggested that these two differences are the determinants of the identity of the cofactor bound by a particular NHase (Kobayashi and Shimizu, 1998).

Fig. 1.6: Alignment of NHase β subunit N-terminal sequences, shaded at 60% conservation.

Fig. 1.7: Alignment of the conserved NHase α subunit sequences containing the characteristic cofactor binding motif VC(T/S)LCSC. The alignment is shaded at 60% conservation.

1.3.9 Gene regulation and expression
Requirements for the production of active NHase in the native organisms vary. Some NHases are produced constitutively, irrespective of the presence or absence of nitriles, amides or metal cofactor ions, whilst others have been shown to be inducible (Table 1.3). Whilst no information appears to be available on the effect of supplementing the
growth media of Fe-type NHase producers with Fe$_{2+}$ ions, experiments have shown that the specific activities of Bacillus sp. BR449, B. smithii and R. rhodochrous J1 NHases are significantly enhanced by addition of 0.1-0.5mM CoCl$_2$ to the growth media (Kobayashi et al., 1991, Padmakumar and Oriel, 1999, Takashima et al., 1998). Investigations with recombinant R. rhodochrous J1 H-NHase have shown that cobalt is required for activity rather than protein expression (Kobayashi et al., 1991, Komeda et al., 1996b), similarly, recombinant Bacillus sp. BR449 NHase is expressed in a largely inactive form in the absence of cobalt ions but can be activated upon incubation in 5µM CoCl$_2$ (Kim and Oriel, 2000, Kim et al., 2001). These findings suggest that cobalt is involved in enhancing protein folding rather than induction of NHase expression (Cowan et al., 1998, Kobayashi et al., 1992b, Kobayashi et al., 1991).

The inducible NHases are generally induced by their amide reaction products rather than by nitriles (Kobayashi and Shimizu, 1998). This is also found to be the case for the NHase-associated amidases. This observation, together with the close location of the structural genes and the absence of transcriptional terminators between them, indicates that amidases and NHases are under the control of the same regulatory elements. It has been proposed that the open reading frames encoding the NHase α and β subunits and the amidase monomer are cotranscribed as a single, polycistronic mRNA (Duran et al., 1993, Nishiyama et al., 1991). Indeed, this has been shown to be the case for the NHase operon of Rhodococcus sp. ACV2 (Bigey et al., 1999).

Early attempts to express recombinant NHase in E. coli often resulted in the formation of inactive inclusion bodies or no expression at all (Kobayashi et al., 1992b, Nishiyama et al., 1991). However, further studies revealed that in many cases, the inclusion of upstream and/or downstream open reading frames in expression constructs significantly enhanced expression of activity. Overproduction of P. putida NHase was found to require co-expression of P14K, the gene for which is found immediately downstream of the NHase β subunit (Wu et al., 1997). Similar proteins, also with β-subunit homology have been found downstream of the NHase genes of Bacillus sp. BR449 and R. rhodochrous J1 (Kim and Oriel, 2000, Komeda et al., 1996a, Komeda et al, 1996b).

Efficient production of recombinant expression of NHases of P. chlororaphis, Rhodococcus sp. N-771 and Rhodococcus sp. R312 has been found to require the coexpression of a gene coding for a 47kDa protein, located downstream of the NHase
genes (Bigey et al., 1999, Nishiyama et al., 1991, Nojiri et al., 1999). Although the precise function of these proteins is not known, it has been suggested that they are involved in incorporation of the cofactor ion into the active site (Bunch, 1998). The observation that their presence is required for the expression of constitutively produced enzymes (Bigey et al., 1999, Nojiri et al., 1999), indicates that they are not involved in regulation of expression. It is noteworthy that the β-homologue proteins are associated with the Co-type NHases whereas the P47K proteins are Fe-type associated.

1.3.9.1 Regulation of the NHase genes of R. rhodochrous J1

The regulation of the R. rhodochrous J1 genes has been particularly well characterised (Kobayashi and Shimizu, 1998). Both H-NHase and L-NHase are induced by crotonamide. However urea or cyclohexanecarboxamide induce either H-NHase or L-NHase genes respectively (Yamada and Kobayashi, 1996).

The H-NHase operon contains two regulatory genes, nhhC and nhhD, both located upstream of the NHase genes. NhhC, which is homologous to the negative regulator AmiC of the P. aeruginosa aliphatic amidase, exhibits a positive regulation in the presence of amide, of NhhD production. Presence of NhhD then induces NHase expression (Kobayashi and Shimizu, 1998).

The 5' upstream region of the L-NHase operon is also required for NHase expression; two genes, nhlD and nhlC serve positive and negative regulatory roles respectively. In the presence of amide, NhlC (also an AmiC homologue) inhibits NhlD repression of NHase production (Komeda et al, 1996b).

An additional gene, nhlF/nhhF is found within both NHase operons and has been shown to produce a cobalt uptake protein that confers cobalt uptake on Rhodococcus strains (Komeda et al., 1997) and is believed to play a role in providing cobalt to the NHase (Kobayashi and Shimizu, 1998). Both operons also contain insertion sequences, which are believed to be responsible for the unique arrangement and regulation of the R. rhodochrous J1 genes (Kobayashi and Shimizu, 1998).
**Bacillus BR449**

![Diagram of Bacillus BR449 gene clusters](image)

**R. rhodochrous J1 H-NHase**

![Diagram of R. rhodochrous J1 H-NHase gene clusters](image)

**R. rhodochrous J1 L-NHase**

![Diagram of R. rhodochrous J1 L-NHase gene clusters](image)

**P. putida 5B**

![Diagram of P. putida 5B gene clusters](image)

**Rhodococcus sp. N-771**

![Diagram of Rhodococcus sp. N-771 gene clusters](image)

**P. chlororaphis**

![Diagram of P. chlororaphis gene clusters](image)

**Rhodococcus sp. R312**

![Diagram of Rhodococcus sp. R312 gene clusters](image)

**Fig. 1.8:** Arrangement of the gene clusters from various NHase producing organisms. The direction and extent of the genes are indicated by arrows, homologous genes are indicated by the same colour shading. Stem loop structures that serve as potential transcriptional terminators are shown as black triangles.
1.4 Amidase

Amidases catalyse the hydrolysis of amides to the corresponding carboxylic acid and ammonia. They exist in all kingdoms of the living world but have been most extensively characterised amongst the bacteria.

1.4.1 Amidase classification

A number of studies of amidase classification (Chebrou et al., 1996, Fournand and Arnaud, 2001) have revealed that the bacterial aliphatic amidases (broadly classed as acylamide amidohydrolase, EC 3.5.1.4) are made up of two types.

The first group, the nitrilase-related family, includes the aliphatic amidases, hydrolysing only short-chain aliphatic amides (Fournand and Arnaud, 2001). The enzymes are typically homohexamers of approximately 230kDa, and contain a Cys166 residue (Pseudomonas aeruginosa amidase numbering), conserved across both nitrilase and amidase. This residue is believed to act as the catalytic nucleophile. The amidases from P. aeruginosa (Ambler et al., 1987), Rhodococcus sp. R312 (Azza et al., 1994), Bacillus stearothermophilus BR338 (Cheong and Oriel, 2000), and Bacillus sp. BR449 (Kim and Oriel, 2000) belong to this group (also known as wide spectrum amidases). The nitrilase-related amidases do not however contain the central conserved GGSS motif common to all other enzymes in the amidase class.

The second group forms part of the GGSS-signature amidase family (Chebrou et al., 1996) and contains the amidases that hydrolyse the mid-chain aliphatic amides, often demonstrating enantioselectivity. The enzymes of this class are homodimers with a native molecular weight of approximately 110kD. With the exception of the Bacillus sp. BR449 amidase, all NHase coupled enzymes belong to this class.

Whilst phylogenetic analyses have indicated that the wide spectrum amidases and the nitrilases seem to be descended from a common ancestor, similar analyses of the enantioselective amidases have shown an evolutionary relationship with aspartic proteinases (Kobayashi et al., 1997), indicating a convergent evolutionary pathway for the amidase activity. Comparisons of the phylogeny of this second class of amidases with the phylogeny of nitrile hydratases have revealed an almost parallel evolution, emphasising their role in nitrile metabolism (Fournand and Arnaud, 2001).
1.4.2 Amidase production

The expression of amidase genes is usually induced by amides. In the case of most NHase associated amidases, amide induction also leads to the co-expression of NHase genes immediately downstream of the amidase structural genes (generally thought to be transcribed as a single polycistronic mRNA) (Sections 1.3.8 and 1.3.9). However, some wide spectrum amidases, such as the NHase-associated amidase of *Bacillus* sp. BR449, are expressed constitutively. *Rhodococcus* sp. R312 expresses at least four different amidases: a wide spectrum amidase, an α-amino amidase, an enantioselective amidase and an adipoamidase (Maestracci *et al.*, 1984, Mayaux *et al.*, 1990, Moreau *et al.*, 1993). The differential expression of these enzymes depends on the growth conditions and choice of inducer. The expression of *Methylophilus methylotrophus* amidase has been found to be negatively regulated by ammonia, a reaction product of amide hydrolysis (Wyborn *et al.*, 1996).

1.4.3 Mechanism of action

Many amidases, including *Rhodococcus* sp. R312 wide spectrum amidase (Maestracci *et al.*, 1986) and *Ps. aeruginosa* wide spectrum amidase (Clarke, 1970) have the ability not only to transfer the acyl moiety of the amide to water, but also to hydroxylamine to form hydroxamates:

Amide hydrolysis:

\[
\text{RCONH}_2 + \text{H}_2\text{O} \rightarrow \text{RCOOH} + \text{NH}_3
\]

Amide transfer:

\[
\text{RCONH}_2 + \text{NH}_2\text{OH} \rightarrow \text{RCONHOH} + \text{NH}_3
\]

Not only does the amide transfer reaction have significant industrial potential (Fournand, 2001), but has greatly facilitated the study of amidase reaction mechanisms. It was demonstrated that amides react with the enzyme to give an acyl-enzyme complex, which is then subject to nucleophilic attack by the hydroxylamine co-substrate (Clarke, 1970, Maestracci, 1986). It is therefore assumed that the mechanism for amide hydrolysis should be the same. Once the acyl-enzyme intermediate is formed, nucleophilic attack leads to the transfer of the acyl group to the co-substrate (water or hydroxylamine), leading to the formation of carboxylates or hydroxymates (Fig. 1.9)
Fig. 1.9: Mechanism of the amidase acyl transfer reaction to: A – hydroxylamine B – water. Redrawn from Maestracci. et al., (1986).

\[ \text{A} \]

\[ \text{B} \]
1.5 Applications of NHases

1.5.1 Commodity chemical production

The enzymatic conversion of acrylonitrile to acrylamide is one of the most significant and successful applications of biotechnology in commodity chemical production and has unarguably demonstrated the commercial viability of NHase (Kobayashi et al., 1992b, Cowan et al., 1998, Thomas et al., 2002). Acrylamide is widely used industrially in coagulators and soil conditioners, for paper treatment and paper sizing, and for adhesives, paints and petroleum recovery agents (Yamada and Kobayashi, 1996). Prior to the introduction of the enzymatic process, a conventional chemical synthesis, using copper salts as a catalyst for the hydration of acrylonitrile was used. However, this process was expensive and inefficient, providing the incentive to search for superior catalysts (Kobayashi et al., 1992b). The industrial bioconversion of acrylonitrile to acrylamide started in 1985 using the Fe-type NHase of *Rhodococcus* sp. N-774, the first-generation strain employed by the Nitto Chemical Industry, Tokyo, Japan (Nagasawa and Yamada, 1989). The catalyst for this process has been upgraded twice since that date, first, in 1988 with the Fe-type NHase of *P. chlororaphis* B23, then again, in 1991 with the Co-type H-NHase of *R. rhodochrous* J1 (Nagasawa and Yamada, 1995). Currently, Mitsubishi Rayon Co., Ltd (Yokohama, Japan) (with whom Nitto Chemical Industry merged in 1998), produces over 20,000 metric tons of acrylamide per year. The process uses a series of fixed-bed reactors containing the immobilised *R. rhodochrous* J1 cells, producing acrylamide at ~99.99% yield. Overall, productivity has increased by over 50% since the days of chemical synthesis, and costs have been reduced dramatically (Thomas et al., 2002).

In 1999, Lonza Guangzhou Fine Chemicals (China) began production of nicotinamide (niacinamide, vitamin B₃) in a process developed by Lonza (Switzerland) (Thomas et al., 2002). The process involves a four-stage chemo-enzymatic synthesis, the starting material 3-methyl-1,5-diaminopentane (the nylon 6,6 byproduct) is catalytically converted into 3-picoline which is ammoxidated into 3-cyanopyridine then finally biocatalytically hydrolysed to nicotinamide using immobilised *R. rhodochrous* J1 cells induced to produce L-NHase (Heveling et al., 1998). The advantages of this process over the original chemical process include low energy uses, a low rate of emissions as well as stochiometric conversion of high concentration of the substrate to the
nicotinamide product at a purity of over 99.5% (Heveling et al., 1998, Shimzu et al., 2001).

A further example of an NHase-catalysed synthesis currently in operation is the production 5-cyanovaleramide (5-CVAM) from adiponitrile (ADN), a starting material for the synthesis of a DuPont herbicide, azafenidin (Thomas et al., 2002). This synthesis was also originally run using a chemical catalyst, however problems with low conversion rates (~25%), and significant production (~20% of converted ADN) to the byproduct adipamide (ADAM) (Fig. 1.10), solvent extraction and catalysts deactivation lead to the development of a biocatalytic process.

**Fig. 1.10:** Hydration of adiponitrile (ADN) to 5-cyanovaleramide (5-CVAM) and adipamide (ADAM), taken from Thomas et al. (2002).

Immobilised *P. chlororaphis* B23 cells are now used in the hydration of adiponitrile to 5-CVAM, the NHase of this organism regioselectively produces 5-CVAM (96% selectivity) from ADN to a yield of 93%.

NHases that show novel stereoselectivity, regiospecificity or substrate specificity may be developed to produce amides that are components of pharmaceutical compounds (Kobayashi and Shimzu, 1998). Such potential developments include the production, in conjunction with amidase, either as whole-cell systems or purified enzymic systems of optically pure amino acids, hydroxy acids and keto acids (Nagasawa and Yamada, 1989). For example, *Rhodococcus* sp. R312 has been shown to hydrolyse lactonitrile to DL-lactic acid (Thompson et al., 1988) and the stereospecific conversion of amino-propionitrile to L-alanine by whole cells of an *Acinetobacter* sp. has also been demonstrated (Macadam and Knowles, 1985)
1.5.2 Bioremediation

As was discussed in Section 1.1.1, there is a significant spread of synthetic nitriles throughout the environment, particularly from industrial wastewaters. Since conventional methods for decreasing the nitrile concentration of these effluents add significant costs to manufacturing processes, novel approaches involving low-cost bioremediation have recently been undertaken (Kobayashi and Shimizu, 1998). Two approaches have been reported for acrylonitrile decontamination of waste streams. The first describes the enzymic degradation of acrylonitrile (using either NHase or Nase) in the aqueous polymer emulsions used in raw rubber and plastic manufacture (Battisel et al., 1997). The second cites the development of a stable activated sludge system, consisting of a consortium of microbial isolates (including NHase-containing bacteria), for the treatment of the toxic waste stream produced from the large-scale manufacture of acrylonitrile (Thompson et al., 1988).

1.6 Thermophiles

Three major divisions are commonly used when classifying organisms in terms of their temperature range of growth (Herbert, 1992): the psychrophiles (-5°C to 20°C); the mesophiles (15°C to 40°C) and the thermophiles (>40°C). The thermophiles have been arbitrarily divided into three sub-categories (Cowan, 1992): moderate thermophiles, with a maximum growth temperature of 60°C; extreme thermophiles, with a maximum growth temperature of 70°C and the hyperthermophiles which have a growth range of ~60°C to a current maximum of 113°C (Blochl et al., 1997). It has been suggested that the upper limit of stability may be imposed by the stability of low molecular weight metabolites and coenzymes and hence is unlikely to be above 150°C (White, 1984).

Almost all the hyperthermophiles are archaeal; the only bacterial representatives are *Thermatoga* and *Aquifex* (Stetter, 1998). However, representatives of all three primary Kingdoms (Bacteria, Archaea and Eukarya) have been found in the moderate and extreme thermophile categories, though the temperature maximum for the eukaryotic thermophiles is typically less than 60°C (Brock and Madigan, 1991).

Thermophilic growth is supported in a range of biotopes. The heat sources may be of geothermal origin: such as hot water springs; geysers; steam vents and submarine
hydrothermal systems, or, they may be of non-geothermal origin: such as solar heated soil; decaying vegetation and heated industrial waters (Kristjansson and Stetter, 1992).

Thermostable proteins can be obtained from any source; however a fundamental requirement for survival in a high temperature environment is macromolecular stability, it is therefore reasonable to assume that thermophiles represent an obvious source of thermostable enzymes as such character will confer their proteins a degree of thermostability (Cowan, 1995).

1.7 Thermostable enzymes

Microorganisms and their enzymes are used as biocatalysts in numerous industrial, medical and environmental applications (Kristjansson, 1989, Coolbear et al., 1992, Herbert, 1992). However, biocatalysts are inherently labile, and their operational stability, in particular their thermostability, is often a limiting factor in many biotechnological applications (Illanes, 1999). Biocatalyst thermostability enables a higher operation temperature; advantageous due to higher reactivity (higher reaction rate and lower diffusion restrictions), higher process yield (increased solubility of substrates and products and favourable equilibrium displacement in endothermic reactions), lower viscosity and a reduced contamination threat (Illanes, 1999). Furthermore, many thermostable proteins have been described with a high level of resistance to degradation by proteolysis and to denaturation by organic solvents, detergents and chaotropic agents such as urea and guanidine hydrochloride (Daniel et al., 1982, Cowan, 1995). For these reasons, thermostable proteins have attracted the attention of both academia and industry.

At biologically relevant temperatures (up to 110°C), most covalent bonds in proteins are stable. The thermodynamic stability of a protein (defined as the free energy difference between the folded and unfolded state) is therefore a balance between large stabilising (non-covalent intramolecular) interactions (in the order of 1,000kJ mol⁻¹ in an average protein) and similarly large destabilising forces (primarily chain conformational entropy) (Daniel et al., 1996, Daniel and Cowan, 2000). Thus ΔG, the free energy difference between the folded and unfolded state, is only of the order of -40kJ.mol⁻¹ (Jaenicke, 1991). Point mutation studies have shown that replacement of a single amino
acid can have a significant effect on stability without any detectable effect on the three-dimensional structure (Daniel and Cowan, 2000). Such effects can be explained by the fact that typical intramolecular interactions (hydrogen bonds, salt bridges and hydrophobic interactions) can contribute between -2 and -25 kJ mol\(^{-1}\). Since the adaption of thermostable proteins is accompanied by only marginal increases in $\Delta G$ (Jaenicke and Bohm, 1998), it appears that this additional stability may be accounted for by only a few additional interactions (Daniel, 1996b, Daniel and Cowan, 2000).

Both theoretical and experimental approaches have been used in the examination of protein thermostability. The comparison of the sequences and tertiary structures of homologous proteins from thermophiles and mesophiles (e.g. Argos, 1979) and, more recently statistical analysis of large, non-redundant sets of protein families (e.g. Vogt, 1997, Kumar, 2000) have formed the basis of theoretical efforts. Experimental analysis has usually involved protein engineering, most commonly in the form of site directed mutagenesis. The data amassed to date suggests that nature relies on no one single mode of protein stabilisation, thus many different, even conflicting hypotheses are presented on the subject (Lehmann, 2001).

Generally, there appears to be an inverse correlation between thermal stability and conformational flexibility and hence, specific activity. Whereas enzyme activity is dependant upon flexibility, a less flexible enzyme will be more stable (Cowan, 1995, Daniel and Cowan, 2000). However, it is believed that at their respective growth temperatures, related proteins from mesophilic and thermophilic organisms will have comparable levels of molecular flexibility (Jaenicke and Bohm, 1998), as a consequence, functionally related proteins from mesophilic and thermophilic organisms tend to possess similar catalytic activities at their respective growth temperature (Cowan, 1995).

A myriad of explanations have been used to explain this increased rigidity and the resultant thermostability of thermophilic proteins (Querol et al., 1996, Ladenstein and Antranikian, 1998). However, large scale statistical analyses (Vogt et al., 1997, Kumar et al., 2000), made possible by the recent rapid increase in available sequences and structures have tended to highlight increases in the number of salt bridges and side-chain hydrogen bonds as being the most consistent factors in protein stabilization.
Factors such as greater hydrophobicity, better internal packing, polar and non-polar contributions have also often been cited as stabilizing (Querol, 1996), though the large scale statistical analyses have indicated that these factors occur with similar frequencies in both mesophilic and thermophilic proteins (Vogt et al, 1997, Karshikoff and Ladenstein, 1998, Kumar et al, 2000). For example, it has frequently been suggested that stabilising substitutions tend to improve the packing efficiency and reduce cavity volume (Vogt and Argos, 1997). However, in a study of a set of 80 non-homologous proteins, Karshikoff and Ladenstein (1998) found that proteins from mesophiles and thermophiles have similar atom packing. Rather, it was concluded that, while thermostable proteins bury hydrophobic moieties more efficiently, this was not accompanied by increasing packing density. It was suggested that the stabilisation was more likely due to an exposure of polar charged groups creating favourable conditions for the formation of stabilising hydrogen bonds and salt bridges at the protein surface (Karshikoff and Ladenstein, 1998).

It also appears that there is significance in the amino acid distributions of thermophilic and mesophilic proteins. In an investigation into the thermal stability of a xylose isomerase, Sriprapundh et al. (2000), observed more proline and less asparagines and glutamate residues; consistent with the observation of Daniel et al. (1996) that the major degradative mechanisms at elevated temperatures are deamidation of asparagines, and succinamide formation at glutamate. More arginines and tyrosines but less cysteine and serine residues have also been reported in thermostable proteins (Kumar et al, 2000, Kumar and Nussinov, 2001). In addition, higher thermostability has been correlated with a larger fraction of residues in $\alpha$ helices and with more arginine and less proline and histidine in those $\alpha$ helices (Lehmann and Wyss, 2001). These residue substitutions within $\alpha$ helices increase thermostability via an increase in the rigidity of the helical pairs in the thermophilic proteins (Ptak, 1998). It has been suggested however, that, until recently, the limited number of sequences available has precluded statistically reliable detection of the significant residue differences above the background of amino acid changes which are evident for taxonomic or functional reasons (Daniel, 1996, Kumar et al., 2000).
1.8 *Bacillus* sp. RAPc8

Screening of a number of microbial isolates from sediment samples taken from several Australian thermal environments for nitrile-degrading activity led to the isolation of a single species that was capable of growth on media containing acetonitrile or propionitrile as the sole nitrogen source (Pereira, 1998, Pereira et al., 1998). Electron microscopic examination revealed that the organism appeared as short rods (0.8-1.2μm width, 2-2.5μm length). Cells were sporulating, motile and formed round, smooth nonpigmented colonies. Reaction to Gram staining was variable. 16S ribosomal sequencing studies indicated that the organism was a *Bacillus* species, most closely related to *Bacillus* sp. (starch negative) (DSM 2349), with 16S rDNA sequence similarity of 99.9%. The isolate was hence designated *Bacillus* sp. RAPc8.

Optical density and nitrile degrading activity studies were used to determine a maximum specific growth temperature of 65°C, (although high growth rates were observed at temperatures between 55 and 70°C) no growth was observed above 75°C or below 37°C. The optimum temperature for production of nitrile-degrading activity was 60°C. The *Bacillus* sp. RAPc8 strain was found to constitutively produce nitrile-degrading activity, chemical inducers such as cobalt, urea, acetamide, valeramide, and benzamide did not increase activity. Thus indicating that amide or nitrile substrate induction does not play a role in the expression of nitrile-degrading activity.

1.9 *Bacillus* sp. RAPc8 NHase

An NHase was purified from *Bacillus* RAPc8 cultures to near homogeneity in four steps with a yield of just less than 3%; no nitrilase activity was detected. The native enzyme had a relative molecular mass of 110,000Da whilst SDS-PAGE showed that the purified NHase produced two bands with relative molecular masses of 29,000Da ±1,500 (β subunit), and 28,000Da ± 1,500 (α subunit), suggesting that *Bacillus* strain RAPc8 NHase is a heterotetramer. Twenty amino acids at the N-terminus of each of the *Bacillus* strain RAPc8 nitrile hydratase subunits were sequenced. A BLAST search of the β subunit revealed distinct homology to previously described NHases (summarised in Fig. 1.11). The RAPc8 β subunit NHase N-terminus displayed the highest sequence identity with that of *R. rhodochrous* J1 H-NHase (65% identity). The lowest identity
with other NHase $\beta$ subunits was approximately 62%. A similar search with the $\alpha$ subunit N-terminus revealed no homology with the NHase $\alpha$ subunits of other organisms. These results are consistent with the previously observed high similarities of $\beta$ subunit and low similarities between $\alpha$ N-terminal regions from other microorganisms, and the commonly held belief that the $\beta$ subunit N-terminus implies a critical structural or functional role (Huang et al., 1997, Pereira, 1998).

![Fig. 1.11: Bacillus strain RAPc8 nitrile hydratase $\beta$ subunit N-terminal sequence BLAST search summary. ■ = 100% Identity, ■ = positive ■ = no homology (After Pereira, 1998).](image)

<table>
<thead>
<tr>
<th>Query sequence</th>
<th>% Identity</th>
<th>% Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus strain RAPc8</td>
<td>MNGIHDVGGMDFGKVMYVK</td>
<td>65%</td>
</tr>
<tr>
<td>R. rhodochrous J1 H-NHase</td>
<td>MDGIHDTGGMTGYGVPVQK</td>
<td>68%</td>
</tr>
<tr>
<td>R. erythropolis</td>
<td>MDGVHDLASVQGFGKV</td>
<td>68%</td>
</tr>
<tr>
<td>P. chlororaphis B23</td>
<td>MDGFHDLGGFQFGKVR</td>
<td>62%</td>
</tr>
<tr>
<td>P. putida</td>
<td>MNGIHDTGGHGYGPV</td>
<td>68%</td>
</tr>
<tr>
<td>Rhodococcus sp.</td>
<td>MNGVFDLGGTDGIIPV</td>
<td>64%</td>
</tr>
<tr>
<td>R. rhodochrous J1 L-NHase</td>
<td>MDGIHDLGQRAGLG</td>
<td>64%</td>
</tr>
</tbody>
</table>

The optimum pH for activity of the purified enzyme was 7.0, with 50% activity levels at 5.1 and 8.7. An apparent temperature optimum of 60°C was determined for the enzyme, although its half-life at this temperature was less than 20 minutes. The enzyme was found to encompass a broad substrate specificity range, albeit constrained to aliphatic nitriles; a particular preference for branched and cyclic aliphatic nitriles was noted. Turnover rates for propionitrile and acetonitrile under optimum reaction conditions were 436 sec$^{-1}$ and 765 sec$^{-1}$ respectively.

### 1.9.1 RAPc8 NHase cloning studies

Degenerate oligonucleotides were designed from the 20 amino acid N-terminal sequences of the $\alpha$ and $\beta$ subunits; the $\alpha$ primer was a 26mer and $\beta$ a 20mer. These oligonucleotides were first used as DIG-labelled probes for Southern hybridisation screening of a lambda library of partially digested genomic RAPc8 DNA: however no positive clones were identified using this technique.
Chapter 1  

Introduction

The oligonucleotides were also used in PCR reactions. The expected product size from such reactions would be ~600bp or ~1300bp, depending on the arrangement of the α and β subunit genes. Four products of 0.7kb, 0.75k, 1kb and 1.5kb were obtained in reactions using *Pfu* DNA polymerase. However, only two of these products, the 0.75kb and 1kb products could be cloned and subsequent sequencing of the clones revealed no significant homology to any known NHase sequences. Further work with these oligonucleotides was therefore abandoned at this stage.

1.10 Aims

The aims of this research were:

1. To clone the genes encoding the amidase and NHase of *Bacillus* sp. RAPc8.

2. To overexpress those genes in an active, soluble form in a suitable host expression system.

3. To apply the knowledge gained form the cloning and expression studies in attempts to alter the substrate specificity to include aromatic nitriles through enzyme engineering.

4. To apply the knowledge gained form the cloning and expression studies in investigations directed at the isolation of novel thermostable nitrile-degrading enzyme systems. 
Chapter 2: Materials and Methods

2.1 Chemicals and reagents

Unless otherwise stated, chemicals of analytical/reagent grade were supplied by Sigma - Aldrich Chemical Company, BDH Ltd., Bio-Rad Laboratories. Difco Laboratories or Oxoid Ltd supplied culture media. DNA and DNA size markers were obtained from Sigma - Aldrich Chemical Company, Roche and New England Biolabs (NEB). Primers were obtained from Amersham Pharmacia Biotech.

2.2 Enzymes

Restriction enzymes were purchased from NEB or Roche. DNA modifying enzymes and polymerases were obtained from Gibco Life Technologies, NEB, Roche and Stratagene. RNase, protease and recombinant amidase were supplied by Sigma – Aldrich Chemical Company.

2.3 Media

Bacterial strains were routinely grown either in Nutrient Broth No. 2 (NB) (Difco) or on Nutrient Agar (NA) (Difco) prepared according to the manufacturer’s instructions, with the exception that *Bacillus* sp. RAPc8 was grown in 2x nutrient broth No. 2. Culture media used in special protocols are listed below.

**SOC media**

SOC medium was used for the recovery of newly transformed *E. coli* strains.

<table>
<thead>
<tr>
<th>Constituent</th>
<th>L⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone</td>
<td>20g</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>5g</td>
</tr>
<tr>
<td>1M NaCl</td>
<td>10ml</td>
</tr>
<tr>
<td>1M KCl</td>
<td>2.5ml</td>
</tr>
</tbody>
</table>
The pH was adjusted to 7.0 with 5M NaOH prior to sterilisation. After allowing the broth to cool to ~55°C, the following filter-sterilised ingredients were added aseptically.

<table>
<thead>
<tr>
<th>Constituent</th>
<th>L.1dH₂O</th>
<th>ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>1M MgCl₂</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1M MgSO₄</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>2M Glucose</td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

NZY+ broth

NZY+ broth was used for the recovery of *E. coli* XL1 Blue cells used in site directed mutagenesis experiments.

<table>
<thead>
<tr>
<th>Constituent</th>
<th>L.1dH₂O</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ amine (casein hydrolysate)</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Yeast extract</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>NaCl</td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>

The pH was adjusted to 7.5 with 5M NaOH prior to sterilisation. After allowing the broth to cool to ~55°C, the following filter-sterilised ingredients were added aseptically.

<table>
<thead>
<tr>
<th>Constituent</th>
<th>L.1dH₂O</th>
<th>ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>1M MgCl₂</td>
<td></td>
<td>12.5</td>
</tr>
<tr>
<td>1M MgSO₄</td>
<td></td>
<td>12.5</td>
</tr>
<tr>
<td>2M Glucose</td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>
Modified Castenholz media

Modified Castenholz media were used as the nitrile containing minimal media for screening and identification of nitrile metabolising strains.

**Constituent** | **L.1 dH₂O**
--- | ---
A: (autoclaved separately) | 
Nitrilotriacetic acid | 100mg
(Added to 500ml dH₂O, adjusted to pH6.5 with 5M KOH before addition of remaining constituents)

Yeast extract | 10mg
FeCl₃ | 2mg
CoCl₂ | 2mg
CaSO₄ | 40mg
MgSO₄ | 140mg

Then adjusted to pH7.5

To prepare Castenholz agar plates, 20g of agar was added to solution A prior to autoclaving.

**B:** (autoclaved separately)
Na₂HPO₄ | 0.2M
KH₂PO₄ | 0.2M
(Adjusted to pH8.2 and added to A to a final phosphate concentration of 20mM)

**C:** (filter sterilized)
Carbon source: 20mM succinate or 0.05% v/v nitrile (acetonitrile or benzonitrile).

**D:** (filter sterilized)
Nitrogen source: 50mM ammonium sulphate or 0.05% v/v nitrile (acetonitrile or benzonitrile).
### 2.4 Buffers and solutions

The buffers and solutions used throughout this work are listed in Table 2.1 below:

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Constituents</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x Agarose gel loading dye</td>
<td>15% Ficoll, 0.2% bromophenol blue, 0.2% xylene cyanol FF</td>
<td></td>
</tr>
<tr>
<td>Ammonia assay reagent A</td>
<td>0.59M phenol, 1mM sodium nitroprusside</td>
<td></td>
</tr>
<tr>
<td>Ammonia assay reagent B</td>
<td>0.11M sodium hypochlorite, 2M sodium hydroxide</td>
<td></td>
</tr>
<tr>
<td>Blotting denaturing buffer</td>
<td>0.5M NaOH, 1.5M NaCl</td>
<td></td>
</tr>
<tr>
<td>Blotting neutralisation buffer</td>
<td>0.5M Tris-HCl, 3.0M NaCl</td>
<td>7.4</td>
</tr>
<tr>
<td>100x Denhardt’s solution</td>
<td>2% BSA, 2% Ficoll, 2% PVP, 3x SSPE, Filter sterilized. Stored at -20°C.</td>
<td></td>
</tr>
<tr>
<td>DIG hybridisation buffer</td>
<td>5x SSC, 0.1% N-lauroylsarcosine (w/v), 0.02% SDS, 1% Blocking reagent (Roche)</td>
<td></td>
</tr>
<tr>
<td>Western transfer buffer</td>
<td>25mM Tris base, 190mM glycine, 20% methanol</td>
<td>8.3</td>
</tr>
</tbody>
</table>
## Materials and methods

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Constituents</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybridisation buffer</td>
<td>6x SSC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5x Denhardt’s</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5% SDS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100μg/ml ss salmon sperm DNA</td>
<td></td>
</tr>
<tr>
<td>1M Potassium phosphate buffer</td>
<td>717ml 1M K$_2$HPO$_4$</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>283ml 1M KH$_2$PO$_4$</td>
<td></td>
</tr>
<tr>
<td>10x SDS PAGE electrode buffer</td>
<td>0.25M Tris-HCl</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>2M glycine</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1% SDS</td>
<td></td>
</tr>
<tr>
<td>2x SDS PAGE sample buffer</td>
<td>125mM Tris-HCl</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>4% SDS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20% glycerol</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10% 2-mercaptoethanol</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.004% bromophenol blue</td>
<td></td>
</tr>
<tr>
<td>SDS PAGE staining solution</td>
<td>0.2% Coomassie Blue R-250</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40% methanol</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10% acetic acid</td>
<td></td>
</tr>
<tr>
<td>SDS PAGE destaining solution</td>
<td>5% methanol</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.5% acetic acid</td>
<td></td>
</tr>
<tr>
<td>20x SSC</td>
<td>3.0M NaCl</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>0.3M Sodium citrate</td>
<td></td>
</tr>
<tr>
<td>20X TAE buffer</td>
<td>2M Tris base</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>25mM EDTA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adjusted with glacial acetic acid</td>
<td></td>
</tr>
<tr>
<td>TE buffer</td>
<td>10mM Tris-HCl</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>1mM EDTA</td>
<td></td>
</tr>
</tbody>
</table>
### 2.5 Bacterial strains

The bacterial strains used in this study are listed in Table 2.2 below:

<table>
<thead>
<tr>
<th>Bacterial strain</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacillus sp. RAPc8</strong></td>
<td>NHase and amidase producer</td>
<td>UCL stock</td>
</tr>
<tr>
<td><strong>E. coli JM107</strong></td>
<td>e14− (McrA−) endA1 gyrA96 (Nalr) thi-1 hsdR17 (rK−mK+) supE44 relA1 Δ(lac-proAB) [F' traD36 proAB lacIqZ ΔM15]</td>
<td>UCL stock</td>
</tr>
<tr>
<td><strong>E. coli BL21pLysS</strong></td>
<td>E. coli B F- dcm+ ompT hsdS(rB− mB−) gal1 (DE3) [pLysS Cam']a</td>
<td>Stratagene</td>
</tr>
<tr>
<td><strong>E. coli TOP10</strong></td>
<td>F− mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 deoR araD139 Δ(ara-leu)7697 galU galK rpsL (Str') endA1 nupG</td>
<td>Invitrogen</td>
</tr>
<tr>
<td><strong>E. coli TOP10F'</strong></td>
<td>F' [lacIqTn10(TetR)] mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (Str') endA1 nupG</td>
<td>Invitrogen</td>
</tr>
<tr>
<td><strong>E. coli XL1-Blue</strong></td>
<td>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacIqZ ΔM15 Tn10 (TetR)]</td>
<td>Stratagene</td>
</tr>
<tr>
<td><strong>R. erythropolis JCM6823</strong></td>
<td>NHase and amidase producer</td>
<td>Japan Collection Microorganisms</td>
</tr>
</tbody>
</table>
## 2.6 Plasmids

The cloning vectors and final expression constructs used in this study are listed in Table 2.3 below:

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC18</td>
<td>High copy number cloning vector. Insertion into multiple cloning site disrupts lacZα - recombinants identified through blue white screening and ampicillin resistance.</td>
<td>UCL stock</td>
</tr>
<tr>
<td>pCR2.1</td>
<td>Supplied as linear with 3’ T-overhangs for direct cloning of Taq amplified DNA fragments. Cloning site disrupts lacZα - recombinants identified through blue white screening and ampicillin resistance.</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pCR-XL-TOPO</td>
<td>Supplied as linear with topoisomerase I covalently bound to 3’ T-overhangs for direct cloning of long Taq amplified DNA fragments. Cloning site disrupts lethal ccdB – non-recombinants are killed on plating. Confers kanamycin resistance.</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pET21a (+)</td>
<td>High-level expression vector. Expression of target genes from T7lac promoter upon induction with IPTG. Confers ampicillin resistance.</td>
<td>Novagen</td>
</tr>
<tr>
<td>pNH223</td>
<td>pET21a (+) derivative carrying amidase gene of <em>Bacillus</em> sp. RAPc8</td>
<td>This work</td>
</tr>
<tr>
<td>pNH461</td>
<td>pET21a (+) derivative carrying NHase α and β subunit genes and partial P14K gene of <em>Bacillus</em> sp. RAPc8</td>
<td>This work</td>
</tr>
<tr>
<td>pNH512</td>
<td>pET21a (+) derivative carrying NHase α and β subunit genes of <em>Bacillus</em> sp. RAPc8</td>
<td>This work</td>
</tr>
<tr>
<td>pW124G</td>
<td>NHase αW124G mutant of pNH512</td>
<td>This work</td>
</tr>
<tr>
<td>pF36L</td>
<td>NHase βF36L mutant of pNH512</td>
<td>This work</td>
</tr>
<tr>
<td>pF52G</td>
<td>NHase βF52G mutant of pNH512</td>
<td>This work</td>
</tr>
<tr>
<td>pY67A</td>
<td>NHase βY67A mutant of pNH512</td>
<td>This work</td>
</tr>
<tr>
<td>pY67E</td>
<td>NHase βY67E mutant of pNH512</td>
<td>This work</td>
</tr>
<tr>
<td>pW76G</td>
<td>NHase βW76G mutant of pNH512</td>
<td>This work</td>
</tr>
</tbody>
</table>
Chapter 2

Materials and methods

2.7 Analytical Procedures

2.7.1 Spectrophotometry
Cell cultures, ammonia assays (Section 2.7.8), protein and DNA samples were routinely quantified by spectrophotometric analysis at 600nm, 600nm, 595nm and 260/280nm respectively using a Beckman DU7500 spectrophotometer.

2.7.2 PicoGreen assay for dsDNA
As the quantity of DNA isolated from thermal sediments was limited, its concentration was determined fluorometrically using the PicoGreen dsDNA quantitation kit (Molecular Probes). The assay was performed as described in the manufacturer’s instructions using a Shimadzu RF-5001 PC spectrofluorophotometer (excitation at 480nm, emission at 570nm). Known concentrations of λ DNA were used to produce a standard curve for each set of assays.

2.7.3 Determination of protein concentration by the Bradford assay
Protein standards were prepared using 0 to 1.4mg/ml bovine serum albumin (BSA) from a 10mg/ml stock. One hundred μL aliquots of protein standards or samples were dispensed into test tubes. Five ml of diluted and filtered protein assay reagent (BioRad) was added to each solution and immediately mixed using a vortex mixer. The solutions were incubated for at least five minutes at room temperature before measuring absorbance at 595nm. Protein sample concentrations were calculated by interpolation of absorbance values against a standard curve (Appendix I.i) determined for the BSA standard solutions.

2.7.4 Agarose gel electrophoresis
DNA fragments were separated by gel electrophoresis in 0.8 - 3% (w/v) agarose gels (depending on size range to be separated), as described by Sambrook et al. (1989). The electrophoresis buffer used was TAE (Table 2.1). Following electrophoresis, gels were stained in TAE buffer containing 0.05μg/ml ethidium bromide. After rinsing in distilled water, DNA bands were visualised and photographed under UV light.

56
2.7.5 SDS-polyacrylamide gel electrophoresis

Separation of proteins by SDS PAGE was performed using a Mighty Small vertical Slab Unit SE 280 (Hoefer). One mm and 1.5mm gels containing 10-15% acrylamide (Tables 2.4 and 2.5) were prepared using the Hoefer multiple gel caster. Protein samples were prepared by suspending in SDS PAGE sample buffer (Table 2.1) and denatured by boiling for three minutes. Samples were loaded onto the gel and run at a constant current of 50mA. Following electrophoresis, gels were stained with Coomassie Blue for a maximum of three hours, and then destained overnight by soaking in a destaining solution (Table 2.1).

Table 2.4 SDS PAGE resolving gel constituents.

<table>
<thead>
<tr>
<th>% Acrylamide</th>
<th>10%</th>
<th>12%</th>
<th>15%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distilled H$_2$O</td>
<td>7.9</td>
<td>6.6</td>
<td>4.6</td>
</tr>
<tr>
<td>1.5 M Tris-HCl, pH 8.8</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>20% (w/v) SDS</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Acrylamide/Bis-acrylamide (30%/0.8% w/v)</td>
<td>6.7</td>
<td>8.0</td>
<td>10</td>
</tr>
<tr>
<td>10% (w/v) ammonium persulfate</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>TEMED</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Total volume</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 2.5 SDS PAGE stacking gel constituents.

<table>
<thead>
<tr>
<th></th>
<th>Volume (ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distilled H$_2$O</td>
<td>2.80</td>
</tr>
<tr>
<td>0.5 M Tris-HCl, pH 6.8</td>
<td>1.25</td>
</tr>
<tr>
<td>20% (w/v) SDS</td>
<td>0.05</td>
</tr>
<tr>
<td>Acrylamide/Bis-acrylamide (30%/0.8% w/v)</td>
<td>0.85</td>
</tr>
<tr>
<td>10% (w/v) ammonium persulfate</td>
<td>0.05</td>
</tr>
<tr>
<td>TEMED</td>
<td>0.005 ml</td>
</tr>
<tr>
<td>Total volume</td>
<td>5.005 ml</td>
</tr>
</tbody>
</table>

2.7.6 Western blotting

Semi-pure samples of NHase were electrophoresed by SDS PAGE as described in Section 2.7.5, with the exception that 2mM mercaptoacetic acid was included in the upper electrode buffer.

A sheet of PVDF membrane (Roche) and 4 sheets of Whatman 3M paper were cut to the size of the gel. The membrane was soaked in methanol, then water, then transfer buffer (Table 5.1) as described in the manufacturers instructions. The Whatman paper
was also soaked in transfer buffer. Two sheets were then placed on the blotting apparatus, followed by the membrane, then the gel (also rinsed in transfer buffer) and the remaining sheets of Whatman paper were overlaid and the apparatus closed. Electroblotting was performed at 200mA for one hour. Transfer was confirmed with the aid of pre-stained markers (SeeBlue – Invitrogen).

Following transfer, the membrane was stained and destained as for SDS PAGE (Section 2.7.5), air-dried and sent for N-terminal sequencing.

2.7.7 Activity assays

For the determination of amidase and NHase activities, both colorimetric and gas chromatographic assays were used.

2.7.7.1 Ammonia detection assay

Conversion of nitriles or amides to the corresponding acids was assayed by quantification of the release of ammonia using a modification of the phenol-hypochlorite ammonia detection method (Fawcett and Scott, 1960).

Unless otherwise specified, the 300μl reaction mixture contained 25mM K2HPO4/KH2PO4 buffer, pH 7.2, 25mM nitrile or amide and known concentration of cell free extract or semi-pure enzyme solution. When assaying recombinant NHase, the reaction mixture was supplemented with an excess of recombinant amidase. All assays were carried out in duplicate at 50°C. One hundred μl aliquots were removed from the reaction mixture at timed intervals and immediately quenched by addition of 350μl of solution A (Table 2.1), and mixed immediately. A 350μl aliquot of solution B was then added to the mixture, which was then incubated for exactly five minutes at 50°C.

The absorbance was then measured at 600nm and the results interpolated against a standard curve prepared using 0 to 2mM ammonium chloride (Appendix I.ii). One unit of enzyme activity was defined as the amount of enzyme that catalysed the release of 1μmole of ammonia per minute under standard assay conditions.

2.7.7.2 Gas chromatography

Where necessary, nitrile-degrading activity was determined by detection of substrate loss using gas chromatography. An RT-Q Plot capillary column (30m x 0.53mm,
Chapter 2 Materials and methods

ChiralDex Ltd.) was with a Perkin-Elmer AutoSystem XL GC for quantification of nitrile substrates.

For all substrates, 1μl aqueous or organic samples were injected, using helium at a pressure of 14.0 PSI as the carrier gas. The oven temperature was held at 130°C for five minutes, then ramped to 230°C, with a rise of 10°C/min and held for 10 minutes. Injector and flame ionisation detector temperatures were set at 220°C.

2.8 Growth and maintenance of cultures

E. coli and R. erythropolis JCM6823 cultures were grown at 37°C and 30°C respectively in sterile liquid NB medium or on NA plates unless otherwise specified. Bacillus sp. RAPc8 was grown at 60°C in sterile 2x NB or on NA plates unless otherwise specified. Liquid cultures were routinely grown at a reciprocation rate of 230-300rpm. When growing a plasmid-containing strain, the growth medium was supplemented with the appropriate antibiotics: ampicillin - 100μg/ml; carbenicillin - 100μg/ml; chloramphenicol - 50μg/ml or kanamycin - 50μg/ml. Cells were harvested by centrifugation at various speeds (not usually more than 4000x g) depending on volume. The viability of bacterial strains was maintained over extended periods of time by storage in glycerol. A 1ml aliquot of an overnight culture was harvested by centrifugation at 3000x g for 3 minutes and resuspended in 20% sterile glycerol. These stocks were kept frozen at -80°C.

2.9 Preparation of competent E. coli cells

2.9.1 Electrocompetent cells

A single, fresh colony was used to inoculate 5ml of nutrient broth and grown overnight at 37°C. The overnight culture was used subsequently to inoculate 1L of NB and grown in the same conditions until the cells reached an OD_600nm of ~ 0.6. Cells were chilled on ice for 20 minutes then harvested by centrifugation at 3000x g, 4°C for 15 minutes. The cell pellets were washed five times in 1L of ice-cold 10% glycerol. The final pellets from these washings were gently resuspended in 4ml of ice-cold 10% glycerol.
and divided into 80μl aliquots, to either be used immediately for electro-transformation or to be stored frozen at -80°C.

### 2.9.2 Chemically competent cells

A single, fresh colony was used to inoculate 5ml of nutrient broth and grown overnight at 37°C. The overnight culture was used subsequently to inoculate 50ml of NB and grown in the same conditions until the cells reached an $\text{OD}_{600\text{nm}}$ of ~ 0.3. Cells were chilled on ice for 20 minutes then harvested by centrifugation at 3000x g, 4°C for 15 minutes. The cell pellets were washed twice in 25ml of ice-cold 0.1M CaCl$_2$. The final pellets from these washings were gently resuspended in 2ml of ice-cold 0.1M CaCl$_2$ and divided into 100μl aliquots and used immediately for transformation.

### 2.10 Transformation of competent *E. coli* cells

Competent cells supplied with commercial kits were transformed according to the manufacturers instructions; cells prepared in the laboratory were transformed using the following protocols.

#### 2.10.1 Electroporation

1. Electrocompetant *E. coli* cell suspensions were thawed on ice.
2. An 80μl aliquot of cell suspension was added to a 2mm electroporation cuvette to which up to 5μl (~5ng DNA) of ligation mixture (salts removed prior to transformation using a QIAquick PCR purification kit) was added and gently mixed with a pipette tip.
3. Cells were shocked once in the electroporator - set at a capacitance of 25 μF, voltage of 2.5 kV and resistance of 200 Ω.
4. One hundred μl room temperature SOC media was added immediately; the suspension was then transferred to a sterile 1.5ml eppendorf tube and incubated with agitation at 37°C for one hour.
5. Transformed cells were plated on appropriate agar plates and grown at 37°C overnight.
2.10.2 CaCl$_2$ transformation

1. Up to 20µl of ligation mixture was added to 100µl of competent cells, mixed gently and incubated on ice for 30 minutes.
2. The mixture was heat shocked for 45 seconds at 42°C, and then placed on ice for two minutes.
3. Two hundred µL SOC media was added, and the suspension incubated with agitation at 37°C for one hour.
4. Transformed cells were plated on appropriate agar plates and grown at 37°C overnight.

2.11 DNA Purification

2.11.1 Genomic DNA isolation

The following protocol was used (with minor variations) for the isolation of genomic DNA from all bacterial strains described in this work.

1. One g of cells (wet weight) was resuspended by gentle agitation in 5mL RNase-lysozyme solution (50µg/ml RNase, 10mg/ml lysozyme in TE buffer, pH 8), and incubated at 30°C until partial lysis was achieved.
2. 1.2mL 0.5M EDTA was added and shaken gently.
3. 0.13mL Proteinase K solution (10mg/ml) was added, the solution mixed gently and incubated at 30°C for 5 minutes.
4. 0.7mL 10% SDS was added, the solution shaken and incubated at 37°C until the solution began to clear (or for a maximum of 2 hours).
5. 5mL buffered phenol was added and shaken for 10 minutes.
6. 5mL 24:1 chloroform:isoamyl alcohol was added and shaken for five minutes. The mixture was then centrifuged at 2500x g for 10 minutes.
7. The top, aqueous layer was removed to a fresh container and steps 5 and 6 repeated twice.
8. The aqueous layer was removed carefully. To this 1/10 volume 3M sodium acetate and 2 volumes isopropanol were added, and the solution gently mixed until the DNA was thoroughly precipitated.
9. DNA was spooled onto a Pasteur pipette and washed in 70% ethanol.
10. DNA was respooled and dissolved overnight at 4°C in 1-2mL TE buffer.

11. DNA purity and concentration were determined by absorbance measurements at 280nm, 260nm and 230nm, according to the method of Sambrook et al. (1989) where one absorbance unit at 260nm is equal to 50μg dsDNA/ml. A DNA solution was considered pure if the A₂₆₀nm to A₂₃₀nm ratio was between 1.8 and 2.3 and the A₂₆₀nm to A₂₈₀nm ratio was between 1.5 and 2.0. (Marmur, 1963).

2.11.2 Plasmid DNA isolation and purification

Cultures of *E. coli* harbouring a plasmid of interest were prepared by inoculation of a 5ml overnight culture of LB, supplemented with the appropriate antibiotic from a fresh colony (glycerol stocks were streaked on agar plates before hand in order to circumvent problems of plasmid instability). Plasmid DNA was prepared from these cultures using QIAprep Spin Miniprep kit (Qiagen) as described in the manufacturer’s instructions.

2.11.3 Environmental DNA isolation

DNA was extracted from thermal sediment samples using the FastDNA spin kit for soil (BIO 101 Inc.) and a Mini-Bead Beater-8 (Biospect Products Ltd.) The extraction was followed according to the manufacturer’s protocol. In order to maximise the quality and quantity of isolated DNA, the bead beating, set at ~1500rpm for 1 minute, was repeated once. Following the final elution, DNA was precipitated by adding 1/10th of the volume of 3M sodium acetate pH 5 and 2 volumes of cold ethanol and incubating at -70°C for one hour. The DNA was recovered by centrifugation at 4°C for 30 minutes at 10000x g. The DNA pellets were washed twice in cold 70% ethanol, dried under vacuum and suspended in 60μl DNase/pyrogen free water (DES).

2.11.4 Sample DNA purification

Homogeneous DNA samples from PCR reactions, restriction digests etc. were purified using the QIAquick PCR Purification kit (Qiagen). Samples containing more than one DNA species were fractionated by agarose gel electrophoresis, and the relevant band(s) excised and purified by means of the QIAquick Gel Extraction kit (Qiagen) as described in the manufacturer’s instructions.
2.12 Polymerase chain reaction (PCR)

Unless otherwise stated, reaction mixtures were made up as detailed below. Reagents were added to the reaction mixtures as listed in Table 2.6. Hot start reactions were incubated for 5 minutes at 95°C before the addition of a polymerase; in all cases the thermo-cycler was allowed to reach 95°C before starting the reaction cycle.

Table 2.6 Typical PCR reaction constituents.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume/50µl reaction</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>dH₂O</td>
<td>To 50µl</td>
<td>-</td>
</tr>
<tr>
<td>10x reaction buffer (As supplied with polymerase)</td>
<td>5µl</td>
<td>1x</td>
</tr>
<tr>
<td>Template DNA</td>
<td>1µl</td>
<td>~1ng/µl</td>
</tr>
<tr>
<td>20μM upstream primer or 20μM upstream primer (degenerate primers)</td>
<td>1µl 3µl</td>
<td>0.4μM 1.2μM</td>
</tr>
<tr>
<td>20μM downstream primer or 20μM downstream primer (degenerate primers)</td>
<td>1µl 3µl</td>
<td>0.4μM 1.2μM</td>
</tr>
<tr>
<td>10mM (each) dNTP mix (Roche)</td>
<td>1µl</td>
<td>0.2mM</td>
</tr>
<tr>
<td>Polymerase (usually 5U µl⁻¹ supplied by Roche)</td>
<td>Variable</td>
<td>2.5U</td>
</tr>
</tbody>
</table>

Details of the amplification cycles and primers used for particular reactions are given in the relevant sections.

2.13 Site-directed mutagenesis

*In vitro* site-directed mutants were generated using the PCR based QuikChange site-directed mutagenesis system (Stratagene). The system enables site-specific mutation of plasmids of up to 8kb by means of two complementary mutagenic primers (sequences detailed in Table 5.1), which are extended during temperature cycling with the high-fidelity *Pfu* DNA polymerase. A mutated plasmid containing staggered nicks is thus generated. Following the temperature cycling, the resultant product is treated with *DpnI*, a restriction enzyme that is specific for methylated and hemimethylated DNA.
Thus non-mutated parental DNA, which was produced by a methylating strain of *E. coli*, is digested, thus selecting for the non-methylated mutated DNA. The protocol was followed according to the manufacturers instructions. Mutants were subsequently transformed into chemically competent *E. coli* XL1-Blue cells (supplied with the kit).

### 2.14 Restriction endonuclease digestion of DNA

Digestions were carried out at the manufacturers recommended temperature according to the method of (Sambrook *et al.*, 1989). Restriction reactions were stopped and prepared for agarose gel electrophoresis (Section 2.7.4) by adding 1/10 volume of 10x DNA agarose gel loading dye (Table 2.1). Alternatively, if restricted DNA was to be used directly in an enzymic reaction, the digest was stopped either by heating at 65°C for 10 minutes or by gel purification (Section 2.11.4).

### 2.15 DNA cloning

#### 2.15.1 Construction of *Bacillus* sp. RAPc8 DNA libraries

##### 2.15.1.1 Size fractionation

In order to create libraries enriched for the appropriate DNA fragments, approximately 50µg of *Bacillus* sp. RAPc8 DNA was digested in an 80µl reaction with either *BamHI* or *HindIII* as described in Section 2.14. The restricted DNA was then size fractionated by agarose gel electrophoresis at ~10V for a period of ~10h. To avoid damage due to exposure to both ethidium bromide and UV light, the electrophoresis technique was modified: 4µl of 1kb ladder was loaded into wells flanking the sample DNA. Following electrophoresis, the flanking lanes were excised and stained in an ethidium bromide solution, visualised under UV light and marked at the appropriate positions (between 2kb and 2.5kb for *HindIII* cut DNA and between 4kb and 5kb for *BamHI* cut DNA). The strips of gel were repositioned next to the original gel; DNA between the marks was excised.

DNA fragments were extracted from the agarose gel using the QIAquick gel extraction kit (Qiagen). The protocol was followed according to the manufacturer’s specification except that the Nal solution containing the agarose and DNA was passed through a
column 4 times before washing in order to maximise recovery. DNA was resuspended in ~20μl elution buffer.

2.15.1.2 Dephosphorylation of DNA
Removal of phosphate residues from the 5'-termini of BamHI or HindIII restricted pUC18 DNA prior to cloning reactions was performed according to the method of Sambrook et al. (1989), using calf intestine alkaline phosphatase (CIP). pUC18 DNA in 20μl of 1x CIP buffer was incubated with 1-2U CIP at 37°C for 2 hours. CIP was inactivated by incubating the reaction mixture at 65°C for 15 minutes. DNA samples were then used directly in ligation reactions.

2.15.1.3 Ligation of DNA
Ligation reactions were carried out according to the method of Sambrook et al. (1989), using T4 DNA ligase. Up to 1μg DNA (quantified by agarose gel electrophoresis), at an insert:vector ratio of approximately 2:1 - 3:1 in 20μl of 1x ligation buffer, was incubated overnight at 16°C with 1-2U T4 DNA ligase. Ligation reactions were either stored frozen at -20°C or desalted by QIAquick PCR purification and used to transform electrocompetent E. coli JM107 (Section 2.10.1).

2.15.2 Multiplex cloning of environmental DNA
2.15.2.1 Size fractionation
As the purification of environmental DNA (detailed in Section 2.11.3) resulted in sheared DNA of ~12kb or less, there was no need for subsequent enzymatic restriction of DNA size for plasmid cloning.

Sheared DNA of less than 1.5kb was excluded from cloning reactions using the agarose gel fractionation method detailed in Section 2.15.1.1, with the exception that 60μl of DNA was added per gel lane and that following electrophoresis, DNA between 1.5kb and 10kb was excised and purified.

2.15.2.2 Blunt-end polishing of DNA ends
The polymerase and 3' to 5' exonuclease activities of Vent DNA polymerase (NEB) facilitated the blunt-end polishing of DNA damaged through shearing, by removal of 3' termini and filling-in of 5'terminal and nicks. Blunt-end polishing was carried out in
150-μl reactions in the presence of 200μM dNTPs (50μM each of dATP, dCTP, dGTP & dTTP) and 1x ThermoPol buffer (NEB). 1U Vent DNA polymerase was added and the mixture incubated at 55°C for 30 minutes. The reaction conditions allowed for primer extension without strand displacement of the DNA template. The reaction was purified using the DNA using the QIAquick PCR purification kit (eluted in 100μl TE).

2.15.2.3 Addition of deoxyadenosine to the 3' termini of blunt-ended DNA
To facilitate subsequent TA cloning, dATP was added to the termini of the polished DNA by incubation with Taq polymerase. 13μl of (10x) PCR standard buffer (final MgCl₂ concentration of 1.5mM), 2.5μl 10mM dATP (0.25mM final concentration), 3μl Taq DNA polymerase and 11.5μL H₂O was added to the DNA solution and incubated at 72°C for 30 minutes.

2.15.2.4 Dephosphorylation of DNA
Following the addition of dATP to the 3' termini of the DNA, removal of phosphate residues from the 5'-termini of DNA was performed using CIP. 6μL of CIP buffer and 4μL CIP (4U/μl) were added to the mixture and incubated at 37°C for 1 hour. The reaction was purified using the QIAquick PCR purification kit and eluted in 30μl TE.

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

pCR-XL-TOPO is supplied as a linearised plasmid with 3'thymine overhangs. Covalently bound to the vector is topoisomerase I, which catalyses the ligation of plasmid and input DNA containing 3' adenosine overhangs (Shuman, 1994). The TOPO-cloning reaction was initiated by adding 10ng pCR-XL-TOPO (1μl) to 4μl of 5'dephosphorylated, 3'A-DNA containing 1μl of the supplied salt solution (diluted 1:4). The mixture was incubated at 22°C for exactly 5 minutes. To quench the cloning reaction 1μl of stop solution was added to the reaction, which was then quickly placed on ice before immediately being used to transform electrocompetant E. coli TOP10 cells (rather than those supplied with the kit) as previously described (Section 2.10.1).
Chapter 2 Materials and methods

The pCR-XL-TOPO vector contains the lethal ccdB gene adjacent to upstream-sequences encoding the C-terminus of the LacZα fragment. Ligation of inserts into pCR-XL-TOPO disrupts expression of the lacZα-ccdB gene fusion permitting growth of recombinant transformants only. Non-recombinants are killed upon plating. Blue/white screening was, therefore, not required (Bernard et al., 1994).

2.15.2 TA cloning
The TA cloning kit (Invitrogen) was used in the cloning of all PCR products. pCR2.1, supplied as a linearised vector with 3’thymine overhangs, facilitates the cloning of Taq amplified PCR products, or of products amplified by a heterogeneous mixture of polymerases (providing Taq is the most abundant species) by virtue of the 3’adeosine overhangs that the enzyme incorporates into the amplicons.

The ligation reaction was carried out as recommended in the manufacturer’s instructions. Ligations were subsequently transformed into chemically competent E. coli TOP10F’ cells (homogeneous PCR products – Chapters 3 and 4) as described in Section 2.10.2 or into electrocompetent cells (heterogeneous PCR library – Chapter 6) as described in Section 2.10.1.

2.16 DNA hybridisation
Two hybridisation techniques were used: non-radioactive DNA hybridisation was used in initial studies using oligonucleotides and PCR products to produce DIG-labelled probes for Southern hybridisation. PCR products were radioactively labelled for use as probes in colony hybridisations.

2.16.1 DNA labelling
2.16.1.1 Digoxigenin-dUTP (DIG) DNA labelling
Non-radioactive DIG-dUTP labelling and enzyme immunoassay detection of oligonucleotides or DNA fragments was carried out using the DIG Oligonucleotide Tailing Kit (terminal transferase based), DIG DNA Labelling and Detection Kit (random primed method) (Roche) according to the protocols supplied with the kits.

### 2.16.1.2 Radiolabelling DNA fragments

DNA was radioactively labelled using the random primed method (Feinberg and Vogelstein, 1983) as described below.

1. Double stranded template DNA was denatured by heating to 95°C for 10 minutes and chilled quickly on ice/NaCl.
2. ~ 25ng denatured DNA was added to a microcentrifuge tube containing 3µl dNTP mix (1+1+1 mixture of 0.5mM dATP, dGTP and dTTP), 2µl 10x hexanucleotide mix (Roche), 5µl [α^32P] dCTP (3000 Ci/mmol), 1µl 2U Klenow enzyme and made up to a final volume of 20µl.
3. The solution was mixed and briefly centrifuged before incubation at 37°C for 1-2 hours.
4. The reaction was quenched by the addition of an equal volume of 0.2M EDTA, pH8.

### 2.16.2 Southern blotting

DNA was transferred from agarose gels to nylon membranes by a method based on that of Sambrook *et al.* (1989) as follows.

1. Following electrophoresis (generally in 0.8% agarose gels), lanes containing molecular weight markers were excised, stained in TAE buffer containing 1µg/ml ethidium bromide and visualised next to a ruler under UV light in order to aid in the subsequent identification of hybridising fragments.
2. Gels containing the DNA to be transferred were gently agitated in 0.25M HCl for 10-15 minutes.
3. Gels were rinsed with distilled water, then covered with denaturing buffer and agitated gently for 40 minutes.
4. Gels were rinsed with distilled water, then covered with neutralisation buffer and agitated gently for 1 hour.
5. During these incubations nylon strips and a stack (approximately 1 inch deep) of Whatman 3MM paper were cut to the same size as the gel. The nylon strips and 2 sheets of Whatman paper were soaked in 2x SSC.
6. An electrophoresis tray was upturned in a large dish containing 20x SSC, a double layer of Whatman paper, soaked in 20x SSC then placed over the tray. The gel block was placed on top of the tray, followed by the nylon strips, then the 2 sheets of Whatman paper soaked in 2x SSC. On top of this was placed the stack of dry Whatman paper and a stack of paper towels cut to the desired dimensions. The stack was covered with a glass plate and a 500g weight. Transfer was allowed to proceed overnight.
7. The following day, the nylon membranes were removed and marked at the bottom right hand corner for orientation purposes.
8. DNA was permanently immobilised to the membrane by UV crosslinking.

2.16.3 Colony lifts
Plasmid libraries to be screened by DNA hybridisation were immobilised to nylon membranes using the following protocol.

1. Duplicates of all plates to be screened were prepared using a replica plating apparatus, taking caution to mark all plates in particular positions to facilitate the location of colonies corresponding to positive hybridisation signals.
2. Nylon membranes cut to the size and shape of the plate were overlaid on the surface of the agar plate. The membranes were removed once damp then
placed, colony side, up on Whatman 3MM paper soaked in 10% SDS for 5 minutes or until the colonies lysed (apparent by a shiny appearance).

3. The membranes were removed, blotted briefly on dry Whatman paper, then place on Whatman paper soaked in denaturing buffer for 5 minutes followed by 5 minutes on neutralisation solution and 5 minutes on 2x SSC.

4. The membrane was then left to air dry completely before UV crosslinking to immobilise the DNA to the filter.

5. Before being used in hybridisations, the filters were submerged in 2x SSC for 5 minutes then washed in pre-wash buffer for 30 minutes at 50°C. Cell debris was then scraped off using a damp, clean tissue. Membranes were not allowed to dry between this step and hybridisation.

2.16.4 DNA hybridisation

Hybridisations were carried out using minor variations of the following protocol:

1. Nylon membranes were prehybridised for $1\frac{1}{2}$ hours at 42°C in approximately 20ml of hybridisation buffer (DIG hybridisation buffer or standard buffer depending on hybridisation type), per 100cm$^2$ membrane.

2. Following this period the prehybridisation buffer was replaced with approximately 2.5ml (DIG hybridisations) or 7ml (standard hybridisations) prewarmed hybridisation buffer per 100cm$^2$ and the labelled probe added to the buffer. Hybridisations were allowed to proceed overnight; temperatures of the specific experiments are detailed in the relevant sections.

3. Membranes were washed for 2 x 5 minutes at room temperature with at least 50ml 2x SSC; 0.1% SDS per 100cm$^2$ membrane then for 2 x 15 minutes (DIG hybridisations) or 2 x $\frac{1}{2}$ hour (standard hybridisations) with a more stringent wash (detailed in the relevant sections) at a specified hybridisation temperature.

4. Hybridising DNA was visualised either by immunological detection according to the protocols supplied with the DIG kits or by autoradiography.

2.17 Protein expression

Prior to expression studies, expression clones were transformed into *E. coli* BL21 pLysS and grown overnight on NA containing chloramphenicol and
ampicillin/carbenicillin. Fresh colonies were used to inoculate 5ml NB supplemented with chloramphenicol and ampicillin/carbenicillin and grown until slightly turbid. The cells were centrifuged at 3000x g for 5 minutes, 4°C, washed once with NB to remove secreted \(\beta\)-lactamase, resuspended in 1ml NB and used to inoculate 500ml NB. Cultures were grown at 37°C, 230rpm until \(A_{600nm} \approx 0.60\). At this point, expression of the recombinant protein was induced by addition of 0.4mM IPTG. When applicable, cobalt was added to concentrations of 0.1-0.5mM 15 minutes prior to induction. After incubating an additional 3-4h, cells were harvested by centrifugation at 3000x g for 15 minutes at 4°C. Cell pellets were washed with 50mM potassium phosphate (\(K_2HPO_4/KH_2PO_4\)) buffer, centrifuged, and resuspended in 25ml 50mM \(K_2HPO_4/KH_2PO_4\) buffer, (pH 7.2) for each pellet derived from 500 ml cell culture.

2.18 Cell free extract preparation

Because of the presence of T7 lysozyme, encoded by pLysS, the cells were easily lysed by one cycle of freezing at -80°C and thawing at 37°C. DNA was sheared by a brief sonication of 5-9 cycles of 5s sonication bursts followed by a break of 15s. During sonication, samples were cooled on iced water. The extract was cleared by centrifugation at 3000x g for 30 min at 4°C. Soluble proteins were recovered in the supernatant following centrifugation. Cell free extracts of *Bacillus* sp. RAPc8 samples were prepared in the same manner, using 10-15 sonication cycles.

2.19 Protein purification

2.19.1 Heat treatment

25ml of cell free extract was sealed in 50ml Falcon tubes and incubated in a water bath at 75°C for 30 minutes for amidase cell free extracts and at 55°C for 45 minutes for NHase cell free extracts. Following incubation, samples were centrifuged at 4000x g for 20 minutes at 4°C. The supernatant was transferred to a fresh 50ml Falcon tube, either to be separately aliquoted and stored at -80°C, or used immediately.
2.19.2 Ion exchange chromatography

25ml samples of heat-treated NHase solutions were concentrated to 1ml using a Centriprep 10 column (cut-off limit 10kDa) (Amicon Inc) according to the manufacturers instructions.

A 1ml Mono-Q HR 5/5 column (fitted to a Perkin-Elmer FPLC) was equilibrated with 25mM K$_2$HPO$_4$/KH$_2$PO$_4$ buffer, pH 7.2. The sample was then loaded onto the column; once absorbance at 280nm reached baseline level a linear elution gradient (0-0.5M NaCl) of up to 50ml at 0.5ml/min was used. The elution buffer was 25mM K$_2$HPO$_4$/KH$_2$PO$_4$ buffer, pH 7.2 containing 0.5M NaCl. 2ml fractions were collected.

2.20 Growth and expression curves

Growth curves and expression profiles for *E. coli* strains expressing recombinant amidase or NHase were determined in 250ml shake flask cultures containing 50ml NB. Growth and expression was carried out as detailed in Section 2.17.

Samples were aseptically removed at 15-30 minute intervals and the OD measured at 600nm to determine growth. The ammonia assay (Section 2.7.8) was used to determine the amidase or NHase activity of cell free extracts.

2.21 NHase characterisation

2.21.1 Thermostability profiles

To determine the thermal stability of recombinant NHase, two methods were used:

The first method measured the loss of NHase activity over time at different temperatures. 500µl samples of semi-pure NHase were sealed in 1.5ml Eppendorf tubes and incubated at 20°C, 37°C, 50°C and 60°C. 1µl aliquots were removed at specific time intervals and residual NHase activity determined using the ammonia assay.
Chapter 2 Material and methods

The second method measured the loss of activity over a thirty-minute incubation period at 30°C, 42°C, 50°C, 55°C, 60°C, 65°C and 70°C. 50μl samples of semi-pure NHase were removed and assayed for residual NHase activity as previously described. All experiments were performed in duplicate.

2.2.1.2 Substrate specificity
Substrate specificity studies were carried out in a 10ml stirred glass cell at 50°C. The reaction was initiated by addition of nitrile substrates dissolved in 30μl dichloromethane (final nitrile concentration: 50mM) to 2970μl of 50mM K₂HPO₄/KH₂PO₄ buffer, containing 3U NHase and an excess of amidase (at least 20U). A 200μl sample was immediately removed and the reaction quenched by addition to 500μl DMSO in a glass vial ready for GC analysis. This sample was taken as the nitrile concentration at time zero. After 30 minutes, a second 200μl sample was removed and quenched as before. Substrate degradation over this period was measured by GC analysis. All experiments were performed in triplicate.

2.2.1.3 Reaction kinetics
Reaction kinetics studies were performed as 2ml reactions in a stirred glass cell containing 50mM K₂HPO₄/KH₂PO₄ buffer, 2U semi-pure NHase and an excess of amidase (at least 20U). Reactions were initiated by addition of a specified volume of acetonitrile or acrylonitrile substrate. 100μl aliquots were removed every 30 seconds and assayed using the ammonia detection assay. All experiments were performed in duplicate.

2.2.1.4 Benzonitrile inhibition studies
The benzonitrile inhibition studies were performed as 2ml reactions in a stirred glass cell containing 50mM K₂HPO₄/KH₂PO₄ buffer, 1mM DTT, 20% glycerol and 60μl NHase solution. Once the temperature of the solution reached 50°C, 10μl dimethylformamide containing a specified concentration of benzonitrile was added, together with 70mM acetonitrile (final concentration). One hundred μl samples were removed at 30-second intervals and the ammonia concentration determined using the ammonia assay. At the concentrations used, the amidase activity was not affected by benzonitrile (data not shown). DMF did have a slight inhibitory affect on NHase.
activity, though this was taken into account in the calculations performed. All experiments were carried out in duplicate.

### 2.22 Sequencing protocols

#### 2.22.1 DNA sequencing

All PCR products, cloned genomic sequences and mutated sequences were purified using the QIAquick PCR purification kit (Section 2.11.2) and dispatched for commercial automated dideoxy sequencing by Oswel DNA, University of Southampton.

#### 2.22.2 N-terminal protein sequencing

Proteins were separated by SDS PAGE, electroblotted onto PVDF and stained as described in Sections 2.7.5 and 2.7.6 and dispatched for commercial automated Edman sequencing by PNAC, University of Cambridge.

### 2.23 Computational techniques

DNA and amino acid sequences were obtained from GenBank, EMBL, DDBJ, SWISSPROT, Protein Research Foundation (PRF), and Protein Data Bank (PDB) databases, searched using either the Entrez retrieval system on the National Centre for Biotechnology Information (NCBI) server, or the sequence retrieval system (SRS) on the European Molecular Biology Laboratory (EMBL) server.

BLAST searches (Altschul et al., 1990) were conducted using the BLAST facility on the NCBI server. Sequence alignments were performed using the CLUSTAL W (Higgins et al., 1994) facility on the EMBL server. Alignments were edited and highlighted using BioEdit (free software - North Carolina State University).

Sequence analysis was performed using the Gene Construction Kit (Textco Inc.), DNASTAR (DNASTAR Inc.) and BioEdit programs. Mutagenic primers were designed with the aid of an online version (found on the John Hopkins Medical Centre website) of the Primer Generator program (Turchin and Lawler, 1999).
Secondary structure prediction was performed using the PHD program (Rost, 1996) on the EMBL server. Homology modelling was conducted using the TURBO FRODO program (AFMB – CNRS, France) on a Silicon Graphics workstation. The assistance of Dr. S. Djordjevic of the Department of Biochemistry and Molecular Biology, University College London in making this facility available is gratefully acknowledged.
Chapter 3: Cloning the nitrile hydratase operon from *Bacillus* sp. RAPc8

3.1 Introduction

Nitrile hydratase (NHase) genes have been cloned from various *Rhodococci* (Bigey *et al.*, 1995, Duran *et al.*, 1993, Ikehata *et al.*, 1989, Kobayashi *et al.*, 1991, Komeda *et al.*, 1996a, 1996b, Mayaux *et al.*, 1990), *P. chlororaphis* (Nishiyama *et al.*, 1991), and from the thermophiles *Ps. thermophila* (Yamaki *et al.*, 1997), *B. smithii* (Takashima *et al.*, 1995) and *Bacillus* sp. BR449 (Kim and Oriel, 2000). In addition, the recent completion of several microbial genomes has revealed probable NHase genes in two legume symbionts: *Mesorhizobium loti* (NCBI Microbial Genomes Annotation Project) and *Sinorhizobium meliloti* (Capela *et al.*, 2001). In each case the $\alpha$ and $\beta$ subunits were coded within two separate, adjacent open reading frames (ORFs), separated by sequences of between 16 and 29 bases, except in *Ps. thermophila*, whose genes overlapped by 4 bases. Typically, the $\alpha$ subunit ORF was located upstream of the $\beta$ subunit ORF. However, in *R. rhodochrous* J1, *Ps. thermophila*, *B. smithii* and *Bacillus* sp. BR449, the $\beta$ subunit gene precedes the $\alpha$ subunit gene. There is a high degree of similarity between the NHase sequences, and a number of strictly conserved regions. In particular all share the characteristic cofactor-binding motif CXLCSC (Huang *et al.*, 1997). With the exception of the NHase operons of *R. rhodochrous* J1, all the enzymes described above are co-expressed with an amidase, the gene for which is generally positioned approximately 100bp upstream from the NHase genes. The H-NHase operon of *R. rhodochrous* J1 contains no amidase gene, whilst the amidase gene of the L-NHase operon is located downstream of the NHase genes. It is probable therefore that the amidase and nitrile hydratase genes are co-expressed from a single polycistronic mRNA, often including additional flanking genes. This hypothesis has been confirmed for the NHase operon of *Rhodococcus* sp. R312 through transcriptional analysis of a mutant derivative of the operon (Bigey *et al.*, 1999).
3.2 Verification of the identity of *Bacillus* sp. RAPc8

Freeze dried samples of *Bacillus* sp. RAPc8 (Pereira, 1998), were rehydrated with sterile LB broth and grown overnight at 60°C on nutrient agar plates. Colonies were restreaked on modified Castenholz agar (Section 2.3) containing acetonitrile as the sole nitrogen source and incubated at 60°C overnight. The formation of distinct colonies confirmed the ability of the organism to grow on media containing a nitrile as sole nitrogen source.

Acetonitrile degrading activity of cell free extracts (Section 2.18) was assessed using the ammonia assay (Section 2.7.1.1) and found to be approximately 1.2U/mg, (data not shown), close to the previously reported level of ~1.5U/mg (Pereira, 1998).

By way of additional confirmation, 16S rDNA was amplified by PCR using primers BSF8, BSF343 and BSF784 (Wilmotte *et al.*, 1993), corresponding to *E. coli* positions 8, 343 and 784. Sequencing of such PCR products gives reliable data of a 1427 nucleotide portion of the 16S rDNA corresponding to *E. coli* positions 30-1457. The highest scoring hits in BLAST searches (Appendix H i) were the sequences of *Bacillus* sp. B3 (Karita *et al.*, 2001) and *B. pallidus* (Rainey, 1994). In a similar study conducted by Dr F. Rainey (DSM) in collaboration with R. A. Pereira, a 1486bp section of *Bacillus* sp. RAPc8 16S rDNA, corresponding to *E. coli* positions 5-1518, was amplified and found to have 99.9% and ≥99.4% sequence identity with the equivalent sequences from DSM 2349 (*Bacillus* sp.) and DSM 3670 (*B. pallidus*) respectively. Comparisons with 16S rDNA sequences of DSM 2349 and DSM 3670 revealed that the 1427bp fragment of this work shared a 99.72% identity with DSM 2349 and 99.86% identity with DSM 3670. Whilst the values obtained in this work and those of Pereira are similar, it should be noted that the four nucleotide differences between the amplified DNA and the DSM 2349 sequence lay within the sequence of the amplicon described by Pereira and that the identity values, despite the differences in length, should therefore be closer. These small differences might however be attributable to errors either of the polymerase or of the sequencing in either work. Therefore given the origin of the sample, and the high rDNA sequence identity with the DSM species, it was concluded that the samples were of the same organism as characterised by Pereira (Pereira, 1998, Pereira *et al.*, 1998).
3.3 Cloning Strategies

Several strategies were employed concurrently in the attempt to isolate the genes encoding the nitrile hydratase of Bacillus sp. RAPc8; most of these strategies had been used previously to clone the NHase genes referenced above. These are described in the following section.

3.3.1 DNA probing using the gene encoding the \(\alpha\) subunit of the nitrile hydratase of Rhodococcus erythropolis.

The high levels of similarity in the amino acid sequences of many of the recently characterised nitrile hydratases facilitates the use of a cloned NHase gene as a probe for Southern hybridisation that screens for NHase genes. The nitrile hydratase genes of \(P.\) chlororaphis B23 (Nishiyama et al., 1991), \(R.\) erythropolis (Duran et al., 1993), and \(R.\) rhodochrous J1 (Kobayashi et al., 1991) have all been cloned in this manner using the genes encoding the \(\alpha\) and \(\beta\) subunits of NHase from \(Rhodococcus\) sp. N-774. The results of the \(R.\) rhodochrous J1 cloning experiments suggested that this approach would be worthwhile in the search for the \(Bacillus\) sp. RAPc8 NHase genes. The indications for this were that \(R.\) rhodochrous J1 produces cobalt containing enzymes (cf. the iron containing enzyme of \(Rhodococcus\) sp. N-774), that the \(\alpha\) and \(\beta\) subunit genes are found in an opposite organisation to those of \(Rhodococcus\) sp. N-774 and finally that the DNA sequence identity between the two sets of genes was not particularly high – only 60% (L-NHase) and 57% (H-NHase) with the \(\alpha\) subunits and 51% (L-NHase) and 48% (H-NHase) with the \(\beta\) subunit. On the assumption that the \(B.\) smithii and \(Bacillus\) sp. RAPc8 NHase gene sequences were similar and, given that the sequence identity of the \(B.\) smithii \(\alpha\) subunit with those of \(Rhodococcus\) sp. N-774 and \(R.\) erythropolis (52% and 51% respectively) was in a similar range, it was decided to use the \(\alpha\) subunit gene of \(R.\) erythropolis as a probe for hybridisation against RAPc8 DNA.
3.3.1.1 PCR cloning of the R. erythropolis $\alpha$ subunit gene.

*R. erythropolis* JCM 6823 was obtained from the Japan Collection of Micro-organisms and grown according to their recommendations. DNA from the organism was prepared as detailed in Section 2.11.1. Primers were designed to encompass the entire $\alpha$ subunit gene (Fig. 3.1), producing a 674bp product (Fig 3.2). The program of cycles used to amplify the gene was as follows:

1 cycle of 94°C - 5 minutes

30 cycles of

- 94°C - 30 seconds
- 40°C - 30 seconds
- 72°C - 1 minute

1 cycle of 72°C - 7 minutes

3.3.1.2 Hybridisation with the amplified product

The amplified alpha subunit was DIG-dUTP labelled (Section 2.16.1.1) and used to hybridise against Southern blots of digested *Bacillus* sp. RAPc8 DNA. As the experiments involved hybridisations of DNA from two distinct genera, they were performed at low stringency (hybridisation at 42°C overnight, subsequent washes using a minimum salt concentration of 0.25xSSC; 0.1%SDS, also at 42°C). A photograph of a typical probing experiment following the detection procedure is shown in Fig 3.3. Whilst strong signals were observed on the positive controls of *HindIII* or *BamHI* digested *R. erythropolis* DNA (demonstrating that the probe was adequately labelled) no hybridisation was detected against the RAPc8 DNA despite the very low stringency conditions. The investigation was therefore not pursued further.
**Fig. 3.1:** A: Organisation of the NHase operon of *R. erythropolis*, coding regions are represented by black arrows, primer positions shown in blue. B: Nucleotide sequence of the primers used.

**A**

*R. erythropolis NHase operon*  
(2830bp)

![Diagram of NHase operon with arrows indicating coding regions and primer positions.]

**B**

REALPHAF: atg tca gta aac atc gag cac ac  
(23mer)

REALPHAR: gca aga tgc tgt act cca toc at  
(23mer)

**Fig. 3.2:** PCR amplification of *R. erythropolis* JCM6823 NHase α subunit gene.  
Lanes: 1-1kb ladder; 2 - PCR reaction.
Fig. 3.3: Southern hybridisation of the DIG labelled *Rhodococcus* NHase α subunit gene against *R. erythropolis* and *Bacillus* sp. RAPc8 genomic DNA. Lanes: 1 - DIG-labelled molecular weight marker; 2 - *R. erythropolis* DNA cut with *BamHI*; 3 - *R. erythropolis* DNA cut with *EcoRI*; 4 - *Bacillus* sp. RAPc8 DNA cut with *BamHI*; 5 - *Bacillus* sp. RAPc8 DNA cut with *EcoRI*.

3.3.2 Designing oligonucleotides for use in PCR and hybridisation studies

As discussed in Section 1.3.8, NHase genes have been successfully isolated using oligonucleotides designed from amino acid sequence as either probes in hybridisation studies or as primers to produce PCR products for subsequent use as probes. Previous attempts to use oligonucleotides derived from the N-terminal sequences of the α and β subunits of the enzyme to clone the NHase genes of *Bacillus* sp. RAPc8 have been unsuccessful (Pereira, 1998). Although the oligonucleotides described in the following section were originally designed for use in PCR, they were also used as probes for Southern hybridisation. Degenerate oligonucleotides, corresponding to highly conserved regions in the NHase α subunit family were designed using multiple alignments (Fig 3.4) of the enzymes previously described (Section 1.3.8), whilst the elucidation of the N-terminal sequences of *Bacillus* sp. RAPc8 NHase α and β subunits (Pereira, 1998), enabled the design of oligonucleotides more specific to this enzyme.
**Fig. 3.4:** Alignment of the α subunit sequences of 10 recently cloned NHases (Duran et al., 1993, Hashimoto et al., 1994, Kobayashi et al., 1991, Mayaux et al., 1991, Mayaux, 1990, Nishiyama et al., 1991, Takashima et al., 1995, Wu et al, 1997, Yamaki et al., 1997). The alignment is highlighted at 60% conservation. Regions highlighted in colour denote those chosen as conserved regions with minimal codon degeneracy for the design of oligonucleotides that would bind specifically to NHase α subunit genes.
α Subunit N-terminal sequence

5’M K M D A N E I I S F I Q N S K K T T 3’
  atg aar atg atg gay gcn aay gar ath ath wcn tty ath car aay wcn aar aar acv acv
  C8ALPHAF (24mer)

5’N G I H D V G M D G F G K V M Y V K 3’
  atg aay ggn ath cay gay gtn ggn ggn atg gay ggn tgy ggn aar gtn atg tay gtn aar
  C8BETAF (28mer)
  C8BETAR (28mer)

Conserved sequence 1

5’V A K A W T D P
  gtn gcn aar gcn tgg acv gay ccn
  NHF1 (23mer)
  gtt gtn aar gcc tgg anc gat cc

Conserved sequence 2

5’C T L C S C Y
  gtn tgy acn ytn tgy ycn tgy tay
  NHF2 (24mer)
  gty tgy acn ytr tgy tcv tgy tay
  NHF1 - reverse complement (24mer)
  rta rca bga rca yar ngt rca rac

Conserved sequence 3

5’W D T / S S / A E V R
  gtn tgg gay acv/wsn wsn wsn/gcn gar gtn sgn
  Reverse complement
  ncs nac ytc nga nsw nsw/bgt rtc cca nac
  NHF2 (26mer)
  cg nac ttc igm rct rst rtc cca rac

β Subunit N-terminal sequence

5’M  N G  I H D V G M D G F G K V M Y V K 3’
  atg aay ggn ath cay gay gtn ggn ggn atg gay ggn tgy ggn aar gtn atg tay gtn aar
  C8ALPHAR (24mer)

Conserved sequence 1

5’V A K A W T D P
  gtn gcn aar gcn tgg acv gay ccn
  NHF1 (23mer)
  gtt gtn aar gcc tgg anc gat cc

Conserved sequence 2

5’C T L C S C Y
  gtn tgy acn ytn tgy ycn tgy tay
  NHF2 (24mer)
  gty tgy acn ytr tgy tcv tgy tay
  NHF1 - reverse complement (24mer)
  rta rca bga rca yar ngt rca rac

Conserved sequence 3

5’W D T / S S / A E V R
  gtn tgg gay acv/wsn wsn wsn/gcn gar gtn sgn
  Reverse complement
  ncs nac ytc nga nsw nsw/bgt rtc cca nac
  NHF2 (26mer)
  cg nac ttc igm rct rst rtc cca rac

Fig. 3.5: Degeneracy of sequences used in oligonucleotide design and the sequences of oligonucleotides used as probes and primers. By weighting codon usage towards that of the gene encoding the B. smithii NHase subunits, it was possible to further reduce the degeneracy of the sequences corresponding to conserved regions
3.3.3 Southern hybridisation using degenerate oligonucleotides

The oligonucleotides in Fig. 3.5 were DIG-labelled, and used to probe Southern blots of Bacillus sp. RAPc8 chromosomal DNA digested with BamHI, EcoRI and HindIII. Taking into account the lack of success previously encountered (Pereira, 1998) in probing using primers designed against the N-terminal regions of Bacillus sp. RAPc8 NHase, the C8 primer series was not included in these experiments. Conditions for the hybridisations were optimised to minimise the high background levels of hybridisation that were observed with all three probes in preliminary experiments: hybridisations using sequences NHFl and NHR2 (Fig. 3.5) were carried out at 45°C, final washes were at 36°C, in 0.125x SSC, 0.1% SDS; hybridisation with NHF2 was carried out at 49°C, final wash at 45°C in 0.125x SSC, 0.1% SDS; signals could not be detected at stringencies above these levels. Whilst the NHF1 probe was the least degenerate of the sequences, it also appeared to be the least specific, since in addition to giving significant background hybridisation, up to five signals of approximately equal intensity were observed on each lane (Fig. 3.6). NHR2 also exhibited background hybridisation, though to a lesser extent, and single distinct signals were detected at ~2.2kb in the HindIII digest and ~4.5kb in the BamHI digest (Fig. 3.7). NHF2, corresponding to the cofactor binding region of the subunit (Huang et al., 1997), was clearly the most specific probe, showing no background hybridisation and giving clear, intense signals at ~2.2kb in the HindIII digest and ~4.5kb in the BamHI digest as well as at ~1kb in the EcoRI digest (Fig. 3.8). The fact that the same pattern of bands was observed when the blots were probed with NHF2 and NHR2 lent confidence to the supposition that they were hybridising to the desired NHase sequence.
**Fig. 3.6:** Southern hybridisation of the DIG end-labelled NHFl probe against *Bacillus* sp. RAPc8 genomic DNA. Lanes: 1 - DIG-labelled molecular weight marker; 2 - *Bacillus* sp. RAPc8 DNA cut with *BamHI*; 3 - RAPc8 DNA cut with *Bacillus* sp. RAPc8 DNA cut with *HindIII*.

**Fig. 3.7:** Southern hybridisation of the DIG end-labelled NHR2 probe against *Bacillus* sp. RAPc8 genomic DNA. Lanes: 1 - DIG-labelled molecular weight marker; 2 - *Bacillus* sp. RAPc8 DNA cut with *BamHI*; 3 - RAPc8 DNA cut with *Bacillus* sp. RAPc8 DNA cut with *HindIII*. 
Fig. 3.8: Southern hybridisation of the DIG end-labelled NHF2 probe against *Bacillus* sp. RAPc8 genomic DNA. Lanes: 1 - DIG-labelled molecular weight marker; 2 - *Bacillus* sp. RAPc8 DNA cut with *BamHI*; 3 - *Bacillus* sp. RAPc8 DNA cut with *EcoRI*; 4 - *Bacillus* sp. RAPc8 DNA cut with *HindIII*.

Two libraries of *Bacillus* sp. RAPc8 DNA fragments (generated by digestion with *BamHI* or *HindIII*) were constructed in pUC18 (Section 2.15.1). These libraries were duplicated, transferred by colony lifts to nylon membranes (Section 2.16.3) and screened with the probes described above. Despite optimisation of the washing steps, excessive background signals precluded the identification of positive clones. Approximately 50 plasmids were isolated from randomly selected colonies and screened by Southern hybridisation. However, no positive signals were detected. As further attempts to screen the library in this way would have been both costly and labour intensive, it was decided to attempt to produce a larger probe by PCR with the primers described in Section 3.3.2.
3.3.4 PCR using degenerate primers

Initial PCR trial experiments with all possible primer concentrations were conducted under low stringency conditions (annealing temperature ≤ 45°C; [MgCl₂] ≥ 1.5mM). It was accepted that, as a consequence of the large number of primer species present in each reaction, the initial DNA amplified would be contaminated with the products of non-specific amplification (due mainly to the low annealing temperature). The rationale behind this approach was that reactions that represent the amplification of NHase genes would result in DNA products of the predictable size (Fig. 3.9) amongst the background. Subsequent reactions could be optimised for the amplification of these products. Then, once the desired products had been obtained, they could be cloned into pCR2.1 (Section 2.15.2) for sequence analysis.

The method seemed to prove effective since bands within the desired size range were detected in five reactions using primer pairs: NHF1 + C8BETAR; NHR2 + C8ALPHAF; NHF2 + C8BETAR; NHF2 + NHR2; NHF1 + NHR1 (Fig. 3.5). However the background smearing visualised when these reactions were run on agarose gels was so significant that it was necessary to attempt the optimisation of the reaction as discussed in order to determine if these products were fragments of NHase encoding genes.
**Fig. 3.9:** Schematics outlining the arrangement of the PCR primers, represented in orange on the two possible NHase gene structures: A where the α subunit gene precedes the β and B, where the β subunit gene precedes that of the α subunit. The approximate sizes (in base pairs) of the products of reactions containing each primer pair, given an average length of ~630bp for the α subunit genes and ~675bp for the β subunit genes, are shown beneath the primers in blue.
The primary objective of the optimisation was to eliminate the background smearing and to reduce as much as possible contamination by non-specific amplification.

The causes of background smearing can be primarily attributed to an excess of template DNA, polymerase or MgCl$_2$, whilst non-specific priming may be due to too low an annealing temperature, too many amplification cycles or an excess primer concentration. It is commonly held that amongst the numerous variables involved in a PCR reaction, the most critical when considering optimisation steps are: (1) primer design (2) primer annealing temperature and (3) the concentration of MgCl$_2$ (Koelle, 1996, Innis et al., 1995). Therefore, to simplify optimisation, the concentrations of the template (~1ng.µL$^{-1}$), enzyme (2.5U/50µl) and primers were kept constant, whilst annealing temperature was increased stepwise by 5°C increments, at MgCl$_2$ concentrations of 1, 1.5 and 2mM. A number of other strategies for the prevention of non-specific priming have been described (Finckh and Rolfs, 1998) including ‘Hot Start’ PCR (whereby the reaction is heated to 95°C before the addition of the polymerase) and the addition of helix destabilisers such as DMSO. All PCRs performed in this section of the work were started using the Hot Start method. However, DMSO addition to 5% total volume appeared to have an inhibitory affect on the reaction, a phenomenon that can be attributed to inhibition of the polymerase (Finckh and Rolfs, 1998).

These optimisation experiments involving primers designed against the N-terminal sequences of Bacillus sp. RAPc8 did not succeed in amplifying any products of the correct size, indicating that these products were in fact the result of non-specific amplification. In contrast, there appeared to be a marked increase in the specific products involving primer pairs NHF1 + NHR1 and NHF2 + NHR2. Under the conditions outlined below, a single band of approximately 170bp was formed by the reaction containing primers NHF1 + NHR1, whilst the reaction with primers NHF2 + NHR2 produced an intense band of approximately 180bp, flanked by fainter bands of ~870bp and 120bp.
3.3.4.1 PCR procedure for amplification with NHF1 + NHR1 and NHF2 + NHR2

Reactions were performed in 1.5mM MgCl₂ and incubated for five minutes at 95°C before the addition of *Taq* polymerase to the mixture. The following program of cycles was then initiated:

1 cycle of 94°C - 5 minutes

25 cycles of
- 94°C - 30 seconds
- 55°C - 30 seconds
- 72°C - 30 seconds

1 cycle of 72°C - 7 minutes

An example of the results of the reactions is shown in Fig. 3.10. The choice of a MgCl₂ concentration at the standard level recommended for most PCR applications resulted in effective amplification. Higher concentrations resulted in considerable background smearing whereas at any lower concentration, no amplification was detected. As has been argued elsewhere (Koelle, 1996), the more critical factor for successful optimisation was found to be the annealing temperature, (Fig. 3.11) where an increase of only 5°C greatly improved the specificity of the reaction in addition to a significant reduction of background smearing.
**Fig. 3.10:** PCR amplification using primer pairs NHF1 + NHR1 and NHF2 + NHR2 run on a 3% TAE-agarose gel. Lanes: 1 - φX174 RF DNA/HaeIII fragments; 2 - PCR with NHF1 + NHR1; 3 - PCR with NHF2 + NHR2.

![PCR gel image](image1.jpg)

**Fig. 3.11:** Annealing temperature optimisation of the reaction using primers NHF2 + NHR2. Lanes: 1 - 1kb DNA ladder; 2 - negative control – no template; 3 – PCR reaction at 50°C; 4 – negative control – no *Taq* polymerase; 5 - PCR reaction at 55°C.

![Annealing temperature optimisation gel image](image2.jpg)
These above PCR products were cloned into the pCR2.1 vector using the TA cloning kit (Section 2.15.2) so as to yield two plasmids carrying inserts of ~170bp (designated pNHFl) and 180bp (designated pNHF2) respectively (identified by digestion with EcoRI). The inserts were sequenced (Section 2.22.1) and the results used in translated BLAST homology searches (Section 2.23). The high levels of homology with known NHases revealed by the BLAST searches confirmed the identity of these inserts as overlapping segments of an NHase α subunit gene. A summary of the highest scoring hits is shown in Fig. 3.12. The insert of pNHFl (167bp) corresponded to the portion of the gene coding for residues upstream of the cofactor-binding motif CXLCSC, the translation product of this fragment scoring most highly in the BLAST searches with α subunit of the NHase of a Rhodococcus species (Mayaux et al., 1991). The 179bp insert of pNHF2 correspondingly coded for the section of the protein downstream of the cofactor-binding motif and scored most highly with the α subunit of the low molecular weight NHase of R. rhodochrous J1 (Kobayashi et al., 1991). The sequences shown in Fig. 3.13 reveal discrepancies between the overlapping portion of the two fragments, as silent mutations at the nucleotide level and in the substitution of the final serine of pNHFl for a tyrosine in pNHF2. These silent mutations can probably be attributed to the use of degenerate primers, as several different primer species may have been incorporated into the amplicons, giving rise to slightly different sequences. Furthermore, as the sequencing was performed only on one strand solely for identification purposes, data towards the ends of the sequences is likely to be to a low level of confidence.

Overleaf: Fig. 3.12: Summary of the six highest scoring hits in translated BLAST searches with the inserts of A: pNHFl and B: pNHF2
Chapter 3 Cloning the NHase operon from Bacillus sp. RAPc8

A: NHF1

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Score</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>2e-14</td>
</tr>
<tr>
<td>75</td>
<td>1e-13</td>
</tr>
<tr>
<td>73</td>
<td>8e-13</td>
</tr>
<tr>
<td>73</td>
<td>8e-13</td>
</tr>
<tr>
<td>67</td>
<td>3e-11</td>
</tr>
<tr>
<td>67</td>
<td>3e-11</td>
</tr>
</tbody>
</table>

Alignments

| gi|128170|sp|P29378|NHAA2_RHORH LOW-MOLECULAR WEIGHT COBALT-CONTAINING NITRILE HYDRATASE SUBUNIT
| ALPHA (L-NITRILASE) (L-NHASE) |
| gi|98828|pir|C41326 nitrile hydratase (EC 4.2.1.84) alpha chain, L-type - Rhodococcus rhodochrous (strain J1) |
| gi|49062|emb|CA45712.1 X64360 nitrile hydratase [Rhodococcus rhodochrous] |

Length = 207
Score = 78.2 bits (191), Expect = 2e-14
Identities = 35/56 (62%), Positives = 44/56 (78%)
Frame = -1

Query: RTVAKAWDPADPKRLEDQTELRELGGGLQVHTRVVENTDTVHNVVCTLCSC 6 + VA+AW DP FKQLL D+ + RE+G +GQE + V+ENT TVHN+VCTLCSC
Sbjct: KIVARAWDPADPKRLEDQTELRELGGGLQVHTRVVENTDTVHNVVCTLCSC 113

B: NHF2

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Score</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>2e-19</td>
</tr>
<tr>
<td>94</td>
<td>2e-19</td>
</tr>
<tr>
<td>89</td>
<td>7e-18</td>
</tr>
<tr>
<td>89</td>
<td>7e-18</td>
</tr>
<tr>
<td>85</td>
<td>9e-17</td>
</tr>
<tr>
<td>84</td>
<td>1e-16</td>
</tr>
</tbody>
</table>

Alignments

| gi|2499194|sp|Q53118|NHAA_RHOSO NITRILE HYDRATASE SUBUNIT ALPHA (NITRILASE) (NHASE) |
| gi|152054|gb|AAA26185.1| (M74531) nitrile hydratase alpha subunit [Rhodococcus sp.] |

Length = 199
Score = 93.6 bits (231), Expect(2) = 2e-19
Identities = 40/55 (72%), Positives = 49/55 (88%), Gaps = 1/55 (1%)
Frame = +3

Query: YPWPLLGLPPWYKPYARVVKRPQVL-KEFGLDPDSEIRVVTWSSEVF 182
YPWP+LGLP+WYKPYAR+V+EFR VL +EF LF+S EIR+MDS+SE+R+
Sbjct: YPWPLLGLPPWYKPYARVVKRPQVL-KEFGLDPDSEIRVVTWSSEVF 182
Score = 21.6 bits (44), Expect(2) = 2e-19
Identities = 7/7 (100%), Positives = 7/7 (100%)
Frame = +1

Query: VCTLCSC 21
Sbjct: VCTLCSC 107
Fig 3.13: Sequences of the inserts from A: pNHF1 and B: pNHF2; nucleotide sequence is represented in red, the translation products shown below in black. The arrows indicate the primer positions.

A

| CGTAAGCTGATAGCCCATTGCTTTTAAAGATATGCTAGAGATTTAG | VAKAWTDPAFKQRLLLDESE |
| GACTGAAGGAGCTAGGTACATGCTGGTTTGTTCTAGGAGCAATACTAGG | TRLRELYGYGLQGEHIRV |
| JTA'AATATCAGCATACTAGTTGCTGCTGAACTATGCTTAAGCAC | VENTDTVHNVVVCTLCSS |

B

| CTGACTTGGTTTCAATGTTCCTGCGCACTTTGGGTAC | VCTLCSKYPWPLLGLPPSWY |
| AAACCCGCTTTAAGGCTGCTTGTGTAACAGCCAGCATCAGCTTCTTCATGC | KEPAYRARKVKEPQVLKEF |
| GACATTAGGATCTCCAGATTCAGAGAACTCGGATCTGGGATACAGCTCCGAGTGACG | GLDLPSVESIRVWDTSSEV |
3.3.4.2 Hybridisation with the PCR products

Using both PCR products as templates for DIG DNA labelling, probes were made for Southern analysis of the NHase gene of *Bacillus* sp. RAPc8. Signals were observed in a similar pattern to the Southern hybridisation experiments with the end-labelled primers described in Section 3.3.3, at ~4.5kb in the *BamHI* digest, ~900bp in the *EcoRI* digest, ~2.2kb in the *HindIII* digest. In addition, signals were noted at ~10kb in an *EcoRV* digest and ~8.5kb in a *PvuI* digest (Fig 3.14).

**Fig. 3.14:** Southern blot of RAPc8 genomic DNA probed with the DIG labelled PCR products NHF1 and NHF2 Lanes: M – DIG-labelled molecular weight marker; 1-5: RAPc8 DNA cut with: 1- *BamHI*; 2- *EcoRI*; 3- *EcoRV*; 4 – *HindIII*; 5- *PvuI*

As the hybridising *EcoRV* and *PvuI* fragments were rather large for screening for genes of ~1.2kb, the previously constructed *BamHI* and *HindIII* libraries were screened by colony hybridisation (Section 2.16.3-4) with the DIG-labelled PCR probes. Unfortunately the same non-specific hybridisation as had occurred with the DIG end-labelled primers was observed in these experiments. For this reason it was decided to substitute the DIG-labelled approach with $^{32}$P dCTP labelled probes for the colony screens. The method proved much more sensitive as clear positive signals were
detected from both libraries. Plasmids were isolated from equivalent colonies from duplicate plates. A plasmid from the BamHI library, designated pNHB4 and a plasmid from the HindIII library, designated pNHH16 were analysed through restriction digestion and reprobed with the above PCR products. Fragments yielding positive signals confirmed their identity. The complete plasmid inserts were then sequenced.

It was found that the HindIII fragment, (2260bp) was contained entirely within the 4393bp BamHI fragment, one HindIII site being just 16bp from one end of this longer sequence (Fig. 3.15). Translational analysis of this 4.4kb Bacillus sp. RAPc8 sequence indicated that, as with the NHase operons of R. rhodochrous J1 (Kobayashi et al., 1991) and Ps. thermophila (Yamaki et al., 1997), the β subunit gene is upstream of the α subunit gene, but that the majority of the cloned sequence lay downstream from the NHase genes. Whilst the entire α subunit gene was contained within the fragment, only 347bp (approximately half) of the β subunit gene had apparently been cloned (Fig 3.15). To obtain the full-length sequence of the β gene and as much upstream information as possible, fragments that would contain these sequences were identified with the aid of the restriction site data available (Figs. 3.14, 3.15). From the position of the EcoRI site within the α subunit gene, it was clear that the remainder of the ~900bp hybridising fragment (Fig 3.14) would contain only another ~150bp of the unknown β sequence and was therefore not worth further investigation. As an EcoRV site was found downstream of the NHase genes in pNHB4, the remainder of the NHase operon would almost certainly be identified in a clone of the 10kb EcoRV fragment. However it was thought that Southern hybridisations against genomic DNA cut with NcoI, NdeI, SacI, or XhoI may identify sequences of a more manageable size.
**Fig. 3.15:** Organisation of NHase genes within the insert of pNHB4. Arrows represent the coding sequences of the genes of each subunit. Positions of several of the commonly used enzymes are shown in the diagram. ORFs found downstream of the α subunit are shown later in Fig. 3.22 Bar = 1000 nucleotides.

3.3.5 PCR with primers designed against the amidase gene sequence of *Bacillus* sp. BR449

As library construction and hybridisation screening is a lengthy procedure, a second, faster approach was employed concurrently. BLAST searches (data not shown) with the partial NHase sequences revealed a very high identity with the sequences of the NHase genes of *Bacillus* sp. BR449 (Kim and Oriel, 2000). This became publicly available only following the sequencing of the PCR products described in the previous section. The DNA sequences of the α subunit genes had 99% identity (98% at the amino acid level) and the partial β subunit gene, 99% identity with the corresponding genes of *Bacillus* sp. BR449 (97% at the amino acid level). These exceptionally high levels of sequence identity prompted an attempt to use the sequence information of *Bacillus* sp. BR449 to design primers to clone the remaining unknown sequence. Given that all previously described NHase genes are located downstream of an amidase gene and postulating that the amidase genes of the two organisms would also share a similar level of sequence identity, primers were designed corresponding to the 5' and central regions of the amidase gene sequence of *Bacillus* sp. BR449 and to the 3' region of the
Clone the NHase operon from Bacillus sp. RAPc8 NHase. Using the High Fidelity enzyme mix (Roche), to minimise chances of polymerase error, two amplicons (Fig 3.17) of 643bp and 1787bp were generated with after a PCR reaction using primers described in Fig 3.16 below. The products were cloned and sequenced and, as expected, found to code for the entire amidase gene and the 5' end of the NHase β subunit gene found in pNHβ4. The program of cycles used for both amplifications was as follows:

1 cycle of 94°C - 5 minutes
30 cycles of 94°C - 30 seconds
50°C - 30 seconds
72°C - 1.5 minutes
1 cycle of 72°C - 7 minutes

Fig. 3.16: A: Organisation of the primers AMD1, AMD2 and C8B2 used in the amplification of the amidase and NHase β genes. Bar = 1000 nucleotides.
B: Nucleotide sequence of the primers used.

A:

B:

AMD1: ATG AGA CAC GGG GAT ATT TC
AMD2: GAG CAG CAA ATC ATG ATG GC
C8B2: CAA CGC CAA TCA GCA ACG AC
3.4 Nitrile hydratase operon structure

Translational analysis of the complete 5911bp locus revealed nine putative ORFs, one truncated by the BamHI site (Fig. 3.22). All ORFs had the same polarity, started with an ATG codon and terminated with a TAA codon with the exception of ORF3 (TAG) and ORF8 (TGA). The identities of these potential ORFs were deduced with the aid of BLAST searches (Appendix II.ii – II.xi). Potential Shine-Dalgarno ribosome binding sequences (S.D.) were identified 6-9 nucleotides upstream of the start codons of each potential gene with the exception of ORF8. The consensus S.D. sequence was (A/G)G(A/G)(A/G)G(A/G); the particular sequence of each S.D. site is detailed in a summary of the gene properties in Table 3.2. The overall GC content of the cloned DNA was 40.2%.

ORF1 encoded a 348 amino acid protein (MW 38.6kDa), with 100% identity with the amidase of Bacillus sp. BR449. However, due to six silent nucleotide substitutions the DNA sequences shared a slightly lower identity (99%).
ORF2, located 127bp downstream of the stop codon of the amidase, encoded a protein of 229 amino acid (26.5kDa) the N-terminus of which corresponds precisely to the previously determined N-terminus of the purified NHase β subunit of *Bacillus* sp. RAPc8 (Pereira, 1998). The entire sequence showed significant similarity to those of known NHase β subunits (Figure 3.19 and Table 3.1), particularly that of *Bacillus* sp. BR449 (99% identity).

25bp downstream of ORF2, ORF3 encodes a protein of 216 amino acids (24.6kDa) which, although displaying high similarity to many α NHase sequences (Figure 3.18 and Table 3.1) including a 97% identity match with the α subunit of *Bacillus* sp. BR449, showed absolutely no similarity to the N-terminal sequence deduced from the purified protein (Pereira, 1998). As with the α subunit of *Bacillus* sp. BR449, the sequence contained the motif CTLCSCY (residues 116-121), characteristic of cobalt binding enzymes, as opposed to the CSLCSCT iron-binding motif (Huang *et al.*, 1997).

**Table 3.1:** Sequence identity and similarity of the genes and peptide sequences of *Bacillus* sp. RAPc8 NHase with those recently characterised from 13 microorganisms.

<table>
<thead>
<tr>
<th>NHase ORIGIN</th>
<th>ALPHA SUBUNIT</th>
<th>BETA SUBUNIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DNA identity (%)</td>
<td>aa identity (%)</td>
</tr>
<tr>
<td><em>Bacillus</em> BR449</td>
<td>99</td>
<td>97</td>
</tr>
<tr>
<td><em>B. smithii</em></td>
<td>82</td>
<td>88</td>
</tr>
<tr>
<td><em>P. putida</em></td>
<td>57</td>
<td>55</td>
</tr>
<tr>
<td><em>Klebsiella</em></td>
<td>56</td>
<td>53</td>
</tr>
<tr>
<td><em>Rhod. J1 L</em></td>
<td>53</td>
<td>54</td>
</tr>
<tr>
<td><em>Rhod. sp.</em></td>
<td>?</td>
<td>56</td>
</tr>
<tr>
<td><em>P. thermophila</em></td>
<td>56</td>
<td>59</td>
</tr>
<tr>
<td><em>S. meliloti</em></td>
<td>53</td>
<td>50</td>
</tr>
<tr>
<td><em>M. loti</em></td>
<td>48</td>
<td>69</td>
</tr>
<tr>
<td><em>Rhod. N774</em></td>
<td>51</td>
<td>46</td>
</tr>
<tr>
<td><em>Rhod. R312</em></td>
<td>46</td>
<td>45</td>
</tr>
<tr>
<td><em>R. erythropolis</em></td>
<td>52</td>
<td>44</td>
</tr>
<tr>
<td><em>P. chlororaphis</em></td>
<td>51</td>
<td>45</td>
</tr>
</tbody>
</table>
Fig. 3.18: Alignment of the peptide sequences NHase of the α subunits, including that of *Bacillus* sp. RAPc8. The 11 cobalt containing enzymes are listed first, followed by the iron containing enzymes of *Rhodococcus N774*, *Rhodococcus R312*, *R. erythropolis* and *P. chlororaphis*. Highlighted at 60% conservation, identical residues are in black, similar residues in grey. The cofactor-binding region is highlighted in red.
### Chapter 3 Cloning the NHase operon from Bacillus sp. RAPc8

#### Beta

<table>
<thead>
<tr>
<th>Organism</th>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus c8</td>
<td>PALVKLVEKALLELGSPRVEEYFAPIKRVVKNHFDQ-116</td>
</tr>
<tr>
<td>Bacillus BR449</td>
<td>PALVKLVEKALLELGSPRVEEYFAPIKRVVKNHFDQ-116</td>
</tr>
<tr>
<td>B. smithii</td>
<td>PALVKLVEKALLELGSPRVEEYFAPIKRVVKNHFDQ-116</td>
</tr>
<tr>
<td>P. putida</td>
<td>RTSEQWPDLSAAHQP</td>
</tr>
<tr>
<td>Klebsiella</td>
<td>RTSEQWPDLSAAHQP</td>
</tr>
<tr>
<td>Rhodococcus J1 H</td>
<td>RTSEQWPDLSAAHQP</td>
</tr>
<tr>
<td>Rhodococcus J1 L</td>
<td>RTSEQWPDLSAAHQP</td>
</tr>
<tr>
<td>Rhodococcus sp.</td>
<td>RTSEQWPDLSAAHQP</td>
</tr>
<tr>
<td>Ps. thermophila</td>
<td>RTSEQWPDLSAAHQP</td>
</tr>
<tr>
<td>S. meliloti</td>
<td>RTSEQWPDLSAAHQP</td>
</tr>
<tr>
<td>Rhodococcus N774</td>
<td>RTSEQWPDLSAAHQP</td>
</tr>
<tr>
<td>Rhodococcus R312</td>
<td>RTSEQWPDLSAAHQP</td>
</tr>
<tr>
<td>R. erythropolis</td>
<td>RTSEQWPDLSAAHQP</td>
</tr>
<tr>
<td>P. chlororaphis</td>
<td>RTSEQWPDLSAAHQP</td>
</tr>
</tbody>
</table>

#### Fig. 3.19: Alignment of NHase β subunit peptide sequences, including that of Bacillus sp. RAPc8. The 11 cobalt containing enzymes are listed first, followed by the iron containing enzymes of *Rhodococcus N774*, *Rhodococcus R312*, *R. erythropolis* and *P. chlororaphis*. Highlighted at 60% conservation, identical residues are in black, similar residues in grey.
ORF4 was found only 13bp further downstream and encoded a 14.6kDa protein of 122 amino acids. BLAST searches indicated that it had significant similarities to P12K of *Bacillus* sp. BR449, NhhG (Komeda et al, 1996a) and NhlE (Komeda et al., 1997) of *Rhodococcus* J1, P14K of *P. putida*, and hypothetical proteins of *S. meliloti* and *M. loti* (81%, 33%, 29%, 12%, 25% and 22% respectively), all of which are associated, at least by proximity with the NHase genes of their respective organisms. Interestingly, with the exception of *P. putida* P14K, these proteins also have a low but significant similarity (15-30%) to the N-terminus of most NHase β subunits. The significance of these proteins is discussed in the Section 3.6.2.

Fig. 3.20: Alignment of the translation product of ORF4 with the homologues identified in BLAST searches. Highlighted at 60% conservation, identical residues are in black, similar residues in grey.

A strong (ΔG = -14.6 kcal.mol⁻¹) hairpin structure (Fig. 3.21) that may serve as a ρ independent transcription termination signal (d'Aubenton Carafa et al., 1990) is located 11bp downstream of the termination codon of the *pl4k* gene. The absence of significant structures elsewhere in the sequence and the proximity of the NHase α and β and *pl4k* genes suggests that these are all co-transcribed as a single mRNA which most
likely also contains the amidase sequence. This would be in accordance with the suggestions in the literature regarding the related NHase operons of *Rhodococcus* sp. R312 (Bigey et al., 1999), *Rhodococcus erythropolis* (Duran et al., 1993) and *Pseudomonas chlororaphis* (Nishiyama et al., 1991).

**Fig. 3.21:** Hairpin loop structure found just downstream of the *P14K* gene.

269bp away from the genes of this proposed nitrile hydratase operon, ORFs 5-9 are organised in a manner that suggests that they may be transcribed in a single polycistronic mRNA so as to form part of a separate operon. Furthermore, each gene has significant homology to genes involved in cobalt uptake and linked to cobalamin biosynthesis pathways.

The first gene of this putative operon, ORF5, encodes a 119 amino acid (13.8kDa) protein with similarity to the 2Fe-2S class of ferredoxins (Meyer, 2001): it has a 48% amino acid identity with a hypothetical ferredoxin of *B. halodurans* (Takami et al., 2000) and 39% with a ferredoxin homologue of *B. megaterium*, designated *cbiW*. The *cbiW* gene is the first in the *cbi* (cobalamin biosynthesis) operon of the latter organism (Raux et al., 1998).

ORFs 6-9 encode putative peptides of 252aa (24.7kDa), 100aa (11kDa), 215aa (24.7kDa) and 159aa, the highest scoring BLAST hit in each case being with putative cobalt uptake proteins of the putative cobalamin biosynthesis pathway of *Clostridium perfringens* (Shimizu et al., 2002): ORF6 - 56% identity with *cbiM*; ORF7 - 45% identity with *cbiN*; ORF8 - 34% identity with *cbiQ*; ORF9 - 52% identity over 159aa with *cbiO*. The significance of these putative *cbi* homologues is discussed in Section 3.5.3.
### Table 3.2: Summary of the properties of the genes found in the *Bacillus* sp. RAPc8 Nlase locus

<table>
<thead>
<tr>
<th>GENE</th>
<th>Start</th>
<th>Stop</th>
<th>Codon</th>
<th>S.D. sequence</th>
<th>Length of:</th>
<th>Comparison to reference protein identified through BLAST searches (Appendix II)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gene Peptide</td>
<td>Organism</td>
<td>Gene</td>
</tr>
<tr>
<td>Amidase</td>
<td>1</td>
<td>1047</td>
<td>TAA</td>
<td>TAA</td>
<td>1044 348</td>
<td>\textit{Bacillus} sp. BR449 Amidase</td>
</tr>
<tr>
<td>\textit{nh\beta}</td>
<td>1174</td>
<td>1864</td>
<td>TAA</td>
<td>GGGAGG</td>
<td>687 229</td>
<td>\textit{Bacillus} sp. BR449 \textit{NH\beta}</td>
</tr>
<tr>
<td>\textit{nha}</td>
<td>1890</td>
<td>2540</td>
<td>TAG</td>
<td>AGGAGG</td>
<td>648 216</td>
<td>\textit{Bacillus} sp. BR449 \textit{NHa}</td>
</tr>
<tr>
<td>\textit{p14k}</td>
<td>2551</td>
<td>2919</td>
<td>TAA</td>
<td>AGGAGG</td>
<td>366 122</td>
<td>\textit{Bacillus} sp. BR449 \textit{P12K}</td>
</tr>
<tr>
<td>ORF5</td>
<td>3188</td>
<td>3548</td>
<td>TAA</td>
<td>GGGAGA</td>
<td>357 119</td>
<td>\textit{B. megaterium} \textit{Ferredoxin (cbiW)}</td>
</tr>
<tr>
<td>ORF6</td>
<td>3599</td>
<td>4357</td>
<td>TAA</td>
<td>AGGAGG</td>
<td>756 252</td>
<td>\textit{S. typhimurium} \textit{LT2} \textit{cbiM}</td>
</tr>
<tr>
<td>ORF7</td>
<td>4359</td>
<td>4663</td>
<td>TAA</td>
<td>AGGAGG</td>
<td>300 100</td>
<td>\textit{S. typhimurium} \textit{LT2} \textit{cbiN}</td>
</tr>
<tr>
<td>ORF8</td>
<td>4785</td>
<td>5433</td>
<td>TGA</td>
<td>TGA</td>
<td>645 215</td>
<td>\textit{S. typhimurium} \textit{LT2} \textit{cbiQ}</td>
</tr>
<tr>
<td>ORF9</td>
<td>5433</td>
<td>5911</td>
<td>-</td>
<td>GGAGGA (479) (159)</td>
<td>479 (159)</td>
<td>\textit{S. typhimurium} \textit{LT2} \textit{cbiO}</td>
</tr>
</tbody>
</table>

Table 3.2: Summary of the properties of the genes found in the *Bacillus* sp. RAPc8 Nlase locus
Bacillus sp. RAPc8 NHase gene cluster
(5911 bp)

Fig 3.22: Layout of the Bacillus sp. RAPc8 gene cluster containing the NHase genes. Potential open reading frames are represented by black arrows, the identity of each ORF given below. Bar = 1000 nucleotides
Fig 3.23: Nucleotide sequence (red) and deduced amino acid sequences (black) of the ORFs contained in the *Bacillus* sp. RAPcS NHase gene cluster. Shine Dalgarno sequences (S.D.) are represented by black lines.
3.5 Discussion

3.5.1 Nitrile hydratase and amidase genes

This chapter describes the cloning of the NHase and amidase genes of *Bacillus* sp. RAPc8 using DNA hybridisation and PCR amplification techniques. Whilst the DIG hybridisation procedure used initially provided a fast and safe method of detection, its lack of sensitivity precluded its use in the colony hybridisations. The level of sequence identity with the *R. erythropolis* α gene (Table 3.1) was subsequently found to be equivalent to the identity between the L-NHase genes of *R. rhodochrous* J1 and those used to identify them from *Rhodococcus* sp. N-774 (Fig. 3.18). This suggests that the experiment described in Section 3.3.1.2 may have been more successful had a more sensitive[^P] dCTP labelled probe been used in place of the DIG labelled probe.

3.5.1.1 Lack of similarity with the N-terminus of the native enzyme

The sequence of the nitrile hydratase operon raises concerns over the total lack of similarity between the gene sequence of the α subunit and the N-terminal sequence of the purified native protein. N-terminal sequencing of the recombinant α subunit gene product (Section 4.6.1) subsequently confirmed the sequence predicted by computational translation, therefore leaving three possible explanations for this anomaly: (a) That the gene been cloned from an organism other than that *Bacillus* sp. RAPc8; (b) That, as with *Rhodococcus* sp. J1 the *Bacillus* sp. RAPc8 produces more than one NHase, or (c), that the N-terminal sequence previously reported (Pereira, 1998) was incorrect. Given the similarity of the growth properties, enzyme activity and 16S rDNA sequences reported in this work and that of Pereira, it would seem unlikely that the cloned DNA originated from another organism. The observation of only one hybridising fragment (Sections 3.3.4.1 and 3.4) indicates that only one set of NHase genes is present on the chromosome (the observation of double bands in *EcoRI* and *HindIII* lanes probed with the entire operon is explained simply by the fact that such sites are found within this sequence). It is probable therefore that the N-terminal data of the alpha subunit reported by Pereira is erroneous.

It is possibly significant that the relative molecular weights estimated by Pereira were all approximately 10% greater than those calculated from the predicted amino acid sequence (Table 3.3). Such overestimations however are not uncommon and it is worth noting that a similar 10% overestimation was made of the wide spectrum amidase of
Rhodococcus sp. R312 (Azza, 1994, Chion et al., 1991). Whilst the calculated molecular weights of the NHase subunits remain in agreement with Pereira's prediction of a heterotetrameric enzyme, the decrease in Mw of the amidase subunit more closely fits the hexameric structure of the nitrilase-related amidases than the pentameric structure previously suggested (Pereira, 1998).

Table 3.4: Comparison of the molecular weights of Bacillus sp. RAPc8 amidase and nitrile hydratase subunits and native proteins.

<table>
<thead>
<tr>
<th>Relative molecular weight</th>
<th>NHase</th>
<th>Amidase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Native α subunit</td>
<td>β subunit</td>
</tr>
<tr>
<td>Estimated¹</td>
<td>108 ± 5</td>
<td>28 ± 1.5</td>
</tr>
<tr>
<td>From amino acid sequence²</td>
<td>-</td>
<td>24.6</td>
</tr>
</tbody>
</table>

¹ Estimated chromatographically by Pereira (1998)
² Calculated from the predicted amino acid sequences.

3.5.1.2 High sequence identity with Bacillus sp. BR449 NHase operon

The second important observation is the very high similarity to the NHase and related genes of Bacillus sp. BR449 (Kim and Oriel, 2000). The length of cloned DNA common to both organisms, comprising 2868bp contained just 17 nucleotide differences, corresponded to an identity of 99.41% between the two sequences. Of the six nucleotide substitutions found within the amidase genes, each was silent in terms of the translated peptide sequence. The β gene contained five substitutions (3 amino acid residue substitutions) whilst the α gene contained a single substitution at position 2139, resulting in a P to S substitution in the amino acid sequence, and an insertion at the end of the gene (position 2521) resulting in a change in reading frame from the BR449 gene but an increase in sequence identity with the C-terminus of the α subunit of B. smithii. The pI 4k genes had one substitution at position 2778, resulting in an E to D substitution and a deletion resulting in a stop codon, TAA, 64bp further downstream than that of BR449. Unfortunately an error in the sequencing (which has since been corrected) led
to the belief that this gene terminated with the same sequence as that of the BR449 P12K gene. As a result an incomplete gene was sub-cloned for use in the expression studies (Chapter 4).

The remarkably high sequence identity between the two operons suggests that they originate from very closely related *Bacillus* strains. A comparison of the 16S rDNA of *Bacillus* sp. RAPc8 and the corresponding 485bp from *Bacillus* sp. BR449 (Oriel *et al*., 2001) revealed a 99.79% sequence identity, supporting this conclusion. Though it is likely that *Bacillus* sp. BR449 was isolated from geothermal soil or samples similar to that from which RAPc8 was isolated, its precise origin has not been reported.

The few sequence differences noted above may of course be due to sequencing errors either in the work of Oriel *et al* or in this work. However, both the sequence of BR449 NHase (Kim and Oriel, 2000) and the RAPc8 NHase gene sequences reported here were determined to a high level of confidence. Furthermore, previous studies on a number of NHase-producing *Rhodococcus* species have demonstrated that related organisms contain NHase genes with remarkably high sequence identities (Duran *et al*., 1993, Hashimoto *et al*., 1994, Ikehata *et al*., 1989).

Comparisons of the NHases of *R. erythropolis* and *Rhodococcus* sp. N-774, which share 94.7% and 96.2% amino acid sequence identity for the α and β subunits respectively, demonstrated that a very limited number of amino acid changes can significantly change NHase activity and substrate specificity (Duran *et al*., 1993). Therefore, whilst it is clear that the enzymes produced by the two *Bacilli* are closely related, it would be reasonable to expect variations in catalytic behaviour. However, only one residue difference was found in the central portion of sequence of the α subunit. As this region is widely regarded as the primary determinant of cofactor binding and substrate specificity, it would follow that such catalytic variations may be slight.

The extremely high similarity of the two nucleotide sequences, even in intervening non-coding sequences, suggests that the gene clusters of each organism are almost identical. Therefore it is likely that the two ORFs found upstream of the amidase of BR449 are found also in RAPc8 and that the operons will share identical promoter elements. However, functions for these upstream ORFs have not yet been elucidated. Based on
this data, it was also assumed that the expression of the *Bacillus* sp. RAPc8 NHase gene would be possible under the conditions used outlined by Kim and Oriel (2000).

### 3.5.1.3 Transcript production

The organisation of the amidase, NHase α and β and P12K genes, the similarity of their Shine-Dalgarno sequences and the presence of a stem loop structure just downstream suggests that they are transcribed as a single polycistronic mRNA. It is thought that a similar transcriptional organisation is used by the NHase operons of *Rhodococcus* sp. N-774 (Hashimoto *et al.*, 1994), *Rhodococcus* sp. N-771 (Nojiri *et al.*, 1999), *R. erythropolis* (Duran *et al.*, 1993), *P. chlororaphis* B23 (Nishiyama *et al.*, 1991). This has been confirmed to be the case with the operon of *Rhodococcus* sp. R312 (Bigey *et al.*, 1999). Following the molecular characterisation of the *Bacillus* sp. BR449 NHase operon, Kim and Oriel (2000) presented the hypothesis that the spacing between the NHase and amidase genes and the significant heterologous NHase expression in reverse orientation to the vector promoter suggests the presence of a promoter upstream to the NHase genes that facilitates their expression independently of the amidase. Whilst no promoter elements were detected within the cloned DNA of RAPc8, confirmation of either hypothesis can only be irrefutably demonstrated by transcript analysis.

### 3.5.1.4 RAPc8 amidase belongs to the nitrilase-related family

*Bacillus* sp. RAPc8 amidase, like the identical amidase of *Bacillus* sp. BR449 (Kim and Oriel, 2000), belongs to the nitrilase related amidase family. As was discussed in Section 1.4.1, all other previously described NHase-associated amidases belong to the aspartic proteinase class, hence the anomaly presented by the amidases of these two closely related organisms would seem to suggest that (in this case) the NHase operon was assembled from genes pre-existing with the organism’s genome rather than being result of horizontal gene transfer as has been previously suggested (Chebrou, 1996, Pereira, 1998). Kim and Oriel (2000) also argue the variation in amidase type is further evidence that transcription of *Bacillus* sp. BR449 NHase genes may differ from other NHase operons in so far as the amidase gene is probably not included in the polycistronic mRNA.
3.5.2 Significance of the 14kDa protein

The fourth open reading frame in the cloned gene cluster (Section 3.4), codes for a protein designated P14K, that has significant sequence similarities with the beta subunit homologues P12K (*Bacillus* sp. BR449); NhbG, NhlE, (*R. rhodochrous* J1); and P14K (*Pseudomonas putida* 5B). Although the sequence similarities are not high, these proteins share several common traits: these proteins are encoded immediately downstream of the genes for the corresponding NHase subunits and appear to be part of a polycistronic operon. Furthermore, the NHases with which they are associated are all cobalt containing enzymes.

Whilst the precise function of the β homologues is not known, it has been found that their presence, together with cobalt ions, is essential for functional heterologous expression of *R. rhodochrous* J1 and *P. putida* 5B NHases in *E. coli* (Kobayashi et al., 1991, Komeda et al, 1996a, 1997, Wu et al., 1997). Studies of the *P. putida* P14K protein have shown that it is expressed at a very low level, leading Wu *et al.* (1997) to postulate that it serves as a chaperone or acts in a catalytic capacity. In each case, it has been shown that cobalt is required for activity rather than for expression or correct folding and that its incorporation into the enzyme is not passive. The chaperon/catalyst hypothesis might therefore predict that the β homologues are involved in the integration of cobalt into the active site.

Conversely, Kim and Oriel (2000) found that functional expression of *Bacillus* sp. BR449 NHase in *E. coli* required only the structural genes and the inclusion of cobalt ions in the growth medium. Expression of the P12K protein was not detected when the p12k gene was included in expression constructs, neither did the gene seem to affect NHase expression or activity. As shown in Section 4.4.2, similar results were observed in expression experiments with the *Bacillus* sp. RAPc8 NHase. These results may indicate that the original function of the β homologues of the two *Bacilli* has been lost and is no longer required for active NHase expression.

3.5.3 Significance of the cbi homologues

The molecular genetics of cobalamin (vitamin B12) biosynthesis has been studied in depth in only three organisms: *Bacillus megaterium* (Raux *et al*., 1998); *Pseudomonas denitrificans* (Cameron *et al*., 1989; Crouzet *et al*., 1990a, Crouzet *et al*., 1990b,
Crouzet et al., 1991), and Salmonella typhimurium (Roth et al., 1993). Although gene clustering is evident in all three organisms, the gene orders are quite different, with a number of the individual genes being unique to one organism (Fig 3.24). These genetic differences reflect the different pathways of cobalamin synthesis used by these organisms (Raux et al., 1998).

Translated BLAST searches with the sequences of ORFs 5-9 revealed that each had significant homology to a gene involved in cobalamin biosynthesis. Interestingly however, each homologue was unique to either B. megaterium or S. typhimurium, suggesting that they may not be crucial genes in cobalamin biosynthesis. ORF 5 shared 39% sequence identity with the cbiW gene product of B. megaterium, a ferredoxin homologue that is suggested to act as an oxido-reductant (Raux et al., 1998).

Fig. 3.24: Taken from Raux et al., (1998). A. The order of the cobi genes found in B. megaterium. The presence and direction of the transcription terminators, T1–T6, are also shown. B. The order of the cbi genes found in the S. typhimurium cobi operon. Similar genes are indicated by the dotted lines. Genes represented by shaded arrows are common to S. typhimurium, B. megaterium and P. denitrificans, those represented by black arrows are not found in P. denitrificans whilst un-shaded arrows are unique.

ORFs 6-9 showed significant homology (between 54 and 20% amino acid sequence identity), to the gene products of (in order): cbiM; cbiN; cbiQ and cbiO of S. typhimurium. These genes do not have homologues in P. denitrificans or B. megaterium but most significantly, three are reported to be involved in cobalt uptake.
Throughout the *S. typhimurium* cobalamin biosynthesis operon, these four are the only genes unique to that organism. Whilst the homologous genes in *Bacillus* sp. RAPc8 are found in the same order, unlike their *S. typhimurium* counterparts (which often overlap), there is significant spacing between the genes and only one (potential) functionally associated gene found upstream. Whilst the function of cbiM is unknown, BLAST searches and hydropathy analyses were used to predict that cbiO is an ATP hydrolysing transport protein. Both cbiQ and cbiN, which have hydropathy plots (Fig. 3.25), typical of proteins with multiple membrane spanning domains, may serve as anchor proteins, stabilising cbiO at the membrane.

**Fig. 3.25:** Comparison of Kyte and Doolittle hydrophobicity plots of *S. typhimurium* (blue) **A:** cbiQ and **B:** cbiN with the *Bacillus* sp. RAPc8 homologues (red), potential transmembrane regions are numbered, weaker regions are shown in parentheses.
Further investigation, including cloning of the sequence downstream of that found in the 4.5kb BamHI fragment, is clearly required before grounded predictions can be made as to the function of these genes. However, the proximity of the genes to an operon producing a cobalt-containing enzyme, and the absence of further cobalamin biosynthesis genes upstream does suggest that they are involved in cobalt uptake. This view is further supported by the fact that iron and cobalt transport proteins have been found associated with other NHase operons. NhlF of the \textit{R. rhodochrous} J1 L-NHase gene cluster has been shown to confer cobalt uptake on \textit{Rhodococcus} and \textit{E. coli} hosts and its expression significantly enhances L-NHase activity in \textit{Rhodococcus} transformants in cobalt limiting conditions (Komeda \textit{et al.}, 1997). An 1188bp ORF found in the operons of \textit{Rhodococcus} sp. N-774 and N771, encodes a protein that bears distinct homology to an ATP dependent iron-transporter, magA of \textit{Magnetospirillum} sp. AMB-1 and, interestingly, to cobW (distinct from cbiW) of \textit{P. denitrificans}. This ORF has found to be important for high-level recombinant expression (Hashimoto \textit{et al.}, 1994, Nojiri \textit{et al.}, 1999). If indeed this is the case, then it may be that heterologous co-expression of these genes with the NHase genes would enable the expression of a higher level of NHase activity than presently achieved and would also abolish the need for subsequent heat activation of the enzyme (Kim \textit{et al.}, 2001).
Chapter 4: Expression and characterisation

4.1 Introduction

From the outset of the project, the heterologous expression of the nitrile hydratase genes of *Bacillus* sp. RAPc8 and a comparison of the properties of the recombinant enzyme with those of native protein was a central aim. The discrepancies in molecular weights and N-terminal sequences between the purified NHase and those predicted from the gene sequences meant that a characterisation of the recombinant enzyme was critical for the verification of its identity.

Many early attempts to express cloned NHase genes were inefficient, often resulting in the production of insoluble and inactive inclusion bodies (Ikehata *et al.*, 1989, Mizunashi *et al.*, 1998). Successful heterologous expression was finally enabled through the development of host-vector systems in *R. rhodochrous* strains for the expression of recombinant *Rhodococcus* NHases (Hashimoto *et al.*, 1992) and the discovery that active overexpression could be achieved when additional "activator proteins" were coexpressed from genes flanking the NHase genes (Hashimoto *et al.*, 1994, Nishiyama *et al.*, 1991). The NHases of *R. rhodochrous* J1, *Rhodococcus* sp. ACV2, N-774 and N-771, *P. chlororaphis* B23 and *P. putida* 5B have now been expressed as active enzymes in *E. coli* or *R. rhodochrous* strains in the presence of their respective activators. The expression of *Bacillus* sp. BR449 NHase in *E. coli* (Kim and Oriel, 2000) was unique in that only the structural genes were required for expression of activity. To the author’s knowledge, no reports have yet been made of the expression of the other thermostable NHases of *Ps. thermophilia*, *B. smithii* or *B. pallidus* DAC521.

Only four NHase associated amidases (from *Rhodococcus* sp. R312, *R. rhodochrous* J1, *P. chlororaphis* B23 and *Bacillus* sp. BR449) have been expressed to date. In contrast to the expression of the NHases, amidase expression has proved rather more straightforward, with no requirement for inducers (other than IPTG), or additional proteins.
4.2 Expression strategy

The volume of work previously published on the expression of NHase genes provided a sound basis for this investigation. Every attempt to express a cobalt-containing enzyme has revealed that the addition of cobalt ions to the growth medium is essential for efficient expression of activity; typically 0.1-0.5mM CoCl_2 is used (Mizunashi et al., 1998, Wu et al, 1997).

The high sequence identity of the NHase operon of Bacillus sp. RAPc8 with that of Bacillus sp. BR449 as highlighted in Section 3.5.1.2 indicated that the requirements for expression would be similar and therefore a similar approach should be taken as that described by Kim and Oriel (2000). It was thus assumed that the expression of RAPc8 NHase would not require the presence of flanking genes and that the amidase expression would also be straightforward. Attempts to express the Bacillus sp. BR449 NHase genes under the control of an IPTG-inducible lac promoter resulted in a very weak activity; a constitutive expression system of E. coli DH5α and pBluescriptII as host and expression vector was therefore used instead (Kim and Oriel, 2000). However, the vector inserts that were used in their experiments were restriction fragments of cloned DNA that contained substantial upstream sequences prior to the start codons of the genes for expression; working on the assumption that these sequences may have reduced the transcriptional efficiency of the vector it was decided to attempt expression from an inducible expression vector.

The strategy was to use PCR to amplify the individual genes, which were subsequently cloned into pCR2.1. Restriction sites were engineered into the primer ends to enable subcloning into the expression vector pET21a(+) in the correct orientation and reading frame. When transformed into the appropriate E. coli host, a T7lac promoter in the expression vector facilitates production of the recombinant protein only after induction with IPTG.
4.3 Subcloning of NHase and amidase genes

Three expression constructs were investigated, one containing the amidase gene, one containing the NHase α and β genes and the third comprising the NHase genes together with the downstream P14K gene, (an Ndel site between the amidase and NHase β genes prevented the cloning of the entire operon). As reported in Chapter 3, an error in the interpretation of the initial sequence results meant that the third expression construct did not contain the entire P14K gene, hence the product could not be adequately characterised. A summary of the results of that work is presented with the results of the NHase expression.

Five primers (Fig. 4.1) were designed for the amplification of each set of genes. By engineering restriction sites into the PCR primers, the genes could be inserted in the correct orientation and reading frame just downstream of the ribosome binding site of pET21a(+). Restriction sites and stop codons were included in such a manner so as to avoid the addition of C or N-terminal tags to the expressed proteins. The ATG start codon of the amidase gene and the β gene were used to create an Ndel site for the 5' (sense strand) insertion point whilst EagI (amidase construct), or NotI (NHase construct) sites were added to the ends of the 3' primers (anti-sense strand).

The genes were amplified using the High Fidelity enzyme mix (Roche) so as to avoid mutations as a result of polymerase errors. All reactions were performed under the following conditions:

1 cycle of 94°C - 5 minutes

30 cycles of 94°C - 30 seconds
50°C - 30 seconds
72°C - 1.5 minutes

1 cycle of 72°C - 7 minutes
Fig. 4.1: A: Layout of the NHase operon of *Bacillus* sp. RAPc8 and the organisation of the primers (represented by orange arrows) used for the cloning of the genes for expression studies. Bar = 500 nucleotides. B: Nucleotide sequence of the primers used, sequence in red indicates the restriction site position, sequence in uppercase corresponds to the gene coding sequence.

A

B

NHOP1:  
gct cat ATP AAC GGT ATT CAT GAT GIT GG (NdeI site) (29mer)

NHOP2:  
gca ggg ggt GCT TAT TTA ACA TTC CTG TTT CC (NotI site) (32mer)

NHOP3:  
gca ggg ggg gGC TAA CCT ACC GTA ACT TTA GG (NotI site) (32mer)

AMD3:  
gct cat ATP AGA CAC GGG GAT ATT TC (NdeI site) (26mer)

AMD4:  
gca cgg cgt TAT TCC AGA ATT ACA CAC C (EagI site)
Following PCR, the products were purified using the QIAquick PCR Purification kit and cloned into pCR2.1. Positive clones (pNH22 – amidase gene; pNH46 -α, β and P14K; pNH51 – NHase α and β) were identified by restriction analysis then sequenced to ensure that no mutations had occurred during PCR. The inserts were excised by means of the engineered restriction sites, purified using the QIAquick Gel Extraction kit and cloned into the corresponding sites of pET21a(+) . The ligations were allowed to proceed overnight before being used to transform chemically competent E. coli JM107. Positive recombinants (pNH223 – amidase gene; pNH461 – NHase α, β and P14K; pNH512 – NHase α and β) were re-identified by restriction analysis. Cultures of each clone were used to prepare glycerol stocks for long-term storage at -70°C. Overviews of the cloning and subcloning of the amidase and NHase genes are shown in Figures 4.3 and 4.4.
Fig. 4.3: Schematic of the PCR cloning and subcloning into the pET expression vector of the amidase gene of *Bacillus* sp. RAPc8. Black blocks represent multiple cloning sites; inserted DNA is shown as a red block whilst vector genes are shown as blue arrows.
Fig. 4.4: Schematic of the PCR cloning and subcloning into the pET expression vector of the NHase genes of Bacillus sp. RAPc8. Black blocks represent multiple cloning sites; inserted DNA is shown as a red block whilst vector genes are shown as blue arrows.
4.4 Expression studies

The pET system chosen for these studies enables stringent control of gene expression. The host strain used for expression was *E. coli* (BL21) λDE3 pLysS, a lysogenic strain that contains the gene for T7 RNA polymerase (T7pol), under the control of a lacUV5 promoter. T7 lysozyme, an inhibitor of T7pol, is produced in low quantities by the compatible plasmid pLysS to provide additional stability through the inhibition of basal levels of T7pol produced before induction. The target genes are inserted into the expression vector downstream of a T7lac promoter that combines the strong T7 transcription promoter and the lac operator. The vector also carries the natural promoter and coding sequence for the lac repressor (lacI), enabling the control of expression through a constitutive repression of transcription of both the T7pol gene from the host chromosome and of the target gene by any T7pol that is made, where repression is removed upon the addition of IPTG.

Inoculation and induction of cultures for expression was carried out as described in Section 2.17

4.4.1 Amidase expression

A high level of amidase expression was achieved under standard conditions (Section 2.17), with the specific activity on acetamide reaching a maximum of 37 U/mg ± 2 approximately 3.5 hours after induction. A typical expression profile is shown in Fig. 4.5. The activity level observed compares very favorably with that of the native organism, with a specific activity 40 times higher. Using previous estimate (Pereira, 1998) of 95% purity following a four-step purification of the native enzyme and specific activity of 117 U/mg, it was estimated that the recombinant enzyme comprised approximately 30% of the total soluble protein.
Fig. 4.5: Expression profile of recombinant amidase

![Expression profile graph]

Fig. 4.6: SDS-PAGE analysis of the soluble protein fractions (~10μg) of amidase (arrowed) expressed in *E. coli* (BL21) λDE3 pLysS. Lanes: 1, molecular weight marker; 2, BL21 only; 3, pET21a un-induced; 4, pNH223 un-induced; 5, pET21a induced; 6, pNH223 – 1 hour; 7 pNH223 – 2 hours; 8 pNH223 – 3 hours; 9 pNH223 – 4 hours.
4.4.2 Nitrile hydratase expression

In preliminary studies of NHase expression from constructs pNH461 (data not shown) and pNH512 (Fig. 4.4), very low NHase activities were observed. Plasmid stability experiments revealed that cultures of both strains were prone to overgrowth with bacteria that had lost the expression plasmid. The problem was overcome by using carbenicillin as the selective antibiotic at a concentration of 100µg/ml in place of ampicillin; in addition, the growth medium of the seed cultures was replaced twice prior to inoculation of larger cultures for expression in order to remove secreted β-lactamase.

To investigate the effect of the addition of cobalt ions to the growth medium, a series of expression experiments were performed with varying concentrations of CoCl₂. However initial experiments revealed that the addition of cobalt inhibited the growth of the E. coli host, even at concentrations as low as 0.1 mM (data not shown). For this reason cobalt was not added until 15-20 minutes prior to induction.

Expression from pNH512 in the absence of Co²⁺ resulted in the production of a large quantity of soluble protein (Fig. 4.8) but very low activity. Very similar activities were observed when cobalt was added to cultures at concentrations of 0.1 and 0.5 mM (within the range used by previous workers), (Table 4.1). A maximum level of specific activity (on acetonitrile) of 49 U/mg ± 4 was achieved by supplementation with 0.1 mM CoCl₂ approximately 3 hours after induction. The recombinant protein was estimated to account for at least 30% of the total cell protein by SDS-PAGE analysis (Fig. 4.8). Calculations based on specific activity determined for the native pure protein were not feasible as the reported value was significantly lower than observed for crude extracts of recombinant protein.

Attempts to express the NHase genes from cells containing pNH461 (containing the NHase α and β subunit genes and the attenuated P14K gene) resulted in very low levels of activity, reaching a maximum of ~5 U/mg 2 hours after induction. The addition of cobalt to the growth medium made no effect on the level of activity expressed. When crude extracts were analysed by SDS-PAGE, it was not possible to detect bands corresponding to either the NHase subunits or P14K protein.
Table 4.1: Effect of cobalt addition to cell cultures on NHase activity. Cell extracts of cultures containing pNH512 were prepared 3 hours after induction and assayed for NHase activity on acetonitrile for 10 minutes at 50°C.

<table>
<thead>
<tr>
<th>Co^{2+} added (mM)</th>
<th>0</th>
<th>0.1</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHase activity (U/mg)</td>
<td>5 ± 2</td>
<td>49 ± 4</td>
<td>44 ± 4</td>
</tr>
</tbody>
</table>

Fig. 4.7: Expression profile of recombinant NHase.
Fig. 4.8: SDS-PAGE analysis of the soluble protein fractions (~10μg) of NHase expressed in *E. coli* (BL21) ΔDE3 pLysS. Lanes: 1 & 14, molecular weight marker; 2, BL21 only; 3, pET21a un-induced; 4, pNH512 un-induced; 5, pET21a induced; 6-9, pNH512, no Co²⁺, 1-4 hours; 10-13, pNH512 + Co²⁺, 1-4 hours. Alpha and beta subunits are indicated by arrows, by convention the smaller subunit is referred to as the α subunit and the larger the β subunit.
4.5 Purification

Many vectors currently in use for heterologous expression in E. coli (including the pET system) enable rapid one-step purification through the addition of an N or C-terminal tag sequence. However as the expression and activity of the recombinant NHase were to be characterised for the first time and it was not known how a tag might affect the physical and functional properties of the enzyme, more traditional purification techniques were used. An advantage of using mesophilic hosts for the expression of proteins from thermophiles is that a rapid and cost effective heat denaturation step can be added early in the purification process to precipitate large proportions of the host protein.

4.5.1 Amidase Purification

As a detailed characterisation of this enzyme was not a primary objective of this work (the amidase was required only for use in the coupled ammonia detection assay), the degree of purity was not of great importance. Nevertheless, a heat treatment step was included in order to minimise any inhibitory factors present in the crude extracts. Heat denaturation trials were conducted at temperatures ranging from 60°C to 80 °C (Fig. 4.9). It was found that samples incubated at 80°C lost activity very rapidly (data not shown) whilst incubations of up to 45 minutes at 75°C resulted in loss of very little activity and 2.8-fold increase in specific activity. Interestingly, an increase in total activity was observed with incubations below 80°C. A similar phenomenon was observed during the purification of the amidase and nitrile hydratase of Bacillus pallidus DAC521 (Cramp, 1997) and is attributed to the removal of uncharacterised inhibitory elements during the purification. Subsequent purifications therefore included a heating step, carried out at 75°C for a maximum of 30 minutes. Estimates from SDS-PAGE and specific activities indicated a purity of ≥70%. A summary of the purification is detailed in Table 4.2.
Fig. 4.9: Heat purification of RAPc8 amidase represented by: A, increase in specific activity; B, % initial activity.

A

B
Table 4.2: Heat purification of RAPc8 Amidase

<table>
<thead>
<tr>
<th>Purification Step</th>
<th>Volume (ml)</th>
<th>Total Protein (mg)</th>
<th>Total Activity (U)</th>
<th>Specific Activity (U/mg)</th>
<th>Yield (%)</th>
<th>Purification (fold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude extract</td>
<td>25</td>
<td>219</td>
<td>6875</td>
<td>31.4</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Heat-treatment</td>
<td>25</td>
<td>96</td>
<td>8456</td>
<td>87.5</td>
<td>123</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Fig. 4.10: SDS-PAGE showing amidase purification. Lanes: 1, molecular weight marker; 2, BL21 only; 3 pNH223 un-induced; 4, amidase crude extract; 5, heat treated amidase.
4.5.2 NHase purification

4.5.2.1 Heat precipitation

Heat precipitation of host proteins was also used as the first stage of NHase purification. However, due to lower molecular stability of this enzyme, it was necessary to operate at lower temperatures. As is shown in Fig. 4.11, the optimum temperature for heat treatment of the NHase was 55°C; a 45-minute incubation enabled a nearly two-fold purification with no loss of activity. As with the amidase purification, the heat precipitation appeared to 'activate' the NHase, resulting in an increased total activity.

4.5.2.2 Ion exchange chromatography

Although the basic functional characterisations of NHase did not require purification to homogeneity, the level of purity achieved by heat treatment alone was not adequate. Purification studies of the NHases of *Bacillus* sp. RAPc8 (Pereira, 1998) and *B. pallidus* DAC521 (Cramp, 1997) both used an anion exchange chromatography stage as the initial step, as it was found that this method achieved the highest purification factor. For this reason, a similar approach was adopted for purification of the recombinant NHase, with the knowledge from previous linear salt elution gradients that NHase eluted at approximately 0.2M NaCl.

A pre-packed 1ml Mono Q HR 5/5 column was used to purify the NHase. 40mg of the heat-treated extract was concentrated to 2ml in an Amicon Centricprep 10 concentrator (10kDa cut-off) and applied to the column (preequilibrated with 25mM potassium phosphate buffer, pH 7.2) at a flow rate of ~0.5ml/min (used for all steps). The column was washed with 50ml starting buffer before eluting with a linear salt gradient (0M NaCl to 0.5M NaCl in 25mM potassium phosphate buffer, pH 7.2). The eluate was collected in 2ml fractions and assayed for activity using the ammonia assay (Section 2.7.8). Four highly active fractions 8-11 were collected (Fig. 4.12). Fraction 9 displayed the highest activity at 244U/mg, equivalent to a 2.6 fold increase in activity from heat-treated extracts (5 fold increase from crude extracts). The overall yield of the step was 42% and was not optimised, since the objective was essentially to obtain enough protein for basic functional comparisons with the native enzyme.
Fig. 4.11: Heat purification of RAPc8 NHase represented by: **A**, increase in specific activity; **B**, % initial activity.
Fig. 4.12: SDS-PAGE showing fractions from NHase ion exchange chromatography. Lanes: 1 and 10, molecular weight marker; 2-9, 15 μl of fractions 6 to 13.

Table 4.3: Purification of RAPc8 NHase. Values in parentheses relate to the activity of the initial crude extract.

<table>
<thead>
<tr>
<th>Purification Step</th>
<th>Vol. (ml)</th>
<th>Total Protein (mg)</th>
<th>Total Activity (U)</th>
<th>Specific Activity (U/mg)</th>
<th>Yield (%)</th>
<th>Purification (fold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude extract</td>
<td>25</td>
<td>281</td>
<td>13770</td>
<td>49</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Heat-treatment</td>
<td>25</td>
<td>171</td>
<td>16070</td>
<td>94</td>
<td>117</td>
<td>1.9</td>
</tr>
<tr>
<td>Treated extract</td>
<td>2</td>
<td>40</td>
<td>3760</td>
<td>94</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Mono Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction 8</td>
<td>2</td>
<td>5.3</td>
<td>583</td>
<td>110</td>
<td>16 (18)</td>
<td>1.2 (2.3)</td>
</tr>
<tr>
<td>Fraction 9</td>
<td>2</td>
<td>0.9</td>
<td>220</td>
<td>244</td>
<td>6 (7)</td>
<td>2.6 (5)</td>
</tr>
<tr>
<td>Fraction 10</td>
<td>2</td>
<td>3.6</td>
<td>432</td>
<td>120</td>
<td>12 (13)</td>
<td>1.3 (2.5)</td>
</tr>
<tr>
<td>Fraction 11</td>
<td>2</td>
<td>3.1</td>
<td>341</td>
<td>110</td>
<td>9 (11)</td>
<td>1.2 (2.3)</td>
</tr>
<tr>
<td>Fractions 8-11 combined</td>
<td>8</td>
<td>12.9</td>
<td>1576</td>
<td>122</td>
<td>42 (49)</td>
<td>1.3 (2.5)</td>
</tr>
</tbody>
</table>
4.6 Characterisation of recombinant Nitrile hydratase

4.6.1 N-terminal sequence

Aliquots of the purest NHase fraction were electrophoresed (SDS-PAGE), transferred to PVDF membrane (Section 2.7.6) and stained with Coomassie Blue. The N-terminal sequences were determined by Mike Weldon (PNAC, University of Cambridge). The results were in complete agreement with the gene sequences except that the N-terminus methionine was not observed for the α subunit, probably as a result of post-translational processing (conversely, sequencing of the NHase subunits of BR449 did not detect the N-terminal methionine of the β subunit (Kim and Oriel, 2000)).

Fig. 4.13: Comparison of the deduced NHase α and β subunit N-terminal sequence with those of the recombinant protein and of the native organisms Bacillus sp. RAPc8 and B. pallidus DAC521 (Cramp, 1997) as determined by Edman degradation.

<table>
<thead>
<tr>
<th>Translated α gene</th>
<th>Recombinant NHase α</th>
<th>RAPc8 NHase α</th>
<th>B. pallidus NHase α</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNKIHDVGGMDGFGKVMYVK</td>
<td>MNGIH</td>
<td>MKKMMDANEISSFIQNSKKT</td>
<td>--------------KNTIDPRF</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Translated β gene</th>
<th>Recombinant NHase β</th>
<th>RAPc8 NHase β</th>
<th>B. pallidus NHase β</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNGIH</td>
<td>MNGIHDV</td>
<td>MNGIHDVGGMDGFGKVMYVK</td>
<td>MNGIXDVGGM</td>
</tr>
</tbody>
</table>
4.6.2 Nitrile hydratase thermostability

Comparisons of NHase stability both as the pure enzyme and as a crude extract have indicated that the enzyme is significantly stabilised by intracellular components (Cramp and Cowan, 1999). Investigations of the thermostability of the recombinant NHase were therefore carried out using the purest fraction (number 9) from the ion-exchange column.

The results are presented in two formats, the first showing the decrease in activity over time of samples incubated at temperatures ranging from 20°C to 60°C (Fig. 4.14). The half-life of the enzyme at each temperature was calculated from first order natural logarithm plots of residual activity versus time. These plots demonstrated that whereas at lower temperatures the stability of the recombinant enzyme was analogous to that of the native enzyme, the former was nearly 4-fold more stable at 60°C (Table 4.4).

For comparison with the thermostability data of the NHases of *B. smithii* (Takashima *et al.*, 1998) and *Ps. thermophila* (Yamaki *et al.*, 1997), the effect of temperature was plotted as % residual activity against temperature for samples incubated for 30 minutes at temperatures between 30°C and 70°C (Fig. 4.15). From the plot, it is seen that the enzyme is fully stable for 30 minutes at 55°C and moderately stable up to approximately 60°C. Estimates from this graph infer that the NHase has a half-life of ~1.8 hours at 60°C – nearly double the estimate from the first set of data. It must be stressed that this is a crude method of half-life prediction as it effectively relies on only one data point, and as the half-life predictions of other NHases given in Table 4.4 are made in this manner, they should be interpreted with caution.
Fig. 4.14: Thermostability profile for the purified NHase as defined by loss of activity over time at various temperatures.

Fig. 4.15: Thermostability of recombinant NHase as defined by residual activity after a 30-minute incubation at specific temperatures.
Table 4.4: Comparison of the thermostability of recombinant and native NHase of *Bacillus* sp. RAPc8 with other bacterial nitrile hydratases. (After Pereira, 1998).

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>Native RAPc8 NHase</th>
<th>Recombinant NHase</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>16 minutes (+/- 4)</td>
<td>1 hour</td>
</tr>
<tr>
<td>50</td>
<td>2.5 hours</td>
<td>4.6 hours</td>
</tr>
<tr>
<td>37</td>
<td>45 hours</td>
<td>41 hours</td>
</tr>
<tr>
<td>20</td>
<td>80 hours</td>
<td>89 hours</td>
</tr>
<tr>
<td>4</td>
<td>&gt;1 month</td>
<td>&gt;2 months</td>
</tr>
</tbody>
</table>

Organism | Half-life of Activity$^1$
---|------------------
*Corynebacterium pseudodiphtericum* | 65 min at 20°C (Li *et al.*, 1992)
*Pseudomonas chlororaphis* B23 | 11 min at 30°C (Nagasawa *et al.*, 1987)
*Pseudomonas putida* | 27 min at 50°C (Payne *et al.*, 1997)
*Rhodococcus rhodochrous* J1$^2$ | 58 min at 60°C (Nagasawa *et al.*, 1991)
*B. pallidus* | 51 min at 50°C (Cramp, 1997)
| 7.3 min at 60°C (Cramp, 1997)
*B. smithii* | 1.5 hrs at 55°C (Takashima *et al.*, 1998)
*Pseudonocardia thermophila*$^3$ | 13 h at 60°C (Yamaki *et al.*, 1997)

$^1$ Recalculated from author’s data assuming logarithmic activity loss.
$^2$ Stabilised by presence of 44mM n-butyrte acid
$^3$ Stabilised by addition of 34mM n-butyric acid
4.6.3 Nitrile hydratase specificity
The recombinant enzyme was compared with the native RAPc8 NHase in terms of activity towards various nitriles. Using the purest fraction from the ion-exchange chromatography, relative activity was determined over a hydrolysis period of 30 minutes using 50mM nitrile substrate under standard conditions; residual substrate was subsequently determined by GC analysis. It should be noted that as this assay method relies on substrate utilisation rather than product generation, it is inherently insensitive, with a maximum sensitivity of ±5%. Activity values must therefore be interpreted with caution, particularly the data concerning the less readily degraded nitriles.

The recombinant NHase accepted a broad range of nitriles as substrates (Table 4.5) and the specificity observed was effectively very similar to that reported for the native enzyme (Pereira, 1998). However it is notable that whilst similar activity to the native enzyme was observed with methacrylonitrile, no activity could be detected on the larger branched aliphatic nitriles. The preferred substrate was acetonitrile, followed by valeronitrile; the reverse was observed with the native enzyme.

The NHase readily degraded dinitriles, although the short hydrocarbon chain substrate malononitrile was inactive as a substrate. It is possible that the second cyano group of short chain substrates impinges on activity through steric hindrance (Pereira, 1998). Specificity was almost totally restricted to aliphatic nitriles, of the aromatic nitriles tested; only 3-cyanopyridine was accepted as a substrate. There appears to be no clear correlation between activity and carbon chain length. However, it is important to note that only relatively short hydrocarbon chained substrates were tested. It has been shown that the NHase of *Arthrobacter* J-1, which shares a similar substrate specificity profile (data not shown), nitriles of greater than five carbon-atom chain length are inactive as substrates (Asano et al., 1982).
**Table 4.5:** Comparison of the substrate specificities of the both the recombinant RAPc8 NHase and the native enzyme with other bacterial NHases.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Structure</th>
<th>Relative activity(^1) (%) ±5 % for:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Recombinant RAPc8 NHase</td>
</tr>
<tr>
<td>Acetonitrile</td>
<td>CH(_3)CN</td>
<td>100</td>
</tr>
<tr>
<td>Chloroacetonitrile</td>
<td>CICH(_2)CN</td>
<td>34</td>
</tr>
<tr>
<td>Acrylonitrile</td>
<td>CH=CHCN</td>
<td>67</td>
</tr>
<tr>
<td>Propionitrile</td>
<td>CH(CH(_3))CN</td>
<td>26</td>
</tr>
<tr>
<td>Methacrylonitrile</td>
<td>CN=CHCN</td>
<td>60</td>
</tr>
<tr>
<td>Butyronitrile</td>
<td>CH(_2)CHCN</td>
<td>46</td>
</tr>
<tr>
<td>Isobutyronitrile</td>
<td>CH(_2)CHCH=CHCN</td>
<td>0</td>
</tr>
<tr>
<td>Valeronitrile</td>
<td>CH(_2)CNCN</td>
<td>89</td>
</tr>
<tr>
<td>Isovaleronitrile</td>
<td>CH(_2)CNCN</td>
<td>0</td>
</tr>
<tr>
<td>Cis, trans-crotonitrile</td>
<td>CH(_2)=CHCN</td>
<td>79</td>
</tr>
<tr>
<td>Benzylicyanide</td>
<td>C(_6)H(_5)=C(_2)H(_2)CN</td>
<td>0</td>
</tr>
<tr>
<td>Benzonitrile</td>
<td>C(_6)H(_5)CN</td>
<td>0</td>
</tr>
<tr>
<td>3-cyanopyridine</td>
<td>NC-CN</td>
<td>53</td>
</tr>
<tr>
<td>Malononitrile</td>
<td>NC-CN</td>
<td>0</td>
</tr>
<tr>
<td>Glutaronitrile</td>
<td>NC-CN</td>
<td>70</td>
</tr>
<tr>
<td>Adiponitrile</td>
<td>NC-CN</td>
<td>20</td>
</tr>
</tbody>
</table>


\(^2\) Specificity of *B. pallidus* DAC521 NHase expressed as % substrate (50mM initial concentration) utilised over a 30 minute hydrolysis period.
4.6.4 Nitrile hydratase reaction kinetics

For comparison with the native enzyme, reaction kinetics were determined (Section 2.21.3) for two of the most readily degraded substrates: acetonitrile and acrylonitrile. Kinetic constants (Table 4.6) were estimated as averages of values calculated from Hanes, Eadie Hofstee and Lineweaver Burk plots (Figs. 4.16 and 4.17).

The recombinant enzyme displayed a similar affinity for both acetonitrile and acrylonitrile with $K_m$ values of 11.1 mM and 11.9 mM respectively, compared with the values of 8 mM and 11 mM for the native enzyme (Pereira, 1998). The reaction rates were also of a similar order with $k_{cat}$ values of 930 s$^{-1}$ and 369 s$^{-1}$ for acetonitrile and acrylonitrile respectively; cf. 746 s$^{-1}$ and 484 s$^{-1}$ for the native enzyme.

Table 4.6: Comparison of the kinetic data of the native and recombinant RAPc8 NHase on acetonitrile and acrylonitrile substrates.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_m$ (mM)</th>
<th>$V_{max}$ (U/mg)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetonitrile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recombinant NHase</td>
<td>11.1</td>
<td>546</td>
<td>930</td>
</tr>
<tr>
<td>Native NHase</td>
<td>8.8</td>
<td>407</td>
<td>746</td>
</tr>
<tr>
<td>Acrylonitrile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recombinant NHase</td>
<td>11.9</td>
<td>217</td>
<td>369</td>
</tr>
<tr>
<td>Native NHase</td>
<td>11</td>
<td>264</td>
<td>484</td>
</tr>
</tbody>
</table>

$^1$Enzyme concentration in assay = 1.1 μg/ml

$^2$k$_{cat}$ for recombinant enzyme calculated assuming a molecular mass of 102.2 kDa ($\alpha_2\beta_2$ heterotetramer where $\alpha = 24.6$ kDa and $\beta = 25.4$ kDa) and a single active site per heterotetramer.

$k_{cat}$ for native enzyme calculated assuming a molecular mass of 110 kDa as estimated from gel filtration chromatography and a single active site per heterotetramer.
Fig. 4.16: Kinetics of acetonitrile hydrolysis represented by: A Lineweaver-Burke plot; B Hanes plot and C Eadie-Hofstee plot.

A

B

C
Fig. 4.17: Kinetics of acrylonitrile hydrolysis represented by: A Lineweaver-Burke plot; B Hanes plot and C Eadie-Hofstee plot.

A

\[ V = \frac{K_m}{S} + \frac{V_{\text{max}}}{K_m} \] \[ y = 0.0533x + 0.0047 \] \[ R^2 = 0.991 \]

B

\[ V = \frac{K_m}{S} + \frac{V_{\text{max}}}{K_m} \] \[ y = 0.0044x + 0.0572 \] \[ R^2 = 0.9616 \]

C

\[ V = \frac{K_m}{S} + \frac{V_{\text{max}}}{K_m} \] \[ y = -0.0867x + 18.319 \] \[ R^2 = 0.9211 \]
4.6.5 Benzonitrile inhibition

It has been demonstrated in a previous study that benzonitrile acts as a potent inhibitor ($K_i \sim 1.3 \text{mM}$) of acetonitrile hydrolysis by *B. pallidus* DAC521 NHase (Cramp, 1997). Assuming a similar reaction mechanism, the effect of benzonitrile on the activity of RAPc8 NHase was investigated. For continuity with the previous investigation, the $K_i$ for benzonitrile was taken as the concentration required to reduce the reaction rate by 50% (assuming that at 70mM acetonitrile, the initial rate was equal to $V_{\text{max}}$), rather than the classical approach of estimation from Dixon plots.

Using this approach, a benzonitrile $K_i$ of 2.25mM was determined for the recombinant enzyme (Fig. 4.18).

**Fig. 4.18:** Benzonitrile inhibition of acetonitrile hydrolysis by semi-pure recombinant RAPc8 NHase.
4.7 Discussion

4.7.1 Expression studies

Both the NHase and amidase enzymes were actively expressed without the need for co-expression of additional genes, in keeping with previous findings (Pereira, 1998) that the expression of nitrile degrading activity by the native *Bacillus* sp. RAPc8 was constitutive. The requirements for expression were similar to those reported for the expression of the *Bacillus* sp. BR449 genes (Kim and Oriel, 2000). Minor differences between the two studies may be attributed to the different host-vector systems used. The single major disparity concerns the expression of the downstream gene. Kim and Oriel (2000) observed that no protein corresponding to the P12K gene was detected by SDS-PAGE analysis and that NHase expression was not affected by the presence of the gene in expression constructs. It may be of significance that the entire, shorter P12K gene of BR449 was contained within the relevant expression constructs whereas pNH461 contained a truncated P14K gene. A possible explanation is that the truncated gene decreased transcript production and hence protein synthesis. It is also feasible however that the use of carbenicillin did not cure the plasmid instability problem as it did with pNH512 (Section 4.4.2). Clearly this work should be repeated with the complete gene sequence to elucidate the function, if any, of P14K. However this was not done due to time constraints.

4.7.2 Cobalt requirements

The results obtained in this work were consistent with previous findings that presence of cobalt is required for the expression of NHase activity rather than protein production (Kim and Oriel, 2000, Komeda et al., 1996b, Mizunashi et al., 1998). Whilst direct comparison cannot be made with the results of Kim and Oriel (2000) due to their use of acrylonitrile as the standard substrate as, (c.f. acetonitrile in this study), it is of interest to note that a five-fold increase in added cobalt concentration from 0.1mM to 0.5mM resulted in an increase in specific activity of the recombinant BR449 NHase from 23U/mg to 43U/mg, whereas the same increase in cobalt concentration apparently had no apparent effect on the activity of the RAPc8 enzyme. Furthermore, the addition of cobalt to cell free extracts had no effect on the activity of the RAPc8 enzyme (data not shown), whereas it significantly enhanced the activity of BR449 NHase. As results from studies of both enzymes have indicated that an excess of inactive enzyme was
expressed (Kim et al., 2001), the disparity is likely to be due to a difference in the cobalt uptake capacity of the host strains rather than a difference in expression levels.

4.7.3 Purification
A heat purification step proved to be a very convenient and efficient strategy for the purification of both recombinant enzymes, in each case resulting in an increase in the yield of activity – most likely attributed to removal of inhibitory factors within the crude extracts. The high level of purity achieved by this single step for the amidase will facilitate a simple purification protocol in further studies.

The further purification of NHase by anion exchange resulted in very low yields but also in a low purification factor (Table 4.3), whilst the level of purity achieved was adequate for the comparisons described here, it would be desirable to purify the enzyme to homogeneity in subsequent investigations. It is noteworthy that Phenyl-Sepharose hydrophobic interaction chromatography (HIC) has been successfully employed as the primary or secondary stage in the purification of three thermophilic Bacillus NHases (Cramp and Cowan, 1999, Kim et al., 2001, Takashima et al., 1998). The NHase of Bacillus sp. BR449 was purified to homogeneity in a single chromatographic step following a 30-minute heat-treatment (Kim et al., 2001). Whilst Pereira (1998) noted that attempts to purify the native enzyme by HIC were very inefficient, it is significant that a sodium sulphate salt solution was used as an eluent rather than ammonium sulphate and that the purification was of both amidase and NHase.

From SDS-PAGE gels of the partially purified protein (Fig. 4.12), it was estimated that the fraction with the highest specific activity (fraction 9), was approximately 75% pure. This would translate to an activity of ~325U/mg for homogeneous protein, indicating that the enzyme was expressed to only ~15% of the total cell protein. Electrophoresis of crude extracts (Fig. 4.8), indicated that expression levels were significantly higher. It might be concluded therefore that some protein was being expressed in an inactive form. It was observed that the heat-treatment increased the activity of the samples up to 120% however this does not fully explain the observations. Kim et al. (2001) reported that the incubation of crude extracts of BR449 NHase at 50°C in 50μM CoCl₂ increased activity over 50-fold and concluded that expressed NHase was inactive due to poor cobalt uptake of the host cells and that the subsequent incubation with cobalt ions
‘activated’ the inactive enzyme. Preliminary, qualitative experiments (data not shown) suggested that a similar activation was possible with the recombinant RAPc8 NHase, which would explain the inconsistency between the activity and the quantity of expressed protein.

The specific activity recorded for the semi-pure NHase in this work was up to 24 fold higher than that recorded for the pure native protein. As the same assay method was used in both cases, this is unlikely to be due to experimental deviation. It is probable that the difference is due to the absence of additional cobalt in the media used by Pereira (1998a). Similar results have been observed with the production of native NHase by Bacillus sp. BR449 and Bacillus smithii SC-J05-1, where 22-fold and 160-fold activity increases were detected respectively when cobalt ions were added to the media (Padmakumar and Oriel, 1999, Takashima et al., 1996).

4.7.4 Characterisation
4.7.4.1 Thermostability

The results of the thermostability investigations suggested that the recombinant NHase had a significantly higher stability at elevated temperatures than the native enzyme. This difference is unlikely to be due to stabilisation by contaminating proteins, since previous experiments (Pereira, 1998) were also performed with semi-pure enzyme. While differences in enzyme concentration may account for an apparent increase in stability, this factor is unlikely to be responsible for such a significant difference. Furthermore, as no details were available concerning the NHase concentration used in the previous study (Pereira, 1998), direct comparison with the results reported in this work is not possible.

As would be expected of an enzyme from a thermophile, RAPc8 NHase is considerably more thermostable than most of its mesophilic counterparts. The predicted half-life of R. rhodochrous J1 NHase is in the same order as that of RAPc8. However, the predicted half-lives of the mesophilic enzymes are only approximations, and the stability of the J1 enzyme was assessed in the presence of the stabilising agent n-butyric acid (Nagasawa et al., 1991). The recombinant RAPc8 enzyme was apparently more stable at elevated temperatures than the enzymes of Bacillus pallidus DAC521 and Bacillus smithii SC-J05-1 (80% activity after 30 min incubation at 55°C) (Cramp and
Cowan, 1999, Takashima et al., 1998) and approximately equivalent to the NHase of *Pseudonocardia thermophila* - although this enzyme was also stabilised by n-butyric acid (Yamaki et al., 1997). No data is yet available on the stability of the purified BR449 NHase. The crude extract was shown to be stable for prolonged periods of incubation at 50°C (Kim and Oriel, 2000).

### 4.7.4.2 Substrate specificity

The specificity experiments confirmed previous observations (Pereira, 1998) that the *Bacillus* sp. RAPc8 NHase catalyses the hydrolysis of a broader range of nitriles than many NHases; linear, branched, saturated and unsaturated aliphatic nitriles as well as the hetero-aromatic nitrile 3-cyanopyridine were all substrates for the enzyme (Table 4.5). There is generally a high similarity of the specificity profiles of the native enzyme and recombinant enzyme (Pereira, 1998). However, significant differences were observed between the activities of the two enzymes on the branched nitriles isobutyronitrile and isovaleronitrile. As both enzymes hydrolysed methacrylonitrile whereas the recombinant enzyme did not hydrolyse isobutyronitrile (which is structurally almost identical), it is possible that this lack of observed activity was due to experimental error or to the inaccuracy of the method used (Section 4.6.3). However, as the experiments were performed in triplicate, the data recorded should be reliable. Similar results have been observed in a number of studies on the specificities of several NHases (Table 4.5) and have shown that most have little or no activity on branched nitriles (Langdahl et al., 1996, Nagasawa et al., 1987, Nagasawa et al., 1991, Wieser et al., 1998). This may be due to steric hindrance caused by the branched carbon atom. However, it is noted that the methyl side chain proximal to the C≡N moiety of methacrylonitrile clearly does not impinge significantly on catalytic activity.

The crotononitrile cis and trans isomers were catalysed at different rates. However, the difference in rates was sufficiently small that it would probably not be possible to exploit the enzyme for the synthesis of pure amide isomers. Further work is required in order to deduce whether the enzyme displays such limited selectivity towards other isomeric nitriles. Currently, there is no data of the activity of RAPc8 NHase on enantiomeric nitriles; only the NHases of *A. tumafaciens* (Bauer et al., 1994) and *P. putida* 5B (Wu et al, 1997) have so far been reported to selectively hydrate enantiomeric nitriles.
In agreement with previous findings (Pereira, 1998), the recombinant enzyme displays no activity towards the aromatic nitriles, benzonitrile and benzyl cyanide. However, a relatively high activity on 3-cyanopyridine was observed (no information is available on the activity of the native enzyme on this substrate). It should be stressed that pyridines are heteroaromatic compounds and, as a result of the lone pair of electrons of the nitrogen atom within the aromatic ring (Fig. 4.19), have different physical and chemical properties to homoaromatic compounds. The relative activity of the recombinant NHase on 3-cyanopyridine is analogous to that of the Co-NHase of \textit{R. rhodochrous} J1 L-NHase (Wieser \textit{et al}, 1998) and that of the Fe-NHase of \textit{R. erythropolis} JCM6823 (Duran \textit{et al}, 1993). These findings seem to discredit the claim that Co-NHases are aromatic-specific whilst Fe-NHases are aliphatic-specific (Kobayashi and Shimizu, 1998, Miyanaga \textit{et al}, 2001).

\textbf{Fig. 4.19:} Diagram of the arrangement of the electron orbitals of pyridine.

\begin{center}
\includegraphics[width=0.5\textwidth]{pyridine_orbitals.png}
\end{center}

\textbf{4.7.4.3 Reaction Kinetics}

The substrate affinities and reaction rates of the recombinant enzyme for both acetonitrile and acrylonitrile were very similar to those displayed by the native enzyme (Pereira, 1998). Furthermore, the ratio of constants between acetonitrile and acrylonitrile remained approximately the same in each study. There are variations between the catalytic constants determined in this work and that of Pereira, but these are most likely due to the accuracy of the assay method and the difference in the purity of
the enzyme solutions used. The data recorded in this section therefore provide strong
evidence that, despite the discrepancies between the α-subunit N-terminal sequences
and the substrate specificities, the recombinant enzyme is essentially identical to the
previously described native enzyme (Pereira, 1998, Pereira et al., 1998).

The affinity for acrylonitrile is within the range reported for mesophilic enzymes:
*R. rhodochrous* J1 H-NHase - 1.89 mM (Nagasawa et al., 1991), *R. rhodochrous* J1
L-NHase - 1.89 mM (Wieser et al., 1998), *Rhodococcus* sp. R312 - 16.7 mM
(Nagasawa et al., 1986) and *P. chlororaphis* B23 - 34.6 mM (Nagasawa et al., 1987).
However, comparisons of kcat values: *R. rhodochrous* J1 H-NHase- 14660 s⁻¹ (high kcat
value due to 10 active sites/enzyme molecule [α₁₀β₀] compared with two/enzyme
molecule for the other enzymes [α₂β₂]), *R. rhodochrous* J1 L-NHase - 1518 s⁻¹, *P.
chlororaphis* B23 - 2480 s⁻¹ and *Rhodococcus* sp. R312 - 2080 s⁻¹, highlights a
substantially lower reaction rate. This is consistent with the generally accepted view
that thermal stabilisation of proteins leads to decreases in protein chain flexibility and
hence catalytic activity (Sterner and Liebl, 2001). However, stabilisation may not affect
the architecture of the active site where its conformation is crucial for substrate
recognition, K_M is not altered greatly.

It must be stressed that the enzyme solutions used for these assays were only semi-pure
and whilst this is unlikely to have a significant impact on recorded K_M values, a
significant rise in turnover number using the fully purified enzyme would be expected.
Furthermore, depending on the substrate binding properties of the inactive Co-
apoenzyme, Co activation (Section 4.7.2) may also give rise to an increase in reaction
rate.

4.7.4.4 Benzonitrile inhibition

The inhibition studies demonstrated that, as with that of *B. pallidus* DAC521 (Cramp,
1997), the NHase of *Bacillus* sp. RAPc8 was significantly inhibited by benzonitrile. In
fact, the enzyme appeared to have a higher affinity for benzonitrile than either of the
active substrates acetonitrile or acrylonitrile (K_i = 2.25mM cf. 11.1mM and 11.9mM
respectively). Characterisation of the inhibition mechanism of the *B. pallidus* DAC521
enzyme indicated that the inhibition was irreversible; a similar phenomenon might be
expected for RAPc8 NHase. Inhibition of NHase by nitriles and substrate analogues
has been observed previously with several enzymes; for example, *Rhodococcus* sp. R312 had a very high affinity for isobutyronitrile, acting as a competitive inhibitor for propionitrile: $K_i = 5.4 \mu M$ (Nagasawa *et al.*, 1986), while KCN acted as a non-competitive inhibitor for the L-NHase and a potent competitive inhibitor for the H-NHase of *Rhodochrous* J1 ($K_i = 0.2 \text{mM}$ and $0.75 \mu M$ respectively), (Wieser *et al.*, 1998).

The inhibition data, together with the observation of NHase activity on 3-cyanopyridine, suggests that the physico-chemical properties of the homo-aromatic nitriles result in an incorrect non-functional binding at the active site. It may be the case that this is due to aromatic interaction with the aromatic residues that line the entrance to the active site (Nagashima *et al.*, 1998). The observation (Cramp, 1997) that the inhibition is irreversible would support this hypothesis. If this is the case, then the activity on 3-cyanopyridine may be explained by interference with the non-functional interaction by the sp$^2$ orbital (Fig. 4.19) of the heteroaromatic nitrogen atom.
Chapter 5: Enzyme engineering

5.1 Introduction
As outlined in the previous chapter, the NHase of *Bacillus* sp. RAPc8 shows no activity on aromatic nitriles such as benzonitrile or benzyl cyanide. It was, however, demonstrated that benzonitrile acts as a potent inhibitor of the enzyme. This chapter describes an experimental program aimed at altering the substrate specificity of RAPc8 NHase to include aromatic nitriles. Nakasako *et al.* (1999) proposed that the lack of aromatic specificity of *Rhodococcus* sp. N771 NHase was due to the narrow entrance to the active site from the bulk solvent (Section 1.3.6). However, the activity of RAPc8 NHase on 3-cyanopyridine indicates that for this enzyme at least, this is not the case.

The working hypothesis that shaped this investigation stemmed from the observed benzonitrile inhibition: it was proposed that interaction between the π-clouds of the aromatic rings of substrate analogues and aromatic protein residues prevented correct positioning within the active site. The hypothesis was supported by the observation that 3-cyanopyridine, a heteroaromatic compound, is a substrate for the enzyme (Section 4.7.3.2). The approach taken therefore was to identify aromatic amino acids proximal to the active site and replace them by site-specific mutagenesis, with small aliphatic amino acids (either glycine or alanine).

5.2 Homology modelling
In order to determine which aromatic residues, if any were responsible for the putative interaction with aromatic substrate analogues, it was necessary to identify residues in the proximity of the active site. At the time of this investigation, only two known NHase structures existed: that of *Rhodococcus* sp. R312 – resolved to 2.65Å (Huang *et al.*, 1997) and that of the NO inactivated *Rhodococcus* sp. N-771 NHase (which has an identical amino acid sequence), resolved to 1.7Å (Nagashima *et al.*, 1998). The residues lining the active sites of these enzymes have been determined and include aromatic residues αW118Ω (αW117 in *Rhodococcus* sp. N-771), βY37, βY67, βY72 and βY76,
The Greek letter notation refers to the specific subunit four of which are conserved in the RAPc8 NHase. Furthermore, Nakasako et al. (1999) identified a cluster of 10 aromatic residues (αW43, αF47, αF51, αW117, βF28, βW32, βY37, βF41, βY72, and βY73), which are believed to stabilise the channel from the bulk solvent to the active site (see Fig. 5.1). Of these, eight are conserved in the RAPc8 NHase and five are strictly conserved throughout all known NHases.

Despite this detailed information, and the fact that RAPc8 NHase shares 45% identity with the alpha subunit and 32% identity with the β subunit amino acid sequences of these enzymes – above the 30% ‘twilight zone’ where structural predictions from sequence alignments become unreliable (Terradot et al., 2001), it was decided to construct a preliminary model of the enzyme’s structure for further investigation. Typically, when no experimental structure is available, homology modelling represents a starting point for investigation of structure-function relationships (Combet et al., 2002). Such a model enables a more precise method of identification of key residues in and around the active site than sequence comparison, not least through the ability to manipulate the structure in terms of spatial orientation, virtual mutagenesis etc.

The structural model described in this Section was generated by protein homology modelling (Section 2.23), using the crystal structure of Rhodococcus sp. R312 (Huang et al., 1997) as a template (obtained from the Protein Database; ID 1AHJ). The template residues were systematically substituted with the corresponding residues of the RAPc8 enzyme, using the alignments shown in Fig. 5.1 as a reference.

The alignments highlight three major deviations of the RAPc8 sequences from those of the R312 enzyme: these are the N-terminal regions of both the α and β subunit sequences and a 19 amino acid ‘insertion’ in the RAPc8 β subunit at position 104. The general consensus is that protein structures are generally better conserved than are their sequences (Rossmann and Johnson, 1989) and that major differences within the homologous structures usually correspond to deletions and insertions of residues that do not disturb the essential core of the protein fold (Terradot, 2001). It was therefore decided to ignore these regions, leaving the original R312 sequence and structure in
place. This decision was based on the assumption that these variations occurred in external loops.

The R312 crystal structure demonstrated that the N-terminal regions of each subunit are located on the exterior of the structure and that the region corresponding to the 19 amino acid insertion in the β subunit is also found on the solvent-exposed surface of the enzyme. Secondary structure prediction of the corresponding RAPc8 sequences using the PHD program (Rost, 1996) (Section 2.23) was used to validate this assumption (Fig. 5.2). The data indicated that the first 21 residues of the α subunit and the first 28 residues of the β subunit were not involved in secondary structure folds. However, a short (7 residue) α helix was predicted within the 19 amino acid insertion of the β subunit (positions 116-122). In the recently determined structure of the Co-type NHase of *Ps. thermophila* (Miyanaga *et al.*, 2001), a similar insertion (residues β95-β138) containing a 14 residue α helix was found to be the most significant deviation from the Fe-type structures (Huang *et al.*, 1997, Nagashima *et al.*, 1998).

Following these replacements, the core of the model was visually inspected to minimise unacceptable steric interactions between local regions.

A more accurate model could have been developed at this stage by using specialist software to determine the energetic quality of the protein structure. However time constraints precluded such procedures, and it was assumed that, although crude, the model would nevertheless be of value in the identification of target residues.

Six aromatic residues (Figs. 5.3-5.7), all in close proximity to the cobalt ion (exact distances were not determined), were found to line the active site cavity and the channel from the bulk solvent. Only one residue, αW124, came from the α subunit; the remainder, βF36; βF52; βY67; βY72 and βW76, originated from the β subunit.
<table>
<thead>
<tr>
<th>ENZYME</th>
<th>SEQUENCE</th>
<th>H1</th>
<th>H2</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHAA_PSEPU</td>
<td>KLEB_A</td>
<td>NHAI_RRH</td>
<td>PTHERM_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOSO</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NH2_RRL</td>
<td>BR449_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RAPC8_A</td>
<td>BSMITII_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RN-774_A</td>
<td>NHAA_RHOER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP312_A</td>
<td>NHAA_PSECL</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>ENZYME</th>
<th>SEQUENCE</th>
<th>H3</th>
<th>H4</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHAA_PSEPU</td>
<td>KLEB_A</td>
<td>NHAI_RRH</td>
<td>PTHERM_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOSO</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NH2_RRL</td>
<td>BR449_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RAPC8_A</td>
<td>BSMITII_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RN-774_A</td>
<td>NHAA_RHOER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP312_A</td>
<td>NHAA_PSECL</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>ENZYME</th>
<th>SEQUENCE</th>
<th>H5</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHAA_PSEPU</td>
<td>KLEB_A</td>
<td>NHAI_RRH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOSO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NH2_RRL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BR449_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RAPC8_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BSMITII_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RN-774_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP312_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_PSECL</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>ENZYME</th>
<th>SEQUENCE</th>
<th>H6</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHAA_PSEPU</td>
<td>KLEB_A</td>
<td>NHAI_RRH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOSO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NH2_RRL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BR449_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RAPC8_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BSMITII_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RN-774_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP312_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_PSECL</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>ENZYME</th>
<th>SEQUENCE</th>
<th>H7</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHAA_PSEPU</td>
<td>KLEB_A</td>
<td>NHAI_RRH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOSO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NH2_RRL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BR449_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RAPC8_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BSMITII_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RN-774_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP312_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_PSECL</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>ENZYME</th>
<th>SEQUENCE</th>
<th>H8</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHAA_PSEPU</td>
<td>KLEB_A</td>
<td>NHAI_RRH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOSO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NH2_RRL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BR449_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RAPC8_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BSMITII_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RN-774_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP312_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_PSECL</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>ENZYME</th>
<th>SEQUENCE</th>
<th>H9</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHAA_PSEPU</td>
<td>KLEB_A</td>
<td>NHAI_RRH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOSO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NH2_RRL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BR449_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RAPC8_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BSMITII_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RN-774_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP312_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_PSECL</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>ENZYME</th>
<th>SEQUENCE</th>
<th>H10</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHAA_PSEPU</td>
<td>KLEB_A</td>
<td>NHAI_RRH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOSO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NH2_RRL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BR449_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RAPC8_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BSMITII_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RN-774_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_RHOER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP312_A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHAA_PSECL</td>
</tr>
</tbody>
</table>
Previous pages: Fig. 5.1: Amino acid alignments of the α and β subunits of the NHases of (in order): P. putida (Fallon et al., 1997); Klebsiella sp. (Pooru et al., 1994); Rhodococcus sp. (Mayaux et al., 1991); Ps. thermophila (Yamaki et al., 1997); R. rhodochrous J1 L-NHase (Kobayashi et al., 1991); Bacillus sp. BR449 (Kim and Oriel, 2000); Bacillus sp. RAPc8 (this work); B. smithii (Takashima et al., 1995); R. rhodochrous J1 H-NHase (Kobayashi et al., 1991); Rhodococcus sp. R312 (Mayaux et al., 1990); Rhodococcus sp. N-774 (Hashimoto et al., 1994); R. erythropolis JCM 6823 (Duran et al., 1993) and P. chlororaphis B23 (Nishiyama et al., 1991) used to create the homology model. Residues coloured according to hydrophobicity. The significant deviations of the RAPc8 sequence from that of the R312 sequence that were omitted from the modelling are highlighted in yellow. The secondary structure elements of Rhodococcus sp. R312 are shown on the first line, letters H and S represent α-helix and β-sheet respectively. The symbols + and ^ on the bottom denote the positions indicated by Nakasako et al (1999) of the cluster of hydrophobic residues found around the active site and the residues forming the entrance channel from the bulk solvent, respectively.
**Fig. 5.2:** Output of the PHD secondary structure prediction program (Rost, 1996) for RAPc8 NHase sequences: A - α subunit N-terminal sequence; B - β subunit N-terminal sequence and C - the insertion found at residues β105-β123.

**Legend**

- **AA**: amino acid sequence
- **PHD_sec**: PHD predicted secondary structure: H=helix, blank=other (loop)
- **Rel_sec**: reliability index for PHDsec prediction (0=low to 9=high)
- **SUB_sec**: subset of the PHDsec prediction, for all residues with an expected average accuracy > 82% NOTE: for this subset the following symbols are used: L represents a loop and . means that no prediction is made for this residue, as the reliability is: Rel < 5
Fig. 5.3: Space filled model centred on the active site region of the β subunit of RAPc8 NHase. The cobalt ion is represented by a red sphere; the aromatic residues βF36, βF52, βY67, βY72 and βW76 are highlighted in purple. The N-terminal loop domain that wraps around the α subunit can be seen to the right hand side of the image.

Fig. 5.4: Space filled model centred on the active site region of the α subunit of RAPc8 NHase. The cobalt ion is represented by a red sphere; the aromatic residue αW124 is highlighted in purple.
Fig. 5.5: Space filled model of the RAPc8 NHase holoenzyme, centred on the channel to the active site. The α subunit is shown in yellow, the β subunit in blue. The aromatic residues αW124, βF36, βF52, βY67, βY72 and βW76 are highlighted in purple.

Fig. 5.6: Space filled model of the RAPc8 NHase holoenzyme from the reverse orientation of Fig. 5.5, centred on the active site region. The α subunit is shown in yellow, the β subunit in blue. The cobalt ion is represented by a red sphere; the aromatic residues αW124, βF36, βF52 and βW76 are highlighted in purple.
Fig. 5.7: Space filled model of an αβ dimer of RAPc8 NHase. The α subunit is shown in yellow, the β subunit in blue; aromatic residues within the active site of the enzyme are highlighted in purple. The figure clearly illustrates the interaction of the β subunit N-terminal loop with the “globular domain” of the α subunit, and of the α subunit “arm domain” wrapped around the “helix domain” of the β subunit (Nakasako et al. 1999).
5.3 Site directed mutagenesis

The aim of the second phase of the investigation was to replace each of the six aromatic residues described in the previous Section. It is preferable in mutagenesis experiments directed at alteration of activity or specificity to replace with residues that will not alter the fold or stability of the enzyme. This approach increases the confidence that any observed changes are due to the properties of the residues involved rather than variations in structural conformation. Bordo and Argos (1991) created a series of exchange matrices that provide guidelines for the 'safe' amino acid substitutions least likely to disturb the protein structure. The preferred substitutions for the aromatic residues are generally to either other aromatic residues or to the larger aliphatic, non-polar residues. The approach taken in this investigation however was to replace aromatic residues with small aliphatic residues: either alanine or glycine, on the assumption that small, uncharged residues would be least likely to disturb the active site conformation. Nevertheless, the guidelines for safe substitutions were followed to the extent that in each case the favoured residue of G or A was used (i.e. F to G; Y to A and W to A or G). In addition to the glycine and alanine replacements, βF36 and βY67 was also substituted for lysine and glutamic acid respectively. These changes were based on the observation that these charged residues are found at the equivalent positions in the *R. rhodochrous* J1 L-NHase (βL34) and H-NHase (βL34 and βE66), both of which accept aromatic nitriles as substrates.

The mutations that gave rise to these amino acid substitutions were introduced by means of the Quik-change site-directed mutagenesis kit (Stratagene). The system eliminates the need to generate single-stranded DNA and enabled the mutagenesis to be performed directly on the double-stranded pNH512 expression construct. The Quik-change strategy follows a PCR-based approach (Section 2.13, Fig. 5.8) where site-specific mutations are incorporated via two complementary mutagenic primers into newly synthesised DNA during the amplification. Novel restriction sites were introduced with the mutations via these primers in order to enable rapid screening of the products. All primers were designed with the aid of the Primer Generator program (Section 2.23); the details of these primers are shown in Table 5.1 and Fig. 5.9.
Chapter 5
Enzyme engineering

Step 1
Plasmid Preparation

Gene in plasmid with target site (○) for mutation

Step 2
Temperature Cycling

Denature the plasmid and anneal the oligonucleotide primers (△) containing the desired mutation (×)

Mutagenic primers

Using the nonstrand-displacing action of PfuTurbo DNA polymerase, extend and incorporate the mutagenic primers resulting in nicked circular strands

Step 3
Digestion

Mutagenic plasmid (contains nicked circular strands)

Digest the methylated, nonmutated parental DNA template with Dpn I

Step 4
Transformation

Transform the circular, nicked dsDNA into XL1-Blue supercompetent cells

After transformation, the XL1-Blue supercompetent cells repair the nicks in the mutated plasmid

LEGEND

Parental DNA plasmid
Mutagenic primer
Mutated DNA plasmid

Fig. 5.8: Schematic of the Quik-change mutagenesis procedure. (Taken from Quik-change catalogue, Stratagene)
Table 5.1: Details of the mutagenic primers used, including primer sequence and the identities of the novel restriction sites created. Only the sense primer is shown in each case; the antisense primers were the reverse complement of the displayed sequences. Nucleotides that gave rise to the mutations and/or novel restriction sites are highlighted in bold and underlined.
Fig. 5.9: Restriction map of the RAPc8 NHase genes, indicating the recognition sites of *Fau*I, *Fok*I, *Hae*III, *Nco*I, *Nde*I and *Xma*I for the wild type enzyme (blue) and those engineered into the mutants (magenta). Black lines represent the coding sequences. The residues targeted for mutation are indicated at the bottom of the diagram.

Following mutagenesis and transformation (Section 2.13), the mutated NHase genes were amplified by PCR and treated with the appropriate restriction enzyme in order to screen for mutated genes. Figures 5.9 – 5.14 demonstrate the success of six of the eight attempted reactions. These mutations were subsequently verified by DNA sequencing (data not shown). Despite several repeated attempts, it was not possible to create the mutants pF36G or pY72A. It may be necessary to redesign the mutagenic primers if these mutations are attempted in the future.
**Fig. 5.10:** Confirmation of the mutation of αW124 to glycine by treatment with *Xma*I. Lanes: 1 - 1kb ladder; 2 - 1366bp amplicon containing non-mutated NHase α and β genes – uncut; 3 – mutated sequence giving two fragments of ~1kb and ~0.3kb.

![Image of gel showing bands for Fig. 5.10](image)

**Fig. 5.11:** Confirmation of the mutation of βF36 to leucine by treatment with *Hae*III. Lanes: 1 – 1000bp ladder; 2 – 1366bp restricted amplicon containing non-mutated NHase α and β genes giving three bands at ~220bp, 280bp and 880bp; 3 – mutated sequence showing three fragments of ~100bp, 280bp and ~880bp.

![Image of gel showing bands for Fig. 5.11](image)
Fig. 5.12: Confirmation of the mutation of βF52 to glycine by treatment with HaeIII. Lanes: 1 – φX174 DNA-HaeIII digest; 2 – 1366bp restricted amplicon containing non-mutated NHase α and β genes giving three bands at ~220bp, 280bp and 880bp; 3 - mutated sequence giving four fragments of ~70bp, 150bp, 280bp and ~880bp.

Fig. 5.13: Confirmation of the mutation of βY67 to alanine by treatment with FokI. Lanes: 1 – φX174 DNA-HaeIII digest; 2 – mutated sequence showing five fragments of ~40-50bp, ~120bp, ~240bp, ~310bp and ~610bp; 3 - 1366bp restricted amplicon containing non-mutated NHase α and β genes giving five bands at ~40bp, ~120bp, ~290bp, ~310bp and ~610bp.
Fig. 5.14: Confirmation of the mutation of βY67 to glutamic acid by treatment with FokI. Lanes: 1 - 1366bp restricted amplicon containing non-mutated NHase α and β genes showing four bands at ~40bp, ~115bp, ~300bp, and ~610bp; 2 – φX174 DNA-HaeIII digest; 3 – mutated sequence giving five fragments of ~40-50bp, ~120bp, ~240bp, ~310bp and ~610bp.

Fig. 5.15: Confirmation of the mutation of βW76 to glycine by treatment with NcoI. Lanes: 1 - 1kb ladder; 2 – 1366bp amplicon containing non-mutated NHase α and β genes – uncut; 3 – mutated sequence giving two fragments of ~1.15kb and ~0.23kb.
5.4 Expression of mutants

The mutated plasmids were isolated from the E. coli XL1-Blue host strains (Section 2.5) by minipreps (Section 2.11.2) and used to transform the expression host E. coli BL21 pLysS (DE3). Expression of the mutants was then carried out under the conditions determined to be optimal for expression of the wild type enzyme (Section 4.7.1), i.e.: growth at 37°C, induction at mid-log growth phase (OD₆₀₀nm ~0.6) with 0.4mM IPTG in the presence of 0.1mM CoCl₂, followed by harvesting after approximately 3.5 hours. Due to time constraints, these conditions were not further optimised for expression of the mutants.

The approximate expression level of each mutant was analysed by SDS PAGE (Fig. 5.15). Activity on acetonitrile (Table 5.2) was determined using the ammonia assay (Section 2.7.8) while activity on benzonitrile was determined by GC analysis (Section 2.7.9). Separate samples were subjected to a 30-minute, 55°C heat-treatment step as was conducted for the purification of the wild type enzyme (Section 4.5.2.1).

Both expression levels and activities varied significantly between the mutants, although all mutant enzymes were expressed at lower protein titres than the wild type enzyme (Fig. 5.16). Although a crude method of quantification, the SDS-PAGE gel nevertheless aids in interpretation of the specific activity values. Mutants βF52G and βY67A appear to be expressed at the highest levels. These were samples also demonstrated the highest activities on acetonitrile (72U/mg and 85U/mg respectively). It is also noteworthy that their activities were significantly higher than those determined crude extracts of the for the wild type enzyme (49U/mg). Fig. 5.15 indicates that the βW76G mutant was expressed to the lowest level, although this was not reflected in its specific activity value (59u/mg), which was analogous to that of the wild type enzyme. No benzonitrile activity was detected for any of the mutants.

The heat denaturation step resulted in a decrease in specific activity for all samples except βF36L and βY67A (which exhibited 1.7 and 1.5 fold purification factors respectively), indicating that the thermal stability of the mutant enzymes had been compromised by these mutations.
Fig. 5.16: SDS PAGE gel of cell free extracts of the expressed mutant proteins. 10μg of each sample was loaded. Lanes: 1 – molecular weight marker; 2 – BL21 only; 3 – wt enzyme; 4 - βF36L; 5 - βF52G; 6 - βY67A; 7 - βY67E; 8 - βW76G and 9 - αW124G.

Table 5.2: Specific activities (on acetonitrile) of each mutant of the recombinant RAPc8 NHase before and after heat treatment.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Specific activity (U/mg)</th>
<th>Specific activity after heat treatment (U/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>49</td>
<td>94</td>
</tr>
<tr>
<td>βF36L</td>
<td>51</td>
<td>86</td>
</tr>
<tr>
<td>βF52G</td>
<td>72</td>
<td>34</td>
</tr>
<tr>
<td>βY67A</td>
<td>85</td>
<td>125</td>
</tr>
<tr>
<td>βY67E</td>
<td>29</td>
<td>2</td>
</tr>
<tr>
<td>βW76G</td>
<td>59</td>
<td>27</td>
</tr>
<tr>
<td>αW124G</td>
<td>26</td>
<td>3</td>
</tr>
</tbody>
</table>
5.5 Thermostability

As the heat-treatment step demonstrated, the thermal stability of several of the mutants had been significantly compromised. A brief investigation into the thermostability of each sample was therefore carried out (Section 2.21.1).

The thermostability profile was plotted as the percentage of the initial activity remaining after a 30-minute incubation at temperatures ranging from 4°C to 70°C (Fig. 5.16). Comparison with the wild type enzyme indicated that all mutations resulted in some of loss of stability, though to varying extents. The most stable of these mutants was βF36L, which only showed ~20% loss of activity at 55°C (wild type was 100% active at this temperature), and was reduced to 50% initial activity at ~58°C – approximately 7°C lower than the wild type enzyme. The two most active mutants: βY67A and βF52G were also reasonably stable, both retaining at least 50% of their initial activities at 50 °C. The mutant βY67E displayed the most substantial reduction in stability, only ~50% activity remained following a 30-minute incubation at 45°C. Calculation of half-lives from these plots was not attempted due to the low levels of accuracy, as discussed in Section 4.6.2.

Fig. 5.17: Thermostability profile of wild type and mutated NHases as defined by residual activity after a 30-minute incubation at specific temperatures.
5.6 Benzonitrile Inhibition

As none of the mutations resulted in detectable levels of benzonitrile hydrating activity, a brief investigation was carried out to determine their effects on the inhibition of acetonitrile hydrating activity by benzonitrile. The experiments were carried out in the same manner as described for the wild type enzyme (Section 4.6.5), with $k_i$ being taken as the benzonitrile concentration required to reduce the rate, $v$ (assumed to be equal to $V_{\text{max}}$ at 70mM acetonitrile) to half the uninhibited rate. For this approach, one must assume that $K_M$ for each mutant is close to the 11.1mM value of the wild type enzyme. In fact, this is not necessarily a valid assumption as the $K_M$ values of active-site mutants may vary significantly. However, time constraints precluded the empirical determination of $K_M$ values for each of the mutants.

The data from these experiments is shown in Figs. 5.17-5.22, and the $k_i$ values are shown in Table 5.3. There is considerable variation between the estimated $k_i$ values: mutation of βF36L does not have a great impact, whereas βF52G actually resulted in an increase in benzonitrile inhibition. The greatest effect is shown by the βW76 to glycine mutation, which results in an increase in $k_i$ to 31.5mM; βY67E and αW124G also show significant decreases in affinity for benzonitrile.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Benzonitrile $K_i$ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type enzyme</td>
<td>~2.3</td>
</tr>
<tr>
<td>βF36L</td>
<td>~2.5</td>
</tr>
<tr>
<td>βF52G</td>
<td>~0.5</td>
</tr>
<tr>
<td>βY67A</td>
<td>~4</td>
</tr>
<tr>
<td>βY67E</td>
<td>~9</td>
</tr>
<tr>
<td>βW76G</td>
<td>~32</td>
</tr>
<tr>
<td>αW124G</td>
<td>~5.5</td>
</tr>
</tbody>
</table>

Table 5.3: Estimated benzonitrile inhibition constants for each of the RAPc8 NHase mutants.
Fig. 5.18: Benzonitrile inhibition of mutant βF36L

Fig. 5.19: Benzonitrile inhibition of mutant βF52G
Fig. 5.20: Benzonitrile inhibition of mutant βY67A

Fig. 5.21: Benzonitrile inhibition of mutant βY67E
Fig. 5.22: Benzonitrile inhibition of mutant βYW76G

![Graph showing benzonitrile inhibition of mutant βYW76G.]

Fig. 5.23: Benzonitrile inhibition of mutant αW124G

![Graph showing benzonitrile inhibition of mutant αW124G.]

Chapter 5

Enzyme engineering

175
5.7 Discussion

To the author's knowledge, this is only the second report of mutagenic analysis of a NHase, being the first involving a cobalt NHase and also the first attempt to alter the substrate specificity on a NHase. In an investigation of catalytic mechanism, Piersma et al. (2000) created a mutant of *Rhodococcus* sp. N-771, replacing the highly conserved β subunit arginine 56 with lysine. It was concluded that this residue was essential for both catalysis and stability of the enzyme and proposed that the residue was involved in the fine-tuning of activity through its positional influence on the modified cysteine residues (positions α112 and α114, Section 1.3.4).

The overall aim of the work presented in this chapter was to specifically alter the substrate preference of RAPc8 NHase through rational design. In the final event, this aim was not met, although the work has produced some insights into the catalytic importance of the residues targeted and into how the work should be continued (discussed in Sections 5.7.2 and 5.7.3).

5.7.1 Homology modelling

It was unfortunate that the *Ps. thermophila* NHase structure was released (Miyanaga et al., 2001) only following this investigation. The α and β subunits of this enzyme share 59% and 40% sequence identity, respectively, with the RAPc8 NHase, significantly higher than with the R312 subunits. Furthermore, as thermostable Co-type enzymes, RAPc8 and *Ps. thermophila* NHases might be expected to have highly conserved structures. However, Miyanaga et al. (2001) showed that the active site structure of the *Ps. thermophila* NHase was almost superimposable with that of the R312 NHase – indicating that the modelling investigation was a valid approach, as the RAPc8 enzyme, having such high sequence identity with the *Ps. thermophila* enzyme, would be expected to be similarly super-imposable.

Miyanaga et al. (2001) discussed the importance of a phenylalanine residue at position β51 in the *Ps. thermophila* NHase (which corresponds to βF55 in RAPc8 NHase) in substrate recognition. If the structures of the two enzymes were closely related, then it would appear that this residue has been overlooked in the modelling investigation.
Chapter 5

Enzyme engineering

It was also noted (Miyanaga et al., 2001) that the most significant deviation of the *Ps. thermophila* structure from that of R312 NHase (Huang et al., 1997) was found in a section of the β subunit (residues 95-138); this section corresponds to the large insertion in the alignment shown in Fig. 5.1 (RAPc8 residues β105-β123). In the *Ps. thermophila* structure, this region contains an α helix (residues 111-125) that interacts with an α helix from the α subunit (residues 36-49). It has been proposed that this interaction contributes to the thermostability of the enzyme (Miyanaga et al., 2001). However, the β subunits of *R. rhodochrous J1* L-NHase (Kobayashi et al., 1991) and *Rhodococcus* sp. NHase (Mayaux et al., 1991), neither of which enzyme have particularly high thermal stabilities, have very similar sequences in the region of the *Ps. thermophila* insertion. In fact, all Co-type NHases contain an insertion of 13-21 residues within this region, suggesting that this sequence relates to cofactor preference rather than stability. It was noted of the model in this study that many of the residue substitutions (of the RAPc8 enzyme with respect to the R312 NHase) were located on the external surfaces of the α helices (data not shown). It has been previously reported that a common stabilisation mechanism of thermostable proteins is the replacement of external residues on helices with amino acids with strong helical propensities, potentially increasing stability through increased numbers of H-bonds, salt bridges and side chain-side chain interactions (Facchiano et al., 1998).

5.7.2 Expression and characterisation of the mutants

As discussed in Section 5.4, the expression levels of the site-specific mutants varied considerably, and all appeared to be expressed at lower concentrations (with respect to total cell protein) than the wild type enzyme. At this stage, and without optimization of expression, conclusions cannot be made as to the reasons for these differences. The method used to approximate expression levels was relatively inaccurate, since ideally calculations from activity levels of cell extracts and purified protein would be used. It may be the case that the introduced mutations lowered the efficiency of translation of the transcripts and hence resulted in lower protein concentration. In comparison, it was found that mutation of *Rhodococcus* sp. R312 amidase resulted in expression of higher levels of activity (cf. the wild type enzyme) due to increased translation levels (Bigey et al., 1999).

177
Each of the RAPc8 NHase mutants displayed a decrease in thermostability, though to varying degrees, perhaps reflecting the relative importance of each residue in stabilization of the enzyme. Several factors may be involved in this destabilization:

1. The mutations may sterically disrupt the overall conformation of the protein, though the choice of 'safe' residue substitutions (Bordo and Argos, 1991) reduces this possibility. In the case of the βY67 mutations, βY67A is very much more stable than βY67E. It might be concluded that this difference was caused by disruptions due to the insertion of an unpaired charged group into a hydrophobic region.

2. Reduction of the number of hydrogen bond in the active center. Nakasako et al. (1999) reported the stabilization of this cavity in Rhodococcus sp. N771 NHase by 20 hydration water molecules, and that the residues corresponding to αW124, βF36 and βY67 of RAPc8 NHase all form a hydrogen bond with one of these molecules. The substantial reduction in stability of αW124G may be explained by the fact that if αW124 forms a hydrogen bond with a water molecule directly associated with the cofactor-liganding αS114, as in the case of the N771 enzyme, then the replacement of this residue would remove an H-bond critical for stability of the enzyme.

3. It may be the case that the enzyme is stabilized in part by aromatic interaction between several of these aromatic residues. It would follow then that replacement of one of these residues would destabilize the protein.

4. As discussed in Section 1.7, there is evidence that increased thermostability can, in part, be attributed to an increase in the density of molecular packing (Vogt and Argos, 1997). It follows then, that replacement of a bulky aromatic side chain with a smaller side chain such as alanine or glycine will increase void space and hence detrimentally affect thermal stability.
5.7.2.1 Mutant αW124G
This tryptophan residue is strictly conserved throughout all NHases – indicating structural and/or catalytic significance. It is not surprising therefore that replacement severely affected both activity and stability. It may be the case that the low activity is due to the reduction in thermal stability rather than impedance of catalytic activity. The slight rise in $K_i$ for benzonitrile may be due to a reduction in affinity for both substrates and substrate analogues.

5.7.2.2 Mutant βF36L
This position is not strongly conserved among NHases, and its replacement with leucine – a common natural substitution (Bordo and Argos, 1991) resulted in almost no change in activity, stability or benzonitrile inhibition. These results are consistent with the view that this is not a catalytically vital residue.

5.7.2.3 Mutant βF52G
This position is also not highly conserved amongst NHases. Mutation of this residue to glycine resulted in a significant increase in specific activity compared with the wild type. Although the mutant protein was expressed at higher levels than other mutants, it appeared from the SDS PAGE gel (Fig 5.15) that it was expressed at lower levels than the wild type enzyme. Therefore it seems unlikely that this increase in activity is due to an increase in expression level. Nakasako et al. (1999) proposed that a valine residue in the equivalent position in *Rhodococcus* sp. N-771 NHase may act as a ‘gate’ for substrates entering the active site cavity. If this is the case for the RAPc8 enzyme, then replacement of βF52 with the smaller glycine residue may increase activity by enabling more rapid substrate turnover. Another explanation for the observed increase is simply that a greater quantity of cobalt was available for holoenzyme formation (Section 4.7.1.1), or that the mutant bound Co$^{2+}$ more readily than the wt. However, this increase in activity is juxtaposed with a significant reduction in thermal stability, which may limit its viability as a useful catalyst. As yet there is no explanation for the marked increase in benzonitrile inhibition.
5.7.2.4 Mutants βY67A and βY67E

Tyrosine β67 is strongly (though not strictly) conserved amongst NHases, which would indicate that it is a structurally or catalytically significant residue. However, results of the substitution in this investigation indicate that this is not the case. The activity of the βY67A mutant on acetonitrile was almost two fold greater than that of the wild type enzyme, and although mutation did affect stability under the conditions tested, the reduction in thermostability was slight (thermostability studies indicated at least 50% initial activity remained after 30 min incubation at 55°C). βY67A was expressed to higher levels than most other mutants (Fig. 5.15) although this does not fully explain the increase in specific activity, as the level of expression appeared to be lower than that of the wild type. Whilst it may be that cofactor uptake or binding parameters were altered, the characteristics of this mutant make it a very interesting target for further investigation.

There was a reduction in benzonitrile inhibition (cf. wild type NHase), though the decrease in affinity for benzonitrile was so minimal that it is unlikely that this residue is involved in the aromatic interaction proposed to cause the inhibition.

The properties of the βY67E mutant were markedly different from the alanine mutant, the former showing significant decreases in both stability and activity on acetonitrile. Although the $K_i$ for benzonitrile was decreased to $\sim 8.8$ mM, the difference between this value and that for βY67A clearly demonstrates that the fall is due to the properties of the new residue rather than removal of the aromatic residue.

In nature, the substitution of tyrosine for glutamic acid is generally avoided (Bordo and Argos, 1991), most likely because of the disruption that the insertion of an unpaired charged group into a hydrophobic position is likely to cause. This study highlights the importance of careful selection of replacement residues in mutagenesis experiments.
5.7.2.5 Mutant βW76G

The replacement of this residue had little effect on specific activity; in fact a slight rise from that of the wild type enzyme was observed, despite its apparently lower expression level (Fig. 5.15). There was a substantial drop in stability; this may be due to replacement of a residue involved in stabilisation through aromatic interaction (this residue is conserved as either tyrosine or tryptophan throughout all NHases).

The most interesting characteristic of this mutant was the drastic decrease in benzonitrile inhibition ($K_i \sim 31.5\text{mM}$) cf. the wild type enzyme. Assuming that the aromatic interaction hypothesis is correct, this result suggests that βW76 is implicated in the inhibition process. However, despite this drop in $K_i$, the mutant still displayed no activity on benzonitrile. A worthwhile future investigation would be the replacement of βW76 with leucine, a more common substitution than tryptophan to glycine (Bordo and Argos, 1991), in an attempt to minimise the observed loss in stability.

5.7.3 Conclusions

None of the six mutations tested in this study led to the generation of hydratase activity on benzonitrile. It is, however, not surprising that the first attempts to alter specificity through rational design were not successful. Despite the enormous collective understanding of protein structure and enzyme mechanisms, the nature of modifications required to alter catalytic activity often remain elusive (Dufour, 1995). Nevertheless, the investigations described in this chapter represent a important addition to the body of knowledge of NHase activity in general and to the NHases of the thermophilic *Bacilli* in particular.

It has been demonstrated that although involved in the active site architecture, none of the residues investigated are essential for catalysis, though several do appear to have a major influence on catalytic activity. While benzonitrile inhibition was drastically decreased in the βW76G mutant, it was not eliminated. Furthermore it is important to remember that the inhibition assays cannot be considered accurate (Section 5.6), it is therefore not possible at this stage to confirm or discredit the aromatic interaction hypothesis. It would be of value at this stage to purify and thoroughly characterise the kinetic parameters of this mutant. Virtual mutagenesis as well as substrate docking experiments on the homology model may also be of use in explaining the results.
Further mutagenic investigations might involve replacement of βF52, βW76 and αW124 with leucine, the most common non-aromatic substitution for tryptophan and phenylalanine (Bordo and Argos, 1991). These mutations may provide greater structural and thermal stability than those created in this work. Subsequently, double or triple mutations, in particular, combinations of the most active mutants might be attempted in order to delete benzonitrile inhibition and facilitate activity on benzonitrile.
Chapter 6: The search for novel nitrile-metabolising enzymes

6.1 Introduction

Although a significant number of nitrile-degrading enzymes have been characterised to date the successfully applied to industrial scale preparations (Kobayashi et al., 1992b, Yamada and Kobayashi, 1996), the volume of ongoing research in this area and the repeated upgrades with increasingly efficient enzymes in the acrylamide process illustrate the incentive to find ever more potent biocatalysts. The classic approach to enzyme discovery has been to enrich, isolate and screen microorganisms for a desired trait. However, phylogenetic studies over the past decade have indicated that up to 99% of the microorganisms existing in nature cannot be cultivated by standard techniques (Hugenholtz and Pace, 1996, Ward et al., 1990), suggesting that a large fraction of the protein diversity is missed by the traditional screening techniques. The genomes of the total microbiota found in nature, collectively termed the metagenome (Handelsman et al., 1998), contain a huge and virtually untapped natural resource. Advances in molecular techniques have provided powerful tools to enable greater access to this information.

Following the extraction of DNA from environmental samples, there are essentially two strategies for the isolation of catabolic genes directly from the metagenome: PCR-based protocols and vector-cloning based protocols. PCR methods using either degenerate primers (multiplex PCR) or exact sequences for more specific amplification have been applied to the cloning of full-length β-ketoacyl synthase (KSB) genes (Seow et al., 1997) and NHase genes (Precigou et al., 2001) from metagenomic DNA. The alternative method is to clone and directly express the metagenomic DNA: environmental DNA libraries are created and clones positive for the desired trait are identified through screening assays based on target functionality. The earliest reported study of this kind describes cloning and expression of a chitinase in a λ phagemid library (Cottrell et al., 1999). Many variations on this theme have since been reported; vectors from pUC-based plasmids (Henne et al., 1999) to bacterial artificial
chromosomes (BAC) (Rondon et al., 2000) being used to express diverse genes from secondary metabolite pathways (Brady et al., 2001) to ion transporters (Majernik et al., 2001). The attraction of this latter method over PCR cloning is that it is entirely culture independent: whilst PCR cloning is limited by the requirement of a degree of prior knowledge of the gene sequence - usually based on analogous genes of previously cultured organisms, expression screening may be performed ‘blind’. Furthermore the versatility of the approach has been demonstrated by the application of different assays to isolate different activities from a single library (Henne et al., 2000, Rondon, 2000).

The aims of the study described in this chapter were to characterise the NHase gene content of a geothermal sediment sample and also to clone functional, full-length nitrile degrading genes from metagenomic DNA samples. The techniques used for DNA isolation and library construction were based on protocols developed by a previous member of the author’s laboratory (Wilkinson, 2001).

This work was carried out in conjunction with Ilhem Diboun, and was sponsored by Mitsubishi-Rayon Corporation.

6.2 DNA Extraction

DNA isolation from environmental sources can be based either on recovery of bacterial cells and subsequent lysis or on direct lysis of cells in the sample followed by DNA purification. Generally the cell extraction method is not used for soil and sediment samples as it is time consuming and not necessarily any more efficient than the direct lysis approach (Steffan et al., 1988). The FastDNA spin kit (BIO101), a direct lysis protocol, was used to prepare metagenomic DNA. This choice was influenced by Wilkinson’s findings that despite lower DNA yields and higher cost than manual methods, the process had the advantages of being rapid, simple to scale up, reproducible and would generate DNA free of the humic acids often present in environmental DNA samples that are known to inhibit many subsequent manipulations (Tebbe and Vahjen, 1993, Tsai and Olson, 1992a, 1992b).

DNA was extracted from hydrothermal sediment samples collected aseptically from a number of locations in and around New Zealand hot pools. DNA concentrations were
determined fluorometrically using the PicoGreen dsDNA quantitation kit (Molecular Probes) due to the limited quantity available from these preparations. Typical yields for the extractions were in the order of 3.75 µg of DNA per g of soil: within the range described by Wilkinson (2001) (up to 9.9 µg per g wet-sediment), and broadly similar to other reported values (Henne, 1999). DNA isolated in this manner tended to be sheared, with a size range from around 200bp up to approximately 12kb. The majority of the sample was over 3kb in length. A sample, WHA30 (isolated at 62°C, pH 8.02) that appeared to yield the highest DNA concentrations was chosen for further work.

**Fig. 6.1:** Typical result of environmental DNA extraction from sample WHA30. Lanes: 1 and 3, 1kb ladder; 2, 1µl WHA30 DNA. These results were also presented in Miss Diboun’s final year B.Sc. project, Department of Biochemistry and Molecular Biology, University College London.
6.3 PCR analysis of WHA30 environmental DNA sample.

In order to confirm the presence of nitrile hydratase genes within the sample and to assess the diversity of these genes a multiplex PCR protocol was developed. Since the efficacy of primers NHFl, NHF2, NHRI and NHR2 in the identification of NHase genes had been previously demonstrated (Section 3.4.1), it was decided to modify the original reactions using these primers for multiplex PCR. However, as the aim of the work was to include an assessment of the diversity of NHase genes, it was important that the primers used were specific to both cobalt and iron containing enzymes. Preliminary PCR investigations using DNA from the Fe-type NHase producing organism *R. erythropolis* JCM6823 and the Co-type NHase producing organism *Bacillus* sp. RAPc8 as templates demonstrated that only primer pair NHFl-NHRI was specific to both iron and cobalt binding NHase gene sequences (Fig. 6.2).

**Fig. 6.2:** Amplification of NHase fragments using primers NHFl and NHRI. Lanes: 1, φX174 DNA-HaeIII Digest; 2, PCR with RAPc8 DNA; 3, PCR with *R. erythropolis* JCM6823 DNA.
Chapter 6

The search for novel nitrile-metabolising enzymes

On the basis of these preliminary positive control experiments, only primer pair NHF1-NHR1 was used to amplify NHase DNA fragments from soil samples. Despite comprehensive optimisation of the PCR (T\textsubscript{annealing} and [MgCl\textsubscript{2}] – data not shown), initial reactions using this primer pair showed a feint smear of DNA around 100-300bp when electrophoresed on a 1.5% agarose gel. On the assumption that the reaction had amplified target sequences together with non-specific fragments, DNA between 150 and 200bp was excised from the gel and used as the template for a second PCR reaction under identical conditions with the exception of a 5°C rise in annealing temperature. This time a stronger, clearer band was detected at approximately 170bp.

PCR reactions contained ~10ng of template DNA, 1.2μmol of each primer, 0.2mM dNTP's, 2.5μl of (10x) PCR standard buffer (final MgCl\textsubscript{2} concentration of 1.5mM) and 1.5U Taq polymerase (Roche), in a 25μl reaction, overlaid with mineral oil. Reaction conditions were as follows:

- 94°C – 5 minutes 1 cycle
- 94°C – 30 seconds
- 50°C – 45 seconds 30 cycles
- 72°C – 45 seconds
- 72°C – 7 minutes 1 cycle

The product from this reaction was purified using the QIAquick PCR purification kit (Qiagen) and subsequently used to create a PCR library by cloning into the pCR2.1 vector and transforming into electrocompetent E. coli TOP10 (Section 2.15.2). The PCR library was grown at 37°C overnight. Up to 30 colonies were picked and grown up overnight and their plasmids isolated. Inserts from the plasmids were amplified by PCR using the conditions described above.

PCR products of approximately 170bp (~70% of products) were restriction digested directly using 4-base cutters AluI, MspI and HinfI in order to assess the diversity of NHase species represented within the library through their differing restriction patterns. Up to 7 different restriction patterns were observed. However, sequence analysis of representatives selected randomly from each of these groups showed that many of the
products were the result of non-specific priming. Two groups, however were made up of sequences with distinct homology to previously described nitrile hydratase sequences. The first group produced fragments of approximately 25bp, 50bp and 100bp when cut with Alul and 50bp and 120bp when cut with HinfI; the other group produced fragments of approximately 70bp and 100bp when cut with Alul and 50bp and 120bp when cut with HinfI. Sequencing of a number of randomly selected inserts from these groups revealed six unique sequences. Translated BLAST searches indicated that each of these sequences could be classified with a high level of confidence to the nitrile hydratase gene family. All six sequences have the highest identities with the NHase sequences of the thermophilic Bacillus species (table 6.1). The DNA sequences and deduced amino acid sequences of each of the inserts are shown as alignments in Fig. 6.3. Poor sequence results were obtained for the insert of pILHL, despite two sequencing attempts. In the alignment of the DNA sequences, a two base pair insertion in the sequences of four of the amplicons (pILHC3; pILHL; pILHH and pILHQ) is highlighted. This insertion disrupts the reading frame of the sequences, giving rise to a nonsense amino acid sequence. This apparent frameshift is presumably due to an error in the amplification reaction.

Table 6.1: Percentage identity and similarity of the DNA and translation products of the six amplicons with the equivalent DNA and amino acid sequences of Bacillus sp. BR449 and Bacillus sp. RAPc8. Data for pILHC3 and pILHL have also been presented in Miss Diboun’s final year B.Sc. project, Department of Biochemistry and Molecular Biology, University College London.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>DNA Identity</th>
<th>Amino acid Identity</th>
<th>Amino acid Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>pILHC3</td>
<td>89</td>
<td>93</td>
<td>96</td>
</tr>
<tr>
<td>pILHG</td>
<td>93</td>
<td>96</td>
<td>100</td>
</tr>
<tr>
<td>pILHH</td>
<td>89</td>
<td>91</td>
<td>98</td>
</tr>
<tr>
<td>pILHL</td>
<td>82</td>
<td>91</td>
<td>98</td>
</tr>
<tr>
<td>pILHQ</td>
<td>90</td>
<td>91</td>
<td>98</td>
</tr>
<tr>
<td>pILHR</td>
<td>92</td>
<td>96</td>
<td>98</td>
</tr>
</tbody>
</table>
Fig. 6.3: A, Alignment of DNA sequences of the six PCR sequences against the corresponding sequences of *B. smithii*, *Bacillus* sp. BR449 and *Bacillus* sp. c8. B, Alignment of the deduced amino acid sequences of the six PCR sequences.
6.6.2 Library Construction

The ability to reproducibly construct relatively large libraries via the TOPO-TA cloning method was clearly demonstrated in this work, with results similar to those of Wilkinson (2001). The cloning efficiency was 2.5-fold lower than reported by Wilkinson (2001), though this is likely to be due to the lower insert DNA concentrations used in this investigation. A higher efficiency may have been achieved if extracted DNA was further concentrated and fewer ligation reactions had been performed per extraction. The substantial shearing caused by the extraction technique used in this work also limited the choice of vector. As the DNA was a maximum of 12kb in size, production of BAC or cosmid vector libraries was not possible. Such vectors enable the cloning of large DNA fragments, which might have increased the likelihood of identifying positive, full length genes as the probability of cloning a full length gene increases with increasing insert size, as does the probability that the insert will contain regulatory elements that may be required for functional expression. Similarly, restriction digestion of such (relatively) small fragments for use in cohesive-end cloning was likely to produce a high proportion of heterogeneous and hence uncloneable termini (Wilkinson, 2001).

Table 6.2: Comparison of the library with the environmental expression libraries produced by Wilkinson (2001) and Henne et al. (1999).

<table>
<thead>
<tr>
<th>Vector</th>
<th>Cloning Efficiency (x10^6 cfu/µg vector)</th>
<th>% Recombinant</th>
<th>Insert Size (kb)</th>
<th>Independent Clones (^1)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCR-XL-TOPO</td>
<td>0.56</td>
<td>80</td>
<td>~4.2</td>
<td>110000</td>
<td>This work</td>
</tr>
<tr>
<td>pCR-XL-TOPO</td>
<td>1.4</td>
<td>84</td>
<td>~5</td>
<td>37000 (^2)</td>
<td>(Wilkinson, 2001)</td>
</tr>
<tr>
<td>pBluescript</td>
<td>0.017 (^3)</td>
<td>80</td>
<td>~6.5</td>
<td>930000</td>
<td>(Henne, 1999)</td>
</tr>
</tbody>
</table>

\(^1\) Defined as the number of recombinants prior to library amplification.

\(^2\) This library was built from only three ligations and hence is substantially smaller than the library described in this investigation.

\(^3\) Estimated from the cloning information provided by the investigators - assuming that 50ng of vector was used per ligation at an insert-to-vector molar ratio of 3:1.
Fig. 6.4: (Taken from Wilkinson, 2001) Overview of environmental library construction using pCR-XL-TOPO.
Unamplified, the final library, built from 22 ligations (2 pCR-XL-TOPO kits) possessed approximately 110,000 transformants (a cloning efficiency of $0.56 \times 10^6$ cfu/µg vector DNA). Restriction analysis of randomly selected clones suggested that 80% of plasmids contained an insert, suggesting that the library contained approximately 88,000 independent recombinant clones. Insert sizes of recombinants ranged from ~1.5kb to >10kb (Fig. 6.5), with an average insert size of 4.2kb, representing a total of 370Mbp of cloned DNA. Given an average of 1kb per gene, the library might contain 370000 genes.

Fig. 6.5: EcoRI-digested plasmids recovered en masse from the library. Plasmid DNA was isolated by a miniprep of the amplified library. Lanes: 1, 1kb marker; 2, Plasmid DNA digested with EcoRI. Linear plasmid DNA indicated by the arrow at ~3.5kb.
6.5 Library screening

A complementation-based method was used to screen for functional NHase genes within the library, based on the supposition that if these genes were expressed, they would be able to support growth on media containing a nitrile as the sole nitrogen source. Since *E. coli* does not express amidase, the method requires both NHase and amidase to be expressed functionally. The screening experiments in this work were performed on a modified Castenholz media that has been previously used successfully in our laboratory to isolate organisms producing nitrile-degrading genes (Pereira, 1998).

An adaptation of the Clarke and Carbon (Clarke, 1976) equation for calculating the required size of DNA libraries was used to estimate the number of recombinants needed to be screened in order to have a particular probability of screening all clones within the library:

\[
N = \frac{\ln(1 - p)}{\ln\left(1 - \frac{x}{y}\right)}
\]

where,

- \(N\) = the number of agar plates to be screened
- \(p\) = the probability of screening all independent transformants
- \(x\) = the cell density per agar plate (5000 cfu)
- \(y\) = the number of independent transformants (88,000)

For each screening assay, 50 140mm plates were used at a density of ~5000 cfu/plate (Section 2.3): giving a probability of ~95% of assaying all independent clones present in the library. As the strain of *E. coli* used (TOP 10) contained a deletion preventing leucine biosynthesis \([\Delta(ara-leu)7697]\), leucine was added to the media at a concentration of 0.5mg/l, sufficient to support growth on Castenholz-nitrogen+ plates, though not sufficient to enable growth on nitrogen free media. Benzonitrile or acetonitrile were added at concentrations just below the toxicity threshold that was determined for these cells (5mM and 800mM respectively). The plated library was incubated for 4 days at 37°C, and checked daily for growth.
Chapter 6 The search for novel nitrile-metabolising enzymes

After ~18hrs, many pinprick colonies appeared on the plates. However, over a 4-day period they did not visibly increase in size. The colonies from 10 of these plates were isolated, washed and re-plated on the nitrogen-free agar at a dilution of ~1000cfu/plate. No growth was observed on the plates over the same 4-day period, suggesting that the initial growth was supported by residual nitrogen either from the media used to amplify the library or from dead cells present during the plating.

6.6 Discussion

6.6.1 PCR detection of NHase genes

The oligonucleotides used in this investigation have been shown to be capable of selective amplification of NHase genes in environmental samples. It is clear, however that there is scope for further improvement of the procedure. The approach taken of using a low annealing temperature for initial reactions, followed by a second round of PCR at an increased temperature served to enrich the template DNA for potential target sequences but would also have increased the likelihood of false positives through mispriming and non-specific amplification.

Although the study successfully confirmed the presence of NHase genes within the sample tested, the method appeared biased towards the Bacillus species Co-type NHase genes. The high sequence identity of the NHase fragments amplified in this work with the previously described Bacillus species NHase genes raises the question of how widely distributed the thermostable NHase genes are. While it may be the case that the primers have a bias toward NHase genes of the Bacilli over those of other organisms, it is significant that four of the five thermostable NHases described so far (Takashima et al., 1995, Yamaki et al., 1997, Cramp et al., 1997, Pereira et al., 1998, Kim and Oriel, 2000) are produced by thermophilic Bacillus species. Furthermore, it should be noted that all thermophilic NHases described to date have been of the cobalt-containing family. It may then be that the results presented in this study are in fact representative of the true diversity of the NHase producing community of WHA30.

The fragments produced span a section of the gene that codes for the cofactor liganding residues in and around the active site of the enzyme. Thus, a single residue substitution in this region may have a marked effect on the specificity of the enzyme. It is perhaps
notable that the majority of the differences between the sequences lie within the primer-binding regions and that they may therefore be artifacts of the degenerate primers, in that only one or two unique sequences were amplified and the differences are due to mismatched priming or mutations that occurred during the reaction. This argument is supported by presence of an identical insertion (assumed to be an error of the PCR) found in the sequences of pILHC3; pILHL; pILHH and pILHQ.

Following the completion of this investigation, a similar study based on DNA isolated from biotopes more favourable to mesophilic organisms was published (Precigou, 2001). Using similar techniques, the group has identified novel sequences with particular homology to the NHases of *R. rhodochrous* J1 (H-NHase) and *P. putida*. In their investigations, Precigou *et al.* (2001) considered cobalt and iron NHases separately. Using primers corresponding exactly to the NHases operons of *R. erythropolis* (Fe-NHase) or *R. rhodochrous* J1 (Co-NHase), the group was able to amplify full-length NHase gene sequences and express them in *E. coli*, though the activities observed were low (similar attempts to amplify full length genes using the NHOP series of primers [Section 4.3] were made in this investigation but without success). Figure 6.6 below shows a schematic outlining the positions of the primers used by Precigou *et al.* (2001) in comparison to those used in this work. The high levels of sequence identity between the amplicons described in this report and the previously described thermophilic Bacilli NHase gene sequences suggests that the use of non-degenerate primers may be a valuable approach in further searches for novel thermostable NHases. The obvious limitation of this approach however is that it is aimed at the isolation of genes that, by the nature of the technique, are required to be highly similar to those already discovered. Indeed, the amino acid sequences reported by Precigou *et al.* (2001) were ≥ 90% identical to the *Rhodococcus* or *Putida* homologues. However, whilst useful for the detection and acquisition of NHase genes, this approach is unlikely to give a true representation of the diversity of the NHase genes. It appears then that in order for such a technique to be sufficiently selective, one must compromise the degree of enzyme variation expected. This is the major reason why the expression screening strategy is superior when searching for a novel enzyme with a defined function.
**Fig. 6.6:** Schematic diagram illustrating the organisation of the nitrile hydratase operons of *R. erythropolis* and *R. rhodochrous* J1 (H-NHase) and the positioning of primers used in this work (orange) and those used by Precigou *et al.* (2001) (green).
6.6.2 Library Construction

The ability to reproducibly construct relatively large libraries via the TOPO-TA cloning method was clearly demonstrated in this work, with results similar to those of Wilkinson (2001). The cloning efficiency was 2.5-fold lower than reported by Wilkinson (2001), though this is likely to be due to the lower insert DNA concentrations used in this investigation. A higher efficiency may have been achieved if extracted DNA was further concentrated and fewer ligation reactions had been performed per extraction. The substantial shearing caused by the extraction technique used in this work also limited the choice of vector. As the DNA was a maximum of 12kb in size, production of BAC or cosmid vector libraries was not possible. Such vectors enable the cloning of large DNA fragments, which might have increased the likelihood of identifying positive, full length genes as the probability of cloning a full length gene increases with increasing insert size, as does the probability that the insert will contain regulatory elements that may be required for functional expression. Similarly, restriction digestion of such (relatively) small fragments for use in cohesive-end cloning was likely to produce a high proportion of heterogeneous and hence uncloneable termini (Wilkinson, 2001).

Table 6.2: Comparison of the library with the environmental expression libraries produced by Wilkinson (2001) and Henne et al. (1999).

<table>
<thead>
<tr>
<th>Vector</th>
<th>Cloning Efficiency</th>
<th>% Recombinants</th>
<th>Insert Size (kb)</th>
<th>Independent Clones</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(x10^6 cfu/µg vector)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCR-XL-TOPO</td>
<td>0.56</td>
<td>80</td>
<td>~4.2</td>
<td>110000</td>
<td>This work</td>
</tr>
<tr>
<td>pCR-XL-TOPO</td>
<td>1.4</td>
<td>84</td>
<td>~5</td>
<td>37000^2</td>
<td>(Wilkinson, 2001)</td>
</tr>
<tr>
<td>pBluescript</td>
<td>0.017^3</td>
<td>80</td>
<td>~6.5</td>
<td>930000</td>
<td>(Henne, 1999)</td>
</tr>
</tbody>
</table>

1 Defined as the number of recombinants prior to library amplification.
2 This library was built from only three ligations and hence is substantially smaller than the library described in this investigation.
3 Estimated from the cloning information provided by the investigators - assuming that 50ng of vector was used per ligation at an insert-to-vector molar ratio of 3:1.
6.6.3 Library Screening

Previous cloning experiments, involving complementation screening for growth on minimal media containing acetamide as the sole nitrogen and/or carbon source led to the cloning of the amidase genes of *Pseudomonas aeruginosa* (Drew, 1984) and *Rhodococcus* sp R312 (Azza et al., 1994). This in turn suggested that the method could be reliably adapted to screening for novel NHase genes. However, no growth of the expression library was supported on either acetonitrile or benzonitrile containing plates in this investigation.

The simplest explanation for the absence of growth was that no nitrile degrading genes were present in the library. However even if the genes were present, expression and detection may not always be straightforward. The main factors which govern the screening of any recombinant enzyme are: i) the level of gene expression, and ii) enzyme assay sensitivity (Short and Mathur, 1998).

Heterologous expression of genes cloned in this library may be initiated from the promoter sequences of either the pCR-XL-TOPO or the cloned DNA, if present. Translation of the foreign transcript from the vector ribosome-binding site, however, requires that the target gene be fused to *lacZα* fragment encoded by the vector in the proper reading frame. If heterologous expression is under the control of a cloned promoter, the *E. coli* TOP10 host transcriptional and translational machinery must be able to recognise these sequences. It is also noteworthy that the biosynthesis of NHases often depends on specific growth conditions (amide inducers, for example) (Duran et al., 1993) and if the cloned genes require these specific inducers, they may be overlooked. However, the thermophilic NHases characterised to date have been produced constitutively (Cowan et al., 1998, Padmakumar and Oriel, 1999, Takashima et al., 1998). Assuming that the required conditions for expression are available, one would expect a detectable level of expression from the library irrespective of the origin of the promoter as the vector used is a derivative of the pUC cloning vectors (Vieira and Messing, 1982) and hence can be maintained stably at high copy number in the *E. coli* host (although precise copy numbers will depend on the nature of the insert).
Chapter 6

The search for novel nitrile-metabolising enzymes

A chief concern when considering the sensitivity of the complementation screening is that the method requires enzymes that normally function at temperatures above 50°C to support growth at only 37°C. It is also important to take into consideration factors such as substrate specificity. It is unfortunate that there was no adequate positive control available for the isolation of an amidase-NHase operon. It would be useful to determine the promoter region of the *Bacillus* sp. RAPc8 NHase operon or that of any other thermostable NHase operon in order to include the entire operon, with promoter in a clone: such a clone could be used as a positive control to assess the viability of the complementation screening approach.

It must be stressed that this was only the first stage in screening. The frequency of detecting expressed proteins may potentially be increased by increasing the number of different assays performed (Wilkinson, 2001), in this case there are a number of options available:

1. Complementation screening on amides for amidase clones. As NHase genes are usually associated with amidase, they may potentially be isolated by ‘walking’ along amidase positive clones.

2. Through the development of screening assays that do not require expressed genes to support growth. For example, Nessler’s reagent (Fisher scientific) has been used for detection of amidase-released ammonia by bacterial colonies, enabling the cloning of a wide spectrum amidase (Cheong and Oriel, 2000).

3. Hybridisation with known NHase gene sequences.
Chapter 7: General discussion

The primary objective of the research presented in this thesis was to clone the genes encoding the previously described NHase of *Bacillus* sp. RAPc8 (Pereira, 1998) and to overexpress the recombinant protein in an active form in a suitable expression system. This was accomplished using standard degenerate PCR and DNA hybridisation techniques, the primary investigations facilitated through the availability of sequence information derived from the cloning of related NHase genes from a variety of mesophilic and thermophilic bacteria.

The work presented in Chapter 3 is the fourth report to date describing the cloning of a (partial) NHase operon from a thermophilic organism. Other than the arrangement of the α and β subunit genes, little information is available regarding the layout of the NHase operons of *B. smithii* (Takashima et al., 1996) or *Ps. thermophila* (Yamaki et al., 1997) whilst the cloned NHase operon sections of *Bacillus* sp. BR449 and *Bacillus* sp. RAPc8 span opposing sections of the locus. Nevertheless, studying the thermophile NHase genes as a whole reveals consistencies in their structure and organization: all are arranged in β-α order and all contain the VCTLCSCY motif in the α subunit, characteristic of Co-type NHases. This second observation would seem to support the proposal that the Co-type enzymes are generally more thermostable than their Fe-type counterparts (Section 1.3.7, Kobayashi and Shimizu, 1998). Although it is not possible at this stage to rule out the existence of thermophiles producing Fe-type NHase, the data collected so far indicates that this is not likely to be the case. It has been previously noted that sequence identities between NHases are highly conserved (Pereira, 1998). The observations of this work show that this is particularly so amongst the thermostable NHases (the lowest sequence identity being 64% between the β subunits of *Bacillus* sp. RAPc8 and *Ps. Thermophila*). Interestingly, this group also exhibits high sequence similarity with the H-NHase of *R. rhodochrous* J1, which, despite originating from a mesophilic organism, is remarkably thermostable (Wieser et al., 1998). Thus, the collective sequence data of the thermostable NHase present a useful tool for investigation of the strategies used by these enzymes for increased thermostability over
their mesophilic homologues. Furthermore, the information will be valuable for designing primers and/or probes for the isolation of novel NHases.

In contrast to most attempts to over-express recombinant NHases in soluble and active form, expression of *Bacillus* sp. RAPc8 NHase, as with that of *Bacillus* sp. BR449 (Kim and Oriel, 2000) was relatively straightforward, requiring only the presence of the cloned structural genes in expression constructs. From the characterisation studies, it may be concluded that the thermostability and catalytic characteristics of the purified native protein were successfully reproduced in the recombinant NHase. The system generated in this study enables facile production of large quantities of recombinant protein which, following the design of an efficient purification protocol, will facilitate further characterisation of the enzyme (activity on novel substrates, structural studies etc.).

As yet, no concerted efforts have been made to co-express the RAPc8 P14K gene. In the case of *Bacillus* sp. BR449, attempts to co-express the downstream P12K gene with NHase resulted in little change in expression levels and no protein corresponding to P12K (Kim and Oriel, 2000). These are the first reports concerning expression of soluble NHase without the co-expression of activator proteins and supplementation of growth media with cofactor ions (Kim and Oriel, 2000, Kim et al., 2001). Although no data are currently available concerning the expression of *B. smithii* or *Ps. thermophila* NHases, it appears that the thermostable *Bacillus* NHases form a new sub-group within the Co-type NHases.

Despite the high expression levels of the RAPc8 enzyme, much of the soluble protein was apparently inactive, possibly due to problems associated with the availability of cobalt to the enzyme (Section 4.7.2). Given the significant homology of the RAPc8 P14K protein to *P. putida* P14K, which is essential for expression of high activity levels and has been implicated in the introduction cobalt into the enzyme, an important investigation for the near future would be a repeated attempt to co-express this protein with the NHase.

At this stage, comparisons of estimated kinetic constants ($K_m$ and $k_{cat}$) indicate that RAPc8 NHase cannot compete with the industrial catalyst *R. rhodochrous* J1 NHase in
terms of efficiency for commercial production of acrylamide. However, it may be the case that the presence of inactive Co-NHase apoenzyme distorted these calculations: repeated experiments containing only active NHase holoenzyme may reveal a higher affinity and turnover number for acrylonitrile. Furthermore, it was found in this work that the RAPc8 enzyme has significant activity on 3-cyanopyridine (a starting material for production of nicotinamide). Thus it would be of value to determine kinetic data for this hydrolysis in order to establish whether the enzyme may have industrial potential in this valuable biotransformation.

Drawing on the data generated in the cloning, expression and characterisation studies, the research diverged into two separate investigations. The aim of the first investigation (presented in Chapter 5) was to address the problem of the enzyme's lack of activity towards (homo)-aromatic nitriles. A rational design approach was taken to address this problem, where site-specific mutations were used to replace aromatic residues within the active site cavity (identified through homology modelling). Although the overall aim (to generate a mutant with detectable activity on benzonitrile) was not realised in this research, the results of the preliminary characterisation of the mutants seem to support the aromatic interaction hypothesis of benzonitrile inhibition. Residue βW76 has been implicated in this interaction. Before proceeding with further mutagenic studies (which may include the generation of double or triple mutants of the key aromatic residues), it would be most useful to perform a more detailed characterisation of the benzonitrile inhibition phenomenon, both of the wild type enzyme and of the relevant mutants. Such characterisations should include detailed studies of the kinetics of inhibition and of differences in the absorbance spectra in the presence and absence of benzonitrile for wild type and mutant enzymes.

There exist a number of alternative routes to achieving the aim of engineering a thermostable NHase active on homoaromatic nitriles. Such routes include directed evolution approaches such as gene shuffling and error prone PCR, exposure of the native parent strain to mutagens under selective pressure (for example utilisation of the nitrile as sole carbon or nitrogen source) or even the engineering of a mesophilic aromatic-specific NHase for enhanced thermal stability.
The final section of this thesis describes the development of culture-independent techniques and the preliminary investigations aimed at the isolation of novel NHase genes using both PCR based and plasmid based cloning techniques. Unfortunately, despite the relatively large size of the environmental library produced in this work, the complementation based screening did not identify any recombinants capable of growth on the media used. It is possible that NHase genes present within the library were missed using this approach and that more powerful and sensitive techniques such as DNA hybridisation may be more suitable for future screening of this or similar libraries.

The multiplex PCR procedure has proved a sensitive method for the detection of NHase genes within environmental samples. Whilst the investigation was not of a scale large enough to enable significant quantitative observations regarding the distribution of NHase groups within the sample tested, tentative qualitative interpretation of the results supports previous statements within this thesis that the thermostable NHases all belong to the Co-type family and that the *Bacillus* genus dominates the community of thermophilic NHase producing microorganisms. Although the procedure does not yield full-length NHase genes, it may be possible to adapt the experiments, using semi-nested PCR in order to isolate full-length genes.
References


204
References


References


References


208


http://info.med.yale.edu/mbb/koelle/protocols/protocol_degenerate_PCR.htm


References


References


References


References


Appendices

Appendix I Standard curves

Li: BioRad Bradford assay standard curve

\[ y = -0.2869x^2 + 1.2321x + 0.0101 \]
\[ R^2 = 0.9998 \]

I.ii: Ammonia assay standard curve

\[ y = 1.3728x \]
\[ R^2 = 0.9876 \]
Appendix II Blast searches

II. BLAST search with the 1427bp PCR fragment of Bacillus sp. RAPc8

16SrDNA

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Query</th>
<th>gi</th>
<th>12957158</th>
<th>dbj</th>
<th>AB040122.1</th>
<th>AB040122 Bacillus sp. B-3 gene f...</th>
<th>2829</th>
<th>0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>131101</td>
<td>emb</td>
<td>226930.1</td>
<td>BF165RRAA B.pallidus gene for 16S r...</td>
<td>2813</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>131101</td>
<td>emb</td>
<td>226929.1</td>
<td>BF165RRAA Bacillus sp. gene for 16S ...</td>
<td>2795</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>511106</td>
<td>emb</td>
<td>226931.1</td>
<td>BT16SRNC B.thermoalkalophilus gene...</td>
<td>2680</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>8575827</td>
<td>gb</td>
<td>AF276306.1</td>
<td>AF276306 Geobacillus subterraneus ...</td>
<td>2189</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>1350045</td>
<td>gb</td>
<td>AF391974.1</td>
<td>AF391974 Thermal soil bacterium Y...</td>
<td>2187</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>8575828</td>
<td>gb</td>
<td>AF276307.1</td>
<td>AF276307 Geobacillus subterraneus ...</td>
<td>2181</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>13624837</td>
<td>emb</td>
<td>AJ306648.1</td>
<td>16SSP306G 48 Uncultured Saccharococ...</td>
<td>2167</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>720954</td>
<td>dbj</td>
<td>AB021197.1</td>
<td>AB021197 Bacillus thermoglucosida...</td>
<td>2155</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>1402727</td>
<td>gb</td>
<td>U59630.1</td>
<td>B.thermoglucosidicus gene for 16S ...</td>
<td>2139</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>213977</td>
<td>emb</td>
<td>226937.1</td>
<td>BDS16SRRNB B.dentihominis gene for ...</td>
<td>2139</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>5006251</td>
<td>dbj</td>
<td>AB028234.1</td>
<td>AB028234 Bacillus thermodenitrificans...</td>
<td>2129</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>535100</td>
<td>emb</td>
<td>X70430.1</td>
<td>ST16SR S.thermosphaera gene for 16S ...</td>
<td>2123</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>397920</td>
<td>gb</td>
<td>LS5227.1</td>
<td>LS5227 Saccharococcus thermostilb...</td>
<td>2123</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

Alignments

>gi|12957158|dbj|AB040122.1|AB040122 Bacillus sp. B-3 gene for 165 rRNA, partial sequence

Length = 1474

Score = 2829 bits (1427), Expect = 0.0

Identities = 1427/1427 (100%)

Strand = Plus / Plus

Query: 1 ggcggcggtcttaaatcactcaagtctcgagcgcccagggactgtgtctcatttaggta 60

| Subject: 9 gcgggcgggtcttaaatcactcaagtctcgagcgcccagggactgtgtctcatttaggta 68 |
|---|---|

Query: 61 ggcggcgccggcgtgtaaascagtggcaccctctgacctgattgataactctggaga 120

| Subject: 69 gcgggcggcggcgtgtaaascagtggcaccctctgacctgattgataactctggaga 128 |
|---|---|

Query: 121 aacccgagtactataccgggataascaccaaaaccgcatctgttttctgttggattagggctt 180

| Subject: 129 aacccgagtactataccgggataascaccaaaaccgcatctgttttctgttggattagggctt 188 |
|---|---|

Query: 181 ttagcgtctacgctcaggtgaggcccgccgctcattagctgttggattgtaaccgccct 240

| Subject: 189 ttagcgtctacgctcaggtgaggcccgccgctcattagctgttggattgtaaccgccct 248 |
|---|---|

Query: 241 ccaacggacacagtctggcagtaccgaggctgaggggtgactgccccactggactgacac 300

| Subject: 249 ccaacggacacagtctggcagtaccgaggctgaggggtgactgccccactggactgacac 308 |
|---|---|

Query: 301 ggcgccgactttcctacggagcagccaggattgaaatcctccgcaatagcgagaaatgtcgc 360

| Subject: 309 ggcgccgactttcctacggagcagccaggattgaaatcctccgcaatagcgagaaatgtcgc 368 |
|---|---|

Query: 361 ggcgccgactttcctacggagcagccaggattgaaatcctccgcaatagcgagaaatgtcgc 420

| Subject: 369 ggcgccgactttcctacggagcagccaggattgaaatcctccgcaatagcgagaaatgtcgc 428 |
|---|---|

Query: 421 aacacgtcgctggcactagcgccagttctgtctgtcactggtcaggggtgactgccccactggactgacac 480

| Subject: 429 aacacgtcgctggcactagcgccagttctgtctgtcactggtcaggggtgactgccccactggactgacac 488 |
|---|---|
Query: 481 actacgtgccccgcggtatcactgtaaggtgcaagtgttcgcggaaattatggg 540
Sbjct: 489 actacgtgccccgcggtatcactgtaaggtgcaagtgttcgcggaaattatggg 548

Query: 541 gtaaagcgcgagggctttcctaaacgctgcctaatcactggtgcggcagcagggag 600
Sbjct: 549 gtaaagcgcgagggctttcctaaacgctgcctaatcactggtgcggcagcagggag 608

Query: 601 cgccattggaaacctggggacttggtcagtcagaggagggagccagatccacgcgttagc 660
Sbjct: 609 cgccattggaaacctggggacttggtcagtcagaggagggagccagatccacgcgttagc 668

Query: 661 gttgaattcggtagagtggagagggagacacgcgtctgtctgcgttac 720
Sbjct: 669 gttgaattcggtagagtggagagggagacacgcgtctgtctgcgttac 728

Query: 721 ctagcgtgagggcgaagcgctgtgagagaatgctatccggtatc 780
Sbjct: 729 ctagcgtgagggcgaagcgctgtgagagaatgctatccggtatc 788

Query: 781 cgtaaagaagttgatctgtagtcagagttggttagctaccccttttagcgtaagcaaaccgc 840
Sbjct: 789 cgtaaagaagttgatctgtagtcagagttggttagctaccccttttagcgtaagcaaaccgc 848

Query: 841 attaagcactccctgggagatcgccgcaaggcttaacaaagttggagggg 900
Sbjct: 849 attaagcactccctgggagatcgccgcaaggcttaacaaagttggagggg 908

Query: 901 accgcccaacgcgtgagcctttgtttataagctacacacccagcaggg 960
Sbjct: 909 accgcccaacgcgtgagcctttgtttataagctacacacccagcaggg 968

Query: 961 ttcggacatcccctgcaccaaccccttagagataggccccgttcctcgggacaggtgaca 1020
Sbjct: 969 ttcggacatcccctgcaccaaccccttagagataggccccgttcctcgggacaggtgaca 1028

Query: 1021 gttgctagtctcgtgtctgtagtctgtattgtgtcagtttggttaagtgctc 1080
Sbjct: 1029 gttgctagtctcgtgtctgtagtctgtattgtgtcagtttggttaagtgctc 1088

Query: 1081 cgccacctttgacccggatttgagctacactttcgggacctgtgacgtcggga 1140
Sbjct: 1089 cgccacctttgacccggatttgagctacactttcgggacctgtgacgtcggga 1148

Query: 1141 aaagtccagggagaggctgttgggagctaatcattatgcctgcctagttg 1200
Sbjct: 1149 aaagtccagggagaggctgttgggagctaatcattatgcctgcctagttg 1208

Query: 1201 cacgtgcctactggtgggtcaacagccccagcagggagggctgtggagc 1260
Sbjct: 1209 cacgtgcctactggtgggtcaacagccccagcagggagggctgtggagc 1268

Query: 1261 aacacaccttcagctcgtgatagctacagctgtcctagtgacagccagaaatc 1320
Sbjct: 1269 aacacaccttcagctcgtgatagctacagctgtcctagtgacagccagaaatc 1328

Query: 1321 taacgcgcaatccgcatcgcgctgtggacagctgttgccttgtaaccacccgcgtc 1380
Sbjct: 1329 taacgcgcaatccgcatcgcgctgtggacagctgttgccttgtaaccacccgcgtc 1388

Query: 1381 acacacgcagagttgtttaacaccccgagcgctgtggataacaccctc 1427
Sbjct: 1389 acacacgcagagttgtttaacaccccgagcgctgtggataacaccctc 1435

Appendices
### Appendices

#### II.ii: Translated BLAST search with the ORF1 (amidase) amino acid sequence

<table>
<thead>
<tr>
<th>Score (bits)</th>
<th>E Value</th>
<th>Sequence producing significant alignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>699</td>
<td>0.0</td>
<td>gi[1739759]gb</td>
</tr>
<tr>
<td>641</td>
<td>0.0</td>
<td>gi[16492287]gb</td>
</tr>
<tr>
<td>589</td>
<td>e-167</td>
<td>gi[1776081]gb</td>
</tr>
<tr>
<td>588</td>
<td>e-167</td>
<td>gi[15598562]ref</td>
</tr>
<tr>
<td>586</td>
<td>e-166</td>
<td>gi[1505961]gb</td>
</tr>
<tr>
<td>572</td>
<td>e-162</td>
<td>gi[1456198]sp</td>
</tr>
<tr>
<td>547</td>
<td>e-155</td>
<td>gi[1564492]ref</td>
</tr>
<tr>
<td>543</td>
<td>e-154</td>
<td>gi[1557915]emb</td>
</tr>
<tr>
<td>182</td>
<td>4e-45</td>
<td>gi[15645852]ref</td>
</tr>
</tbody>
</table>

#### Alignments

<table>
<thead>
<tr>
<th>Score (bits)</th>
<th>E Value</th>
<th>Query</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>699</td>
<td>0.0</td>
<td>gi[1739759]gb</td>
<td>AAF69000.1</td>
</tr>
<tr>
<td>641</td>
<td>0.0</td>
<td>gi[16492287]gb</td>
<td>AAF14257.1</td>
</tr>
<tr>
<td>589</td>
<td>e-167</td>
<td>gi[1776081]gb</td>
<td>pei[12674]</td>
</tr>
<tr>
<td>588</td>
<td>e-167</td>
<td>gi[15598562]ref</td>
<td>NP_252056.1</td>
</tr>
<tr>
<td>586</td>
<td>e-166</td>
<td>gi[1505961]gb</td>
<td>AA255697.1</td>
</tr>
<tr>
<td>572</td>
<td>e-162</td>
<td>gi[1456198]sp</td>
<td>NC_01360</td>
</tr>
<tr>
<td>547</td>
<td>e-155</td>
<td>gi[1564492]ref</td>
<td>NP_207092.1</td>
</tr>
<tr>
<td>543</td>
<td>e-154</td>
<td>gi[1557915]emb</td>
<td>CAF72332.1</td>
</tr>
<tr>
<td>182</td>
<td>4e-45</td>
<td>gi[15645852]ref</td>
<td>NP_208030.1</td>
</tr>
</tbody>
</table>
Appendices

Query: 301 PFDFYRTWLDACKERENKVIKRTIVTGCPTAPIQGIPNEHGKTEIGV 348
PPFDRYWVDAEKACRENEIKETRTSGCTAECPIQGIPNEHGKTEAV 348
Sbjct: 301 PFDFYRTWLDACKERENKVIKRTIVTGCPTAPIQGIPNEHGKTE 348

Score = 589 bits (1518), Expect = e-167
Identities = 269/345 (77%), Positives = 301/345 (86%)

Query: 1 MRHGDSHHSDTGVAVVVNNPKRLHTKAEVITNIAKADAMVQKMGLQPDLLVFPEY 60
MRHGDSHS+DTVG+AVVVKNPKRLHTAEV++NA++KA+DMVQKMGLQPDLLVFPEY
Sbjct: 1 MRHGDSHHSDTGVAVVVNNPKRLHTKAEVITNIAKADAMVQKMGLQPDLLVFPEY 60

Query: 61 STMGMYQDMPFATAASPIGEATAAEACKDADTWVSLITGEKEDMHPKAPNTLV 120
S GMHY EM TA +IGEAIF A+CKA+ WVSFLTGE+H+HP KAPNTLV
Sbjct: 61 SLQGMYQDMDPFAEAVAPFGEATIAPFTRKANWVSFLTGHEEHPKAPNTLV 120

Query: 121 LNNKGEIVQYKRIIIPWCPQIMYPKQQMMKAMMKAWANNYCTTVANAGFDGSYSFGHHAIFFGD 240
LNNKGEIVQYKRIIIPWCPQIMYPKQQMMKAMMKAWANNYCTTVANAGFDGSYSFGHHAIFFGD
Sbjct: 121 LNNKGEIVQYKRIIIPWCPQIMYPKQQMMKAMMKAWANNYCTTVANAGFDGSYSFGHHAIFFGD 240

Query: 241 RTLGECGTEENGIQYAEVSISQIRDFKNAQSHNLKILHRGTYCLNSGEGDRAFAC 300
RTLGECGTEENGIQYAEVSISQIRDFKNAQSHNLKILHRGTYCLNSGEGDRAFAC
Sbjct: 241 RTLGECGTEENGIQYAEVSISQIRDFKNAQSHNLKILHRGTYCLNSGEGDRAFAC 300

Query: 301 PFEFYRTWLDACKERENKVIKRTIVTGCPTAPIQGIPNEHGKTE 348
PPFDRYWVDAEKACRENEIKETRTSGCTAECPIQGIPNEHGKTEAV 348
Sbjct: 301 PFEFYRTWLDACKERENKVIKRTIVTGCPTAPIQGIPNEHGKTE 348

Score = 588 bits (1515), Expect = e-167
Identities = 269/345 (77%), Positives = 301/345 (86%)

Query: 1 MRHGDSHHSDTGVAVVVNNPKRLHTKAEVITNIAKADAMVQKMGLQPDLLVFPEY 60
MRHGDSHS+DTVG+AVVVKNPKRLHTAEV++NA++KA+DMVQKMGLQPDLLVFPEY
Sbjct: 1 MRHGDSHHSDTGVAVVVNNPKRLHTKAEVITNIAKADAMVQKMGLQPDLLVFPEY 60

Query: 61 STMGMYQDMPFATAASPIGEATAAEACKDADTWVSLITGEKEDMHPKAPNTLV 120
S GMHY EM TA +IGEAIF A+CKA+ WVSFLTGE+H+HP KAPNTLV
Sbjct: 61 SLQGMYQDMDPFAEAVAPFGEATIAPFTRKANWVSFLTGHEEHPKAPNTLV 120

Query: 121 LNNKGEIVQYKRIIIPWCPQIMYPKQQMMKAMMKAWANNYCTTVANAGFDGSYSFGHHAIFFGD 240
LNNKGEIVQYKRIIIPWCPQIMYPKQQMMKAMMKAWANNYCTTVANAGFDGSYSFGHHAIFFGD
Sbjct: 121 LNNKGEIVQYKRIIIPWCPQIMYPKQQMMKAMMKAWANNYCTTVANAGFDGSYSFGHHAIFFGD 240

Query: 241 RTLGECGTEENGIQYAEVSISQIRDFKNAQSHNLKILHRGTYCLNSGEGDRAFAC 300
RTLGECGTEENGIQYAEVSISQIRDFKNAQSHNLKILHRGTYCLNSGEGDRAFAC
Sbjct: 241 RTLGECGTEENGIQYAEVSISQIRDFKNAQSHNLKILHRGTYCLNSGEGDRAFAC 300

Query: 301 PFEFYRTWLDACKERENKVIKRTIVTGCPTAPIQGIPNEHGKTE 348
PPFDRYWVDAEKACRENEIKETRTSGCTAECPIQGIPNEHGKTEAV 348
Sbjct: 301 PFEFYRTWLDACKERENKVIKRTIVTGCPTAPIQGIPNEHGKTE 348

>gi|177606|pir|A26741 amida [EC 3.5.1.4], aliphatic - Pseudomonas aeruginosa
Length = 346

Score = 589 bits (1518), Expect = e-167
Identities = 269/345 (77%), Positives = 301/345 (86%)

Query: 1 MRHGDSHHSDTGVAVVVNNPKRLHTKAEVITNIAKADAMVQKMGLQPDLLVFPEY 60
MRHGDSHS+DTVG+AVVVKNPKRLHTAEV++NA++KA+DMVQKMGLQPDLLVFPEY
Sbjct: 1 MRHGDSHHSDTGVAVVVNNPKRLHTKAEVITNIAKADAMVQKMGLQPDLLVFPEY 60

Query: 61 STMGMYQDMPFATAASPIGEATAAEACKDADTWVSLITGEKEDMHPKAPNTLV 120
S GMHY EM TA +IGEAIF A+CKA+ WVSFLTGE+H+HP KAPNTLV
Sbjct: 61 SLQGMYQDMDPFAEAVAPFGEATIAPFTRKANWVSFLTGHEEHPKAPNTLV 120

Query: 121 LNNKGEIVQYKRIIIPWCPQIMYPKQQMMKAMMKAWANNYCTTVANAGFDGSYSFGHHAIFFGD 240
LNNKGEIVQYKRIIIPWCPQIMYPKQQMMKAMMKAWANNYCTTVANAGFDGSYSFGHHAIFFGD
Sbjct: 121 LNNKGEIVQYKRIIIPWCPQIMYPKQQMMKAMMKAWANNYCTTVANAGFDGSYSFGHHAIFFGD 240

Query: 241 RTLGECGTEENGIQYAEVSISQIRDFKNAQSHNLKILHRGTYCLNSGEGDRAFAC 300
RTLGECGTEENGIQYAEVSISQIRDFKNAQSHNLKILHRGTYCLNSGEGDRAFAC
Sbjct: 241 RTLGECGTEENGIQYAEVSISQIRDFKNAQSHNLKILHRGTYCLNSGEGDRAFAC 300

Query: 301 PFEFYRTWLDACKERENKVIKRTIVTGCPTAPIQGIPNEHGKTE 348
PPFDRYWVDAEKACRENEIKETRTSGCTAECPIQGIPNEHGKTEAV 348
Sbjct: 301 PFEFYRTWLDACKERENKVIKRTIVTGCPTAPIQGIPNEHGKTE 348

>gi|15599562|ref|NP_252056.1 (NC 002516) aliphatic amidaase [Pseudomonas aeruginosa]
gi|128200873|sp|PI144391|ALAM_FPEAE Aliphatic amidaase
gi|11347411|pir|HE3222 aliphatic amidaase PA3366 [imported] - Pseudomonas aeruginosa
[strain PA01]
gi|9945051|gb|AAG0754.1|AE004759_1 (AE004759) aliphatic amidaase [Pseudomonas aeruginosa]
Length = 346

Score = 588 bits (1515), Expect = e-167
Identities = 269/345 (77%), Positives = 301/345 (86%)

Query: 1 MRHGDSHHSDTGVAVVVNNPKRLHTKAEVITNIAKADAMVQKMGLQPDLLVFPEY 60
MRHGDSHS+DTVG+AVVVKNPKRLHTAEV++NA++KA+DMVQKMGLQPDLLVFPEY
Sbjct: 1 MRHGDSHHSDTGVAVVVNNPKRLHTKAEVITNIAKADAMVQKMGLQPDLLVFPEY 60

Query: 61 STMGMYQDMPFATAASPIGEATAAEACKDADTWVSLITGEKEDMHPKAPNTLV 120
S GMHY EM TA +IGEAIF A+CKA+ WVSFLTGE+H+HP KAPNTLV
Sbjct: 61 SLQGMYQDMDPFAEAVAPFGEATIAPFTRKANWVSFLTGHEEHPKAPNTLV 120

Query: 121 LNNKGEIVQYKRIIIPWCPQIMYPKQQMMKAMMKAWANNYCTTVANAGFDGSYSFGHHAIFFGD 240
LNNKGEIVQYKRIIIPWCPQIMYPKQQMMKAMMKAWANNYCTTVANAGFDGSYSFGHHAIFFGD
Sbjct: 121 LNNKGEIVQYKRIIIPWCPQIMYPKQQMMKAMMKAWANNYCTTVANAGFDGSYSFGHHAIFFGD 240

Query: 241 RTLGECGTEENGIQYAEVSISQIRDFKNAQSHNLKILHRGTYCLNSGEGDRAFAC 300
RTLGECGTEENGIQYAEVSISQIRDFKNAQSHNLKILHRGTYCLNSGEGDRAFAC
Sbjct: 241 RTLGECGTEENGIQYAEVSISQIRDFKNAQSHNLKILHRGTYCLNSGEGDRAFAC 300

Query: 301 PFEFYRTWLDACKERENKVIKRTIVTGCPTAPIQGIPNEHGKTE 348
PPFDRYWVDAEKACRENEIKETRTSGCTAECPIQGIPNEHGKTEAV 348
Sbjct: 301 PFEFYRTWLDACKERENKVIKRTIVTGCPTAPIQGIPNEHGKTE 348
II.iii: Translated BLAST search with the ORF2 (NHase β) sequence

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Query</th>
<th>Subject</th>
<th>Score</th>
<th>E (bits) Value</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>gb</td>
<td>AA69001.1</td>
<td>AF257486 1 (AF257488) nitrile hydratase beta ...</td>
<td>458 e-128</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>Q53117</td>
<td>NHAB_RHOGO NITRILE HYDRATASE SUBUNIT BETA (NITRILASE)</td>
<td>154 e-37</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P29379</td>
<td>NHB_RHOR LOW-MOLECULAR WEIGHT COBALT-CONTAINING...</td>
<td>150 e-36</td>
<td></td>
</tr>
<tr>
<td>ref</td>
<td>NP 38621.1</td>
<td>(NC 003047) PROBABLE NITRILE HYDRATASE SUB...</td>
<td>149 e-35</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P97052</td>
<td>NHAB_PSEP NITRILE HYDRATASE SUBUNIT BETA (NITRILASE)</td>
<td>145 e-34</td>
<td></td>
</tr>
<tr>
<td>piri</td>
<td>SI5799</td>
<td>nitrile hydratase (EC 4.2.1.64) beta chain, H-...</td>
<td>132 e-30</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>Q11220</td>
<td>NHAB_RHOR HIGH-MOLECULAR WEIGHT COBALT-CONTAININ...</td>
<td>132 e-30</td>
<td></td>
</tr>
<tr>
<td>piri</td>
<td>S15416</td>
<td>nitrile hydratase (EC 4.2.1.64) chain beta - Rh...</td>
<td>99 e-20</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P3449</td>
<td>NHAB_RHOER NITRILE HYDRATASE SUBUNIT BETA (NITRILASE)</td>
<td>97 e-20</td>
<td></td>
</tr>
<tr>
<td>pdb</td>
<td>IAKJB</td>
<td>Chain B, Nitrilase -gi</td>
<td>314475</td>
<td>pdb</td>
</tr>
<tr>
<td>piri</td>
<td>N00370</td>
<td>nitrile hydratase (EC 4.2.1.64) beta chain - Rh...</td>
<td>96 e-19</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P7763</td>
<td>Nitrile hydratase subunit beta (Nitrilase)</td>
<td>91 e-18</td>
<td></td>
</tr>
<tr>
<td>dbj</td>
<td>BAA28227.1</td>
<td>(AB007999) thiocyanate hydrolase alpha subu...</td>
<td>37 0.11</td>
<td></td>
</tr>
<tr>
<td>dbj</td>
<td>BAA11045.1</td>
<td>(D67027) nitrilase-b homologue [Rhodococcus s...</td>
<td>33 1.7</td>
<td></td>
</tr>
<tr>
<td>piri</td>
<td>SI1772</td>
<td>nitrile hydratase - Rhodococcus rhodochrous</td>
<td>33 2.6</td>
<td></td>
</tr>
<tr>
<td>ref</td>
<td>INF 386212.1</td>
<td>(NC_003047) PROBABLE NITRILASE SUBUNIT BETA ...</td>
<td>32 3.4</td>
<td></td>
</tr>
</tbody>
</table>

Alignments

>gb|AA69001.1|AF257486 1 (AF257488) nitrile hydratase beta subunit [Bacillus sp. BR449] Length - 229

Score = 458 bits (117%), Expect = e-128
Identification = 226/229 (99%), Positives = 229/229 (99%)

Query: 1 MNGIHDVGGMDGFGKVMYVKEEEDIYFTHDWERLAFGLVAGCMAQGLGMKAFDEF RIE 60
Subject: 1 MNGIHDVGGMDGFGKVMYVKEEEDIYFTHDWERLAFGLVAGCMAQGLGMKAFDEF RIE 60

Query: 121 IVEKALYGLGSPREGASPRFKVGERKTNKINHPTGTRFFPRVPARDKYQVGYIDE YGAHV 180
Subject: 121 IVEKALYGLGSPREGASPRFKVGERKTNKINHPTGTRFFPRVPARDKYQVGYIDE YGAHV 180

Query: 101 FPDDAHRKGENPQLYLRVFEAEEILGQKQDSDYIDLWSEMPVFV 229
Subject: 101 FPDDAHRKGENPQLYLRVFEAEEILGQKQDSDYIDLWSEMPVFV 229

>sp|P33117|NHAB_RHOGO NITRILE HYDRATASE SUBUNIT BETA (NITRILASE) (NHASE) piri|BI41327|nitrile hydratase (EC 4.2.1.64) beta chain - Rhodococcus sp gb|AA69001.1|AF257486 1 (AF257488) Nitrilase beta subunit [Rhodococcus sp.] Length - 235

Score = 154 bits (38%), Expect = e-37
Identification = 87/232 (37%), Positives = 125/232 (53%), Gaps = 8/232 (3%)

Query: 1 MNGIHDVGGMDGFGKVMYVKEEEDIYFTHDWERLAFGLVAGCMAQGLGMKAFDEF RIE 60
Subject: 1 MNGIHDVGGMDGFGKVMYVKEEEDIYFTHDWERLAFGLVAGCMAQGLGMKAFDEF RIE 60

Query: 120 KLVKALYGLGSPREIASPRFKVGERKTNKINHPTGTRFFPRVPARDKYQVGYIDE YGAHV 179
Subject: 120 KLVKALYGLGSPREIASPRFKVGERKTNKINHPTGTRFFPRVPARDKYQVGYIDE YGAHV 179

Query: 180 VFDDAHRKGENPQLYLRVFEAEEILGQKQDSDYIDLWSEMPVFV 228
Subject: 180 VFDDAHRKGENPQLYLRVFEAEEILGQKQDSDYIDLWSEMPVFV 228
Appendices

>sp|P29379|NH22_RHOR Lowe-MOLECULAR WEIGHT COBALT-CONTAINING NITRILE HYDRATASE SUBUNIT BETA (L-NITRILASE) (L-NHASE)

pir|S91712| nitrile hydratase (EC 4.2.1.84) beta chain, L-type - Rhodococcus rhodochrous (strain J1)

emb|CAA45711.11 (X64360) nitrile hydratase [Rhodococcus rhodochrous]

Length = 226

Score = 150 bits (379), Expect = 8e-36
Identities = 85/229 (37%), Positives = 126/229 (54%), Gaps = 7/229 (3%)

Query: MNGIHDVGGMDGFGKVMYVKEEEDIYFTHDWERLAGVACMAQLQGKADFREFRIGE 60
Subject: MDGIHD+GG G+G + E ++ F DWER + MA G DE R E

Query: IMRVPVDYITSSYGHMITAVANYGVTDOLEDKELDTEVFLLKPDVTKIRREDFAVLYK 120
Subject: QIPFHDIYSYEHHHAMAHIKEAEQIFDSDELDRTQYMDHPSDPTTPQLPQOLVE

Score = 149 bits (375), Expect = 2e-35
Identities = 92/227 (40%), Positives = 122/227 (53%), Gaps = 12/227 (5%)

Query: VFPDDAAHRKGENPQYLYRVRFAEELWGYQDSV -- YIDLWESYMPP 226
Subject: VFPD A GE+P++LY VR + ++ E + Y P

Score = 145 bits (367), Expect = 2e-34
Identities = 90/230 (39%), Positives = 123/230 (53%), Gaps = 16/230 (6%)

Query: IFPRPVG4DQGFGKVMYVKEEEDIYFTHDWERLAGVACMAQLQGKADFREFRIGE 58
Subject: MNGD+GG GFG V EE++ YF +WE+ A G L AG M G W E R E

Score = 150 bits (379), Expect = 8e-36
Identities = 85/229 (37%), Positives = 126/229 (54%), Gaps = 7/229 (3%)

Query: IMRVPVDYITSSYGHM ITAVANYGVTDOLEDKELDTEVFLLKPDVTKIRREDFAVLYK 120
Subject: QIPFHDIYSYEHHHAMAHIKEAEQIFDSDELDRTQYMDHPSDPTTPQLPQOLVE

Score = 149 bits (375), Expect = 2e-35
Identities = 92/227 (40%), Positives = 122/227 (53%), Gaps = 12/227 (5%)

Query: VFPDDAAHRKGENPQYLYRVRFAEELWGYQDSV -- YIDLWESYMPP 226
Subject: VFPD A GE+P++LY VR + ++ E + Y P

Score = 145 bits (367), Expect = 2e-34
Identities = 90/230 (39%), Positives = 123/230 (53%), Gaps = 16/230 (6%)

Query: IFPRPVG4DQGFGKVMYVKEEEDIYFTHDWERLAGVACMAQLQGKADFREFRIGE 58
Subject: MNGD+GG GFG V EE++ YF +WE+ A G L AG M G W E R E

>ref|NP_366212.1 (NC_003047) PROBABLE NITRILE HYDRATASE SUBUNIT BETA PROTEIN [Sinorhizobium mellotii]

emb|CAC46665.1 (AL591789) PROBABLE NITRILE HYDRATASE SUBUNIT BETA PROTEIN [Sinorhizobium mellotii]

Length = 219

Score = 149 bits (375), Expect = 2e-35
Identities = 92/227 (40%), Positives = 122/227 (53%), Gaps = 12/227 (5%)

Query: VFPDDAAHRKGENPQYLYRVRFAEELWGYQDSV -- YIDLWESYMPP 226
Subject: VFPD A GE+P++LY VR + ++ E + Y P

Score = 145 bits (367), Expect = 2e-34
Identities = 90/230 (39%), Positives = 123/230 (53%), Gaps = 16/230 (6%)

Query: IFPRPVG4DQGFGKVMYVKEEEDIYFTHDWERLAGVACMAQLQGKADFREFRIGE 58
Subject: MNGD+GG GFG V EE++ YF +WE+ A G L AG M G W E R E

>ref|NP_103216.1 (NC_002678) nitrile hydratase beta subunit [Mesorhizobium loti]

emb|BAB48802.1 (AP002997) nitrile hydratase beta subunit [Mesorhizobium loti]

Length = 219

Score = 149 bits (375), Expect = 2e-35
Identities = 92/227 (40%), Positives = 122/227 (53%), Gaps = 12/227 (5%)

Query: VFPDDAAHRKGENPQYLYRVRFAEELWGYQDSV -- YIDLWESYMPP 226
Subject: VFPD A GE+P++LY VR + ++ E + Y P

Score = 145 bits (367), Expect = 2e-34
Identities = 90/230 (39%), Positives = 123/230 (53%), Gaps = 16/230 (6%)

Query: IFPRPVG4DQGFGKVMYVKEEEDIYFTHDWERLAGVACMAQLQGKADFREFRIGE 58
Subject: MNGD+GG GFG V EE++ YF +WE+ A G L AG M G W E R E

>ref|NP_103216.1 (NC_002678) nitrile hydratase beta subunit [Mesorhizobium loti]

emb|BAB48802.1 (AP002997) nitrile hydratase beta subunit [Mesorhizobium loti]

Length = 219

Score = 149 bits (375), Expect = 2e-35
Identities = 92/227 (40%), Positives = 122/227 (53%), Gaps = 12/227 (5%)

Query: VFPDDAAHRKGENPQYLYRVRFAEELWGYQDSV -- YIDLWESYMPP 226
Subject: VFPD A GE+P++LY VR + ++ E + Y P

Score = 145 bits (367), Expect = 2e-34
Identities = 90/230 (39%), Positives = 123/230 (53%), Gaps = 16/230 (6%)

Query: IFPRPVG4DQGFGKVMYVKEEEDIYFTHDWERLAGVACMAQLQGKADFREFRIGE 58
Subject: MNGD+GG GFG V EE++ YF +WE+ A G L AG M G W E R E

221
## II.iv: Translated BLAST search with the ORF3 (NHase α) sequence

### Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Score</th>
<th>E Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>398</td>
<td>e-110</td>
<td></td>
</tr>
<tr>
<td>247</td>
<td>6e-65</td>
<td></td>
</tr>
<tr>
<td>246</td>
<td>1e-64</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>1e-62</td>
<td></td>
</tr>
<tr>
<td>228</td>
<td>2e-59</td>
<td></td>
</tr>
<tr>
<td>222</td>
<td>2e-57</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>2e-64</td>
<td></td>
</tr>
<tr>
<td>237</td>
<td>6e-62</td>
<td></td>
</tr>
<tr>
<td>228</td>
<td>2e-59</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>2e-64</td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>237</td>
<td>6e-62</td>
<td></td>
</tr>
<tr>
<td>228</td>
<td>2e-59</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>2e-64</td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>237</td>
<td>6e-62</td>
<td></td>
</tr>
<tr>
<td>228</td>
<td>2e-59</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>2e-64</td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>237</td>
<td>6e-62</td>
<td></td>
</tr>
<tr>
<td>228</td>
<td>2e-59</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>2e-64</td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>237</td>
<td>6e-62</td>
<td></td>
</tr>
<tr>
<td>228</td>
<td>2e-59</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>2e-64</td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>237</td>
<td>6e-62</td>
<td></td>
</tr>
<tr>
<td>228</td>
<td>2e-59</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>2e-64</td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>237</td>
<td>6e-62</td>
<td></td>
</tr>
<tr>
<td>228</td>
<td>2e-59</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>2e-64</td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>3e-62</td>
<td></td>
</tr>
<tr>
<td>237</td>
<td>6e-62</td>
<td></td>
</tr>
<tr>
<td>228</td>
<td>2e-59</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>3e-62</td>
<td></td>
</tr>
</tbody>
</table>

### Alignments

<table>
<thead>
<tr>
<th>Query</th>
<th>Score</th>
<th>E Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gbl</td>
<td>AA69002.1</td>
<td>AF257489</td>
<td>398</td>
</tr>
<tr>
<td>sp</td>
<td>P29378</td>
<td>NHA2 RHORH LOW-MOLECULAR WEIGHT COBALT-CONTAINING NITRILE HYDRATASE SUBUNIT ALPHA (NHase α)</td>
<td>247</td>
</tr>
<tr>
<td>sp</td>
<td>Q53118</td>
<td>NHA4 RHOSO NITRILE HYDRATASE SUBUNIT ALPHA (NHase α)</td>
<td>246</td>
</tr>
<tr>
<td>pim</td>
<td>L54800</td>
<td>nitrile hydratase (EC 4.2.1.84) alpha chain</td>
<td>239</td>
</tr>
<tr>
<td>sp</td>
<td>P21719</td>
<td>NHA2 RHORH HIGH-MOLECULAR WEIGHT COBALT-CONTAINING NITRILE HYDRATASE SUBUNIT ALPHA (NHase α)</td>
<td>237</td>
</tr>
<tr>
<td>ref</td>
<td>NP 36023.1</td>
<td>NITRILE HYDRATASE SUBUNIT ALPHA (NHase α)</td>
<td>228</td>
</tr>
<tr>
<td>pdb</td>
<td>1AUHJA</td>
<td>Chain A, Nitrile Hydratase</td>
<td>199</td>
</tr>
<tr>
<td>sp</td>
<td>P13448</td>
<td>NITRILE HYDRATASE SUBUNIT ALPHA (NHase α)</td>
<td>197</td>
</tr>
<tr>
<td>gbl</td>
<td>AA26478.1</td>
<td>nitrile hydratase alpha subunit, NHase alpha</td>
<td>197</td>
</tr>
<tr>
<td>pim</td>
<td>ZM7065</td>
<td>NITRILE HYDRATASE (EC 4.2.1.84) alpha chain</td>
<td>179</td>
</tr>
<tr>
<td>pdb</td>
<td>1BBA2</td>
<td>NITRILE HYDRATASE SUBUNIT ALPHA (NHase α)</td>
<td>174</td>
</tr>
<tr>
<td>pdb</td>
<td>1BBA2</td>
<td>NITRILE HYDRATASE SUBUNIT ALPHA (NHase α)</td>
<td>174</td>
</tr>
</tbody>
</table>

### Score: 398 bits (1023), Expect: e-110

- Identities: 210/213 (98%)
- Positives: 211/213 (98%)

### Query: 1 MTIDQKNTNIDPRFPHHHPRPQSFWEARAKALESLLIEKGHLSSDAIERVIKHYEHELGP 60

### Subject: 1 MTIDQKNTNIDPRFPHHHPRPQSFWEARAKALESLLIEKGHLSSDAIERVIKHYEHELGP 60

### Query: 61 MNGAKWAKAWTDPAFKQRLLEDSETVLRELGYYGLQGEHIRWENTDTVHNVWCTLCS 120

### Subject: 61 MNGAKWAKAWTDPAFKQRLLEDSETVLRELGYYGLQGEHIRWENTDTVHNVWCTLCS 120

### Query: 121 CYPWPLLGLPSWNYKEPYARAVYKPEQVRKFLGKLDSVEIRVWDDSSEIRFMWLVQ 180

### Subject: 121 CYPWPLLGLPSWNYKEPYARAVYKPEQVRKFLGKLDSVEIRVWDDSSEIRFMWLVQ 180

### Query: 181 RPEGTEGMTEEELAKLVTRDSMIGVAKIEPPKV 213

### Subject: 181 RPEGTEGMTEEELAKLVTRDSMIGVAKIEP+K 213

### Query: 222
Appendices

> sp|Q53118|NHAA RHODO NITRILE HYDRATASE SUBUNIT ALPHA (NITRILASE) (NHASE)
gb|AAA26185.11 (M74531) nitrile hydratase alpha subunit [Rhodococcus sp.]
Length = 199
Score = 246 bits (628), Expect = 1e-64
Identities = 120/186 (64%), Positives = 153/186 (81%), Gaps = 1/186 (0%)
Query: 27 ARAKALESLIEKGLHLSSDAIERVKHYEHELPGMPNGAKVAKWTDFAPA Q RL L D ED SET 86
Sbjct: 13 ARKALESLIEKGLHLSSDAIERVKHYEHELPGMPNGAKVAKWTDFAPAK LR L D ED SET 86
Query: 147 PRQVL-KFEGDLLPSVEIRWDSSEII RMPVFQRPEGTEGEMTSEEELAKLVTRDMSIGV 205
Sbjct: 133 PTYVLSEEENYHLPSTEIRIMDTSSEM R Y W L P F Q R P E G T E G W S E Q L A E L V T R D M S I G V 192
Query: 206 AKIEPP 211
Sbjct: 193 GPVKTP 197

>sp|C41326|nitrile hydratase (EC 4.2.1.84) alpha chain - Rhodococcus sp
Length = 198
Score = 246 bits (627), Expect = 1e-64
Identities = 120/186 (64%), Positives = 153/186 (81%), Gaps = 1/186 (0%)
Query: 27 ARAKALESLIEKGLHLSSDAIERVKHYEHELPGMPNGAKVAKWTDFAPA Q RL L D ED SET 86
Sbjct: 13 ARKALESLIEKGLHLSSDAIERVKHYEHELPGMPNGAKVAKWTDFAPAR L D ED SET 86
Query: 147 PRQVL-KFEGDLLPSVEIRWDSSEII RMPVFQRPEGTEGEMTSEEELAKLVTRDMSIGV 205
Sbjct: 133 PTYVLSEEENYHLPSTEIRIMDTSSEM R Y W L P F Q R P E G T E G W S E Q L A E L V T R D M S I G V 192
Query: 206 AKIEPP 211
Sbjct: 192 GPVKTP 197

>sp|P97051|NHAA PSEPU NITRILE HYDRATASE SUBUNIT ALPHA (NITRILASE) (NHASE)
gb|AAC18418.11 (U89363) nitrile hydratase alpha subunit [Pseudomonas putida]
Length = 210
Score = 239 bits (611), Expect = 1e-62
Identities = 119/192 (61%), Positives = 150/192 (77%), Gaps = 4/192 (2%)
Query: 133 WYKPFAYRARVKEPLVRQVLKFEGLDLLPSVEIRWDSSEII RMPVFQRPEGTEGEMTSEEELAKLVTRDMSIGV 192
Sbjct: 129 WYKPFAYRARVKEPLVRQVLKFEGLLLPSVEIRWDSSEII RMPVFQRPEGTEGEMTSEEELAKLVTRDMSIGV 192
Query: 193 LAKLVTRDMSIGV 204
Sbjct: 188 LAKLVTRDMSIGV
Sbjct: 188 LAKLVTRDMSIGV
II.v: Translated BLAST search with the ORF4 (P14K) sequence

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Query</th>
<th>Score</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gb</td>
<td>AAF69003.1</td>
<td>AF257490 1 (AF257490) P12K [Bacillus sp. BR449]</td>
</tr>
<tr>
<td>dbj</td>
<td>BAA12062.1</td>
<td>(D83695) nitrile hydratase b-subunit homolog</td>
</tr>
<tr>
<td>dbj</td>
<td>BAA11045.1</td>
<td>(D67027) nitrilehydratase-b homologue [Rhodococcus rhodochrous]</td>
</tr>
<tr>
<td>ref</td>
<td>NP 386211.1</td>
<td>(NC_003047) CONSERVED HYPOTHETICAL PROTEIN</td>
</tr>
<tr>
<td>ref</td>
<td>NP 103015.1</td>
<td>(NC_002678) unknown protein [Mesorhizobium]</td>
</tr>
<tr>
<td>gb</td>
<td>AAF69003.1</td>
<td>AF257488 1 (AF257488) nitrile hydratase beta</td>
</tr>
<tr>
<td>sp</td>
<td>P121220</td>
<td>NHBI RHORJ HIGH-MOLECULAR WEIGHT COBALT-CONTAINING</td>
</tr>
<tr>
<td>sp</td>
<td>P55479</td>
<td>nitrile hydratase (EC 4.2.1.84) beta chain, F+</td>
</tr>
<tr>
<td>sp</td>
<td>P55317</td>
<td>NHAB RHOD NITRILE HYDRATASE SUBUNIT BETA (NITRIL)</td>
</tr>
<tr>
<td>gb</td>
<td>AAC14240.1</td>
<td>(U89363) P14K [Pseudomonas putida]</td>
</tr>
<tr>
<td>ref</td>
<td>NP 103016.1</td>
<td>(NC_002678) nitrile hydratase beta subunit</td>
</tr>
<tr>
<td>sp</td>
<td>P57052</td>
<td>NHAB PSEP4 NITRILE HYDRATASE SUBUNIT BETA (NITRIL)</td>
</tr>
<tr>
<td>sp</td>
<td>P29379</td>
<td>RHOR LH-MOLECULAR WEIGHT COBALT-CONTAINING</td>
</tr>
<tr>
<td>ref</td>
<td>NP 386212.1</td>
<td>(NC_003047) PROBABLE NITRILE HYDRATASE SUB</td>
</tr>
</tbody>
</table>

Alignments

>gb|AAF69003.1|AF257490 1 (AF257490) P12K [Bacillus sp. BR449]
Length = 101

Score = 176 bits (447), Expect = 3e-44
Identities = 100/101 (99%), Positives = 101/101 (99%)

Query: MKSCENQPNESLLANMSEEVAPPRKNGELEFQEPWERRSFGMTLALYEEKLYSSWEDFRS

Sbjct: MKSCENQPNESLLANMSEEVAP ---------------PRKNGELEFQEPWERRSFGMTLALYEEKLYSSWEDFRS

>dbj|BAA12062.1| (D83695) nitrile hydratase b-subunit homolog [Rhodococcus rhodochrous]
Length = 148

Score = 75.5 bits (184), Expect = 1e-13
Identities = 43/120 (35%), Positives = 67/120 (55%), Gaps = 9/120 (7%)

Query: MKSCENQPNESLLANMSEEVAP-------PRKNGELEFQEPWERRSFGMTLALYEEKLYS

Sbjct: MPRNGELEFQEPWERRSFGMTLALYEEKLYS53

>dbj|BAA11045.1| (D67027) nitrile hydratase-b homologue [Rhodococcus rhodochrous]
Length = 104

Score = 67.8 bits (164), Expect = 2e-11
Identities = 53/88 (59%), Positives = 53/88 (59%), Gaps = 8/88 (9%)

Query: APPRNGELEFQEPWERRSFGMTLALYEEKLYS80

Sbjct: APPRNGELEFQEPWERRSFGMTLALYEEKLYS80

Appel 2007
Appendices

> ref|NP 386211.11 (NC_003047) CONSERVED HYPOTHETICAL PROTEIN [Sinorhizobium melloi]  
emb|CAC46694.11 [AL591789] CONSERVED HYPOTHETICAL PROTEIN [Sinorhizobium melloi]  
Length = 128

Score = 52.0 bits (123), Expect = 1e-06  
Identities = 32/96 (33%), Positives = 54/96 (55%), Gaps = 8/96 (8%)

Query: 1 MKSCENQDINESLAMSEAVAPPKNGELEQFQWERRSPGQTMLALYEEKLYSWEDFRSK + + + P  
Sbjct: 4 LRSTFALPSRPLLASAS---LPKSEREGPVPFAFPQAVAFMTVLRLHEQGVF--SWSEWA 9 5

Query: 61 RLIEEIKG Q KQNSWNY YEH W LA AL RLV 96  
Sbjct: 60 ALSAEL---YKPGRRADGTD---Y YD CNV AALSRLVIE 91

> ref|NP 103015.11 (NC_002678) unknown protein [Mesorhizobium loti]  
db|BAB40801.11 [AP002997] unknown protein [Mesorhizobium loti]  
Length = 89

Score = 46.2 bits (108), Expect = 6e-05  
Identities = 28/65 (43%), Positives = 39/65 (59%), Gaps = 5/65 (8%)

Query: 42 MTALYEEKLYSWEDFRSRLIEEIKG Q KQNSWNY YEH W LA AL RLV 101  
Sbjct: 1 MTVALHDKGLF--SWSEWADALSAEVK---KPGAAAGHD---Y YH W LA LESLASKGLAA 55

Query: 102 KRDVD 106  
Sbjct: 56 KSDV D60

> sb|AAP69501.11 |AF257488_1 (AF257488) nitrile hydratase beta subunit [Bacillus sp. BR449]  
Length = 229

Score = 45.4 bits (106), Expect = 9e-05  
Identities = 26/91 (28%), Positives = 43/91 (53%), Gaps = 8/91 (8%)

Query: 24 RKNGELEQFQWERRSPGQTMLALYEEKLYSWEDFRSK + + + F  
Sbjct: 20 KEEEDIYFTHDMERLAFGLVAGCMAQGLGMKAFDEFRI-------GIELMRPVDYLTSY 72

Query: 83 YEH W LA AL RLV 113  
Sbjct: 73 YGHWATVAYNLVDPTGLDEKELDTEVF103

> sp|P21220|HHP1 RHOD HIGH-MOLECULAR WEIGHT COBALT-CONTAINING NITRILE HYDRATASE SUBUNIT BETA (H-NITRILASE) (H-NHASE)  
pir|S9713 nitrile hydratase (EC 4.2.1.84) beta chain, H-type - Rhodococcus rhodochrous (strain J1)  
emb|CAA45709.11 [X64359] nitrile hydratase [Rhodococcus rhodochrous]  
db|BAI11043.11 [D47027] nitrile hydratase b-subunit [Rhodococcus rhodochrous]  
Length = 229

Score = 40.0 bits (92), Expect = 0.004  
Identities = 27/94 (28%), Positives = 46/94 (48%), Gaps = 5/94 (5%)

Query: 23 PRNGELEQFQWERRSPGQTMLALYEEKLYSWEDFRSK + + + + + + + SW D + SR  
Sbjct: 17 FYQQKUEPPFYHEDGWGLTSLLTILMLGKI---SWSQ---M R S F P ESROMQNYVNEIRM---SY 71

Query: 83 YEH W LA AL RLV 116  
Sbjct: 72 YTHMLSAERLVLVDKILSTEEKRGRVQIELEG105

225
### II.vi: Translated BLAST search with the ORF5 sequence

<table>
<thead>
<tr>
<th>Sequence producing significant alignments:</th>
<th>Score</th>
<th>E Value</th>
<th>Identity</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref</td>
<td>NP</td>
<td>241233.1</td>
<td>(NC_002570) BH0367-unknown conserved protein...</td>
<td>136</td>
</tr>
<tr>
<td>sp</td>
<td>087688</td>
<td>CB1W_BACME</td>
<td>Putative 2Fe-2S ferredoxin [gml</td>
<td>1135651</td>
</tr>
<tr>
<td>ref</td>
<td>NP</td>
<td>444222.1</td>
<td>(NC_002607) Ferredoxin [Halobacterium sp. ...</td>
<td>60</td>
</tr>
<tr>
<td>gb</td>
<td>AA199441.1</td>
<td>(AE005068) Vng1561c [Halobacterium sp. NRC-1]</td>
<td>60</td>
<td>3e-09</td>
</tr>
<tr>
<td>ref</td>
<td>NP</td>
<td>521258.1</td>
<td>(NC_003296) PUTATIVE FERREDOXIN PROTEIN [R...</td>
<td>55</td>
</tr>
<tr>
<td>emb</td>
<td>C3C2920.1</td>
<td>(AJ312124) HymB protein [Eubacterium acidamin...</td>
<td>52</td>
<td>1e-06</td>
</tr>
<tr>
<td>ref</td>
<td>NP</td>
<td>277826.1</td>
<td>(NC_000853) NADP-reducing hydrogenase, sub...</td>
<td>52</td>
</tr>
<tr>
<td>sp</td>
<td>P072241</td>
<td>FER2 CLOPA Ferredoxin, 2Fe-2S (2FeCpFd) &gt; gi</td>
<td>165694</td>
<td>51</td>
</tr>
<tr>
<td>sp</td>
<td>P28202</td>
<td>FER5_AZOTI Ferredoxin, 2Fe-2s (Shethna protein 1)</td>
<td>51</td>
<td>2e-06</td>
</tr>
<tr>
<td>ref</td>
<td>NP</td>
<td>229225.1</td>
<td>(NC_000683) Fe-hydrogenase, subunit beta [...</td>
<td>39</td>
</tr>
<tr>
<td>emb</td>
<td>CA37873.1</td>
<td>(Y13471) hoxF [Synechococcus sp. PCC 6301]</td>
<td>37</td>
<td>0.035</td>
</tr>
<tr>
<td>ref</td>
<td>NP</td>
<td>484640.1</td>
<td>(NC_003272) unknown protein [Nostoc sp. PC...</td>
<td>37</td>
</tr>
<tr>
<td>ref</td>
<td>NP</td>
<td>450441.1</td>
<td>(NC_000911) unknown protein [Synechocystis sp...</td>
<td>37</td>
</tr>
<tr>
<td>ref</td>
<td>NP</td>
<td>518450.1</td>
<td>(NC_003295) PUTATIVE FERREDOXIN 2Fe-2S PRO...</td>
<td>36</td>
</tr>
</tbody>
</table>

**Alignments**

<table>
<thead>
<tr>
<th>Query:</th>
<th>Score</th>
<th>E Value</th>
<th>Identity</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 MATWNDFLKLHHLICNGSTMRRAEQTSRIASIQGKLENYIHTTRCRNGCRCKDK</td>
<td>136</td>
<td>4e-32</td>
<td>63/118 (53%)</td>
<td>81/118 (68%)</td>
</tr>
<tr>
<td>61 CVVLYPEGMYKQFRPDSAPVEALSFKVYDPQHSVHHYKNGKF</td>
<td>118</td>
<td>9e-04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61 CVLITYPSGDWYKQVTEKVAHRIVEEHLAGGCSLKESLHYBDQKGF</td>
<td>118</td>
<td>9e-04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62 PVAIVYPSGDWYKQVTEKVAHRIVEEHLAGGCSLKESLHYBDQKGF</td>
<td>118</td>
<td>9e-04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query:</th>
<th>Score</th>
<th>E Value</th>
<th>Identity</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 MATWNDFLKLHHLICNGSTMRRAEQTSRIASIQGKLENYIHTTRCRNGCRCKDK</td>
<td>97</td>
<td>3e-20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61 CVVLYPEGMYKQFRPDSAPVEALSFKVYDPQHSVHHYKNGKF</td>
<td>106</td>
<td>5e-04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62 PVAIVYPSGDWYKQVTEKVAHRIVEEHLAGGCSLKESLHYBDQKGF</td>
<td>108</td>
<td>5e-04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query:</th>
<th>Score</th>
<th>E Value</th>
<th>Identity</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref</td>
<td>NP</td>
<td>444222.1</td>
<td>(NC_002607) Ferredoxin [Halobacterium sp. NRC-1]</td>
<td>397</td>
</tr>
<tr>
<td>sp</td>
<td>087688</td>
<td>CB1W_BACME</td>
<td>Putative 2Fe-2S ferredoxin homolog [imported] - Bacillus megaterium</td>
<td>97</td>
</tr>
<tr>
<td>ref</td>
<td>NP</td>
<td>444222.1</td>
<td>(NC_002607) Ferredoxin [Halobacterium sp. NRC-1]</td>
<td>397</td>
</tr>
<tr>
<td>sp</td>
<td>087688</td>
<td>CB1W_BACME</td>
<td>Putative 2Fe-2S ferredoxin homolog [imported] - Bacillus megaterium</td>
<td>97</td>
</tr>
</tbody>
</table>
Appendices

II.vii: Translated BLAST search with the ORF6 sequence

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Ref</th>
<th>Score (bits)</th>
<th>Value</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref</td>
<td>NP 561108.11 (NC_003366)</td>
<td>233</td>
<td>1e-60</td>
</tr>
<tr>
<td>ref</td>
<td>NP 460968.1 (NC_003197)</td>
<td>231</td>
<td>4e-60</td>
</tr>
<tr>
<td>ref</td>
<td>NP 464729.11 (NC_003210)</td>
<td>218</td>
<td>6e-60</td>
</tr>
<tr>
<td>ref</td>
<td>NP 470504.11 (NC_003212)</td>
<td>218</td>
<td>7e-60</td>
</tr>
<tr>
<td>ref</td>
<td>NP 487983.11 (NC_003272)</td>
<td>211</td>
<td>6e-60</td>
</tr>
<tr>
<td>ref</td>
<td>NP 275273.11 (NC_000916)</td>
<td>200</td>
<td>1e-60</td>
</tr>
<tr>
<td>ref</td>
<td>NP 248084.11 (NC_000909)</td>
<td>193</td>
<td>2e-60</td>
</tr>
<tr>
<td>ref</td>
<td>NP 069562.11 (NC_000917)</td>
<td>177</td>
<td>7e-44</td>
</tr>
<tr>
<td>ref</td>
<td>NP 280413.11 (NC_002607)</td>
<td>153</td>
<td>1e-58</td>
</tr>
<tr>
<td>ref</td>
<td>NP 347993.11 (NC_003030)</td>
<td>137</td>
<td>2e-48</td>
</tr>
<tr>
<td>ref</td>
<td>NP 248577.11 (NC_000909)</td>
<td>61</td>
<td>9e-09</td>
</tr>
</tbody>
</table>
Translated BLAST search with the ORF7 sequence

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Score E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(bits)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>18309175</td>
</tr>
<tr>
<td>gi</td>
<td>13431336</td>
</tr>
<tr>
<td>gi</td>
<td>13431346</td>
</tr>
<tr>
<td>gi</td>
<td>15678159</td>
</tr>
<tr>
<td>gi</td>
<td>17239413</td>
</tr>
<tr>
<td>gi</td>
<td>16765352</td>
</tr>
<tr>
<td>gi</td>
<td>16760899</td>
</tr>
<tr>
<td>gi</td>
<td>16800237</td>
</tr>
<tr>
<td>gi</td>
<td>16803245</td>
</tr>
<tr>
<td>gi</td>
<td>15792088</td>
</tr>
<tr>
<td>gi</td>
<td>13431339</td>
</tr>
<tr>
<td>gi</td>
<td>15669297</td>
</tr>
<tr>
<td>gi</td>
<td>13431338</td>
</tr>
<tr>
<td>gi</td>
<td>15628194</td>
</tr>
</tbody>
</table>

Alignments

<table>
<thead>
<tr>
<th>Query</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>18309175</td>
</tr>
<tr>
<td>gi</td>
<td>18143850</td>
</tr>
</tbody>
</table>

Score = 70.5 bits (17%), Expect = 4e-12
Identities = 39/95 (41%), Positives = 51/95 (53%), Gaps = 2/95 (2%)

Query: 2 KNILLIIVMLGVPILLHOQEGGA--MASGRNQRNPQYFKWFQALWPEPSEGE 56
++ + +LL L+ +P ++ +N EFGA A Q Y+FWM L+EP SG EIE
Subject: 4 RHILMLAVILVAVAVPLAVGGEGGQFGGADDSAGDAITEGYKWFQPLWEPSSGE 63

Query: 57 LESLLFALQXXXXXXXXXXXXXLMMR5GNQGEXGKWK 93
IESLFAHQ ++K K E + K
Subject: 67 LESLLFALQAAIGAGVFGFLGLKGGKKKVNDEVNDK 101

Score = 70.1 bits (17%), Expect = 4e-12
Identities = 37/91 (40%), Positives = 50/91 (54%), Gaps = 5/91 (5%)

Query: 2 KNILLIIVMLGVPILLHOQ-----NACEFGGAMASGRNQRNPQYFKWFQALWPEPSEGE 56
++ +LL + + + P + + + + + + EFGA S Y+FWMQ L+EP SG EIE
Subject: 4 RHILMLAVILVAVAVPLAVGGEGGQFGGADDSAGDAITEGYKWFQPLWEPSSGE 63

Query: 57 IESLLFALQXXXXXXXXXXXXXLMMR5GNQGEXGKWK 93
IESLFAHQ ++ + + + + + +
Subject: 64 IESLLFALQAAIGAGVFGFLGLKGGKKKVNDEVNDK 101

Score = 67.0 bits (162), Expect = 4e-11
Identities = 35/89 (39%), Positives = 49/89 (54%), Gaps = 5/89 (5%)

Query: 2 KNILLIIVMLGVPILLHOQ-----NACEFGGAMASGRNQRNPQYFKWFQALWPEPSEGE 56
++ II + V ++ + P +I + + EFGA S ++ ++ Y+FWMQ +EPSSGE
Subject: 4 RHILMLAVILVAVAVPLAVGGEGGQFGGADDSAGDAITEGYKWFQPLWEPSSGE 63

Query: 57 IESLLFALQXXXXXXXXXXXXXLMMR5GNQGEXGKWK 93
IESLFAHQ ++ + + + + + +
Subject: 64 IESLLFALQAAGGALIIYGFGYGRGESS 94

Appendices
### ILix: Translated BLAST search with the ORF8 sequence

<table>
<thead>
<tr>
<th>Sequences producing significant alignments:</th>
<th>Score</th>
<th>E (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref</td>
<td>NP 561110.11 (NC_003366) probable cobalt transport protein [Clostridium perfringens]</td>
<td>111</td>
</tr>
<tr>
<td>ref</td>
<td>NP 467395.11 (NC_003272) cobalt transport protein</td>
<td>88</td>
</tr>
<tr>
<td>ref</td>
<td>T45264 cobalt transport protein Q homolog [imported]</td>
<td>87</td>
</tr>
<tr>
<td>ref</td>
<td>NP 275275.11 (NC_000916) cobalt transport protein Q [M.</td>
<td>84</td>
</tr>
<tr>
<td>ref</td>
<td>NP 460866.11 (NC_003197) synthesis of vitamin B12 adenos...</td>
<td>60</td>
</tr>
<tr>
<td>ref</td>
<td>Q05981</td>
<td>COBALT TRANSPORT PROTEIN CB</td>
</tr>
<tr>
<td>ref</td>
<td>NP 456575.11 (NC_003198) putative cobalt transport protein</td>
<td>79</td>
</tr>
<tr>
<td>ref</td>
<td>T03542 cobalt transport protein cbIQ - Rhodobacter capsul...</td>
<td>79</td>
</tr>
<tr>
<td>ref</td>
<td>NP 248082.11 (NC_000909) cobalt transport protein cbIQ...</td>
<td>73</td>
</tr>
<tr>
<td>ref</td>
<td>NP 464731.11 (NC_003210) similar to cobalt transport pr...</td>
<td>72</td>
</tr>
<tr>
<td>ref</td>
<td>NP 470506.11 (NC_003212) similar to cobalt transport pr...</td>
<td>72</td>
</tr>
<tr>
<td>ref</td>
<td>NP 069564.11 (NC_000917) cobalt transport protein cbIQ...</td>
<td>69</td>
</tr>
<tr>
<td>ref</td>
<td>NP 280441.11 (NC_002607) cobalt transport protein; CbIQ...</td>
<td>65</td>
</tr>
<tr>
<td>ref</td>
<td>AAU04905.11 (U13043) cobalt transport protein CBIQ [Prop...</td>
<td>60</td>
</tr>
<tr>
<td>ref</td>
<td>T35724 cobalt transport integral membrane protein - St...</td>
<td>57</td>
</tr>
<tr>
<td>ref</td>
<td>NP 442313.11 (NC_000911) unknown protein [Synechocystis...</td>
<td>56</td>
</tr>
<tr>
<td>ref</td>
<td>NP 248579.11 (NC_000909) cobalt transport protein cbIQ...</td>
<td>51</td>
</tr>
<tr>
<td>ref</td>
<td>NP 276817.11 (NC_000918) cobalt transport membrane protein...</td>
<td>55</td>
</tr>
</tbody>
</table>

#### Alignments

<table>
<thead>
<tr>
<th>Query: 61 ISKE SLNRSVNLVFM VLSSVSCLYFLILTTPLTAILQVLRKLKLPPLFIELITLTYRFIF 120</th>
<th>Score</th>
<th>E (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref</td>
<td>NP 561110.11 (NC_003366) probable cobalt transport protein Q [Clostridium perfringens]</td>
<td>111</td>
</tr>
<tr>
<td>ref</td>
<td>NP 561110.11 (NC_003366) probable cobalt transport protein Q [Clostridium perfringens]</td>
<td>111</td>
</tr>
<tr>
<td>ref</td>
<td>NP 561110.11 (NC_003366) probable cobalt transport protein Q [Clostridium perfringens]</td>
<td>111</td>
</tr>
<tr>
<td>ref</td>
<td>NP 561110.11 (NC_003366) probable cobalt transport protein Q [Clostridium perfringens]</td>
<td>111</td>
</tr>
<tr>
<td>ref</td>
<td>NP 561110.11 (NC_003366) probable cobalt transport protein Q [Clostridium perfringens]</td>
<td>111</td>
</tr>
<tr>
<td>ref</td>
<td>NP 561110.11 (NC_003366) probable cobalt transport protein Q [Clostridium perfringens]</td>
<td>111</td>
</tr>
<tr>
<td>ref</td>
<td>NP 561110.11 (NC_003366) probable cobalt transport protein Q [Clostridium perfringens]</td>
<td>111</td>
</tr>
<tr>
<td>ref</td>
<td>NP 561110.11 (NC_003366) probable cobalt transport protein Q [Clostridium perfringens]</td>
<td>111</td>
</tr>
<tr>
<td>ref</td>
<td>NP 561110.11 (NC_003366) probable cobalt transport protein Q [Clostridium perfringens]</td>
<td>111</td>
</tr>
<tr>
<td>ref</td>
<td>NP 561110.11 (NC_003366) probable cobalt transport protein Q [Clostridium perfringens]</td>
<td>111</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query: M STPIW VARIPLSYYM KLLAIPM FFLFSGIVTILISFTDGQTVISDVIW SAHIGDW QIF 57</th>
<th>Score</th>
<th>E (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref</td>
<td>NP 248082.11 (NC_000909) cobalt transport protein cbIQ...</td>
<td>73</td>
</tr>
<tr>
<td>ref</td>
<td>NP 464731.11 (NC_003210) similar to cobalt transport pr...</td>
<td>72</td>
</tr>
<tr>
<td>ref</td>
<td>NP 470506.11 (NC_003212) similar to cobalt transport pr...</td>
<td>72</td>
</tr>
</tbody>
</table>

#### IDentities, Positives, and Gaps

<table>
<thead>
<tr>
<th>Length</th>
<th>Score</th>
<th>E (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>235</td>
<td>111</td>
<td>5e-24</td>
</tr>
<tr>
<td>216</td>
<td>87.8</td>
<td>6e-17</td>
</tr>
<tr>
<td>266</td>
<td>87.8</td>
<td>6e-17</td>
</tr>
</tbody>
</table>

### Appendices
II.x: Translated BLAST search with the ORF9 sequence

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>179</td>
<td>8e-45</td>
</tr>
<tr>
<td>177</td>
<td>5e-44</td>
</tr>
<tr>
<td>165</td>
<td>1e-40</td>
</tr>
<tr>
<td>157</td>
<td>3e-38</td>
</tr>
</tbody>
</table>

Similar to cobalt transport ATP-binding protein O homolog - Methanococcus jannaschii
Length = 279

<table>
<thead>
<tr>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>177</td>
<td>5e-44</td>
</tr>
<tr>
<td>149</td>
<td>6e-35</td>
</tr>
<tr>
<td>147</td>
<td>4e-36</td>
</tr>
</tbody>
</table>

Similar to cobalt transport ATP-binding protein cbiO - Methanococcus jannaschii
Length = 279

Score = 177 bits (448), Expect = 5e-44
Identities = 83/148 (56%), Positives = 111/148 (74%)

Query: 16 YADGITALRNVTILIKQKSKKALLGNNGAGKSTILLHNLGIIRPTNGAIRFKGKEVAYDR
Query: 17 YDGTL La ++XG+ ++LLG KAGKSTILLHNGLRPTG Q + KGK + YD+

Alignments

>ref|NP_561111.1| (NC_003366) probable cobalt transport ATP-binding protein O [Clostridium perfringens] (dbj|BAB79901.1) (AP003185) probable cobalt transport ATP-binding protein O [Clostridium perfringens] (Length = 285)
Score = 179 bits (454), Expect = 8e-45
Identities = 84/162 (51%), Positives = 121/162 (73%)

Query: 2 MHPILFDIDQSLHNCDYGTDIALRNVTILIKQKSKKALLGNNGAGKSTILLHNLGIIRPTN 61
Query: 3 M + +++L++ YDGT L++ + +K+G+ ++LLG NGAGKSTILLH NLGIIRPTG 60

>ref|NP_248081.1| (NC_00909) cobalt transport ATP-binding protein O homolog - Methanococcus jannaschii (pirl|IGE4435) cobalt transport ATP-binding protein O homolog - Methanococcus jannaschii (gb|aab90899.1) (U67551) cobalt transport ATP-binding protein cbiO - Methanococcus jannaschii (Length = 279)
Score = 177 bits (448), Expect = 5e-44
Identities = 83/148 (56%), Positives = 111/148 (74%)

Query: 16 YADGITALRNVTILIKQKSKKALLGNNGAGKSTILLHNLGIIRPTNGAIRFKGKEVAYDR
Query: 17 YDGTL La ++XG+ ++LLG KAGKSTILLHNGLRPTG Q + KGK + YD+

Appendices