Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage. [http://researchspace.auckland.ac.nz/feedback](http://researchspace.auckland.ac.nz/feedback)

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form.
pulA, a thermostable pullulanase
from an extreme thermophile
*Caldocellum saccharolyticum*

Gregory D. Albertson

Department of Cellular and Molecular Biology,

University of Auckland

Thesis submitted in fulfilment of the
requirements for the degree of
Doctor of Philosophy

November 1992
Abstract

The pullulanase gene from *Caldocellum saccharolyticum*, an obligate thermophilic anaerobe, was sequenced and expressed in *E. coli*. Expression and substrate induction studies in *E. coli* showed that while gene expression was substrate inducible and the enzyme was exported into the growth medium in *C. saccharolyticum*, expression was non-inducible in *E. coli* and the enzyme remained in the cytoplasm. The nucleotide sequence of the *pulA* gene was shown to be 2478 basepairs (bp) in length, coding for a protein of 96 kDa. The proposed promoter sequences showed homology to both the standard *E. coli* sequences and the consensus sequences obtained from other *C. saccharolyticum* genes.

The enzyme from the native organism was purified from the growth medium and shown to have a molecular mass of approximately 120 kDa. Periodic acid-Schiffs staining showed that this enzyme was glycosylated and substrate characterisation revealed that the enzyme debranched pullulan to produce only maltotriose, but hydrolysed amylopectin, amylose and β-limit dextran to produce a number of smaller oligosaccharides.

The enzyme was expressed in *E. coli* from its own promoters and was purified from the cytoplasmic fraction. Substrate characterisation revealed that the enzyme debranched pullulan to produce only maltotriose, but had only limited activity on β-limit dextran and amylopectin, and no activity on amylose.

The pullulanase gene was also expressed under the control of a heat-inducible overexpression system in *E. coli* and a copper-inducible expression system in yeast.

Amino acid homology comparisons of the pullulanase to other pullulanase sequences and related enzymes revealed a high degree of homology, particularly around three highly conserved regions. In α-amylases amino acids in these regions are involved in catalytic activity, substrate binding and metal ion binding.
# Overview

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>i</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>xiii</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>xv</td>
</tr>
<tr>
<td>Chapter One</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>Chapter Two</td>
<td>47</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td></td>
</tr>
<tr>
<td>Chapter Three</td>
<td>87</td>
</tr>
<tr>
<td>The Pullulanase expressed from <em>Caldocellum saccharolyticum</em></td>
<td></td>
</tr>
<tr>
<td>Chapter Four</td>
<td>113</td>
</tr>
<tr>
<td>DNA Sequence of the Pullulanase Gene from <em>C. saccharolyticum</em></td>
<td></td>
</tr>
<tr>
<td>Chapter Five</td>
<td>141</td>
</tr>
<tr>
<td>Expression of the Pullulanase in <em>Escherichia coli</em></td>
<td></td>
</tr>
<tr>
<td>Chapter Six</td>
<td>173</td>
</tr>
<tr>
<td>Inducible Enzyme Expression in <em>E. coli</em></td>
<td></td>
</tr>
<tr>
<td>Chapter Seven</td>
<td>187</td>
</tr>
<tr>
<td>Pullulanase Expression Studies using Yeast</td>
<td></td>
</tr>
<tr>
<td>Chapter Eight</td>
<td>191</td>
</tr>
<tr>
<td>Analysis of the Deduced Peptide Sequence of the pullulanase</td>
<td></td>
</tr>
<tr>
<td>Chapter Nine</td>
<td>211</td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
</tr>
<tr>
<td>Appendices</td>
<td>231</td>
</tr>
<tr>
<td>References</td>
<td>289</td>
</tr>
</tbody>
</table>
# Table of Contents

Chapter One  Introduction ........................................................................................................ 1

1.1 Thermophiles .................................................................................................................... 1

1.1.1 Thermophilic organisms ............................................................................................... 1
1.1.1.1 History and introduction ......................................................................................... 1
1.1.1.2 Temperature and environmental range .................................................................... 1
1.1.1.3 Taxonomy ................................................................................................................ 2
1.1.1.4 The molecular basis of thermophily ......................................................................... 3

1.1.2 Uses of thermophilic/thermostable enzymes ................................................................. 5
1.1.2.1 Biotechnological advantages of using thermophilic enzymes ............................... 5
1.1.2.2 Increased enzymatic reaction rates ......................................................................... 6
1.1.2.3 Unique properties of thermophilic enzymes .......................................................... 6
1.1.2.4 Purification of thermophilic enzymes expressed in mesophilic hosts .................. 7

1.1.3 Applications of thermophilic enzymes ........................................................................ 7
1.1.3.1 Enzymatic bleaching of wood pulp ....................................................................... 7
1.1.3.2 Biomass fermentation .............................................................................................. 7
1.1.3.3 Thermophilic enzymes used in laboratory research ............................................... 8

1.1.4 Protein thermostability ................................................................................................. 8
1.1.4.1 History and introduction ......................................................................................... 8
1.1.4.2 Random and site-directed mutagenesis ................................................................ 9
1.1.4.3 Statistical analysis .................................................................................................. 9
1.1.4.4 3-dimensional model systems .............................................................................. 10
1.1.4.5 Chimeric genes ...................................................................................................... 11

1.2 Starch ................................................................................................................................. 12

1.2.1 Chemical structure of starch and related carbohydrates ............................................. 12
1.2.1.1 Amylose ................................................................................................................ 12
1.2.1.2 Amylopectin .......................................................................................................... 13
1.2.1.3 Pullulan .................................................................................................................. 14
1.2.1.4 Cyclodextrin .......................................................................................................... 14

1.2.2 The starch industry ...................................................................................................... 15
1.2.2.1 Economic uses and value of starch products ......................................................... 15
1.2.2.2 Economic uses and value of starch hydrolysis products ....................................... 15
1.2.2.3 Current hydrolysis methods .................................................................................. 16
1.2.2.4 Usefulness of thermophilic enzymes .................................................................... 18

1.2.3 Usefulness of a thermostable debranching enzyme ................................................... 19
1.2.3.1 Theory and practice in amylopectin hydrolysis ...................................................... 19
1.2.3.2 Useful properties of debranching enzymes ............................................................. 21

1.3 Genetic Analysis of Pullulanases and Related Genes ..................................................... 23
1.3.1 *E. coli* maltose regulon ............................................................................................ 23
1.3.2 *Klebsiella* maltose regulon ...................................................................................... 24
1.3.3 Thermophilic enzymes ............................................. 26

1.4 Pullulanase and Related Enzymes .................................... 28

1.4.1 Non-branching enzymes ........................................... 29
  1.4.1.1 α-amylase (E.C. 3.2.1.1) 1,4-α-D-Glucan glucanohydrolase .... 29
  1.4.1.2 β-amylase (E.C. 3.2.1.2) 1,4-α-D-Glucan maltohydrolase ........ 30
  1.4.1.3 Glucoamylase (E.C. 3.2.1.3) 1,4-α-D-Glucan glucohydrolase .... 30
  1.4.1.4 Alpha glucosidase ........................................... 30
  1.4.1.5 "Non-classical" α-amylase .................................... 31

1.4.2 Debranching enzymes ............................................. 32
  1.4.2.1 Type I or true pullulanase .................................... 33
  1.4.2.2 Type II pullulanase .......................................... 34
  1.4.2.3 α-amylase-pullulanase ....................................... 34
  1.4.2.4 Neopullulanase and isopullulanase ........................... 35
  1.4.2.5 Isoamylase (E.C. 3.2.1.68) Glycogen 6-glucanohydrolase .......... 36

1.5 Caldocellum saccharolyticum ....................................... 37
  1.5.1 Caldocellum saccharolyticum bacteria .......................... 37
    1.5.1.1 Growth conditions ......................................... 37
    1.5.1.2 Taxonomic characteristics .................................. 37
  1.5.2 C. saccharolyticum enzymes transferred and expressed in E. coli .... 38
    1.5.2.1 β-glucosidase ............................................. 38
    1.5.2.2 Cellulase .................................................. 38
    1.5.2.3 Hemicellulase ............................................. 39

1.6 Previous Work Done and Aims of Thesis ............................ 41
  1.6.1 Gene isolation and expression in E. coli ....................... 41
  1.6.2 Bal 31 deletions ............................................. 41
  1.6.3 The aims of research for this thesis .......................... 45

Chapter Two  Materials and Methods .................................... 47

2.1 Materials .......................................................... 47
  2.1.1 Chemicals ...................................................... 47
  2.1.2 Radioactivity .................................................. 47
  2.1.3 Enzymes ....................................................... 47
    2.1.3.1 Restriction enzymes ........................................ 47
  2.1.4 Common buffers and general solutions ........................... 48

2.2 Organisms .......................................................... 49
  2.2.1 E. coli strains ................................................ 49
2.2.2 Growth media and conditions................................................. 49
  2.2.2.1 Rich media........................................................................ 51
  2.2.2.2 Minimal media................................................................... 51
  2.2.2.3 Antibiotics........................................................................ 52

2.2.3 Measurement of cell growth.................................................... 52
  2.2.3.1 Optical density..................................................................... 52
  2.2.3.2 Viable cell counts............................................................... 52
  2.2.3.3 Plasmid copy number determination.................................... 53

2.2.4 *Caldocellum saccharolyticum*.................................................. 53
  2.2.5 Yeast strains.......................................................................... 54

2.3 DNA.................................................................................................. 55
  2.3.1 Plasmids.................................................................................. 55
  2.3.2 Small scale plasmid isolation................................................... 55
    2.3.2.1 Standard alkaline lysis method........................................ 55
    2.3.2.2 Phenol/chloroform method.............................................. 56
  2.3.3 Large scale plasmid isolation.................................................. 57
    2.3.3.1 CsCl-EtBr density gradient purification............................ 57
    2.3.3.2 Large-Scale LiCl method.................................................. 57
  2.3.4 Standard agarose gel electrophoresis........................................ 58
  2.3.5 Preparative agarose gel electrophoresis...................................... 58
    2.3.5.1 Separation........................................................................ 58
    2.3.5.2 Recovery of DNA fragments from agarose gels.................... 59
  2.3.6 Ligation and transformation.................................................... 59
    2.3.6.1 Ligation........................................................................... 59
    2.3.6.2 Heat shock transformation............................................... 59
    2.3.6.3 DMSO transformation method............................................ 60
  2.3.7 Radioactively labelled nucleic acid probes.................................... 60
    2.3.7.1 Nick translation................................................................ 60
    2.3.7.2 Random priming............................................................. 61
    2.3.7.3 Gel filtration.................................................................... 61
    2.3.7.4 Scintillation counting....................................................... 61
  2.3.8 Transfer of DNA to charged membranes...................................... 61
    2.3.8.1 Southern blotting - alkaline Transfer.................................. 61
    2.3.8.2 Dot blots......................................................................... 62
    2.3.8.3 Colony hybridisation....................................................... 62
  2.3.9 Hybridisation, washing and autoradiography.................................. 62
    2.3.9.1 Prehybridisation and hybridisation.................................... 62
    2.3.9.2 Stringency calculations................................................... 63
    2.3.9.3 Autoradiography............................................................... 63
    2.3.9.4 Removal of probes......................................................... 64
  2.3.10 Genomic DNA preparations................................................... 64
  2.3.11 Polymerase chain reaction.................................................... 64
2.3.12 Yeast DNA methods ............................................. 65
2.3.12.1 Yeast transformations ..................................... 65
2.3.12.2 Plasmid isolation from yeast ............................. 65

2.4 Sequencing .......................................................... 66

2.4.1 Single stranded DNA isolation ................................. 66
2.4.1.1 M13 phage ..................................................... 66
2.4.1.2 Phagemids .................................................... 66

2.4.2 Oligonucleotide primers ........................................ 66

2.4.3 Sequencing methods ............................................ 68
2.4.3.1 Single strand sequencing .................................. 68
2.4.3.2 Plasmid sequencing ........................................ 68

2.4.4 Acrylamide sequencing gels .................................... 68

2.4.5 Sequencing library .............................................. 69
2.4.5.1 Pseudorandom library ....................................... 69
2.4.5.2 Random library ............................................... 69

2.5 Enzyme Purification ............................................... 70

2.5.1 Isolation from E. coli ........................................... 70
2.5.1.1 Cell harvesting ............................................... 70
2.5.1.2 Heat purification ............................................ 70
2.5.1.3 Ammonium Sulphate Precipitation ......................... 70
2.5.1.4 Enzyme location within the E. coli cell ................. 71

2.5.2 Isolation from Caldoccellum saccharolyticum ............... 73

2.5.3 Isolation from yeast ............................................. 73

2.5.4 Chromatography .................................................. 74
2.5.4.1 Dialysis and desalting ..................................... 74
2.5.4.2 Ion exchange ................................................ 74
2.5.4.3 Gel filtration ............................................... 74
2.5.4.4 Affinity chromatography ................................ 75

2.5.5 Estimation of total protein concentrations .................. 75

2.6 Enzyme Characteristics ........................................... 77

2.6.1 Enzyme assays ................................................... 77
2.6.1.1 Reaction Buffers .......................................... 77
2.6.1.2 PAHBAH reducing sugar assay ............................ 79
2.6.1.3 Starch-iodine complex ...................................... 79
2.6.1.4 Dye-linked substrates ..................................... 80
2.6.1.5 Substrate preparation ...................................... 80

2.6.2 Protein gel electrophoresis .................................... 80
2.6.2.1 SDS-PAGE .................................................... 80
2.6.2.2 Non-denaturing PAGE ..................................... 81
2.6.2.3 Coomassie Blue Stain ..................................... 81
Chapter Four  DNA Sequence of the Pullulanase Gene from C. saccharolyticum......113

4.1 Introduction ......................................................................................................................113
4.2 Restriction endonuclease map of pNZ1452 and pNZ1038........................................113
4.3 Sequencing of pNZ1452 .................................................................................................117
   4.3.1 Sequencing of previously constructed recombinants .............................................117
   4.3.2 Pseudo random library ............................................................................................117
   4.3.3 Random shear library ..............................................................................................119
   4.3.4 Recombinant plasmids ............................................................................................121
4.4 Analysis of the Sequence of pNZ1452 ..........................................................................122
   4.4.1 Determination of open reading frames (ORF) .......................................................122
   4.4.2 Incomplete gene on pNZ1452 ................................................................................124
   4.4.3 Sequencing the remainder of pNZ1038 ..................................................................126
4.5 Genetic Structural elements of pNZ1038 ....................................................................129
   4.5.1 ORF 1, 2, 3 and 4 gene assignment ........................................................................129
   4.5.2 Repeat sequences ....................................................................................................132
   4.5.3 Terminator sequences ............................................................................................133
   4.5.4 Leader sequence analysis .......................................................................................135
   4.5.5 Codon preference ....................................................................................................136
4.6 Summary .........................................................................................................................139

Chapter Five  Expression of the Pullulanase in Escherichia coli.......................................141

5.1 Summary .........................................................................................................................141
5.2 Enzyme Expression and Cell Growth............................................................................141
   5.2.1 Expression levels ....................................................................................................141
   5.2.2 Substrate induction of enzyme synthesis ..............................................................143
   5.2.3 Dependence of cellular growth phase on enzyme expression ..............................144
   5.2.4 Consequences of pullulanase expression on the E. coli cells .............................144
      5.2.4.1 Growth rate studies .......................................................................................145
      5.2.4.2 Microscopic examination ............................................................................146
      5.2.4.3 Plasmid stability tests ..................................................................................149
   5.2.5 Cellular location of the enzyme ............................................................................150
   5.2.6 Proof that the pullulanase gene in all constructs was expressed from a C. saccharolyticum promoter .................................................................151
5.3 Purification of the C. saccharolyticum Pullulanase Expressed in E. coli153
   5.3.1 Initial purification studies .......................................................................................155
5.3.2 Heat purification and ammonium sulphate precipitation ...............155
5.3.3 Ion Exchange .............................................................................156
5.3.4 Gel filtration ..............................................................................158
5.3.5 Purification using ultrafiltration .......................................................160

5.4 Enzymatic Characterisation .................................................................161

5.4.1 Activity gels and N-terminal sequencing .........................................161

5.4.2 The effect of metal ions and inhibitors on pullulanase activity .............162
  5.4.2.1 Initial studies ...........................................................................162
  5.4.2.2 Calcium ion studies .................................................................163
  5.4.2.3 Monovalent cations .................................................................164

5.4.3 Thermostability and temperature and pH optima of the pullulanase ......165

5.4.4 Substrate characterisation ...............................................................168

5.5 Summary..........................................................................................171

Chapter Six  Inducible Enzyme Expression in E. coli .................................................173

6.1 Introduction ......................................................................................173

6.2 Optimisation of Expression Levels ......................................................173
  6.2.1 Expression plasmid construction ..................................................173
    6.2.1.1 Expression plasmid choice .....................................................173
    6.2.1.2 Structure of pJLA602 ..............................................................175

  6.2.2 Optimisation of induction .............................................................176
    6.2.2.1 28 - 42 heat shift induction ....................................................176
    6.2.2.2 28 - 37 heat shift induction ....................................................178
    6.2.2.3 28 - 42 - 37 heat shift induction .............................................178
    6.2.2.4 Alkaline pH shift induction ....................................................179
    6.2.2.5 Optimal induction procedure ................................................179

  6.2.3 Enzyme location within the cell ....................................................180

  6.2.4 Microscopic examination of heat induced cultures ............................181

6.3 Attempted Purification of Overexpressed Enzyme ....................................183

6.4 Enzymatic Characterisation ..................................................................184

6.5 Summary..........................................................................................185
Chapter Seven Pulullanase Expression Studies using Yeast ..................................................187

7.1 Plasmid Construction ........................................................................................................187

7.2 Enzyme Induction and Expression Studies .....................................................................189

7.3 Plasmid Stability Tests ....................................................................................................190

7.4 Summary ..........................................................................................................................190

Chapter Eight Analysis of the Deduced Peptide Sequence of the *C. saccharolyticum* Pulullanase .................................................................191

8.1 Introduction ........................................................................................................................191

8.2 Comparison of the Putative Peptide Sequence of the *C. saccharolyticum* Pulullanase to Other Related Enzymes .................................................................192

8.3 The Presence of Three Conserved Regions in Pullulanases ........................................195

8.4 Comparison of the Peptide Sequence of the *C. saccharolyticum* Pulullanase to other Pulullanases ........................................................................................................200

8.5 Pullulanase Secondary Structure Predictions ..................................................................204

8.6 Summary ..........................................................................................................................208

Chapter Nine Discussion ......................................................................................................211

9.1 Comparison of the DNA sequence of the pNZ1038 *Caldocellum saccharolyticum* pulullanase to other known *Caldocellum saccharolyticum* genes and the DNA sequences of other related enzymes .................................................................211

9.2 Comparison of the peptide sequence of the *C. saccharolyticum* pulullanase to the peptide sequences of other related enzymes ........................................................................................................214

9.2.1 Comparison of the amino acid sequences .......................................................................214

9.2.2 Differences related to thermostability ...........................................................................214

9.2.3 Differences related to substrate characterisation ..........................................................215

9.3 Enzymatic properties of the *Caldocellum saccharolyticum* pulullanase compared to the enzyme expressed in *E. coli* ....................................................................................217
9.4 Enzymatic properties of the *Caldocellum saccharolyticum* pullulanase compared to the enzymatic properties of other related enzymes ........................................... 221

9.5 The potential biotechnological applications of the *Caldocellum saccharolyticum* pullulanase ........................................................................................................... 223

9.6 Future work solving expression problems .................................................. 224

9.6.1 Possible methods to improve pullulanase expression in *E. coli*-based expression systems ................................................... 224

9.6.2 Possible methods to improve pullulanase expression in expression systems other than *E. coli* .......................................................... 225

9.7 Future work on the study and modification of enzyme activity .................. 227

Appendices .................................................................................................... 231

One Results of Sequencing pCGN Recombinants ...................................... 231

Two Analysis of pNZ1038 ........................................................................... 234

Three Peptide Comparisons ...................................................................... 241

Four Genetic Analysis of *Caldocellum saccharolyticum* Genes ................ 254

Five Three-dimensional Structure of α-amylase ........................................... 260

Six Possible Errors in Debranching Enzyme Sequences ........................... 268

Seven Frame-shift Detection Algorithm ....................................................... 283

References .................................................................................................... 289
Acknowledgements

I would like to thank my supervisors, Professor Peter Bergquist, for his excellent advice and support during the course of this thesis, and Dr. Richard Gardiner, my associate supervisor. Thank you also to Dr. Ronnie McHale for his useful advice and discussion.

I am grateful to the other members of the Thermophile Genetics group, and other members of the Cell Biology Department as a whole, for creating a friendly atmosphere in which to work. Special thanks go to my fellow PhD student, Gordon Munro, for encouragement and the loan of a hard drive in my hour of need.

The following people from other departments have given time and technical assistance towards the completion of this work. Dr. Dave Christie, Biochemistry Department; Prof. Ted Baker, Massey University and the Thermophile group, Waikato University.

I would like to thank my family and friends for their support and encouragement, especially my wife, Carol, and my daughter, Rebekah, for typing my thesis.

Finally I wish to thank Pacific Enzyme Ltd for a PhD scholarship and funding of consumables for this research.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>adenine</td>
</tr>
<tr>
<td>BCIG</td>
<td>5-bromo-4-chloro-3-indolyl-β-galactopyranoside</td>
</tr>
<tr>
<td>bp</td>
<td>base pair(s)</td>
</tr>
<tr>
<td>C</td>
<td>cytosine</td>
</tr>
<tr>
<td>cpm</td>
<td>counts per minute</td>
</tr>
<tr>
<td>DEAE</td>
<td>O-(Diethylaminoethyl)-cellulose</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulphoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease I</td>
</tr>
<tr>
<td>dpm</td>
<td>disintegrations per minute</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetate</td>
</tr>
<tr>
<td>EtBr</td>
<td>ethidium bromide</td>
</tr>
<tr>
<td>EtOH</td>
<td>ethanol</td>
</tr>
<tr>
<td>G</td>
<td>guanine</td>
</tr>
<tr>
<td>g</td>
<td>acceleration of gravity</td>
</tr>
<tr>
<td>kb</td>
<td>1000 base pairs</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Daltons</td>
</tr>
<tr>
<td>L-broth</td>
<td>Luria broth</td>
</tr>
<tr>
<td>MES</td>
<td>4-morpholineethanesulphonic acid</td>
</tr>
<tr>
<td>min</td>
<td>minutes</td>
</tr>
<tr>
<td>OD&lt;sub&gt;600&lt;/sub&gt;</td>
<td>optical density (absorbance) at 600 nm</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease A</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>T</td>
<td>thymine</td>
</tr>
<tr>
<td>Tm</td>
<td>melting temperature</td>
</tr>
<tr>
<td>Tris</td>
<td>2-amino-2-hydroxymethylpropane-1,3-diol</td>
</tr>
<tr>
<td>Tris-HCl</td>
<td>Tris solution, pH adjusted with HCl</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>U</td>
<td>uracil</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet light</td>
</tr>
<tr>
<td>v/v</td>
<td>volume per volume</td>
</tr>
<tr>
<td>vol.</td>
<td>volume(s)</td>
</tr>
<tr>
<td>w/v</td>
<td>weight per volume</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
</tbody>
</table>

S.I. (Système Internationale (d'Unités)) abbreviations for units and standard notations for chemical elements, formulae, and chemical abbreviations (e.g. M = moles per litre) are used throughout this work.
Chapter One
Introduction

1.1 Thermophiles

1.1.1 Thermophilic organisms

1.1.1.1 History and introduction

Thermophilic organisms (algae) have been known since antiquity and Setchell (1903) isolated thermophilic bacteria from thermal pools in Yellowstone National Park at the turn of the century. However the “thermophilic age” only really began in the late 1960's with the isolation of *Thermus aquaticus* from thermal pools in Yellowstone National Park (Brock and Freeze 1969). Since then, thermophilic bacteria have been isolated worldwide from many geothermally heated natural sources; soil, hot pools (Murao et al. 1988), volcanic lakes (Toda et al. 1988), and deepsea hydrothermal vents (Pledger and Baross 1991) as well as artificial hot water sources, including domestic hot water and industrial waste water (Demharter et al. 1989).

1.1.1.2 Temperature and environmental range

Thermophiles are defined as organisms living at high temperatures, but, as shown in table 1.1, what is high for one class of organisms is not necessarily high for another.

An alternative definition states that thermophiles are those organisms that live above the thermophilic barrier, defined as 55-60°C. This is not an arbitrary, artificial value but is ecologically based, as temperatures lower than 50°C are widespread on Earth, being associated with solar heated habitats, whereas temperatures greater than 55-60°C are almost exclusively associated with geothermal habitats (note also that 60°C is the upper limit for eukaryote life). This, then, defines thermophiles, or more correctly geothermophiles, as prokaryotes growing in geothermal habitats. The only major exception is the interior of compost/manure and mining refuse piles, self-heated (normally up to 70°C) by microbial growth. Thermophiles can be further subdivided into three arbitrary groups based on their optimum growth temperature; moderate thermophiles (45-60°C), extreme thermophiles (60-80°C) and hyperthermophiles (>80°C) (Reviewed in Brock 1986).

Although geothermal habitats are worldwide in distribution, they only cover very small localised areas called thermal basins. These occur at the junction of tectonic plates, for example, in New Zealand, the Pacific Plate slides beneath the Indo-Australian Plate creating the thermal area of the central North Island (Suggate 1978). The geology of thermal basins produces two distinct micro-environments based on pH, which are a reflection of the major buffering components of natural
thermal springs; sulphuric acid (pH 2-4) and carbonate/bicarbonate (pH 7-10). Temperatures within a thermal basin (range 55-100°C) vary not only between pools but also within pools, due to cooling by other non-thermal ground water sources (Rinehart 1980). Deepsea geothermal vents are the other naturally occurring thermal source, their temperatures are higher (up to 400°C) since they are pressurised due to their depth, and their springs have a greater salt concentration due to the added effect of the sea water. As gas solubility decreases with increasing temperature, it is not surprising that most thermophiles are anaerobic, with the genus Thermus being a notable exception (Stetter et al. 1990).

<table>
<thead>
<tr>
<th>Group</th>
<th>Approximate upper temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals</td>
<td></td>
</tr>
<tr>
<td>Fish</td>
<td>38</td>
</tr>
<tr>
<td>Insects</td>
<td>45-50</td>
</tr>
<tr>
<td>Crustaceans</td>
<td>49-50</td>
</tr>
<tr>
<td>Plants</td>
<td></td>
</tr>
<tr>
<td>Vascular plants</td>
<td>45</td>
</tr>
<tr>
<td>Mosses</td>
<td>50</td>
</tr>
<tr>
<td>Eucaryotic microorganisms</td>
<td></td>
</tr>
<tr>
<td>Protozoa</td>
<td>56</td>
</tr>
<tr>
<td>Algae</td>
<td>55-60</td>
</tr>
<tr>
<td>Fungi</td>
<td>60-62</td>
</tr>
<tr>
<td>Procaryotic microorganisms</td>
<td></td>
</tr>
<tr>
<td>Cyanobacteria (blue-green algae)</td>
<td>70-73</td>
</tr>
<tr>
<td>Photosynthetic bacteria</td>
<td>70-73</td>
</tr>
<tr>
<td>Chemolithotrophic bacteria</td>
<td>&gt; 100</td>
</tr>
<tr>
<td>Heterotrophic bacteria</td>
<td>&gt; 100</td>
</tr>
</tbody>
</table>

§ Modified from Brock (1978)

1.1.1.3 Taxonomy

The taxonomy of thermophiles is still largely unresolved at the genus and species levels, as is the case for most bacteria. Classical taxonomic methods, based on phenotypical traits, have failed to produce a definitive phylogeny (reviewed in VanDemark and Batzing 1987). Historically, there has been a tendency with thermophilic eubacteria to classify anaerobic spore formers as Clostridium sp.
aerobes as Thermus sp. and those similar to Bacillus subtilis as Bacillus sp. (normally as Bacillus stearothermophilus).

In 1976 Woese and his co-workers proposed a new method of phylogenetic analysis based on 16S ribosomal RNA sequences (Woese et al. 1976). A later refinement proposed that all living organisms are classified into three domains (Eucarya, Bacteria and Archaea), with thermophilic bacteria being present in two of them (Woese et al. 1990). Since the phylogenetic tree produced from 16S RNA is based on the genetic sequences of an essential cellular component, it is less influenced by the environment or random mutagenesis. In both the Bacteria and Archaea domains, thermophilic bacteria are present as a subset in all branches, i.e. they are closer to their mesophilic cousins than to other thermophilic bacteria (Brock 1985). Some taxonomic examples are given in table 1.2.

<table>
<thead>
<tr>
<th>Genus/Species</th>
<th>Temperature Optimum</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archaea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sulfolobus</td>
<td>75-85°C</td>
<td>aerobic, sulphur oxidisers</td>
</tr>
<tr>
<td>Desulfurococcus</td>
<td>85°C</td>
<td>anaerobic, sulphur reducers</td>
</tr>
<tr>
<td>Methanothermus fervidus</td>
<td>85°C</td>
<td>anaerobic, methanogens</td>
</tr>
<tr>
<td>Pyrodictium</td>
<td>105°C</td>
<td>anaerobic, sulphur reducers</td>
</tr>
<tr>
<td>Bacteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thermotoga maritima</td>
<td>80°C</td>
<td>anaerobic, heterotrophic</td>
</tr>
<tr>
<td>Clostridium thermocellum</td>
<td>60-65°C</td>
<td>anaerobic, cellulolytic</td>
</tr>
<tr>
<td>Thermus aquaticus</td>
<td>70-80°C</td>
<td>aerobic, heterotrophic</td>
</tr>
<tr>
<td>Bacillus stearothermophilus</td>
<td>55-60°C</td>
<td>aerobic/anaerobic, heterotrophic</td>
</tr>
<tr>
<td>Caldocellum saccharolyticum</td>
<td>68°C</td>
<td>anaerobic, cellulolytic</td>
</tr>
</tbody>
</table>

Table 1.2 Examples of thermophilic bacteria.  


1.1.1.4 The molecular basis of thermophily

Although proteins are the most heat labile cellular components of mesophilic bacteria, the plasma membrane is considered the most critical. It must possess an inherent thermostability or high temperature denaturation could cause cellular rupture (Brock 1967). The molecular basis of
thermophily with respect to cell membranes, nucleic acids and other factors is summarised below, and protein thermostability is discussed in section 1.1.4.

In all thermophiles examined the plasma membrane is enriched in glycolipids, which may have a stabilising role. Thermophilic and acidophilic eubacteria, such as Bacillus acidocaldarius, incorporate cyclized triterpenes in their membranes. These compounds add stability by providing structural rigidity, decreasing phase transitions of lipids, and enhancing membrane viscosity (Langworthy and Pond 1986).

Many unique lipids and lipid arrangements have been discovered in thermophiles. For example, unusual diglycerol tetraether lipids are found in thermoacidophilic archaeabacteria and form either a lipid 'monolayer' or covalently-condensed bilayer that protects the membrane at high temperatures (Thurl and Schäfer 1988).

As the 'melting point' of DNA rises with increasing G+C content, one would expect thermophilic DNA to contain a higher proportion of G-C base pairs than DNA from mesophiles. Although the G+C contents of most thermophilic bacteria are higher than most mesophilic bacteria, (e.g. Thermus aquaticus at 66-69% G+C, compared to E. coli at 45-48% (Riley and Krawiec 1987)) there are a number of exceptions that contain a lower G+C % ratio than mesophiles, for example, Clostridium thermocellum (~40%, growth temp 65°C), “Caldocellum saccharolyticum” (~34% growth temp 70°C) and Dictyoglomus thermophilum (~29%, growth temp 73°C) (Yagüe et al. 1990, Donnison et al. 1989, Fukusumi et al. 1988 respectively).

As the melting points of mesophilic genomic DNA (e.g. E. coli: 87-90°C) are already above the temperature optima of many thermophiles, genomic G+C content and thermostability do not appear to be related. However the hyperthermophiles (e.g. Pyrodictium occultum, growth temperature 110°C, G+C content 62%, Stetter 1986) have a high G+C genomic ratio, implying that at these elevated temperatures the stabilising effect of a high G+C content may be required.

In thermophilic bacteria tRNA is stabilised by a combination of specific nucleotide modifications such as thiolations (which are known to strengthen three-dimensional helical stacking interactions) and an increased number of G-C base pairs (extra G-C hydrogen bond) in the double-stranded RNA regions, as compared to mesophilic bacteria (Haas et al. 1989).

Another factor possibly contributing to stability are polyamines that have been hypothesised to be important factors in the stabilisation of nucleic acid synthesis, protein synthesis, and cell division (Tabor & Tabor 1984). Mesophilic bacteria are known to contain only two polyamines, putrescine (a diamine) and spermidine (a triamine). Thermophiles contain a variety of long chain polyamines, for example Thermus thermophilus contains more than 12 distinct polyamines, eight of which, including thermine & thermospermine (tetramines), are novel (Hamana et al. 1991). A novel pentamine,
thermopentamine, has also been identified (Hamana et al. 1990). The addition of either spermine (a common tetramine), thermospermine or thermine was required for in vitro translation using a cell-free extract of *T. thermophilus* at physiological temperatures (65°C) (Oshima 1986).

In conclusion, for thermophiles to be able to grow, all cellular components need to be stable and functional at elevated temperatures. Thermostability depends on many mechanisms, normally the combination of many small cellular changes. These range from intrinsic macromolecular stability to the stabilising effects of other ‘factors’ and cellular modifications. Most of the changes that allow growth at high temperatures are rather subtle, and cellular processes in thermophiles appear not to differ greatly from those in mesophiles.

### 1.1.2 Uses of thermophilic/thermostable enzymes

Nomenclature note: While all thermophilic enzymes are thermostable there are a number of enzymes produced by mesophilic organisms that are also thermostable; hence the term *thermophilic* refers to enzymes from thermophiles and the term *thermostable* to enzymes that are stable at elevated temperatures (see section 1.1.4).

#### 1.1.2.1 Biotechnological advantages of using thermophilic enzymes

Many biotechnological advantages of using thermophilic organisms and their enzymes have been proposed (Hartley and Payton 1983, Sonnleitner and Fiechter 1983, Slapack et al. 1987). The most common are summarised below in table 1.3.

<table>
<thead>
<tr>
<th>Table 1.3 Potential merits of thermophilic fermentation$^8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Higher specific growth rates than mesophiles meaning faster fermentations</td>
</tr>
<tr>
<td>(2) Low cellular growth yields leading to high product/substrate ratios</td>
</tr>
<tr>
<td>(3) Thermophiles can generally withstand quite wide temperature and pH variations</td>
</tr>
<tr>
<td>(4) Reduced energy costs - cheaper to heat fermenters than to cool them</td>
</tr>
<tr>
<td>(5) Cheaper distillation costs for volatile products such as ethanol</td>
</tr>
<tr>
<td>(6) High temperature decrease viscosity - better diffusion and less energy required for mixing</td>
</tr>
<tr>
<td>(7) Utilisation of wide range of substrates (starch, cellulose, hemicellulose) by many thermophiles</td>
</tr>
<tr>
<td>(8) Sterility requirements are not as stringent.</td>
</tr>
</tbody>
</table>

$^8$ Modified from Slapack et al. (1987).
Many of these claimed advantages of thermophilic fermentation over mesophilic are more theoretical than realistic. The classical example of "reduced risk of contamination", is true if only considering mesophiles, as spore contamination (normally *Bacillus stearothermophilus* or *Clostridium thermocellum*, Hyun *et al.* 1983) is a major problem in thermophilic fermentations. Except for the few cases (normally multi-stage multi-enzyme reactions) where thermophilic organisms are required for growth (e.g. biomass fermentation), the real biotechnological advantage comes from the intrinsic properties of the enzymes themselves (i.e. the peptide sequence) not their host organism. Note: recent work on protein folding has shown that other factors can effect the final enzyme properties (this is discussed further in section 9.3).

1.1.2.2 Increased enzymatic reaction rates

As enzymatic reaction rates increase with temperature, increasing the reaction temperature increases the reaction rate. Theoretically the reaction rate can increase indefinitely but in practice a temperature optimum is reached as enzymatic denaturation reduces the overall reaction rate. As thermophilic and thermostable enzymes have higher denaturation temperatures, which give correspondingly higher temperatures of maximum enzyme activity, their reaction rates may be faster.

Other advantages of a higher reaction rate include reduced viscosity (e.g. starch), increased substrate solubility and continuous distillation, under reduced vacuum, of volatile end products (e.g. ethanol and acetone) from industrial scale fermentations (Wiegel 1980).

1.1.2.3 Unique properties of thermophilic enzymes

Currently, industrial enzymes are produced by hyper-producing mesophilic mutants at levels far higher than any thermophilic organism. The industrial competitiveness of thermophilic enzymes comes from their unique properties.

Novel and/or dual enzyme characteristics have been reported from a number of thermophiles. These include transferase activity (i.e. XynA from *C. saccharolyticum*) (see Section 1.5.2.3), and bifunctional enzymes, for example CelB from *C. saccharolyticum* (see section 1.5.2.2) and α-amylase-pullulanase from *Clostridium thermohydrosulfuricum* (Melasniemi 1987a).

A number of cleaning processes that involve a combination of chemicals and enzymes (usually proteases) need to use sub-optimal conditions, normally a reduced chemical strength, to prevent denaturation or inactivation of the enzymes. Thermophilic enzymes are more resistant to denaturing environments and chemicals than the corresponding mesophilic enzymes, possibly due to the harsh environmental conditions of thermophilic habitats (Owusu and Cowan 1990). Thus "full strength" chemicals can be used in combination with proteases more resistant to detergents and/or organic solvents (Cowan *et al.* 1982).
1.1.2.4 Purification of thermophilic enzymes expressed in mesophilic hosts.

One effective method of increasing enzyme production levels is to transfer thermophilic genes into mesophilic hosts, by genetic engineering. For enzymes expressed at low levels this gene transfer can increase enzyme production (Murray 1987). As most mesophilic proteins are non-thermostable, it is possible to purify the thermophilic enzyme by heat denaturation and consequent precipitation of the mesophilic proteins. Although not a complete purification procedure, this method is sufficient for many applications (Patchett et al. 1989, Takesawa et al. 1990).

1.1.3 Applications of thermophilic enzymes

Historically, most research on the commercial application of thermophilic enzymes centred on their use in producing alternative energy sources (Esser and Schmidt 1982, see section 1.1.3.2). Although these strategies were technically feasible, they were generally significantly more expensive than the energy source they were designed to replace and this was compounded by the additional decrease in oil prices during the late 1980's (Zacchi et al. 1988). Although the starch industry uses thermostable enzymes (see section 1.2), these are from mesophilic sources, while the only real commercialisation of thermophilic enzymes has been for use in laboratory research (see section 1.1.3.3). Another potential commercial application is the enzymatic bleaching of wood pulp and is discussed in section 1.1.3.1.

1.1.3.1 Enzymatic bleaching of wood pulp

The current industrial methods used to bleach paper pulp involve the use of “aggressive” oxidising chemicals (e.g. Cl₂ and SO₂) to bleach and extract lignin (brown coloured component of wood), which results in toxic by-products. Enzymatic bleaching of pulp involves removal of lignin (which is attached to cellulose fibres via hemicellulose) by enzymatic hydrolysis of the hemicellulolytic polysaccharides (Paice et al. 1988).

Hemicellulolytic enzymes are produced by a number of thermophilic bacteria (Donnison et al. 1989) and mesophilic fungi (Senior et al. 1991). Economically, chemical oxidation is still cheaper, but proposed changes to pollution laws caused by environmental concerns will favour enzymatic bleaching (Reid and Paice 1990).

1.1.3.2 Biomass fermentation

The purpose of biomass fermentation is to both eliminate waste (Romero et al. 1988, Clanet et al. 1988) and to produce commercially valuable end products. These products are normally volatile organic solvents (e.g. ethanol, methanol, butanol and acetic acid, Kurose et al. 1986), or combustible
gases (hydrogen and methane, Kondratieva and Gogotov 1983). This two-fold purpose greatly enhances both the economical and the social benefit (i.e. political will) of eliminating waste and using renewable resources.

1.1.3.3 Thermophilic enzymes used in laboratory research

The laboratory applications of thermophilic enzymes in most cases are due to their intrinsic thermostability, the major exceptions are restriction endonucleases where the uniqueness of their recognition site is the reason for their use. The greatest application of thermostable enzymes has been the use of *Thermus aquaticus* polymerase I, which made the polymerase chain reaction (PCR) practical. This method has not only revolutionised molecular biology but has also found applications in every field of biology (reviewed by Innis et al. 1990). Many restriction enzymes with unique recognition sites have been isolated from thermophilic sources. Of those commercially available, most are from thermophilic *Bacillus* sp. (e.g. *B. caldolyticus* BclI, and various strains of *B. stearothermophilus* BsaI, BsaAI, BsaBI, BsaJI, BsmAI, BsrI, BssHII, BsrBI, BstEII, BstNI, BstUI, BstXl and BstYI) and *Thermus* sp. (e.g. *T. aquaticus* TaqI, *T. filiformis* TfII and *T. thermophilus* Tth111I) (NEB 1990).

1.1.4 Protein thermostability

1.1.4.1 History and Introduction

Originally, three theories were proposed to explain the apparent thermostability of thermophilic proteins/enzymes. The first of these was based on the erroneous belief that thermophilic enzymes were not thermostable but appeared to be so because of a faster protein synthesis capability, thus allowing an effective enzyme concentration to be maintained.

The second theory proposed that thermostability was due to the presence of “stabilising co-factors” in the bacteria that prevented thermal denaturation of the enzyme. Studies on purified thermophilic enzymes have shown that in common with mesophilic enzymes “stabilising co-factors” (e.g. metal ions and polar alcohols) have only a minor role in increasing thermostability (approximately 5-10°C, Gianfreda and Scarfi 1991). Later work has shown that in some thermophilic proteins, amino acids are modified, which increases their thermostability (Maras et al. 1992).

The final theory stated that thermophilic enzymes are intrinsically thermostable due to their primary amino acid sequence. This theory was shown to be correct by experiments where thermophilic genes were transferred to mesophilic hosts and expressed as thermostable enzymes (see section 1.1.2 and 1.1.3).
Having established that thermophilic proteins were intrinsically thermostable, investigation shifted toward the mechanisms of thermostability. The experimental approaches taken have included random and site-directed mutagenesis (see section 1.1.4.2), statistical analysis (see section 1.1.4.3), tertiary structure of model systems (see section 1.1.4.4) and chimeric gene construction (see section 1.1.4.5).

1.1.4.2 Random and site-directed mutagenesis

Mutagenesis studies using random mutagenesis have involved replacing amino acids in a random fashion either individually or in groups and determining the effect on protein thermostability. Most changes have had either a negative effect or no effect at all, for example, out of 66 mutants produced by random chemical mutagenesis of E. coli trpA gene (Lim et al. 1992) 18 mutants had reduced thermostability whereas only 6 showed increased thermostability (+1.4 to +7.6°C). Using site-directed mutagenesis, amino acid residues that were thought to destabilise protein structure (i.e. α-helixes or salt bridges) were replaced with residues that stabilised these protein structural elements (Strehlow and Baldwin 1989). Although these changes stabilised the individual protein structural element, there was normally minimal overall change, or even a slight decrease in total protein stability, due to the disruptions caused to the rest of the protein. For example Ganter and Plückthun (1990) substituted glycine residues with alanine in helical regions of glyceraldehyde-3-phosphate dehydrogenase and found only one mutation that resulted in an overall increase in stability. Where mutations have been found to be individually effective, it has been possible to combine them and increase thermostability in an additive fashion (Matsumura et al. 1986, Garvey and Matthews 1989).

1.1.4.3 Statistical analysis

Statistical analysis has involved comparison of the amino acid sequences of evolutionarily well-conserved proteins in thermophiles and mesophiles. The major limitation of this method was the requirement of a large number of amino acid sequences.

This method has been successfully applied to L-lactate dehydrogenase (Zulli et al. 1987 and Zulli et al. 1990) and 3-isopropylmalate dehydrogenase (Imada et al. 1991). By determining which amino acid residues were different over the entire protein it was possible to formulate some general rules regarding protein thermostability. Using the amino acid sequences of glyceraldehyde-3-phosphate dehydrogenase, lactate dehydrogenase and ferredoxin from thermophiles and mesophiles, Argos et al. (1979) formulated the following observed replacement rules (figure 1.1).
Figure 1.1 Direction of observed preferred exchanges

![Diagram showing amino acid exchanges](image)

Arrows point from the mesophilic to the thermophilic protein. Numbers indicate the ranking of the significance for the given exchange. (Adapted from Argos et al. 1979).

These changes caused several effects, (i) increased volume of amino acids in thermophilic proteins resulting in a more compact structure, (ii) stabilised alpha helical structures and (iii) decreased external hydrophobicity and increased internal hydrophobicity.

Using regression analysis, Watanabe et al. (1991) have shown that for certain enzyme groups the proline content increases with increases in thermostability.

1.1.4.4 3-dimensional model systems

This method determined the contribution of individual protein secondary structural elements to overall protein thermostability (i.e. α-helices, β-sheets, salt bridges and disulphide bonds). Small proteins of known 3-dimensional structure (i.e. T4 lysozyme, Matthews and his co-workers; and λ repressor protein, Stearman et al. 1988) were used as model systems. This knowledge of the 3-dimensional structure allowed the precise study of the effects that different amino acid replacements had on individual protein secondary structural elements.

As expected, Matthews and his co-workers found that certain amino acid changes increased thermostability both singly and cumulatively. These changes included those that stabilised α-helices (Nicholson et al. 1988), increased the number of salt bridges (Dao-pin et al. 1991), added disulphide bonds (Matsumura et al. 1989), and added internal hydrophobic residues (Eriksson et al. 1992).

Previous studies (Matthews 1987) had shown that if the overall 3-dimensional structure was altered...
1.1.4.5 Chimeric genes

Chimeric genes, and hence proteins, may be constructed from both mesophilic and thermophilic DNA of the same gene from related species. The classic example is the work of Suzuki et al. (1989) using chimeric genes of two Bacillus α-amylase genes (the non-thermostable B. amylooliquefaciens and the thermostable B. licheniformis). As these genes shared considerable DNA sequence homology it was possible to exchange segments from each gene using restriction endonucleases. Analysis of the temperature stability of each hybrid enzyme from each chimeric gene showed that two regions of the sequence were important in the determination of thermostability. A glutamine at position 178 (region I) and residues 255-270 (region II) determined the relative thermostability of the α-amylases. The following changes, that followed Argos' rules for substitution (see figure 1.1), were shown to affect thermostability both independently and additively. Deletion of arginine-176 and glycine-177 in region I and substitution of alanine for lysine-269 and aspartic acid for asparagine-266 in region II, by site-directed oligonucleotide mutagenesis of the B. amylooliquefaciens α-amylase gene, enhanced thermostability.

Of considerable interest in this study was the finding that the hybrid B. amylooliquefaciens enzyme was as stable as the α-amylase from B. licheniformis with respect to irreversible denaturation at 90°C, although its temperature optimum was unchanged (65°C) and the hybrid enzymes were still susceptible to reversible inactivation at temperatures above 65°C. Thus only one facet of thermostability was altered.

Chemical studies have shown the mechanism of irreversible thermoinactivation to be a two-step process. Peptide unfolding (reversible denaturation) allows deamidation of asparagine/glutamine residues resulting in irreversible thermoinactivation (Ahern and Klibanov 1988). Using a statistical analysis method, Klibanov and his co-workers (Tomazic and Klibanov 1988a,b) proposed that the thermostability differences between the α-amylase genes of B. amylooliquefaciens and B. licheniformis were due to the presence of the amino acid residues lysine-385, lysine-88 and/or lysine-253 in B. licheniformis, and that these amino-acid residues formed salt bridges that reduced peptide unfolding. The experimental work of Suzuki has shown that this conclusion is incorrect.

In a similar experiment the C-terminal region (42 amino acids) of glyceraldehyde-3-phosphate from Methanobacterium bryantii (mesophile) was exchanged for the C-terminal region from the thermophile Methanothermus fervidus with a resulting 9°C increase in thermostability (Biro et al. 1990).
1.2 Starch

1.2.1 Chemical structure of starch and related carbohydrates

In plants, starch is stored in a crystalline structure known as a starch granule. This granule is composed of two distinct molecules, amylose (see section 1.2.1.1) and amylopectin (see section 1.2.1.2). Not only does the granule size vary between plant species and cultivars, but the ratio of amylose to amylopectin also changes. Wildtype species normally have approximately 20-30% amylose. For commercial purposes special cereal cultivars have been developed with amylose percentage compositions ranging from nearly 0% -75% (Shannon and Garwood 1984).

1.2.1.1 Amylose

Amylose molecules are long, unbranched chains of about a hundred to several thousand α-D-glucopyranose units linked by α-1,4-glycosidic bonds (see figure 1.2). X-ray diffraction studies have shown that amylose molecules in solution have a helical structure with a diameter of 1.3 nm and six successive glucose residues per turn. Amylose molecules are soluble in hot water but the resulting solution is unstable and eventually spontaneous precipitation (known as retrogradation) occurs. Retrogradation is irreversible and arises because of the tendency of the long amylose molecules to align themselves side by side by hydrogen bonding, forming insoluble aggregates.

Figure 1.2 Amylose structure

Introduction
1.2.1.2 Amylopectin

Amylopectin molecules have highly branched tree-like structures (see figure 1.3) containing up to 50,000 α-D-glucopyranose residues that are predominantly linked by α-1,4-glycosidic linkages. However, at each branch point an α-D-glucopyranose residue is also involved in an α-1,6-glycosidic linkage (via its C-6 hydroxyl oxygen atom to C-1 of the first glucose residue of the branch). The α-1,6-glycosidic linkages constitute about 5% of the total glycosidic linkages in amylopectin molecules. Amylopectin chains are classified as either A, B, or C chains and each amylopectin molecule has only one reducing end (Hizukuri and Maehara 1990, Bertoft 1991).

Figure 1.3 Amylopectin structure

![Amylopectin Structure](image)

C-chains contain the reducing end of the molecule and are linked to the reducing end of B-chains. B-chains (average length 40-60) carry one or more chains as branches, whereas A-chains (average length 11-20) have no branches.
1.2.1.3 Pullulan

Pullulan is a linear water soluble α-glucan polymer consisting mostly of 1,6-α-D-maltotriose and some 1,6-α-D-maltotetraose repeating units (Catley 1979). It is produced from starch sources by the fungi Pullularia pullulans and Aureobasidium pullulans as long chains of about 480 units (Yuen 1974). The solubility of pullulan is higher than that of amylose and amylopectin, with lower viscosity due to the additional flexibility that the α-1,6 bonds confer to the polymer. It has several potential applications in cosmetics, diet foods, medicine and manufacturing industries as a flocculant, adhesive and binder. Being biodegradable, it can be used as a substitute for petroleum-based packaging and plastic containers (Mulchandani et al. 1989). A schematic view of the structure of pullulan is shown in figure 1.4.

**Figure 1.4 Structure of pullulan**

![Structure of pullulan](image)

Symbols: ○ glucose unit; ● glucose unit with reducing end; — α-1,4-bond; ↓ α-1,6-bond

1.2.1.4 Cyclodextrin

Cyclodextrins (Schardinger or cycloamylose) are rings formed by α-1,4-glucosidically bound glucose units, with the most common ring sizes being 6, 7 and 8 glucose units, known as α-, β-, and γ-cyclodextrins respectively. On circularisation, a hydrophobic cavity is formed which is able to form inclusion complexes with a number of organic and inorganic molecules.
1.2.2 The starch industry

1.2.2.1 Economic uses and value of starch products.

Starch has always played an important part in the human diet. Historians connect the emergence of plant cultivation, and hence starch usage, with the transformation of man from a hunter-gatherer to a farmer. Although originally only a food source (flour), the ancient Egyptians used starch size in the production of papyrus paper. Starch was a whitening agent for cloth and hair in the Roman civilisation, and a linen stiffener in Northern Europe during the middle ages. In 1811 it was discovered by Kirchoff that dilute acid could hydrolyse starch to produce a sweet substance, later shown to be oligosaccharides and glucose. Today, world-wide, most starch is still used as human and animal food (reviewed by French 1975).

The properties that make starch industrially useful are its particulate nature, its high viscosity on gelatinisation and its gelling characteristics. These properties make it possible for starch to be used in various forms; as intact granules or powder (for example a dusting agent), as a swollen starch granule in food systems where high viscosities are required, as a molecule in solution (for example, in manufacture of paper products, as a flocculant and an aid in pigment retention), and as a dried film in paper coating or in sizing of textile fibres (reviewed by Radley 1976). Although these properties are inherent in natural starch, chemical modification by either oxidation, esterification or etherification can improve them (Rutenburg and Solarek 1984).

Since starch-based polymers are an abundant, renewable, biodegradable product, they can be used as a replacement for petroleum-based polymers in many applications (e.g. a filler for PVC and PVA plastics). Also linked to the availability and price of petroleum fuels has been their use as a feedstock for production of liquid fuels by hydrolysis and fermentation (de Menezes 1982). Although industrially feasible, the cost is prohibitive when compared to the current petroleum products, the potential yield from cellulase and hemicellulase-based fermentations and increasing world food shortages (Parisi 1983).

1.2.2.2 Economic uses and value of starch hydrolysis products

Sweet substances have always been a highly valued commodity, with the desire for sweet tastes being an inherent human trait. Since honey and fruit (both fructose based) were the only readily available naturally occurring sources, this demand led to the commercial growth and extraction of sucrose from sugar beet and sugar cane in the seventeenth and eighteenth centuries. The present day corn wet milling industry began with the discovery that acid hydrolysis of starch leads to a glucose syrup. The major product was a 42 DE (dextrose equivalent) acid converted corn syrup, which, although useful in many applications, lacked the sweetness of sucrose due to incomplete hydrolysis. When attempts were made to acid convert corn syrup to a higher degree of hydrolysis, the resulting
products were often highly discoloured and contained bitter by-products from acid mediated side-reactions (Lloyd and Nelson 1984).

In the 1940's Dale and Langlois patented the use of amylase (principally α-amylase but containing contaminating β-amylase and α-glucoamylase) from *Aspergillus oryzae* (Taka amylase) for the enzymatic hydrolysis of starch either singly or in combination with acid hydrolysis to produce a wide variety of saccharide mixtures. This process resulted in the economically viable production of crystalline glucose for use in the food industry by the 1960's (Robyt 1984).

Syrups from corn starch hydrolysis are used industrially as added sweeteners. For the manufacturer to increase the final sweetness of his products, he has to either add more syrup or use a "sweeter" sugar. The relative sweetness of different sugars is well documented (Palmer 1975). On the basis of an arbitrary value of 100 relative sweetness for sucrose, the corresponding values for maltose, glucose and fructose are 32, 74 and 173 respectively. As corn syrups are predominantly glucose and maltose, and their relative sweetnesses are significantly less than sucrose, this severely limits commercial utilisation since the addition of more syrup not only adds more calories and is more expensive, but most products are already at maximum sugar concentration.

As glucose and fructose are structural isomers (along with many other monosaccharides), industrial scale isomerisation of glucose to fructose was seen as a way of producing a "sweeter" sugar. Initially, chemical isomerisation based on acid catalysis was attempted but all attempts failed to produce fructose of sufficient quantities and purity. Although glucose isomerase does not exist in nature it was found that D-xylene-ketone isomerase (E.C. 5.3.1.5) not only converted D-xylene to D-xylulose, but also D-glucose to D-fructose to a limited extent (Yamanaka 1968). The discovery and immobilisation of heat stable isomerases from *Bacillus coagulans* allowed the commercial production of HFCS (high fructose corn syrup) (Zittan *et al*. 1975, Oestergaard and Knudsen 1976).

The economic value of starch hydrolysis products is measured in billions of dollars per annum (estimated at over US$3 billion in 1982) with HFCS constituting about 2/3 of the starch hydrolysis market (a volume of 7 million tonnes) or one third of the total sweetener market in 1984 (Coker and Venkatasubramanian 1987). These production levels led to the development of an allied industry producing carbohydrases. The consumption of carbohydrases in the U.S.A in 1985 was valued at US$69 million (including glucose isomerase US$28 million, glucoamylase US$21 million, α-amylase US$11 million, World Biotech 1986).

1.2.2.3 Current hydrolysis methods

While it might seem more logical to use an unbranched feedstock (e.g. amylose), amylopectin is preferred as it is more soluble and does not retrograde (see section 1.2.1.1). The current method for the production of HFCS using starch with a high amylopectin content is shown schematically in figure 1.5.
In some specialist applications where high purity maltose is required, β-amylase (normally from *Bacillus* sp.) is used before liquefaction to produce maltose and a beta-limit dextrin. For lower grade applications, thermostable fungal α-amylases from *Aspergillus* sp. or *Actinomycetes* sp. can be used due to their transglycosylation reaction mechanism that results in maltose as the major hydrolysis product after prolonged incubation, as shown in figure 1.6 and table 1.4.
**Figure 1.6 Production of maltose syrups**

Starch Suspension
30%-40%

- Liquefication (bacterial α-amylase)
  - Saccharification (β–amylase or fungal α-amylase)
    - High maltose syrups
- Liquefication (bacterial α-amylase)
  - Saccharification (β–amylase or fungal α-amylase, pullulanase or isoamylase)
    - Extremely high maltose syrups
- Liquefication (acid/bacteria α-amylase)
  - Saccharification (β–amylase or fungal α-amylase, glucoamylase)
    - High conversion syrups

Modified from Saha and Zeikus 1987.

**Table 1.4 The composition of maltose syrups (%) saccharide, carbohydrate basis**

<table>
<thead>
<tr>
<th>Saccharide</th>
<th>High maltose syrups</th>
<th>Extreme (extra) high maltose syrups</th>
<th>High conversion syrups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>0.5-3</td>
<td>1.5-2</td>
<td>35-43</td>
</tr>
<tr>
<td>Maltose</td>
<td>45-60</td>
<td>70-85</td>
<td>30-47</td>
</tr>
<tr>
<td>Maltotriose</td>
<td>10-25</td>
<td>8-21</td>
<td>8-15</td>
</tr>
<tr>
<td>Higher Saccharides</td>
<td>Balance</td>
<td>Balance</td>
<td>Balance</td>
</tr>
<tr>
<td>DE</td>
<td>35-50</td>
<td>45-60</td>
<td>60-70</td>
</tr>
</tbody>
</table>

Modified from Saha and Zeikus 1987

**1.2.2.4 Usefulness of thermophilic enzymes**

As previously mentioned all the products required from starch hydrolysis can be produced with existing technology and enzymes. Before a thermophilic or thermostable enzyme could be used it would need to show either a significant cost saving in terms of time, raw materials, plant construction or running expenses, or an improvement in final product quality. This is demonstrated by the replacement of the original α-amylase (Taka amylase) with the thermostable *Bacillus subtilis* α-amylase, which in turn was replaced by the extremely thermostable *Bacillus licheniformis* α-amylase that was able to operate at 105°C (Madsen *et al.* 1973).
As starch solubility increases and viscosity decreases significantly at higher temperatures, the major cost advantages of a thermostable enzyme would be the double benefit of lower energy costs of cooling reactor vessels and higher enzymatic reaction rates. These advantages would be considerably reduced if all hydrolysis enzymes were not thermostable.

1.2.3 Usefulness of a thermostable debranching enzyme

1.2.3.1 Theory and practice in amyllopectin hydrolysis

Incomplete hydrolysis, caused by the branched (α-1,6-bonds) structure of amyllopectin, is the major limiting factor in starch hydrolysis. Enzymatic liquefaction of amyllopectin by α-amylase results in hydrolysis of internal α-1,4-bonds, producing a maltodextrin mixture (DE 5) of mostly linear chains with about 25% branched dextrins (see Figure 1.7) (Atkins and Kennedy 1985). These branched dextrins have to be either removed by precipitation/filtration (wasteful in raw material and costly in production expenses and time) or debranched (hydrolysis of α-1,6-bonds) (Allen and Dawson 1975).

Figure 1.7 Amylopectin hydrolysis

- Amylopectin
- Branched dextrins
- Glucose

Modified from Jensen and Norman 1984.
Enzymes that can hydrolyse \( \alpha-1,6 \)-bonds include glucoamylase, at a very slow rate (Meyers and Wasserman 1985); isoamylase, faster but not to completion; and pullulanase, fast and to completion. As glucoamylase is already used for saccharification of maltodextrins, it can be used for the slow hydrolysis of \( \alpha-1,6 \) bonds, but it is more efficient to use a pullulanase or isoamylase in tandem with a saccharifying enzyme (glucoamylase, \( \alpha \)-amylose or \( \beta \)-amyrase) (Slominska and Maczynski 1985), as well as giving a faster rate of hydrolysis this addition increases the final product purity, as shown in figure 1.8. The greatest advantage of this increase in product purity and final degree of hydrolysis (only a few percentage points) is in the rate of enzymatic isomerisation of glucose to fructose. The rate of isomerisation is effectively dependent on the concentration of inhibition products, i.e. isomaltose and other reversion sugars.

Figure 1.8 Effect of pullulanase on amylopectin hydrolysis

Currenty there are two debranching enzymes commercially available, an isoamylase from *Pseudomonas amylofera* SB15 (Japan) (Harada 1984) and a pullulanase from *Bacillus acidopullulyticus*, tradename Promozyme\textsuperscript{TM} 200L, marketed by Novo Industri A/S (Denmark) (Jensen and Norman 1984). Both have been shown to increase the degree of hydrolysis of starch (amylopectin) and increase the purity of the final product. The pullulanase also has the advantage of being
moderately thermostable, with a temperature optimum of 50-55°C. Another proposed hydrolysis strategy is to use a pullulanase together with a thermostable β-amylase to produce a high maltose conversion syrup (DE 45-60, 35-43% glucose, 30-47% maltose) (Shiraishi et al. 1987, Saha and Zeikus 1990).

1.2.3.2 Useful properties of debranching enzymes

With the potential application of debranching enzymes (especially thermophilic), there are a number of references in the literature to enzymes that can either hydrolyse pullulan or debranch amylopectin (see section 1.4 for enzymatic substrate classification). As neopullulanase and isopullulanase preferentially hydrolyse α-1,4-bonds over α-1,6-bonds and only hydrolyse starch to a limited extent, their potential in starch hydrolysis is limited. Isoamylase debranches amylopectin incompletely and it is also of limited use until a thermostable enzyme has been found. Hybrid enzymes that combine both α-1,6-and α-1,4-hydrolysis have been reported from many thermophilic organisms (see section 1.4). Although these hybrid enzymes have been purified to homogeneity (determined by HPLC or SDS-PAGE), none have been crystallised, so it is possible that they still contain some contaminating amylase. In one case, Melasniemi and Paloheimo (1989) have purified, expressed in E. coli and sequenced, a totally new hybrid enzyme, α-amylase-pullulanase. This enzyme has a substrate profile identical to the enzyme isolated from the native host. Amylopectin hydrolysis using these hybrid enzymes is not efficient either singly or with other amylases, since they have a higher affinity for α-1,4-bonds than α-1,6-bonds. Also, their rate of hydrolysis is slower than other amylases, effectively protecting the α-1,4-bonds from the amylases. Hence, when they are used in tandem hydrolysis, they decrease the overall hydrolysis reaction rate. Compared to pullulanases with α-1,4-activity, a true pullulanase (defined as enzymatically similar to Klebsiella pullulanase, see section 1.4) would be an ideal enzyme for amylopectin debranching, this is especially the case with a thermophilic pullulanase.
**Figure 1.9 Enzymes used in the starch industry**

- **Starch Slurry**
  - pH 6.5-7.0
  - 30-40% w/w solids
  - \( \text{Ca}^{2+} \) added

- **Bacillus subtilis**
  - \( \alpha \)-amylase

- **Cook 5 min, 140°C**
  - **Saccharify 55°C, pH 4.5, 24-48 h**
    - Maltose syrups, DE 40-48

- **Cook 7 min, 105°C**
  - Dextrinise 95°C 1 h
    - **Aspergillus oryzae**
      - Fungal \( \alpha \)-amylase
      - Cereal \( \beta \)-amylase
    - **Aspergillus niger**
      - Amyloglucosidase
    - **Aspergillus oryzae**
      - Fungal \( \alpha \)-amylase and
      - Amyloglucosidase

- **Saccharify 60°C, pH 4.0, 48-96 h**
  - Dextrose (glucose syrup) DE 96-98
  - Crystalline glucose
  - High conversion syrups (maltose + glucose) DE 58-68
  - Immobilised glucose isomerase reactor
  - 'Isoglucose' glucose + fructose

- **Maltodextrin products, DE 3-13**
- **Maltodextrin products, DE 10-20**
1.3 Genetic Analysis of Pullulanases and Related Genes

1.3.1 *E. coli* maltose regulon

Research carried out on the *E. coli* maltose regulon began in the late 1940's with the discovery that amylomaltase (*mal*Q) is only expressed in cells grown on maltose. Since then 10 genes in 5 operons have been detected, sequenced and shown to be involved in the transport of maltodextrins into the cytoplasm and their hydrolysis therein. The expression levels of all operons are positively regulated by the concentration of maltodextrins in the media, and catabolite repressed by glucose (through cyclic AMP-binding protein) (Reviewed by Schwartz 1987). It is assumed that the *malT* gene product exists in two distinct conformations, *MalT*(a) (activated) which stimulate expression of all the maltose operons, whereas *MalT*(i) (inducible) can not. When maltodextrins are present *MalT*(i) is comformationly changed into *MalT*(a) which binds to the "promoter" region of each *mal* operon stimulating transcription (see figure 1.10).

**Figure 1.10 E. coli maltose regulon**

Mutation studies, deletion studies and consensus sequence analysis of the promoter region of all maltose operons have shown that a highly conserved six base pair sequence, 5'-GGAGGA-3'
(maltose box) is the site where the MalT(a) protein acts. This maltose box is located 34-35 base pairs upstream of the transcriptional start point. As with most positively regulated genes, a strong -10 RNA polymerase promoter and a weak -35 RNA polymerase promoter are present in all maltose operons (Richet and Raibaud 1989). As more promoter regions have been sequenced, later work has proposed a larger maltose box, 5'-GGGAGCTAGG-3' (Vidal-Ingigliardi et al. 1991). All MalT(a) binding sites determined so far are 70-100% homologous to this new consensus. Transcription studies in vitro with purified MalT(a) have shown that maltotriose is the only dextrin that induces MalT activation, whereas all maltodextrins, including maltose, act as inducers in vivo (Richet and Raibaud 1991). This result implies that there are other genes involved in down regulation of the maltose system and transport into the cell, three of which have been isolated and sequenced (malI (repressor), malX (phosphotransferase) and malY (regulator of malK and malT). Although not transcriptionally activated by MalT(a), expression of all three is regulated by the malI gene product (Reidl et al. 1989, Reidl and Boos 1991).

### 1.3.2 Klebsiella maltose regulon

A closely related Enterobacteriaceae genus, Klebsiella, possesses a maltose regulon of similar genetic structure that allows full complementation of Klebsiella maltose genes when expressed in E. coli. The genes of each operon show sequence similarity of up to 90% (Dahl et al. 1989, Francoz et al. 1990, Schneider et al. 1989, Bloch and Raibaud 1986).

In addition to the standard operons of the maltose regulon, Klebsiella possesses two extra operons, one coding for the pullulanase enzyme (pulA) and the other for those genes involved in its secretion and export from the cell. Murooka transferred a 5.0 kb piece of genomic DNA (Klebsiella aerogenes W70) into E. coli that expressed pullulanase on maltose induction but the enzyme remained in the cytoplasm (Takizawa and Murooka 1984, Takizawa and Murooka 1985). Pugsley then showed that a 19.2 kb fragment of genomic DNA (from Klebsiella oxytoca UNF5023, formerly Klebsiella pneumoniae) was needed for maltose inducible pullulanase expression, modification and cell-surface exposition (Chapon and Raibaud 1985, Pugsley et al. 1986). Transposon mutagenesis confirmed the location of the pullulanase gene and showed that the remaining DNA was required for pullulanase expression (d’Enfert et al. 1987).

Both Murooka (Katsuragi et al. 1987) and Pugsley (Kornacker and Pugsley 1990) have sequenced their respective pullulanase gene (pulA). They are nearly identical, with the leader sequence being the only major sequence difference. The reported sequence differences are discussed in appendix 6. Pugsley has also sequenced the other genes of the pullulanase export operon (Pugsley and Reyss 1990,
Reyss and Pugsley (1990). On sequencing, two divergent operons were detected, one coding for the pullulanase enzyme PulA and a gene of unknown function (*pulB*), and the second for 13 genes *pulC-pulO*. All 13 genes of the second operon (13 kb) are essential for pullulanase secretion and exposition on the cell surface (see figure 1.11).

**Figure 1.11** Characterisation of *Klebsiella pullulanase* genes

<table>
<thead>
<tr>
<th>Transcripts</th>
<th>Stems-loops and direct repeats</th>
<th>Predicted Protein (kDa)</th>
</tr>
</thead>
</table>

Shown are the organization of the pul transcription units, the location of palindromes, and direct repeats located at the 3' end of *pulO*. The location of the two MalT-dependent promoters (●) are shown. Transposon insertions have shown that all genes (except for *pulB*) are required for pullulanase exposition on the cell surface.

Experiments with *E. coli* secretion-negative mutants have shown that six *sec* genes (*secA, secB, secD, secE, secF, secY*) from the general export pathway are required for processing the prepullulanase signal peptide. Another pullulanase secretion gene, *pulS*, has been detected by transposon mutagenesis. Although not part of the maltose regulon and not MalT(*a*) induced, it is required for pullulanase exposition in *E. coli* (d'Enfert and Pugsley 1989, Pugsley *et al.* 1990, Pugsley *et al.* 1991) (see figure 1.12).
Figure 1.12 *Klebsiella* pullulanase export

Step 1: Export  Step 2: Exposition  Step 3: Release

Medium

Spontaneous?

Fatty acids  Outer Membrane

Cytoplasm

Periplasm

Inner Membrane


Model showing the location of pullulanase secretion proteins. The depicted association to form a trans-envelope complex is hypothetical.

Modified from Pugsley et al. 1991

1.3.3 Thermophilic enzymes

Many extracellular glucolytic and amylolytic enzymes have been identified from thermophilic organisms, see section 1.4 (Koch *et al.* 1987, Plant *et al.* 1987a, Goldberg and Edwards 1990), and a number have been expressed (Kitamoto *et al.* 1988, Honda *et al.* 1988, Haeckel and Bahl 1989) and sequenced (Nakajima *et al.* 1985, Kawazu *et al.* 1987), but not a single homolog to the maltose or pullulanase enzyme export pathway or oligosaccharide import pathway has been detected. Homologs
must exist as most thermophilic glucoalytic enzymes are positively regulated by substrate induction. Putative maltose and glucose permease have been proposed for Clostridium thermosulfurogenes and Clostridium thermohydrosulfuricum respectively (Hyun et al. 1985)

Enzymes shown to be substrate regulated are α-amylases with amylose (Brown et al. 1990), β-amylase with soluble starch (Nipkow et al. 1989), glucoamylase and pullulanase (Hyun and Zeikus 1985a) and pullulanase with starch (Odibo and Obi 1990). The level of induction is as high as tenfold above basal level. Enzymatic expression also shows a characteristic glucose repression profile. In Clostridium thermohydrosulfuricum, it has been shown that the ratio of amylases (α-amylase, pullulanase and β-glucosidase) cell-associated or exported into the medium, varies depending on the starch source. With the less soluble “soluble starches” nearly all the α-amylases and pullulanases are cell-associated, whereas using more “soluble starches” (e.g. Zulkowsky or pullulan) the ratio is approximately 1:10 associated/exported (Melasniemi 1987b).
1.4 Pullulanase and Related Enzymes

Starch polymers are the major energy storage system in the plant kingdom, although not as abundant as the structural polymers of cellulose and hemicellulose. Hence most organisms possess an efficient enzymatic system for hydrolysing starch to utilisable saccharides (see figure 1.13).

Figure 1.13 Hydrolysis patterns of some amylolytic enzymes

Symbols: • glucose unit;
         ○ glucose unit with reducing end;
         -- α-1,4-bond;  ↓ α-1,6-bond

Modified from Vihinen and Mäntsälä 1989
Enzymes which hydrolyse starch can be divided into three groups based on their distribution throughout nature. The first group are apparently ubiquitous to all organisms (e.g. α-amylase and α-glucosidase), while the second are common within a particular group (e.g. β-amylase is found predominantly in plants, and glucoamylase in fungi). Those occurring in only a few species (e.g. pullulanases from *Klebsiella* sp, isoamylases from *Pseudomonas* sp and cyclodextrin glycosyltransferases from *Bacillus* sp.) form the third group.

Another more practical division can be made on the mode of enzyme action. Endoenzymes hydrolyse internal glycosidic bonds in an apparently random fashion, whereas exoenzymes hydrolyse from the non-reducing end releasing constant end products and resulting in successively shorter starch molecules. A further subdivision can be made according to enzymatic bond specificity and final hydrolysis products, e.g. debranching enzymes can hydrolyse α-1,6 bonds (Marshall 1975).

### 1.4.1 Non-branching enzymes

#### 1.4.1.1 α-amylase (E.C. 3.2.1.1) 1,4-α-D-Glucan glucohydrolase

α-amylases occur widely in bacteria, fungi and animals. They are endo-acting enzymes that specifically cleave 1,4-α-D glycosidic bonds in amylose, amylopectin, glycogen and long-chained α-1,4-oligosaccharides, releasing oligosaccharides in the α-configuration. They are unable to hydrolyse α-1,6-branch points in amylopectin and glycogen, but are able to by-pass them since they are endo-acting. Hydrolysis of starch causes a rapid decrease in solution viscosity as the high Mr polysaccharides are hydrolysed forming oligosaccharides. These oligosaccharides are in turn hydrolysed to smaller saccharides but at a substantially slower rate (Saito 1973). α-amylases are divided into two categories according to the degree of final substrate hydrolysis, saccharifying α-amylases hydrolyse 50-60% of the bonds in starch (DE 50-60) whereas liquefying α-amylases hydrolyse 30-40% (DE 30-40) (Cowan 1991).

The average molecular mass of α-amylases is approximately 50 kDa, with a pH activity range of 3.5-7 and a temperature optimum of 35-90°C. Many α-amylases are calcium metallo-enzymes containing at least one calcium atom per molecule of enzyme while others have been shown to be stabilised by calcium ions (Yutani 1976).

A large number of α-amylases have been isolated and characterised. A recent review of microbial α-amylases lists 208 (Vihinen and Mäntsälä 1989) and a number of these have been sequenced, with 47 deposited in the Genebank DNA sequence database (Release #70).
1.4.1.2 β-amylase (E.C. 3.2.1.2) 1,4-α-D-Glucan maltohydrolase

β-amylase releases maltose by endohydrolisis of the penultimate 1,4-α-D glycosidic bond, from the non-reducing end of α-1,4-glucans. On unbranched substrates (i.e. containing only α-1,4-bonds) maltose is the major final product of hydrolysis, but on branched substrates (i.e. also contain α-1,6-bonds) the major final hydrolysis products are maltose and a β-limit dextrin, since β-amylase cannot hydrolyse 1,6-α-D glycosidic bonds, and cannot by-pass them as it is an endohydrolising enzyme. Although widespread in plants, only a few microbial examples are known (Hyun and Zeikus 1985b, Kitamoto et al. 1988).

1.4.1.3 Glucoamylase (E.C. 3.2.1.3) 1,4-α-D-Glucan glucohydrolase

Glucoamylase successively hydrolyses the terminal glucose residue from the non-reducing end of α-1,4-glucans releasing β-D-glucose as a final hydrolysis product. It is able to hydrolyse both 1,4-α-D and 1,6-α-D glycosidic bonds, although the rate of hydrolysis of α-1,6-D bonds is extremely slow (Ghosh et al. 1991). This enzyme is distributed widely throughout fungi with only a few known bacterial species (Pretorius and Lambrechts 1991).

1.4.1.4 Alpha glucosidase

Glucosidases are ubiquitous in nature and are involved in the hydrolysis of oligosaccharides resulting from the digestion of polysaccharides. They release either α-D-glucose, if endo-acting, or smaller oligosaccharides if exo-acting. Included in this group are enzymes that hydrolyse α-1,6-bonds, those that hydrolyse α-1,4-bonds and those that have dual activity. Glucosidases may be intracellular or extracellular, and some are capable of both hydrolytic and transglycosylation reactions. They display a wide pH optimum (3-7.5) and a temperature optimum and thermostability related to the growth temperature of their host organism. Based on their substrate profile and mode of enzymatic hydrolysis the following enzyme classes are recognised (IUB Enzyme Nomenclature 1984);

Oligo-1,6-glucosidase (E.C. 3.2.1.10). Hydrolyses 1,6-α-D-glucosidic linkages in isomaltose and dextrins.

Dextranase (E.C. 3.2.1.11). Endohydrolyses 1,6-α-D-glucosidic linkages in dextran.

α-Glucosidase (E.C. 3.2.1.20). Hydrolyses terminal non-reducing 1,4-linked α-D-glucose residues.

Glucan 1,6-α-glucosidase (E.C. 3.2.1.70). Hydrolyses successive glucose residues from 1,6-α-D-glucans and derived oligosaccharides.

Glucan 1,6-α-isomaltosidase (E.C. 3.2.1.94). Hydrolyses 1,6-α-D-glucosidic linkages in polysaccharides removing successive isomaltose units from the non-reducing end.

Dextran 1,6-α-isomaltotriosidase (E.C. 3.2.1.95). Hydrolyses 1,6-α-D-glucosidic linkages in dextrans removing successive isomaltotriose units from the non-reducing end.

Introduction
In Table 1.5 is a summary of some representative characterised glucosidases.

Table 1.5 Properties of Alpha glucosidases

<table>
<thead>
<tr>
<th>Strain</th>
<th>Growth Temp °C</th>
<th>Cellular Location</th>
<th>pH Opt</th>
<th>Temp Opt °C</th>
<th>Thermo-stability</th>
<th>Mr kDa</th>
<th>Inhibitors</th>
<th>Miscellaneous</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Bacillus caldolyticus</em> C2 (E.C. 3.2.1.20)</td>
<td>-</td>
<td>I</td>
<td>7.5</td>
<td>65-68</td>
<td>&gt;1h at 65°C</td>
<td>55</td>
<td>Ca²⁺</td>
<td>Recombinant</td>
<td>Krohn and Lindsay 1991</td>
</tr>
<tr>
<td><em>Bacillus thermoglucosidasius</em> KP1006 (E.C. 3.2.1.10)</td>
<td>-</td>
<td>I</td>
<td>6.8</td>
<td>60</td>
<td>-</td>
<td>60</td>
<td>-</td>
<td>-</td>
<td>Watanabe <em>et al.</em> 1989</td>
</tr>
<tr>
<td><em>Streptococcus mutans</em> (E.C. 3.2.1.11)</td>
<td>37</td>
<td>I</td>
<td>6.0</td>
<td>37</td>
<td>-</td>
<td>62.103</td>
<td>-</td>
<td>Sequenced</td>
<td>Russell and Ferretti 1990</td>
</tr>
<tr>
<td><em>Bacillus cereus</em> ATCC7064 (E.C. 3.2.1.10)</td>
<td>30</td>
<td>I</td>
<td>7.5</td>
<td>35</td>
<td>-</td>
<td>66.010</td>
<td>-</td>
<td>Sequenced and expressed in <em>E.coli</em></td>
<td>Watanabe <em>et al.</em> 1990</td>
</tr>
<tr>
<td><em>Bacteroides thetaiotaomicron</em> 95-1</td>
<td>37</td>
<td>I</td>
<td>6.0</td>
<td>37</td>
<td>-</td>
<td>70</td>
<td>-</td>
<td>-</td>
<td>Smith and Salyers 1991</td>
</tr>
<tr>
<td><em>Lipomyces starkeyi</em> (E.C. 3.2.1.11)</td>
<td>-</td>
<td>E</td>
<td>5.0</td>
<td>55</td>
<td>50min at 50°C</td>
<td>65-74</td>
<td>-</td>
<td>-</td>
<td>Koenig and Day 1989</td>
</tr>
</tbody>
</table>

I = intracellular; E = extracellular; The accuracy of the Mr given depends on the method used. Where the gene has been sequenced the Mr is that deduced from the DNA sequence, whereas in other cases Mr was determined by either SDS-PAGE or gel filtration with a lower degree of accuracy; Where given, thermostability is a measure of halftime at stated temperature.

1.4.1.5 "Non-classical" α-amylase

A number of enzymes have been isolated that are not true α-amylases (in the classical sense) but are enzymatically similar. These include amylases that produce, as the final end-product of digestion, a specific length α-oligosaccharide (3-6 residues). Enzymatic action is by exo-hydrolysis, sometimes followed by transferase activity, to form the longer oligosaccharides, which are subsequently hydrolysed (Fogarty and Kelly 1990).

Enzymes have also been isolated that are enzymatically similar to α-amylases but show no sequence homology (*Dictyoglomus thermophilum*, Fukusumi *et al.* 1988). A third group are high Mr amylases with α-amylase like activity e.g. the 125 kDa α-amylase from *Micrococcus* sp. 207 (Kimura and Horikoshi 1990a), and the 150 kDa α-amylase from *Lactobacillus amylovorus* NRRL B-4540 (Imam *et al.* 1991).
1.4.2 Debranching enzymes

Originally there were only two known endo-acting debranching enzymes, a pullulanase from *Klebsiella* sp. and an isoamylase from *Pseudomonas* sp. They were distinguishable by the ability of pullulanase, unlike isoamylase, to hydrolyse pullulan (Yokobayashi *et al.* 1973). As referred to in section 1.2.3 there is a potential industrial use for these enzymes (particularly thermostable ones) for the debranching of amylopectin. Hence a number of enzymes that can hydrolyse pullulan have been detected, sequenced and characterised from thermophilic sources (Hyun and Zeikus 1985a, Koch *et al.* 1987, Plant *et al.* 1987a, and Madi *et al.* 1987).

**Figure 1.14 Action pattern of enzymes degrading pullulan**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>End Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucoamylase</td>
<td>Glucose</td>
</tr>
<tr>
<td>Pullulanase</td>
<td>Maltotriose</td>
</tr>
<tr>
<td>IsoPullulanase</td>
<td>Isopanose</td>
</tr>
<tr>
<td>NeoPullulanase</td>
<td>Neopanose</td>
</tr>
</tbody>
</table>

Symbols: ○ glucose unit; ● glucose unit with reducing end; — α-1,4-bond; ↓ α-1,6-bond

Modified from Vihinen and Mäntsälä 1989
Section 1.4.2.4 discusses enzymes that exohydrolyse pullulan by hydrolysing the internal α-1,4-bonds (isopullulanase and neopullulanase). Enzymes that exohydrolyse pullulan by hydrolysis of α-1,6-bonds (pullulanases; E.C. 3.2.1.41 α-Dextrin 6-glucanohydrolase), forming maltotriose, have been further subdivided into two groups, Type I pullulanase and Type II pullulanase. Type I or true pullulanases are those with no α-1,4 activity (see section 1.4.2.1), whereas type II pullulanases have some α-1,4 activity (see section 1.4.2.2). Analysis of the enzymatic characteristics of type II pullulanases has shown that a distinct group, the α-amylase-pullulanases, have combined α-amylase and pullulanase activities and are hence discussed independently in Section 1.4.2.3).

1.4.2.1 Type I or true pullulanase

The characteristic type I pullulanase is the Klebsiella pullulanase (see Section 1.3.2). It has been extensively characterised and the DNA sequence of the pulA gene has been determined. It has been shown to be a polypeptide of 1090 amino acids with a Mr of 118,097 daltons (Kornacker and Pugsley 1990). Table 1.6 summarises the enzymatic properties of known true pullulanases.

### Table 1.6 Properties of true pullulanases

<table>
<thead>
<tr>
<th>Strain</th>
<th>Growth temp °C</th>
<th>Cellular location</th>
<th>pH opt</th>
<th>Temp opt °C</th>
<th>thermostability</th>
<th>Mr kDa</th>
<th>Inhibitors</th>
<th>Miscellaneous</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus stearothermophilus TRS128</td>
<td>60</td>
<td>E</td>
<td>6</td>
<td>65</td>
<td>5 min at 70°C</td>
<td>74,974</td>
<td>Cu²⁺ and Zn²⁺</td>
<td>sequenced and expressed in E. coli</td>
<td>Kuriki et al. 1988a</td>
</tr>
<tr>
<td>Thermus sp. AMD-33</td>
<td>-</td>
<td>E</td>
<td>5.5-5.7</td>
<td>70</td>
<td>-</td>
<td>-</td>
<td>Cu²⁺ and Zn²⁺</td>
<td>sequenced and expressed in E. coli</td>
<td>Sashihara et al. 1988</td>
</tr>
<tr>
<td>Micrococcus sp 207</td>
<td>18</td>
<td>E</td>
<td>8.0</td>
<td>50</td>
<td>-</td>
<td>120</td>
<td>Cu²⁺ and Zn²⁺</td>
<td>-</td>
<td>Kimura and Horikoshi 1990b</td>
</tr>
<tr>
<td>Bacillus acidopullulyticus</td>
<td>55</td>
<td>E</td>
<td>5.0</td>
<td>60</td>
<td>-</td>
<td>115-116</td>
<td>commercial enzyme</td>
<td>-</td>
<td>Kusano et al. 1988</td>
</tr>
<tr>
<td>Bacteroides thetaiotaomicron 5482A</td>
<td>37</td>
<td>I</td>
<td>6.5</td>
<td>37</td>
<td>-</td>
<td>70-77</td>
<td>-</td>
<td>-</td>
<td>Smith and Salyers 1989</td>
</tr>
<tr>
<td>Klebsiella sp.</td>
<td>30</td>
<td>E</td>
<td>5.6</td>
<td>50</td>
<td>-</td>
<td>118</td>
<td>Cyclo-dextrin</td>
<td>sequenced and expressed in E. coli</td>
<td>Eisele et al. 1972</td>
</tr>
</tbody>
</table>

I = intracellular; E = extracellular; The accuracy of the Mr given depends on the method used. Where the gene has been sequenced the Mr is that deduced from the DNA sequence, whereas in other cases Mr was determined by either SDS-PAGE or gel filtration with a lower degree of accuracy; Where given, thermostability is a measure of half-life at stated temperature.
1.4.2.2 Type II pullulanase

All type II pullulanases isolated so far have been from thermophilic organisms, where they appear to be the predominant subclass. Characteristics of this subclass of enzymes, compared to Klebsiella pullulanases, are relatively low Mr (<100 kDa), and a predominance of pullulanase activity with some hydrolysis of α-1,4 bonds of long-chained glucose units, especially amylopectin. In one case a single active site has been detected and shown to be responsible for both activities (Plant et al. 1987b). A summary of the data on all type II pullulanases is given in table 1.7.

Table 1.7 Properties of pullulanases with α-1,4 activity (Pullulanase Type II)

<table>
<thead>
<tr>
<th>strain</th>
<th>Growth Temp ºC</th>
<th>Cellular Location</th>
<th>pH opt</th>
<th>Temp opt ºC</th>
<th>thermo-stability</th>
<th>Mr kDa</th>
<th>Inhibitors</th>
<th>Miscellaneous</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Thermus aquaticus</em> YT-1</td>
<td>70</td>
<td>E</td>
<td>6.5</td>
<td>85</td>
<td>4.5h at 95ºC</td>
<td>83</td>
<td>-</td>
<td>-</td>
<td>Plant et al. 1986</td>
</tr>
<tr>
<td><em>Thermoanaerobium brockii</em></td>
<td>56</td>
<td>E</td>
<td>5.0</td>
<td>60</td>
<td>-</td>
<td>105</td>
<td>-</td>
<td>Expressed in B. Subtilis and E. coli</td>
<td>Coleman et al. 1987</td>
</tr>
<tr>
<td><em>Thermoanaerobium</em> Tok6-B1</td>
<td>70</td>
<td>E</td>
<td>5.5</td>
<td>70</td>
<td>17 h at 85ºC</td>
<td>120</td>
<td>-</td>
<td>-</td>
<td>Plant et al. 1987c</td>
</tr>
<tr>
<td><em>Bacillus flavogradarius</em> KP1228</td>
<td>60</td>
<td>I</td>
<td>6.3</td>
<td>80</td>
<td>-</td>
<td>65</td>
<td>-</td>
<td>-</td>
<td>Suzuki et al. 1991</td>
</tr>
</tbody>
</table>

I = intracellular; E = extracellular; The accuracy of the Mr given depends on the method used. Where the gene has been sequenced the Mr is that deduced from the DNA sequence, whereas in other cases Mr was determined by either SDS-PAGE or gel filtration with a lower degree of accuracy; Where given, thermostability is a measure of half-life at stated temperature.

1.4.2.3 α-amylase-pullulanase

Table 1.8 Properties of α-amylase-pullulanases

<table>
<thead>
<tr>
<th>strain</th>
<th>Growth Temp ºC</th>
<th>Cellular Location</th>
<th>pH opt</th>
<th>Temp opt ºC</th>
<th>thermo-stability</th>
<th>Mr kDa</th>
<th>Inhibitors</th>
<th>Miscellaneous</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Bacillus subtilis</em> Tu</td>
<td>30</td>
<td>E</td>
<td>7-7.5</td>
<td>50</td>
<td>-</td>
<td>450</td>
<td>-</td>
<td>-</td>
<td>Takasaki 1987</td>
</tr>
<tr>
<td><em>Bacillus circulans</em> F-2</td>
<td>37</td>
<td>E</td>
<td>7</td>
<td>50</td>
<td>-</td>
<td>220</td>
<td>-</td>
<td>-</td>
<td>Sata et al. 1989</td>
</tr>
<tr>
<td><em>Bacillus</em> sp. 3183</td>
<td>60</td>
<td>E</td>
<td>6</td>
<td>75</td>
<td>-</td>
<td>160</td>
<td>-</td>
<td>-</td>
<td>Shen et al. 1990</td>
</tr>
<tr>
<td><em>Clostridium thermohydrosulfuricum</em> E 101-69</td>
<td>70</td>
<td>E</td>
<td>5.6</td>
<td>85-90</td>
<td>30 min at 80</td>
<td>165.6</td>
<td>-</td>
<td>sequenced and expressed in E. coli</td>
<td>Melasniemi and Paloheimo 1989</td>
</tr>
<tr>
<td><em>Thermoanaerobacter</em> B6A</td>
<td>60</td>
<td>E and I</td>
<td>5</td>
<td>75</td>
<td>45 min at 75ºC</td>
<td>450</td>
<td>cyclo-dextrins</td>
<td>endo-acting</td>
<td>Saha et al. 1990</td>
</tr>
</tbody>
</table>

I = intracellular; E = extracellular; The accuracy of the Mr given depends on the method used. Where the gene has been sequenced the Mr is that deduced from the DNA sequence, whereas in other cases Mr was determined by either SDS-PAGE or gel filtration with a lower degree of accuracy; Where given, thermostability is a measure of half-life at stated temperature.

Introduction
The definitive enzyme for this subclass is Melasniemi's α-amylase-pullulanase from *Clostridium thermohydrosulfuricum*. It has been purified, enzymatically characterised (Melasniemi 1987a), sequenced (Melasniemi *et al.* 1990) and expressed in *E. coli* (Melasniemi and Paloheimo 1989). The characteristics of this subclass are a single high Mr enzyme (>150 kDa) with only one active site, but with both α-amylase and pullulanase type activities. A summary of the data on all α-amylase-pullulanases is given in table 1.8.

### 1.4.2.4 Neopullulanase and isopullulanase

Neopullulanases and isopullulanases (E.C. 3.2.1.57) are a group of enzymes that can hydrolyse pullulan. They hydrolyse the α-1,4-D glycosidic bond rather than the α-1,6-D bond, releasing panose and isopanose respectively. They show an inability to hydrolyse amylopectin or glycogen, but with prolonged incubation they can hydrolyse oligosaccharides, including panose and isopanose, to smaller saccharides (e.g. maltose and glucose). There are only a few known examples and the properties of these enzymes are summarised in table 1.9. The neopullulanase from *Bacillus stearothermophilus* TRS40 has been expressed in *B. subtilis* (Kuriki *et al.* 1988b) and sequenced (Kuriki and Imanaka 1989).

### Table 1.9 Properties of neopullulanases and isopullulanase

<table>
<thead>
<tr>
<th>Strain</th>
<th>Growth Temp °C</th>
<th>Cellular location</th>
<th>pH opt</th>
<th>Temp opt °C</th>
<th>Thermostability</th>
<th>Mr kDa</th>
<th>Inhibitors</th>
<th>Miscellaneous</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isopullulanase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Aspergillus niger</em> ATCC 9642</td>
<td>30</td>
<td>I</td>
<td>3.5-4.0</td>
<td>40</td>
<td>-</td>
<td>-</td>
<td>only known isopullulanase</td>
<td>Sakano <em>et al.</em> 1971</td>
<td></td>
</tr>
<tr>
<td>Neopullulanase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bacteroides thetaiotaomicron</em> 95-1</td>
<td>37</td>
<td>I</td>
<td>6.5</td>
<td>37</td>
<td>-</td>
<td>70</td>
<td>-</td>
<td>-</td>
<td>Smith and Salyers 1991</td>
</tr>
<tr>
<td><em>Bacillus stearothermophilus</em> TRS40</td>
<td>60</td>
<td>I</td>
<td>6.0</td>
<td>60</td>
<td>-</td>
<td>69.144</td>
<td>-</td>
<td>Sequenced and expressed in <em>B. subtilis</em></td>
<td>Imanaka and Kuriki 1989</td>
</tr>
</tbody>
</table>

I = intracellular; E = extracellular; The accuracy of the Mr given depends on the method used. Where the gene has been sequenced the Mr is that deduced from the DNA sequence, whereas in other cases Mr was determined by either SDS-PAGE or gel filtration with a lower degree of accuracy; Where given, thermostability is a measure of half-life at stated temperature.
1.4.2.5 Isoamylase (E.C. 3.2.1.68) Glycogen 6-glucanohydrolase

Isoamylase hydrolyses α-1,6-D glycosidic linkages of glycogen and amylopectin as well as various dextrins and oligosaccharides. It is distinguishable from pullulanase by its inability to hydrolyse pullulan, only partial debranching of amylopectin but complete debranching of glycogen. Studies on branched oligosaccharides have shown that the enzyme requires a minimum side chain size of four α-1,4-linked glucose units. This explains its inability to hydrolyse pullulan or completely debranch amylopectin, restricting its industrial applications when compared with pullulanase (reviewed in Harada 1984).

Although isoamylases have been detected (i.e. glycogen debranching activity) in Cytophaga sp., Saccharomyces cerevisiae and Flavobacterium sp., the isoamylase from Pseudomonas sp. is the only one to have been extensively studied. Three different groups have independently isolated the isoamylase from Pseudomonas [P. amylofera SB-15 (Amemura et al. 1988), P. amylofera JD210 (Chen et al. 1990) and Pseudomonas sp SMP1. (Tognoni et al. 1989)]. All have been expressed in E. coli, sequenced and purified, with the three genes appearing essentially identical (see Appendix 6). Table 1.10 summarises the enzymatic properties of known isoamylases.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Growth Temp°C</th>
<th>Cellular location</th>
<th>pH opt</th>
<th>Temp opt °C</th>
<th>Thermostability</th>
<th>Mr kDa</th>
<th>Inhibitors</th>
<th>Miscellaneous</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudomonas sp.</td>
<td>30</td>
<td>E</td>
<td>3.4</td>
<td>55</td>
<td>15m at 60°C</td>
<td>83.656</td>
<td>-</td>
<td>-</td>
<td>Yokobayashi et al. 1973</td>
</tr>
<tr>
<td>Cytophaga sp.</td>
<td>-</td>
<td>E</td>
<td>5.5</td>
<td>40</td>
<td>-</td>
<td>120</td>
<td>-</td>
<td>-</td>
<td>Gunja-Smith et al. 1970</td>
</tr>
<tr>
<td>Flavobacterium sp.</td>
<td>28</td>
<td>E</td>
<td>5.6</td>
<td>40</td>
<td>-</td>
<td>121</td>
<td>-</td>
<td>-</td>
<td>Sato and Park 1980</td>
</tr>
<tr>
<td>Saccharomyces sp.</td>
<td>-</td>
<td>I</td>
<td>6.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Membrane bound</td>
<td>Gunja et al. 1961</td>
</tr>
</tbody>
</table>

I = intracellular; E = extracellular; The accuracy of the Mr given depends on the method used. Where the gene has been sequenced the Mr is that deduced from the DNA sequence, whereas in other cases Mr was determined by either SDS-PAGE or gel filtration with a lower degree of accuracy; Where given, thermostability is a measure of half-life at stated temperature.

Table 1.10 Properties of isoamylases

Introduction
1.5 *Caldocellum saccharolyticum*

Extremely thermophilic anaerobic bacteria (growth temperatures greater than 75°C), that were capable of degrading cellulolytic material, were isolated from randomly selected sites in the thermal region of the central North Island of New Zealand as part of an alternative energy fuels program. One such isolate TP8.T.6.3.3.1 (shortened to TP8.T) was isolated by an avicel enrichment culture from a thermal pool in the Taupo region (*Sissons et al.* 1987). This isolate was classified as *Caldocellum saccharolyticum* gen. nov. and sp. nov.

1.5.1 *Caldocellum saccharolyticum* bacteria

1.5.1.1 Growth conditions

*C. saccharolyticum* is an obligatorily anaerobic bacterium, and was isolated from the decomposed end of a *Pinus radiata* log in a low ionic strength neutral pH thermal pool (78°C). When grown in a laboratory fermenter it has a temperature optimum of 68-72°C and a range of 60-80°C. It has a narrow pH range of only 6-8, with optimum growth at pH 7.0. *C. saccharolyticum* can utilise cellulose, xylan and starch as sole carbon sources, and urea as sole nitrogen source (*Donnison et al.* 1989)

1.5.1.2 Taxonomic characteristics

When viewed by light microscopy, *Caldocellum saccharolyticum* appears as non-motile, non-spore forming, non-flagellated, oval-ended rods (0.6 x 4 μm), occurring either singly or in pairs, but by mid-log phase long chains are formed. In common with a number of thermophilic bacteria, it stains gram negative but when viewed by thin section under the electron microscope it shows a gram positive cell wall arrangement (*Reynolds et al.* 1986).

*C. saccharolyticum* has a low G/C % molar DNA ratio (34.5%) as determined by thermal denaturation. *Caldocellum saccharolyticum* genomic DNA hybridises strongly to genomic DNA from two other strains from New Zealand (Tp11.W2 and Rt8.B2) and weakly to three further strains Tp8B.3.2, Tp8A.10.3 and Wai21.W2. However it does not hybridise to three *Clostridia* species (*C. thermohydrosulphuricum, C. perfringens* and *C. butyricum*) or to five glycolytic anaerobes, including *Thermoanaerobacter brockii* (ATCC 31550) (*Donnison et al.* 1986)

Modern taxonomic characterisation based on 16s ribosomal RNA shows that *C. saccharolyticum* is one of the deepest eubacterial branches, clustering with some thermophilic *Clostridia, Thermus* and *Thermotoga* (personal communication D. Saul, P. Bergquist, A. Rodrigo; Department of Cell Biology, Auckland University).
Using pulse field gel electrophoresis the *C. saccharolyticum* genome size was estimated as 2.33-2.43 Mb (personal communication K. Borges, Department of Cell Biology, Auckland University).

### 1.5.2 *C. saccharolyticum* enzymes transferred and expressed in *E. coli*

When transferred to *E. coli* the genes of *C. saccharolyticum* are expressed from their own promoters, unlike most thermophiles (*Thermus* in particular). Due to this ease of expressibility, our laboratory has transferred, sequenced and expressed a number of genes. Although all enzymes tested to date showed expression in *E. coli* from their own promoters, none were exported out of the cell under the direction of their own leader sequence (Patchett *et al.* 1989).

#### 1.5.2.1 β-glucosidase

A genomic library of *C. saccharolyticum* was constructed in bacteriophage λ1059 and screened for β-glucosidase activity by plate assay. A recombinant phage was detected and purified. Fragments of thermophilic DNA generated by restriction enzyme digestion were ligated into pBR322 (Love and Streiff 1987). The smallest plasmid recombinant (pNZ1002) to still express β-glucosidase was sequenced. One Open Reading Frame (ORF) was detected which encoded a β-glucosidase gene (*bgIA*) of 1360 base pairs coding for a 54.4 kDa peptide. This gene was also expressed in *B. subtilis* ([.ove *et al.* 1988]). This enzyme was still thermostable when expressed in a mesophilic host, in common with other thermophilic enzymes. The enzyme characteristics are listed in table 1.11. As the native β-glucosidase has not been purified from *C. saccharolyticum*, it is impossible to determine whether any enzymatic properties are different.

#### 1.5.2.2 Cellulase

From the same genomic λ phage library, cellulase-positive recombinant phages were identified and two genes (*celA* and *celB*) were expressed in *E. coli* (Chamley 1987). The *celB* gene has been sequenced and shown to contain an open reading frame of 3117 base pairs encoding a 118 kDa peptide (Saul *et al.* 1989). Expression studies showed that the CelB enzyme was bifunctional, containing two separate catalytic domains each having a distinct activity (see figure 1.15). Exoglucanase activity was coded for by the amino-terminal domain and endoglucanase activity by the carboxy-terminal domain. There was also significant homology between the deduced peptide sequence of each domain and other cellulases. When expressed in *E. coli*, a number of smaller peptides were detected that hydrolysed the synthetic soluble cellulase indicator substrates CMC and MUC. These peptides were produced from alternative internal translational start sites and/or proteolytic degradation products of the translated gene product (Saul *et al.* 1990).
On sequencing the remainder of λ2A, two other genes were identified, celA (Teo 1992) and manA (Gibbs et al. 1992). These genes were shown to possess a similar genetic organisation as celB. This is shown graphically in figure 1.15.

**Figure 1.15 The genetic organisation of NZPλ2**

![NZPλ2 diagram](image)

Activity = CMC'ase  Xylanase  Mannanase  MUC'ase  CMC'ase

C.D= catalytic domain  B.D= binding domain  P-T box

Sp-Sph I  A-AatII  B-BamHI  H-HindII  Ev-EcoRV  Sm-Smal  Sa-Sau3AI

From Teo 1992

The enzymatic data for the recombinant celB gene product are listed in table 1.11. Like the β-glucosidase, the native enzyme has not been purified from C. saccharolyticum, so it is impossible to determine if any enzymatic properties are different.

**1.5.2.3 Hemicellulase**

Four hemicellulolytic genes [endo-1,4-β-xylanase, β-D-xylosidase, β-acetylxylan esterase (Lüthi et al. 1990a) and β-mannanase (Lüthi et al. 1991)] have been isolated from a C. saccharolyticum genomic λ library and recombinant plasmids have been constructed in E. coli, by a process similar to that described in section 1.5.2.1. The first three were located on the same DNA fragment along with two other ORF's (see figure 1.16) and β-mannanase was located between the two cellulase genes celA and celB. All have been sequenced and found to show significant protein homology to other related enzymes.
Introduction

Figure 1.16 Restriction map of xylan-degrading genes from *Caldocellum saccharolyticum*

[Restriction enzyme map diagram]

- **ORF1** (xynA): xylanase/β-xylanidase, 40.5 kDa
- **ORF2** (xynC): acetyl esterase, 30.6 kDa
- **ORF3/4**: not known, 47.2 kDa
- **ORF5** (xynB): β-xylanidase, 56.4 kDa or 55.5 kDa
- **ORF6**

Restriction enzyme abbreviations:
- B: *BamHI*
- Hc: *HincII*
- Bal: *BglII*
- Pv: *PvuII*
- E: *EcoRI*
- R: *RsaI*
- Ev: *EcoRV*
- S: *SalI*
- Ha: *HaeIII*
- S3: *Sau3AI*
- Hp: *HpaI*
- Sp: *SphI*

Restriction map of thermophilic DNA from *Caldocellum saccharolyticum* coding for xylan-degrading enzymes expressed in *E. coli*. The shaded bar indicates the sequenced part of the *C. saccharolyticum* fragment and is represented as the pBR322 recombinant (modified from Lüthi *et al.* 1990a).

The enzymatic data of the recombinant hemicellulolytic gene products are listed below in Table 1.11. The native enzymes from *C. saccharolyticum* have not been purified, so it is impossible to determine if any enzymatic properties are different in the recombinant enzymes.

### Table 1.11 Recombinant enzyme characteristics

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Gene</th>
<th>Size /kDa</th>
<th>t&lt;sub&gt;opt&lt;/sub&gt;/℃</th>
<th>pH opt</th>
<th>t&lt;sub&gt;1/2&lt;/sub&gt;</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endo/Exocellulase</td>
<td>celB</td>
<td>118</td>
<td>85</td>
<td>6.8</td>
<td>50 min (80℃)</td>
<td>Saul <em>et al.</em> 1990</td>
</tr>
<tr>
<td>β-glucosidase</td>
<td>bgIA</td>
<td>54.4</td>
<td>70</td>
<td>6.8</td>
<td>15 hr (80℃)</td>
<td>Plant <em>et al.</em> 1988</td>
</tr>
<tr>
<td>β-xylanase</td>
<td>xynA</td>
<td>39.9</td>
<td>70</td>
<td>5.5-6</td>
<td>8-9 hr (70℃)</td>
<td>Lüthi <em>et al.</em> 1990b,</td>
</tr>
<tr>
<td>β-xylosidase</td>
<td>xynB</td>
<td>56.4</td>
<td>70</td>
<td>6-6.6</td>
<td>45 min (80℃)</td>
<td>Lüthi and Bergquist 1990</td>
</tr>
<tr>
<td>Acetyl xylan esterase</td>
<td>xynC</td>
<td>30.6</td>
<td>70-75</td>
<td>6</td>
<td>30 hr (80℃)</td>
<td>Lüthi <em>et al.</em> 1990c</td>
</tr>
<tr>
<td>β-mannanase</td>
<td>manA</td>
<td>38.9</td>
<td>80</td>
<td>6</td>
<td>60 min (80℃)</td>
<td>Lüthi <em>et al.</em> 1991</td>
</tr>
</tbody>
</table>
1.6 Previous Work Done and Aims of Thesis

Previous work done in this laboratory is summarised diagrammatically in figures 1.17-1.19.

1.6.1 Gene isolation and expression in E. coli

A recombinant genomic library was constructed from high molecular weight (greater than 40 kb) DNA from Caldocellum saccharolyticum (Love and Streiff 1987). This DNA was shortened to approximately 14-20 kb fragments by partial Sau3AI digestion, ligated into bacteriophage λ1059 and packaged in vitro (Karn et al. 1980). Phage stabs from this library were screened for pullulanase activity using a dye-linked pullulan substrate (see section 2.6.1.4). After incubation at 70°C for 2 hours, pullulanase expression was detected by the formation of a dye-cleared halo around a recombinant phage plaque. After seven such screenings a stable and replatable pullulanase-positive phage was detected. Thermophilic DNA fragments from a Sau3AI digestion of the recombinant phage were ligated into pBR322 (Bolivar et al. 1977) to give a recombinant plasmid expressing pullulanase (pNZ1037), which contained 7.5 kb of thermophilic DNA. A smaller recombinant plasmid expressing pullulanase, pNZ1038, was formed by deletion of a 3.0 kb SphI/SphI fragment from pNZ1037 (experiments performed by D. Love). The pullulanase gene was located within a 4.0 kb region of pNZ1038 by Tn5 transposon mutagenesis (unpublished experiments by R. Fisher).

1.6.2 Bal 31 deletions

In an attempt to further define the position of the pullulanase gene and produce nested deleted deletions suitable for sequencing, two complementary series of deletions were produced from each end of the thermophilic DNA in pNZ1038, using Bal31 exonuclease (Misra 1985). In the first series, pNZ1038 was linearised with EcoRI and incubated with Bal31 for various times. For each time point, samples were taken out and the resulting DNA fragment was backfilled by the action of E. coli DNA polymerase (Klenow fragment), cut with Sall, and ligated to SmaI, Sall cut pCGN566 (Stalker et al. 1988). With the second series of recombinants a similar procedure was used, but Sall was used for linearisation and EcoRI for release, and the final plasmid vector was pCGN565 (multi-cloning site in reverse orientation to pCGN566). The deletions from each series were ordered by size and assayed for pullulanase activity (experiments performed by P. Caughey).

These deletions have since been sequenced and the results of this work are covered in section 4.2 and in Appendix 1.
Figure 1.17 Pullulanase gene isolation

High Mr DNA (40 kb) Extracted on sucrose gradient.

Caldocellum saccharolyticum

Size reduced by Sau 3A1 digestion (12-15 KB), ligated in λ1059, packaged in vitro and plated onto dye-linked pullulan. Pullulanase activity was detected by the formation of a dye-clearing halo.

DNA from this λ recombinant was digested with Sau 3A1 and the fragments ligated into pBR322

pNZ1037 expressed pullulanase activity

Digested with Sphl and self ligated to form pNZ1038

Shown is the method used to isolate the Caldocellum saccharolyticum pullulanase gene. Restriction Enzymes are E=Eco RI, H=Hind III, K=Kpn I, P=Pst I, S=Sal I, S=Sph I

Introduction
Figure 1.18 Exonuclease deletions of pNZ1038.

End repaired and cut with Sal I and ligated into the Sma I, Sal I site of pCGN 566

pCGN 566 pCGN 565 are two plasmids that vary only in the orientation of the multi-cloning site.
Figure 1.19 Screening nested deletions for pullulanase activity

Restriction Enzymes are E=Eco RI, H=Hind III, K=Kpn I, P=Pst I, S=Sal I, S=Sph I

Introduction
1.6.3 The aims of research for this thesis

The aim of this work was to sequence and characterise the pullulanase from *Caldocellum saccharolyticum* since, as shown in section 1.2, there is an industrial application for a thermophilic pullulanase if it can be easily and economically produced, and the genes from *Caldocellum saccharolyticum* are known to be easily expressed in *E. coli*. Hence, the following objectives were addressed in this thesis.

1. To purify and characterise the pullulanase expressed in *Caldocellum saccharolyticum* (see Chapter 3).

2. To determine the nucleotide sequence of the pullulanase gene, *pulA* (see Chapter 4).

3. To purify and characterise the pullulanase expressed in *E. coli* from its own promoter (see Chapter 5).

4. To over express the pullulanase under the control of an inducible promoter (see Chapter 6).

5. To express the pullulanase from an inducible promoter system in yeast (see Chapter 7).

6. To compare the deduced amino acid sequence of PulA to other related enzymes (see Chapter 8).
Chapter Two

Materials and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals were of analytical grade, unless otherwise specified. They were purchased from either Sigma Chemical Co., St Louis, Missouri, U.S.A.; British Drug Houses (BDH) Chemicals Ltd., Poole, England; Boehringer Mannheim GmbH, Mannheim, Germany; or Pharmacia LKB Biotechnology, Uppsala, Sweden.

Note that all ‘%’s used in this chapter refer to weight/volume (w/v), unless otherwise stated. All water used was sterile Millipore ‘MilliQ’, microfiltered, deionised water; unless otherwise stated.

2.1.2 Radioactivity

Radioisotopes were supplied by New England Nuclear Corp., 549 Albany Street, Boston, Massachusetts, U.S.A.

Deoxyadenosine 5'-[α-35S] triphosphate ([α-35S] dATP); 10 μCi μl-1, 400 Ci mmol-1
Deoxycytidine 5'-[α-32P] triphosphate ([α-32P] dCTP); 10 μCi μl-1, 3000 Ci mmol-1
Deoxyadenosine 5'-[γ-32P] triphosphate ([γ-32P] dATP); 5 μCi μl-1, 3000 Ci mmol-1

2.1.3 Enzymes

2.1.3.1 Restriction enzymes

Restriction enzymes were purchased from Bethesda Research Laboratories (BRL), (Gaithersberg, MD 20877 U.S.A.) or Boehringer Mannheim GmbH (Mannheim, Germany). Restriction enzyme digests were carried out as recommended by the manufacturer, using buffers supplied with the enzymes. Small scale digestions (<1 μg) were carried out in 20-40 μl reaction volumes for 2-3 hours at 37°C, whereas large scale digestions were incubated overnight, with correspondingly larger reaction volumes.
<table>
<thead>
<tr>
<th>Enzymes</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> DNA Ligase</td>
<td>BRL</td>
</tr>
<tr>
<td>Klenow fragment of DNA polymerase</td>
<td>BRL</td>
</tr>
<tr>
<td>T4 kinase</td>
<td>Boehringer Mannheim</td>
</tr>
<tr>
<td>Proteinase K</td>
<td>Boehringer Mannheim</td>
</tr>
<tr>
<td>RNase A</td>
<td>BRL</td>
</tr>
<tr>
<td>Sequenase (modified T7 DNA polymerase)</td>
<td>United States Biochemical Co.</td>
</tr>
<tr>
<td>Lysozyme</td>
<td>Sigma</td>
</tr>
<tr>
<td>Alkaline phosphatase</td>
<td>Boehringer Mannheim</td>
</tr>
<tr>
<td>T4 DNA Ligase</td>
<td>Boehringer Mannheim</td>
</tr>
</tbody>
</table>

### 2.1.4 Common buffers and general solutions

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revco medium</td>
<td>42 mM K_2HPO_4, 22 mM KH_2PO_4, 1.7 mM sodium citrate, 0.4 mM MgSO_4, 30% (v/v) glycerol</td>
</tr>
<tr>
<td>SSC</td>
<td>150 mM NaCl, 15 mM tri-sodium citrate</td>
</tr>
<tr>
<td>STET</td>
<td>8% (w/v) sucrose, 5% (v/v) Triton X-100, 50 mM EDTA, 50 mM Tris-HCl, pH 7.5</td>
</tr>
<tr>
<td>TAE</td>
<td>40 mM Tris-HCl, 5 mM sodium acetate, 0.4 mM EDTA, pH 7.8</td>
</tr>
<tr>
<td>TBE</td>
<td>(Borate buffer): 89 mM Tris-HCl, 89 mM boric acid, 2.5 mM EDTA, pH 8.3</td>
</tr>
<tr>
<td>TE</td>
<td>10 mM Tris-HCl pH 8.0, 0.1 mM EDTA.</td>
</tr>
<tr>
<td>TNE</td>
<td>10 mM Tris-HCl pH 8.0, 10 mM EDTA, 10 mM NaCl</td>
</tr>
<tr>
<td>TSS</td>
<td>L-broth with 10% PEG8000, 5% DMSO, 20-50 mM MgCl_2, pH 6.5</td>
</tr>
</tbody>
</table>

All buffers and solutions used were pH adjusted at room temperature, unless otherwise stated, and filter sterilised or autoclaved before use. Other common solutions and reagents were made up according to Maniatis *et al.* (1982).
2.2 Organisms

2.2.1 E. coli strains

All media and solutions used for bacterial manipulation were filter sterilised or autoclaved before use.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Use</th>
<th>Reference/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB2946 (JM 83)</td>
<td>F-, ara, Δ(lac-proAB), thi, strA, ϕ80lacZΔM15</td>
<td>Standard host for plasmid vectors</td>
<td>Vieira and Messing (1982)</td>
</tr>
<tr>
<td>PB2477 (C600)</td>
<td>thi-1, leuB6, thr-1, lacY1, supE44, rK, mK</td>
<td>Standard host for plasmid vectors</td>
<td>Hanahan (1983)</td>
</tr>
<tr>
<td>PB4955 (DH5 α)</td>
<td>hsdR17, supE44, recA1, endA1, Δlac U169(ϕ80lacZΔM15) gyrA96, thi1, relA1</td>
<td>Standard host for plasmid vectors</td>
<td>BRL (1986)</td>
</tr>
<tr>
<td>PB2963 (JM101)</td>
<td>Δ(lac-proAB), thi, SupE44, F[rtaD36, proAB+, lacIq, lacZΔM15]</td>
<td>Host for production of single stranded phage</td>
<td>Messing et al. (1981)</td>
</tr>
<tr>
<td>PB4949 (JM109)</td>
<td>recA1, endA1, gyrA96, thi, hsdR17, SupE44, relA1, Δ(lac-proAB) F[rtaD36, proAB+, lacIq, lacZΔM15]</td>
<td>Host for production of single stranded phage</td>
<td>Promega Corp</td>
</tr>
</tbody>
</table>

Strains from existing Revco stocks were streaked to single colonies on L-plates or antibiotic plates where appropriate, and checked for common markers before use.

2.2.2 Growth media and conditions

E. coli strains were grown at 37°C unless otherwise stated, with the only major exception being strains containing the heat inducible plasmid pJLA602 (see section 6.2). These strains were initially grown at 28°C, to produce sufficient cell mass, and then induced at 42°C.

Normally, strains were grown on "rich" media (an undefined mixture of hydrolysis products that contains an energy source and all amino acids and vitamins), which produced faster growth rates and higher cell density. For specific applications such as marker checking, strains were grown on a defined chemical minimal media.
Table 2.4 E. coli strains used during thesis

<table>
<thead>
<tr>
<th>Strain Name</th>
<th>Plasmid</th>
<th>Markers</th>
<th>Section used</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB4583</td>
<td>pNZ1037</td>
<td>(PB2477), Amp, Pul+</td>
<td>1.6, 4.2, 4.4, 5.2</td>
</tr>
<tr>
<td>PB4584</td>
<td>pNZ1038</td>
<td>(PB2477), Amp, Pul+</td>
<td>1.6, 4.2, 4.4, 4.5, 5.2, 5.3</td>
</tr>
<tr>
<td>PB4715</td>
<td>pNZ1072</td>
<td>(PB2946), Amp</td>
<td>1.6, 3.1</td>
</tr>
<tr>
<td>PB4901-4902</td>
<td>pNZ1451-1452</td>
<td>(PB2946), Cm, Pul+</td>
<td>1.6, 4.2, 4.3, 5.2, 5.3</td>
</tr>
<tr>
<td>PB4903-4918</td>
<td>pNZ1453-1468</td>
<td>(PB2946), Cm;</td>
<td>Appendix 1</td>
</tr>
<tr>
<td>PB4919-4925</td>
<td>pNZ1469-1475</td>
<td>(PB2946), Cm, Pul+</td>
<td>4.3, Appendix 1</td>
</tr>
<tr>
<td>PB4926-4947</td>
<td>pNZ1476-1497</td>
<td>(PB2946), Cm;</td>
<td>Appendix 1</td>
</tr>
<tr>
<td>PB4951</td>
<td>pNZ1501</td>
<td>(PB4949), Amp</td>
<td>4.3</td>
</tr>
<tr>
<td>PB4952</td>
<td>pNZ1502</td>
<td>(PB4949), Amp</td>
<td>4.3</td>
</tr>
<tr>
<td>PB4953</td>
<td>pNZ1501</td>
<td>(PB2963), Amp</td>
<td>4.3</td>
</tr>
<tr>
<td>PB4954</td>
<td>pNZ1502</td>
<td>(PB2963), Amp</td>
<td>4.3</td>
</tr>
<tr>
<td>PB4956</td>
<td>pNZ1503</td>
<td>(PB4949), Amp, Pul+</td>
<td>4.4</td>
</tr>
<tr>
<td>PB4957</td>
<td>pNZ1503</td>
<td>(PB2963), Amp, Pul+</td>
<td>4.4</td>
</tr>
<tr>
<td>PB4958</td>
<td>pNZ1504</td>
<td>(PB4949), Amp, Pul+</td>
<td>4.4</td>
</tr>
<tr>
<td>PB4959</td>
<td>pNZ1504</td>
<td>(PB2963), Amp, Pul+</td>
<td>4.4</td>
</tr>
<tr>
<td>PB4962</td>
<td>pNZ1510</td>
<td>(PB4819), Amp, Pul+</td>
<td>6.2</td>
</tr>
<tr>
<td>PB4963</td>
<td>pNZ1505</td>
<td>(PB2963), Amp, Pul+</td>
<td>6.2</td>
</tr>
<tr>
<td>PB4964</td>
<td>pNZ1506</td>
<td>(PB2963), Amp, Pul+</td>
<td>6.2</td>
</tr>
<tr>
<td>PB4970</td>
<td>pNZ1037</td>
<td>(PB2946), Amp, Pul+</td>
<td>5.2</td>
</tr>
<tr>
<td>PB4971</td>
<td>pNZ1507</td>
<td>(PB2946), Amp, Pul+</td>
<td>6.2, 6.3, 6.4</td>
</tr>
<tr>
<td>PB4972</td>
<td>pNZ1507</td>
<td>(PB4819), Amp, Pul+</td>
<td>6.2</td>
</tr>
<tr>
<td>PB4973</td>
<td>pNZ1038</td>
<td>(PB2946), Amp, Pul+</td>
<td>5.2</td>
</tr>
<tr>
<td>PB4974</td>
<td>pNZ1503</td>
<td>(PB2946), Amp, Pul+</td>
<td>5.2, 5.3, 5.4</td>
</tr>
<tr>
<td>PB4975</td>
<td>pNZ1504</td>
<td>(PB2946), Amp, Pul+</td>
<td>5.2</td>
</tr>
<tr>
<td>PB4976</td>
<td>pNZ1508</td>
<td>(PB2477), Amp, Pul+</td>
<td>7.0</td>
</tr>
</tbody>
</table>

Note that this list excludes recombinants from either of the libraries (random sheared or pseudorandom) generated for sequencing.

All nucleotide positions, unless otherwise stated, are based on pNZ1038 with the EcoRI site beginning at position (-375) and BamHI (+1), i.e. the first C. saccharolyticum nucleotide.

---

Materials and Methods
2.2.2.1 Rich media

The following "rich" media were used for growing *E. coli* strains,

**Luria broth** (L-broth): 1% Bactotryptone, 0.5% yeast extract, 1% NaCl, pH 7.0 (Luria & Burrous, 1957).

2xYT + sucrose: 1% Bactotryptone, 1% Yeast extract, 0.5% NaCl, 0.5% Sucrose.

Solid media was made by the addition of 2% Davis agar to the corresponding liquid media.

2.2.2.2 Minimal media

The base for all minimal media was an inorganic salts buffer, 56 (Monod *et al.* 1951) or half strength (56/2), which was supplemented with an energy source (normally glucose), vitamins (normally B,) and amino acids as required for growth. Solid media was made by the addition of 2% Davis agar to the corresponding liquid media.

**56/2 salts solution:** 20 mM KH₂PO₄, 30 mM Na₂HPO₄, 7.5 mM (NH₄)₂SO₄, 0.4 mM MgSO₄, 0.03 mM Ca(NO₃)₂, 0.5 nM FeSO₄ (Adelberg & Burns, 1960).

<table>
<thead>
<tr>
<th>Table 2.5 Normal supplement concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Supplement</strong></td>
</tr>
<tr>
<td>glucose</td>
</tr>
<tr>
<td>maltose</td>
</tr>
<tr>
<td>glycerol</td>
</tr>
<tr>
<td>casamino acids</td>
</tr>
<tr>
<td>leucine</td>
</tr>
<tr>
<td>proline</td>
</tr>
<tr>
<td>threonine</td>
</tr>
<tr>
<td>tryptophan</td>
</tr>
<tr>
<td>IPTG</td>
</tr>
<tr>
<td>BCIG</td>
</tr>
<tr>
<td>vitamin B1</td>
</tr>
</tbody>
</table>
2.2.2.3 Antibiotics

Where appropriate, antibiotic selection was added to the media to aid plasmid maintenance. Listed below in table 2.6 are the common antibiotics used and their final media concentration.

Table 2.6 Antibiotics used in this thesis

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
<td>25-100 μg ml⁻¹ $</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>100 μg ml⁻¹</td>
</tr>
<tr>
<td>Streptomycin</td>
<td>100 μg ml⁻¹</td>
</tr>
</tbody>
</table>

$ Concentration used depended on plasmid copy number

2.2.3 Measurement of cell growth

Optical density measures and viable counts were used in tandem to measure cell growth.

2.2.3.1 Optical density

Optical density was determined using a spectrophotometer at a wavelength of 600 nm and using the appropriate medium as a blank, with cell density being expressed as absorbance, Abs= OD₆₀₀. With increasing cell density, a higher proportion of the light scatters around the cells rather than being absorbed and hence reaches the detector causing an apparently lower OD reading (Stanier et al. 1977). The point at which this departure from direct proportionality occurs (i.e where the calibration curve becomes nonlinear) varies depending on a number of factors (cuvette volume, light wavelength and slit width) but experimentally a detectable deviation occurs at values > 0.3 OD₆₀₀. To overcome this problem, cell cultures were diluted twofold in the appropriate medium before the absorbance was measured. Their resulting absorbance was multiplied by the dilution factor giving a relative absorbance value. This procedure was repeated until two successive relative absorbance values were identical (i.e. cells were diluted into the linear region of the calibration curve).

2.2.3.2 Viable cell counts

Viable cell counts were performed by repeated serial dilution of 100 μl of cell culture in 9900 μl of 56/2, with 100 μl plated onto the appropriate agar media and incubated at 37°C. Viable cells were
counted on each plate (normally only two plates were countable). The number of Colony Forming Units (CFU) per ml was determined by using the formula shown below.

\[(\text{CFU})\text{ml}^{-1} = 100^D x C x 10\]

where \(D\) is the number of serial dilutions
and \(C\) is the number of colonies.

### 2.2.3.3 Plasmid copy number determination

Relative plasmid copy numbers were determined by measuring the mass of plasmid DNA present in each recombinant. Cells were grown to late log phase (OD\(_{600}\) = 1.0) and the plasmid DNA from 1 ml of cells was isolated by the method described in section 2.3.2.1. The amount of DNA was determined using a modification of the method of Projan et al. (1983) that involved measurement of DNA ethidium bromide fluorescence under short wave UV light (254 nm). The amount of fluorescence measured was determined using a Sony XC77CE CCD video camera and V image processing software (Digital Optics Ltd New Zealand). For maximum sensitivity, this system required small tight bands. Due to this limitation the DNA was digested with restriction endonucleases before running on an agarose gel as in Section 2.3.4. The intensity of fluorescence was determined and evaluated as average pixel brightness for each band. This value was compared to a standard curve that had previously been generated using DNA of known concentration. To standardise the degree of ethidium bromide staining, DNA standards were run on each gel and gels were destained until the average pixel brightness of the standards was identical between gels.

### 2.2.4 Caldocellum saccharolyticum

Since *Caldocellum saccharolyticum* is an obligate anaerobe it was grown under anaerobic conditions. Large scale cultures (4 litres) were supplied by the Thermophile and Microbial Biochemistry & Biotechnology Unit at Waikato University, New Zealand. Small scale cultures used for substrate induction tests were grown using the procedure outlined below.

The standard media used was 2/1 minimal media with carbohydrate supplements.

**2/1 minimal media (1 litre):** 0.75 g KH\(_2\)PO\(_4\), 1.5 g K\(_2\)HPO\(_4\), 0.9 g NH\(_4\)Cl, 0.2 g MgCl\(_2\)\(_6\)H\(_2\)O, 0.9 g NaCl, 0.75 g cysteine, 1.0 g yeast extract, 2.0 g peptone, 1 ml SL-10 trace elements, 1 ml selenium-tungstate, 1 ml Resazurin (0.1%) and 2.0 g carbohydrate supplement. pH adjusted to 7.0 with NaOH.
Dissolved oxygen was removed from the media by boiling for five minutes. While still boiling, the media was dispensed into 10 ml anaerobic tubes and sealed under nitrogen. The anaerobic tubes were kept at room temperature for 24 hours to ensure anaerobic conditions were maintained. Tubes were then inoculated via the rubber syringe port and incubated at 70°C.

### 2.2.5 Yeast strains

Yeast strains used.

**PB 5321**: his3-11, 15; *leu* 2-3, *leu* 2-112; *CUP1*, *Can1*

**PB 4977**: (PB5321) with plasmid pNZ1508 (*LEU2* Amp *Pul*+)

All yeast strains were grown at 28°C in either rich media (YPD) or minimal media (SD) with the appropriate supplements (Campbell 1988). Solid media was made by the addition of 2% Davis agar to the corresponding liquid media.

**YPD media**: 1% Bacto yeast extract, 2% Bacto peptone, 2% glucose.

**SD media**: 0.67% Bacto yeast nitrogen base without amino acids, 2% glucose, plus amino acids as appropriate at 20 μg ml⁻¹.
2.3 DNA

2.3.1 Plasmids

Plasmids used and constructed during the course of this work are listed below. Plasmid DNA was stored frozen in water at -20°C.

Table 2.7 “Standard” laboratory E. coli plasmids used

<table>
<thead>
<tr>
<th>Name</th>
<th>pNZ #</th>
<th>PB #</th>
<th>Type</th>
<th>Markers</th>
<th>Source / Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBR322</td>
<td>-</td>
<td>1375</td>
<td>Standard cloning</td>
<td>Amp Tet</td>
<td>Bolivar et al. 1977</td>
</tr>
<tr>
<td>pCGN565</td>
<td>-</td>
<td>2749</td>
<td>Standard cloning</td>
<td>Cm</td>
<td>Stalker et al. 1988</td>
</tr>
<tr>
<td>pCGN566</td>
<td>-</td>
<td>2750</td>
<td>Standard cloning</td>
<td>Cm</td>
<td>Stalker et al. 1988</td>
</tr>
<tr>
<td>pGEM 7 Zf(+)</td>
<td>4950</td>
<td>4950</td>
<td>Sequencing vector</td>
<td>Amp</td>
<td>Promega Corp</td>
</tr>
<tr>
<td>pGEM 5 Zf(+)</td>
<td>-</td>
<td>-</td>
<td>Sequencing vector</td>
<td>Amp</td>
<td>Promega Corp</td>
</tr>
<tr>
<td>M13mp10</td>
<td>-</td>
<td>-</td>
<td>Sequencing vector</td>
<td></td>
<td>Amersham</td>
</tr>
<tr>
<td>pJLA602</td>
<td>1265</td>
<td>5306</td>
<td>Expression vector</td>
<td>Amp, Ci 857</td>
<td>Medac</td>
</tr>
<tr>
<td>pYEYC5</td>
<td></td>
<td></td>
<td>E. coli Yeast shuttle vector</td>
<td>Amp, LEU2, CUP1B</td>
<td>Macreadie 1990</td>
</tr>
</tbody>
</table>

2.3.2 Small scale plasmid isolation

2.3.2.1 Standard alkaline lysis method

The method of Ish-Horowicz & Burke (1981) was followed, using 1.5 ml overnight cultures, but with the addition of an RNase step (100 μg ml⁻¹, 15 min at 37°C) and phenol/chloroform extraction (1:1 mix) prior to the ethanol precipitation stage. Dried DNA pellets were resuspended in 50 μl of TE; 2-5 μl was used per digest.
Table 2.8 Bacterial plasmids containing the C. saccharolyticum pullulanase gene

<table>
<thead>
<tr>
<th>Name</th>
<th>Construction $^\circ$</th>
<th>Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNZ1037</td>
<td>7.5 kb Sau3AI/Sau3AI genomic fragment ligated into pBR322 (BamHI site)</td>
<td>Amp, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1038</td>
<td>4797 bp Sphi-SphI deletion from pNZ1037</td>
<td>Amp, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1072</td>
<td>Internal 1.839 bp Hind III (838)-Hind III (2677) from pNZ1038</td>
<td>Amp, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1451-1452 (Eco 1-2)</td>
<td>Exonuclease deletion of pNZ1038 (EcoRI-Smal) ligated into pCGN566</td>
<td>Cm, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1453-1468 (Eco 3-18)</td>
<td>Exonuclease deletion of pNZ1038 (EcoRI-Smal) ligated into pCGN565</td>
<td>Cm, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1469-1476 (Sal 1-7)</td>
<td>Exonuclease deletion of pNZ1038 (SalI-Smal) ligated into pCGN565</td>
<td>Cm, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1477-1497 (Sal 8-29)</td>
<td>Exonuclease deletion of pNZ1038 (SalI-Smal) ligated into pCGN565</td>
<td>Cm, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1501</td>
<td>403 bp Dra I (1023)-Mlu I (1426) fragment in (Sma I/Mlu I)</td>
<td>Amp, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1502</td>
<td>444 bp Mlu I (1426) - Dra I (1870) fragment in (Sma I/Mlu I)</td>
<td>Amp, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1503</td>
<td>3371 bp Mlu I (1426) - Sph I (4797) fragment in (SphI/Mlu I)</td>
<td>Amp, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1504</td>
<td>3114 bp Mlu I (1426) - Pst I (4540) fragment in (Pst I/Mlu I)</td>
<td>Amp, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1505</td>
<td>2681 bp Sca I (3956) - Sca I (1271) fragment in (Sma I M13mp10)</td>
<td>Amp, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1506</td>
<td>2681 bp Sca I (1271) - Sca I (3956) fragment in (Sma I M13mp10)</td>
<td>Amp, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1507</td>
<td>2696 bp Eco RI - Bam HI fragment from pNZ1505 in pJLA602 (Eco RI - Bam HI)</td>
<td>Amp, Pul$^+$</td>
</tr>
<tr>
<td>pNZ1508</td>
<td>2877 bp Pvu II - Bam HI fragment from pNZ1505 in pYELC5 (Pvu II - Bam HI)</td>
<td>Amp, Pul$^+$</td>
</tr>
</tbody>
</table>

$^\circ$ Nucleotide positions are based on pNZ1038 with the EcoRI site beginning at position (+375) where the second G of the Bam HI is position 1 (i.e. the first C. saccharolyticum nucleotide).

2.3.2.2 Phenol/chloroform method

As a quick method of screening colonies for inserts, a phenol extraction-cleared lysate method was used (Serghini et al. 1989), yielding DNA that was digestible with most restriction enzymes. Overnight cultures (1.5 ml) were pelleted and resuspended in 50 µl of TNE buffer and 50 µl of...
phenol/chloroform. This procedure lysed the bacteria and denatured the proteins, and, on mixing and centrifugation (25°C, 8000 g, 15 min), the plasmid DNA was partitioned into the aqueous phase. The aqueous phase was removed and the DNA precipitated by the standard EtOH/acetate method.

2.3.3 Large scale plasmid isolation

2.3.3.1 CsCl-EtBr density gradient purification

(I) Plasmid DNA for caesium chloride purification was prepared by a modification of the method of Holmes & Quigley (1981).

Cells (200 ml) from an overnight culture (grown in L-broth with appropriate antibiotic) were harvested by centrifugation (4°C, 3000 g, 10 min) and resuspended in 20 ml of STET buffer (8% sucrose, 2.5% (v/v) Triton X-100, 2 mM EDTA, 50 mM Tris-HCl pH 8.0) containing 1 mg ml⁻¹ lysozyme. This mixture was boiled for 5 min to release the plasmid and the resulting cell debris was removed by centrifugation (4°C, 8000 g, 10 min). The DNA was precipitated from the supernatant by the addition of isopropanol (1 volume, 2 hr, -20°C), recovered by centrifugation (4°C, 8000 g, 25 min) and resuspended in 2.5 ml of TE.

(II) Covalently-closed-circular plasmid DNA's prepared in (I) were purified on CsCl-EtBr density gradients. CsCl was added to DNA solutions to a density of approximately 1.619 g/cm³ (refractive index adjusted to 1.391-1.392), and ethidium bromide added to a final concentration of ~0.5 mg ml⁻¹. Gradients were centrifuged (15°C, 225 000 g, 16 hr) in a Sorvall TV-865 vertical rotor. The plasmid band was then recovered and rebanded in another identical gradient for 6 hours. Ethidium bromide was extracted with butan-2-ol, and the CsCl-DNA solution dialysed against two changes of TE at 4°C.

2.3.3.2 Large-Scale LiCl method

This method was developed from a 'mini-prep' protocol described in Current Protocols in Molecular Biology (Ausubel et al. 1989). 100-200 ml of overnight culture was pelleted and resuspended in 3 ml Tris-Glucose buffer (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA). Alkaline lysis was carried out using the same volume ratios as used for Section 2.3.2.1. The supernatant resulting from the acetate precipitation stage was precipitated with 1 volume of isopropanol at room temperature (15 min), spun down and the dried pellet resuspended in 300 µl of TE. The suspension was treated with DNase-free RNase A (Maniatis et al. 1982) to a concentration of 40 µg ml⁻¹ (37°C, 15 min), mixed with 225 µl of 7M LiCl and left on ice for 20 min. Precipitated RNA and protein were spun down (15 min in microfuge at 4°C) and the resulting supernatant was twice
extracted with phenol/chloroform. The DNA in the supernatant was precipitated (2 volumes of ethanol and 1/10 volume acetate), spun and pelleted, washed with 70% EtOH, dried, and resuspended in 100-200 µl of TE.

2.3.4 Standard agarose gel electrophoresis

Horizontal submarine slab gels were used for routine separation of DNA fragments, using TBE buffer (Maniatis et al. 1982). Depending on the sizes of the fragments to be separated, different concentrations of agarose (0.4-2.0%, Sigma: type II, medium EEO) were used in the TBE buffer. Molecular mass (MW) standards, normally the BRL 1 Kilobase (1 kb) Ladder, were run with all gels. DNA samples and MW standards were mixed with the appropriate amounts of 6x gel loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol F and 40% sucrose) prior to loading into the gel wells (0.4-0.7 cm widths).

Electrophoresis was carried out at room temperature with field strengths of 1-5 V/cm for large slab gels, and 1-10 V/cm for small 'minigels'. Run times were dependent on fragment size and the separation range required, hence varying from 30 min (minigels with small fragments) to overnight (large gels with large fragments). Ethidium bromide solution (1 µg ml⁻¹ in deionised water) was used to stain the gels (20 to 60 min), which were then destained in deionised water. A UV (254 nm) Transilluminator (Ultraviolet Products Inc.) was used to visualise the DNA fragments. Stained gels were photographed through a red R60 filter (Toko) onto Kodak Tri-X pan 35 mm film using a Nikon FM2 with a 50 mm f1.8 Nikkor lens.

2.3.5 Preparative agarose gel electrophoresis

2.3.5.1 Separation

DNA fragments for the production of probes were separated by the following method. Completely digested DNA samples were mixed with 6x gel loading buffer (Section 2.3.4) and loaded into 1.8 cm-wide wells of submarine minigels, set at 4°C. Gels were poured onto thin glass support plates lying within the perspex gel mould, using 0.7 - 2.0% low melting point agarose (BRL) made up in TAE buffer. Gels were pre-run for 10-20 min at 4°C, loaded with samples and markers and electrophoresed (2-3 V/cm, 4°C) in TAE buffer containing 0.5 µg ml⁻¹ ethidium bromide. Run times varied between 1.5 and 4 hours, depending on the separation required. Following electrophoresis gels were carefully removed from submarine minigels on the supporting glass plates, and DNA fragments...
visualised with a hand-held, long-wave (366 nm) UV lamp. Appropriate band(s) were excised from the gels with sterile blades, and remaining gels were photographed (as in 2.3.4) to confirm excision of the correct band(s).

2.3.5.2 Recovery of DNA fragments from agarose gels

DNA fragments separated as in Section 2.3.5.1 were extracted from excised agarose blocks using one of two methods.

**Geneclean purification:** Most fragments to be recovered were purified using a Geneclean kit (Bio 101), according to the manufacturer’s instructions.

**Elution from DEAE membranes:** DNA was recovered from agarose gels by electrophoretic transfer onto DEAE membrane (Diethylaminoethyl cellulose, Schleicher and Schuell) and eluted in high salt. DEAE membranes were prepared by washing in 10 mM EDTA pH 8 for 10 minutes, 0.5 M NaOH for 10 minutes, followed by five 2 minute washes in H₂O. Membranes were stored at 4°C in H₂O. Following electrophoresis a slot was cut in front of the band of interest. A small piece of DEAE membrane was inserted into the slot and electrophoresis was continued for a further 15 minutes. The membrane was rinsed twice with 200 µl of washing buffer (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.1 mM EDTA) and the DNA eluted at 65°C for 30 minutes in buffer (1 M NaCl, 20 mM Tris-HCl pH 8.0, 0.1 mM EDTA). This was repeated and the eluates were pooled. The DNA fragment was phenol/chloroform extracted and precipitated with ethanol (4 volumes).

2.3.6 Ligation and transformation

2.3.6.1 Ligation

Ligation reactions were carried out in 10 µl containing: 50 mM Tris-HCl pH 7.8, 10 mM MgCl₂, 20 mM dithiothreitol, 1 mM ATP and 1 unit of T4 DNA ligase. Ligation of DNA fragments in low melting point agarose required melting of the gel slices at 70°C for 10 minutes prior to ligation reactions and again prior to transformation. The ratio of insert to vector was determined according to the guidelines of Revie *et al.* (1988). Generally a standard quantity of 10 ng of vector with a 10 fold molar excess of insert was found to provide sufficient numbers of transformants. Incubations were at 14°C for 8-16 hours or at room temperature for 2-4 hours.

2.3.6.2 Heat shock transformation

Mid-log phase cultures of cells were centrifuged (0°C, 3000 g, 10 min), resuspended in 100 mM MgCl₂ and immediately re-centrifuged (0°C, 3000 g, 10 min). The cells were resuspended in an equal
volume of 100 mM CaCl₂ and left on ice for 30 min. Finally, cells were pelleted and resuspended in 1/20 volume of 100 mM CaCl₂, 15% glycerol and used immediately or stored at -70°C.

For transformation, 100 μl of competent cells were mixed with 1-10 ng of DNA and chilled on ice for 60 min, followed by a heat-shock (42°C, 2 min). L-broth (1 ml) was immediately added, the cells grown at 37°C for 60 min and then spread onto antibiotic plates to select for transformants. Normally, transformation efficiencies averaged about 1x10⁶ transformants/μg of closed-circular pBR322 plasmid DNA.

2.3.6.3 DMSO transformation method

Competent cells were prepared using the following one step method described by Chung et al. (1989). A 1:100 diluted overnight culture of bacteria was incubated in Luria broth at 37°C with shaking and grown to log phase (2-3 x 10⁸ cells ml⁻¹). Cells were pelleted and then made competent by gently resuspending in TSS buffer, which contains DMSO (dimethyl sulfoxide).

Portions of freshly prepared competent cells (100 μl) were gently mixed with 1-10 ng of DNA in a cold glass tube and left at 4°C for 60 min. L-broth (1 ml) plus 20 mM glucose was immediately added, the cells out grown at 37°C, with shaking, for 1 hour, and spread onto antibiotic plates to select for transformants. Transformation efficiencies averaged about 5x10⁶ transformants/μg of closed-circular pBR322 plasmid DNA.

2.3.7 Radioactively labelled nucleic acid probes

The DNA to be labelled was either ‘DEAE-eluted’ or ‘Geneclean-purified’ restriction fragment, with ³²P-dCTP nucleotide (New England Nuclear) used for both nick translation and random priming experiments. Unincorporated nucleotides were separated from labelled product by centrifugation (25°C, 650 g, 5 min) through 1 ml Sephadex G50 columns. Total counts per minute and percentage incorporation were quantified by measurement of Cherenkov radiation from a portion of the purified labelled probe.

2.3.7.1 Nick translation

Originally, most ³²P-labelled DNA-probes were generated using a BRL nick translation kit (cat. #: 8160SB), and by following a modified version of the protocol supplied with the kit.

DNA to be labelled (50-200 ng) was added to 5 μl of nucleotide mix (solution A2 - 0.2 mM dATP, dTTP, dGTP in buffer), 2 μl [α-³²P]-dCTP (20 mCi, 6.6 pmol), and water added to 45 μl. 5 μl of DNA polymerase/DNase mix (solution C) was added to give a final reaction volume of 50 μl and

Materials and Methods
this mixture was incubated at 15°C for 60 min. The reaction was stopped by the addition of 5 μl of stop buffer (solution D - 300 mM EDTA pH 8.0). The final volume was adjusted to 100 μl with water, and unincorporated nucleotides removed by gel filtration as described in Section 2.3.7.3.

2.3.7.2 Random priming

Random priming later replaced nick translation as the method of choice, since a higher specific activity was consistently obtained. Routinely, 20 ng of template DNA were used to make [α³²P]-dCTP labelled probes. A commercial kit supplied by Bresatec or US Biochemical Co. was used and the reactions carried out according to the manufacturers instructions. A specific activity of 1x10⁷ to 1x10⁹ cpm μg⁻¹ was usually achieved.

2.3.7.3 Gel filtration

'Spin columns' for purification of labelled DNA from unincorporated oligonucleotides were prepared using a modified version of the method of Maniatis et al. (1982).

A 1 ml disposable plastic syringe was plugged with a siliconised glass bead and filled with Sephadex G-50 (medium grade) to a 1 ml settled volume. Excess water was removed by centrifuging (25°C, 600 g, 5 min) in a swinging bucket centrifuge (I.E.C. model PR-2). The syringe was equilibrated with 100 μl volumes of water and centrifuged as above until the volume of water recovered equalled the volume applied. The radioactive probe mix was applied to the column and collected in an eppendorf tube after centrifugation was repeated (25°C, 600 g, 5 min).

2.3.7.4 Scintillation counting

All purified, radioactively labelled DNA fragments (probes) were counted to determine specific activities by liquid scintillation. 1 μl of sample was added to 7 ml of Aqueous Counting Scintillant (ACS II, Amersham) in plastic counting vials, and counted on a Beckman LS 3801 machine. Dpm’s were estimated from the cpm values obtained by assuming 100% counting efficiency for ³²P.

2.3.8 Transfer of DNA to charged membranes

2.3.8.1 Southern blotting - alkaline Transfer

Alkaline transfer to Hybond-N⁺ membrane (Amersham) was performed essentially as described by Chomczynski & Qasba (1984), with the addition of an acid depurination step to improve transfer of large DNA fragments.
The agarose gels (0.7%-1.0%) were run, stained and photographed before soaking in 250 ml of 0.25 M HCl for ~30 min, with occasional gentle agitation, for depurination. After rinsing in deionised water, the gel was placed in 200 ml of Transfer Solution (0.4 M NaOH, 0.6 M NaCl) for ~30 min. Capillary blotting (Southern 1975) to Hybond-N+ was then allowed to occur overnight in a pyrex dish containing ~200-300 ml of Transfer Solution. Following transfer, the membrane was placed in Neutralising Solution (0.5 M Tris-HCl pH 7.0, 1 M NaCl) for 15 min, then blotted dry on paper towels. Transfer efficiency was checked by UV illumination of both gel and membrane. Membranes were either used immediately (prehybridised) or stored moist at 4°C, covered with plastic wrap.

2.3.8.2 Dot blots

Solutions of DNA to be probed were spotted (1-200 ng) onto dry Hybond-N+ membranes, dried at room temperature and treated as described in Section 2.3.8.1.

2.3.8.3 Colony hybridisation

Colony hybridisation to Genescreen-Plus (New England Nuclear) was carried out using the protocol described by Bergquist (1987).

2.3.9 Hybridisation, washing and autoradiography

2.3.9.1 Prehybridisation and hybridisation

Prehybridisation and hybridisation were carried out using a modified protocol from Chomczynski & Qasba (1984), and from a Genescreen protocol booklet (catalog no. NEF 976, NEN products).

**Prehybridisation mixture**: 50% (v/v) formamide (deionised), 50 mM Tris-HCl pH 7.5, 1.0 M NaCl, 2.5% SDS, 50-100 μg ml⁻¹ of heat-denatured, sonicated non-specific DNA.

**Hybridisation mixture**: Same as for prehybridisation mix, but with 1% SDS and 10% PEG 6000.

Hybridisation volumes varied between 10 ml and 50 ml, depending on the number and size of the membranes to be probed. Membranes were placed in prehybridisation mixture and incubated at 70°C for 20-30 min with gentle, occasional agitation. Non-specific DNA (sonicated calf-thymus or fish sperm DNA) was denatured by boiling for 10 min, added to the prehybridisation mixture at final concentration of 50-100 μg ml⁻¹ and incubated at 42°C for 1 hr with gentle occasional agitation.

Materials and Methods
The prehybridisation mixture was removed, and replaced with hybridisation solution and ³²P-labelled probes (prepared by boiling for 30 seconds to denature) at a final radioactive concentration of 1-5x10⁵ dpm ml⁻¹. Membranes for hybridisations were incubated, in plastic containers on a rocker, at 42°C for a minimum of 16 hours.

2.3.9.2 Stringency calculations

For all hybridisations very high stringency (>95%) was used to decrease the number of "false positives". As low ionic strength washes were used this meant that the temperature of the second wash determined the final stringency (Wetmur 1991). The following calculations and assumptions were used to determine the final wash temperature and hence degree of stringency of hybridisation of *Caldocellum saccharolyticum* DNA sequences.

**Assumptions:**

(i) As *C. saccharolyticum* has an average % G+C ratio of 34-35% the melting temperatures (Tₘ’s) of its’ DNA sequences are estimated as: \( Tₘ Δ 60°C \), (R. Grayling, M.Sc. thesis, Auckland University).

(ii) \( Tₘ \) decreases by 1°C for every 1.7% mismatch in hybridising DNA sequences (Caccone *et al.* 1988), i.e \( \% \text{ mismatch} = Tₘ \Delta T_{wash} \), where \( T_{wash} \) is the second wash temperature.

Therefore a 5% (5% / 1.7 = 3°C) mismatch (95% stringency) in *C. saccharolyticum*: \( Tₘ = 60 \) would require a second wash temperature of 57°C.

**Wash Solutions**

2x Primary wash: 2 x SSC, 1% SDS at 42°C for 20 min
1x Secondary wash: 0.1 x SSC, 1% SDS at 57°C for 20 min

2.3.9.3 Autoradiography

Membranes containing bound DNA probes were blotted dry on paper towels after washing, and covered with plastic wrap. Radioactive ink ([α-³⁵S]dATPαS in Indian ink) was spotted onto covered membranes where necessary for orientation. Membranes were then placed in an autoradiography cassette (Kodak X-omatic) with 1 or 2 intensifying screens (Cronex Lightning Plus, Du Pont), and autoradiographed at -70°C with Kodak X-Omat RP X-ray film. Autoradiography times varied from 2 hours to several days. Film development was as recommended by the manufacturer.
2.3.9.4 Removal of probes

Blots were 'stripped' of radioactive probe for re-use by immersion in boiling 0.5% SDS for 10 min. They were allowed to cool to room temperature, blotted dry on paper towels and stored at 4°C in plastic wrap, or re-used after autoradiography (overnight) to check the efficiency of probe removal.

2.3.10 Genomic DNA preparations

Genomic DNA was isolated from \textit{C. saccharolyticum} by modification of the enzyme extraction procedure of Grimberg et al. (1989). One ml of overnight culture was pelleted and washed with 1 ml TNE (10 mM Tris pH 8.0, 10 mM NaCl and 10 mM EDTA) and resuspended in 270 μl of TNE plus 1% (v/v) Triton X-100. The cell wall was removed by the addition of 30 μl of lysozyme (5 mg ml⁻¹) and incubation for 30 min at 37°C. The addition of Proteinase K (1 mg ml⁻¹ final concentration) and incubation for 2 hours at 65°C lysed the cells releasing chromosomal DNA, which was precipitated with 1/10 volume (NH₄)₂OAc and 3 volumes EtOH.

2.3.11 Polymerase chain reaction

The Polymerase Chain Reaction (PCR) as described by Sambrook et al. (1989) was used to synthesise defined DNA fragments. A reaction volume of 50 μl, consisting of PCR buffer (100 mM Tris-HCl (pH 8.8), 500 mM KCl, 2.5 mM Mg²⁺, 0.01% Gelatin), dNTPs (Pharmacia 250 μM), 0.25 units of polymerase (Cetus Corporation AmpliTaq or BRL TaqI polymerase), oligonucleotide primers (50-100 ng), and 1-100 μM of target DNA, was routinely used. Mineral oil (30 μl) was placed on top of the reaction mixture to prevent evaporation during amplification. Unless otherwise stated the cycling times and temperatures shown in table 2.9 were used for all PCR reactions.

<table>
<thead>
<tr>
<th>Number of Cycles</th>
<th>Steps</th>
<th>Temperature/Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-30</td>
<td>Denaturation</td>
<td>94°C, 1 minute</td>
</tr>
<tr>
<td></td>
<td>Annealing</td>
<td>55°C, 1 minute</td>
</tr>
<tr>
<td></td>
<td>Polymerisation</td>
<td>72°C, 1 minute</td>
</tr>
</tbody>
</table>
2.3.12 Yeast DNA methods

2.3.12.1 Yeast transformations

Yeast cells were transformed using a modified lithium acetate heat shock method (Treco and Lundblad 1990). A single colony was grown in 50 ml of YPD media to late log phase (OD$_{600}$= 0.4). The cells were harvested by centrifugation (25°C, 3000 g, 5 min), washed twice in TE buffer, resuspended in lithium acetate solution (0.25 M lithium acetate in TE buffer, pH 8.0) and incubated at 30°C for one hour. The plasmid DNA was mixed with 0.3 ml of competent yeast cells and 0.7 ml of PEG solution (50% PEG4000 in lithium acetate solution) and incubated for a further two hours at 30°C. The mixture was then heat-shocked at 42°C for 5 min, cooled and 100 μl plated onto minimal plates (SD media) with the appropriate amino acid supplement(s) present. Recombinant colonies appeared after 72 hours incubation at 28°C.

2.3.12.2 Plasmid isolation from yeast

Yeast plasmids were isolated by a phenol/chloroform/SDS procedure modified from Polaina and Adam (1991). Stationary phase yeast cells (1.5 ml) were centrifuged (25°C, 3000 g, 5 min) and resuspended in 200 μl of cell breakage buffer (2% Triton X-100, 1% SDS, 100 mM NaCl in TE), 100 mg of glass beads (0.5 mm) and 200 μl of phenol/chloroform. Cells were ruptured by vortexing at high speed for 2 min, and centrifuged (25°C, 8000 g, 15 min) to phase partition the aqueous phase, containing the plasmid, from the phenol phase. The aqueous layer was removed, precipitated at -20°C with 2 volumes ethanol and 1/10 volume sodium acetate and centrifuged (4°C, 8000 g, 15 min). The plasmid DNA pellet was washed in 80% ethanol, dried and resuspended in 20 μl H$_2$O. As the yield of plasmid DNA was substantially less than from an E. coli strain containing a high copy number plasmid, it was impossible to analyse the resulting DNA by restriction fragment analysis. Two methods were used to confirm the authenticity of the plasmid, either “retransformation” back into E. coli (see section 2.3.6) and reisolation (see section 2.3.2), or the polymerase chain reaction was used to amplify a specific target sequence (Sathe et al. 1991).
2.4 Sequencing

2.4.1 Single stranded DNA isolation

Single stranded DNA was prepared from either M13 phage (normally mp10) or the phagemid vectors pGEM 5 fZ(+) and pGEM 7 fZ(+).

2.4.1.1 M13 phage

Single stranded M13 template was isolated for sequencing as follows. Stationary phase PB2963 (JM101) was diluted 1:100 in YT broth. 2 ml portions were dispensed into test tubes, each infected with a single plaque of M13 (Messing 1991) and incubated with vigorous aeration for 6 hours at 37°C. Bacteria were pelleted by centrifugation (25°C, 8000 g, 10-20 s) in an Eppendorf centrifuge, then the supernatant added to a microfuge tube, containing 200 µl of 20% PEG, 2.5M NaCl, and mixed by inversion. The phage was precipitated at room temperature for 15 minutes and centrifuged (25°C, 8000 g, 15 min) before being resuspended in 100 µl TE (10 min on ice). It was then mixed with an equal volume of phenol (TE saturated) and recentrifuged (25°C, 8000 g, 15 min). The aqueous phase (95 µl) was mixed with 250 µl ethanol and 10 µl of 2 M sodium acetate, pH 5.2, and precipitated overnight at -20°C. The precipitate was centrifuged (4°C, 8000 g, 15 min), washed in 80% (v/v) ethanol, dried under vacuum, and the M13 single stranded DNA was resuspended in 30 µl TE.

2.4.1.2 Phagemids

Although the M13 phage system produced larger quantities of DNA and was the preferred sequencing vector, the larger multi-cloning site of the pGEM vectors allowed different fragments to be ligated and sequenced. The pGEM vectors are synthetic vectors based on the M13 phage but without the ability to package single stranded phage, so a co-inoculation with a mutant packaging phage (R408, Promega), referred to as a "Helper phage", was required. The procedure for the isolation of phage was identical to the M13 method except that in the initial inoculum plaques were replaced by F' host bacteria containing the plasmid phagemid and a co-inoculum of helper phage. The resulting culture was grown and the phage isolated as per section 2.4.1.1.

2.4.2 Oligonucleotide primers

Besides the standard oligonucleotide primers used for sequencing (M13 forward and reverse primers) a number of others were synthesised to complete the sequence of pNZ1038. These primers

Materials and Methods
were designed using an energy minimisation program ENRGFIT (See section 2.7) to check for, and reduce, the possibility of secondary binding sites.

<table>
<thead>
<tr>
<th>Number</th>
<th>Sequence (5’-3’)</th>
<th>Coordinates (5’-3’) (pNZ1038)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GGC TAT TCA ATC GGA G</td>
<td>2253-2269 sequence reverse direction</td>
</tr>
<tr>
<td>2</td>
<td>CCA AAA GAT GTC CTC ACC AAC C</td>
<td>2022-2044 sequence forward direction</td>
</tr>
<tr>
<td>3</td>
<td>GAG GTG CGA AAA CGA AAT CCC</td>
<td>2358-2379 sequence reverse direction</td>
</tr>
<tr>
<td>4</td>
<td>TGT GTG ATA TGT GTG ATT GA</td>
<td>2619-2639 sequence forward direction</td>
</tr>
<tr>
<td>5</td>
<td>GAC TCA TGT TCA TCT TTT ACC</td>
<td>2833-2854 sequence reverse direction</td>
</tr>
<tr>
<td>6</td>
<td>GGG ATG TAT ATC TCC AGG GG</td>
<td>3216-3236 sequence reverse direction</td>
</tr>
<tr>
<td>7</td>
<td>GGG AGA TAT ACA TCC CAA G</td>
<td>3221-3239 sequence forward direction</td>
</tr>
<tr>
<td>8</td>
<td>GGT AGT TGT AAT TTT CAC CA</td>
<td>3484-3503 sequence forward direction</td>
</tr>
<tr>
<td>9</td>
<td>CTC AGG CGA AAT TGT CCA GTT G</td>
<td>3688-3709 sequence forward direction</td>
</tr>
<tr>
<td>10</td>
<td>GCA GGA AGT TCA GAT TAT CCG</td>
<td>4303-4323 sequence reverse direction</td>
</tr>
<tr>
<td>11</td>
<td>GTA GCA TTT GCT CAG ATG G</td>
<td>3484-3503 sequence reverse direction</td>
</tr>
<tr>
<td>12</td>
<td>CAC TAT CAT CAA ACT CAC CC</td>
<td>3991-4010 sequence forward direction</td>
</tr>
<tr>
<td>13</td>
<td>GCA AAT TGT CAA CAG GC</td>
<td>4242-4258 sequence forward direction</td>
</tr>
<tr>
<td>14</td>
<td>ACC CCA AAT ATT GCA C</td>
<td>4513-4528 sequence forward direction</td>
</tr>
<tr>
<td>15</td>
<td>GTT GAG GCA ACC TCG GC</td>
<td>4712-4728 sequence forward direction</td>
</tr>
<tr>
<td>16</td>
<td>GGC ATT GTG TTT CCC G</td>
<td>1597-1611 sequence reverse direction</td>
</tr>
</tbody>
</table>

Those primers that provide sequence in the forward direction bind to the bottom strand and produce sequence identical to the top strand, whereas those that provide sequence in the reverse direction bind to the top strand and produce sequence identical to the bottom strand.

Oligonucleotides for sequencing were purified on a 20% denaturing acrylamide gel, viewed under UV light (320 nm) and extracted using the method described by Maniatis et al. (1982).
2.4.3 Sequencing methods

The technique used for the sequencing of DNA was based on the dideoxynucleotide chain termination method described by Sanger et al. (1977). Early sequencing was carried out using Klenow large fragment DNA polymerase with the BRL 1KB® sequencing method, while in later sequencing Sequenase® (USB Chemicals) was employed for chain extension, following the procedures of Tabor and Richardson (1987) and Tsang and Bentley (1988).

2.4.3.1 Single strand sequencing

The single stranded DNA produced in section 2.4.1.1 was sequenced using the standard protocols listed in section 2.4.3 using 7 µl of single stranded DNA template.

2.4.3.2 Plasmid sequencing

Before double stranded DNA (plasmids) could be sequenced, denaturation was required to separate the strands and allow the sequencing primer to bind. Denaturation was carried out using a modified alkaline denaturation technique (Hattori and Sakaki, 1986). Double stranded DNA, 3-5 µg in a volume of 7 µl, was denatured with 2 µl of denaturation buffer (0.2 M NaOH, 0.2 mM EDTA), neutralised with 1 µl of 3 M Na₂OAc, precipitated with 30 µl EtOH, centrifuged (4°C, 8000 g, 15 min) and washed with 70% EtOH. The pelleted DNA was redissolved in 7 µl of H₂O, mixed with 1 µl primer (0.5 pmol/µl) and 2 µl of Sequenase® reaction buffer, and left for 10-20 minutes at room temperature to allow primer-template annealing to take place. From this point standard sequencing protocols were followed.

2.4.4 Acrylamide sequencing gels

For DNA sequencing, standard size (20 cm x 40 cm) 6% polyacrylamide 7 M urea denaturing gels (5.7% acrylamide, 0.3% bis-acrylamide) were prepared in TBE buffer. Samples were denatured (i.e. made single stranded) by heating for 3 minutes at 100°C in 47.5% formamide, 10 mM EDTA, 0.025% bromophenol blue and 0.025% xylene cyanol F. Electrophoresis was carried out at 38 W using an LKB constant current power supply. Wedge gels, which were used to increase the number of readable bases obtainable from one sequencing gel, had a thickness of 0.1 mm (top) to 0.3 mm (bottom). The wedge configuration was achieved by the addition of two extra 0.1 mm spacers at the bottom of the gel. Wedge gels were run for 1-2 hours or until the bromophenol blue dye marker reached the bottom of the gel, allowing up to 300+ bp to be resolved. Increased sequence resolution (up to 450 bp) was obtained by using standard 0.1 mm thickness acrylamide gels run for 3-6 hours.
Following electrophoresis, gels were fixed in 10% methanol/10% acetic acid for 20-40 minutes, transferred onto Whatman 3MM paper and dried under vacuum for 30-60 minutes at 85°C. Sequences were detected by exposure to X-ray film at 4°C for 1-14 days depending on the amount of radioactivity present in the gel.

2.4.5 Sequencing library

Two methods for producing sequencing libraries were used, with a total of 5 libraries produced. The pseudorandom libraries were constructed from fragments generated restriction enzymes, whereas the random libraries were produced from sheared, sonicated DNA fragments.

2.4.5.1 Pseudorandom library

Pseudorandom libraries of thermophilic DNA from the plasmid pNZ1452 were generated by digestion with restriction endonucleases (with 4 base pair recognition sequences) which cut DNA giving blunt ends (statistically, these should cut every 256 base pairs). These DNA fragments were ligated into SmaI cut M13mp10, and the ligation mixture (20 μl) was mixed with JM101 plating bacteria (300 μl) and 5 ml of 0.7% soft agar. This mixture was overlaid onto a minimal agar plate (vitamin B1, glucose, IPTG and BCIG) and grown overnight at 37°C. Recombinant phages were selected and grown as in section 2.4.1.1.

1 IPTG (Isopropyl-thio-β-D-galactoside) an inducer of the lacZ gene (β-galactosidase).

2 BCIG (5-bromo-4-chloro-3-indolyl-β-D-galactoside) a chromogenic substrate which changes colour from clear to blue on hydrolysis of the "sugar dye" bond by β-galactosidase.

2.4.5.2 Random library

Random libraries were constructed in an identical manner to the pseudorandom libraries (see section 2.4.5.1) except that the DNA fragments were generated by sonication, which sheared the DNA into fragments of approximately 250-500 base pairs in length. These fragments were end-repaired using T4-DNA polymerase, Klenow fragment and an excess concentration of deoxynucleotide triphosphates (dNTP's) and incubated for 30 minutes at 37°C, before ligation into SmaI cut M13mp10.
2.5 Enzyme Purification

2.5.1 Isolation from *E. coli*

2.5.1.1 Cell harvesting

*E. coli* strains containing the *Caldocellum saccharolyticum* pullulanase gene were grown overnight at 37°C to the desired cell density, normally to stationary phase. The cells were harvested by centrifugation (4°C, 3000 g, 10 min) and the resulting bacterial pellet was washed twice in 56/2 to decrease the reducing sugar background. The bacterial cells were resuspended in 1/20 volume of sonication buffer (56/2 plus 1% (v/v) Triton X-100).

Two methods of cellular breakage were used, depending on the volume of cells harvested. For small volumes (< 5 ml) sonication was the method of choice (10 x 15 s pulses), whereas for larger volumes a French pressure cell (4 x 10^4 kPa) was used. Serial dilution replating assays were used to determine the degree of cellular lysis (routinely >99%). The samples were centrifuged (0°C, 25 000 g, 30 min) to separate the soluble proteins from the cellular debris.

2.5.1.2 Heat purification

The soluble cell extracts from 2.5.1.1 were further purified by heat denaturation of the *E. coli* cellular proteins, using the principle of increased thermostability of thermophilic proteins compared to the mesophilic *E. coli* cellular proteins. The cell lysates were incubated at the empirically determined optimal conditions of 70°C for 3 hours. On denaturation, the *E. coli* cellular proteins became insoluble and precipitated. They were removed by centrifugation (0°C, 25 000 g, 30 min). Results from a typical heat purification procedure are shown in figure 2.1.

2.5.1.3 Ammonium Sulphate Precipitation

The heat-stable protein supernatant from 2.5.1.2 was further purified by ammonium sulphate fractionation. This involved cooling to 0°C and the addition of solid ammonium sulphate with slow stirring until the desired degree of saturation was obtained, as determined from standard tables (Dawson *et al.* 1986). The solution was maintained at 0°C for one hour with constant stirring and the precipitated protein removed by centrifugation (0°C, 25 000 g, 60 min).
Figure 2.1 The results of a typical heat purification procedure

<table>
<thead>
<tr>
<th>Period of Heat Treatment at 70°C (min)</th>
<th>Protein concentration (mg/ml)</th>
<th>Pullulanase activity (maltotriose µg/min/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>75</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Purification Ratio (Pullulanase activity/protein)

1.7 3.4 13 29 70

2.5.1.4 Enzyme location within the *E. coli* cell

The location of the expressed pullulanase was determined by a procedure modified from Cornelius *et al.* (1982). The flow diagram in figure 2.2 shows the procedures and the fractions obtained. Mannitol was used instead of sucrose due to its lower reducing sugar background.
Figure 2.2 Preparation of cellular fraction for determination of pullulanase location

10 mls of Cell Culture

Centrifugation at 5000 g for 20 mins

Fraction

5 Extracellular

4 Cell associated

3 Periplasmic

2 Membrane

1 Cytoplasmic

Washed x 2
Tris 10 mM pH 8.0 +
100 mM NaCl (2 ml)

Centrifugation at 5000 g for 20 mins

Resuspended in 2 ml of Spheroplast buffer (Mannitol 0.5 M, 1 mM EDTA and lysozyme 1 mg/ml) and incubated at 20°C for 1 hr

Centrifugation at 7000 g for 20 mins

Resuspended in 2 ml TE buffer and spheroplasts disrupted by sonication

Centrifugation at 40 000 g for 60 mins

Cell debris homogenized in 2 ml T.E plus 1% Triton X-100
1000 mM NaCl

Prepericated

Non Extractable
Figure 2.3 A schematic view of enzyme location within the *E. coli* cell

![Diagram of enzyme location within *E. coli* cell]

The position numbering is the same as the fraction produced (method 2.5.1.4)

- Fraction / Position 5 = Enzyme secreted into the growth media
- Fraction / Position 4 = Enzyme cell associated (either with the cellwall or outer membrane)
- Fraction / Position 3 = Enzyme secreted into periplasm
- Fraction / Position 2 = Enzyme insoluble or membrane associated but extractable with NaCl/Triton X-100
- Fraction / Position 1 = Enzyme soluble in the cytoplasm
- Fraction / Position 0 = Enzyme insoluble or membrane associated but nonextractable with NaCl/Triton X-100

### 2.5.2 Isolation from *Caldocellum saccharolyticum*

Since the pullulanase from *C. saccharolyticum* was exported into the culture medium during growth, it was isolated from the culture medium by (NH₄)₂SO₄ precipitation (30-70% cut) (see section 2.5.1.3). Cell fractions were produced by the same procedure as section 2.5.1.4.

### 2.5.3 Isolation from yeast

Protein extracts were prepared by mechanical breakage of yeast cells by vortexing in the presence of glass beads, according to the method of Dunn and Wobbe 1990.
2.5.4 Chromatography

2.5.4.1 Dialysis and desalting

Protein samples were dialysed against the appropriate buffer at 4°C across a cellulose dialysis membrane with a nominal molecular mass protein retention of 12 000 Da. Desalting was performed at room temperature using a Pharmacia PD-10 column containing Sephadex G-25 (void volume = 2.5 ml, molecular mass exclusion = 5000 Da).

2.5.4.2 Ion exchange

Ion exchange chromatography separates protein mixtures, at a particular pH, on the basis of net content density and surface distribution of ionised groups. After initial tests (see section 5.3) it was determined that anion exchangers at pH 8.0(-4°C) efficiently bound the pullulanase, which could then be eluted using a 0-500 mM NaCl gradient.

Cellulose-based ion exchange media were used instead of dextran-based media due to the possible binding interaction between the α-1,6-bonds of the dextrans and the pullulanase. For large scale preparation, the cellulose-based cation exchange media DEAE Sephacel (Pharmacia) was used (column size 10 x 2.6 cm), due to its beadlike structure improving column resolution and simplifying column packing, and its reduced column compression at high salt concentrations reducing flowrates.

2.5.4.3 Gel filtration

Gel filtration chromatography separates proteins on the basis of their hydrated molecular radius, which for globular proteins is proportional to their non-denatured molecular mass. Using protein standards of known molecular mass it is possible to generate a standard molecular mass calibration curve for a particular column and hence determine an apparent molecular mass of an unknown protein. Two gel filtration media were routinely used;

**Sephrose 6B** (Pharmacia), an agarose based medium with a molecular mass fractionation range of 1 x 10^4 - 4 x 10^6 Da, packed into a 50 cm x 2 cm column.

**Sephacryl S-200 Superfine** (Pharmacia), an allyl dextran covalently cross-linked to acrylamide, with a molecular mass fractionation range of 5 x 10^3 - 2.5 x 10^5 Da, packed into a 60 cm x 1.6 cm column.

Unless otherwise stated, all gel columns were run at 4°C using as gel buffer 20 mM Tris/HCl, pH 8.0(-4°C) and 100 mM NaCl. The Sephadex G-type (Pharmacia) dextran based medium was not used due to the possible binding interaction between its α-1,6-bonds and the pullulanase.

Materials and Methods
2.5.4.4 Affinity chromatography

The principle of pullulanase enzyme binding to various substrates was used for affinity chromatography. The substrates tested were amylose, amilopectin, pullulan, cross-linked amylose, cross-linked amilopectin and cross-linked pullulan (Leloup et al. 1991). Amylose powder (Sigma) gave the best result due to its high binding capacity and relatively fast flowrate. The conditions of adsorption (without substrate digestion) were determined empirically to be low ionic strength, MES/NaOH buffer (10 mM, pH 6.2) and 4°C. To elute the protein (deadsorption) it is suggested in the literature that the addition of a competing ligand to the column wash buffer should be employed (Saha et al. 1988). A number of potential ligands were tested under varying conditions. These potential ligands included glucose, maltotriose, pullulan and branched dextrins produced by α-amylase partial digestion of amilopectin. However, none of the above eluted the adsorbed enzyme from the column.

As the enzyme-substrate binding in affinity chromatography is essentially an ionic or hydrogen bond interaction, it is possible to disrupt the binding by increasing the buffer ionic strength. This effect was achieved by the addition of a NaCl gradient (0.0 - 2.0 M), and thus the protein was eluted. Only small column volumes were required for affinity chromatography due to the high selective specificity of the enzyme-ligand binding. Chromatography was performed using either small columns (Pharmacia K9/15, 1 ml (100 μg) of adsorbent), Pasteur pipettes (50 μl of adsorbent) with step-gradient elution, or batch adsorption using eppendorf tubes (50 μl of adsorbent) with centrifugation and supernatant removal.

2.5.5 Estimation of total protein concentrations

An estimation of total protein concentrations in all samples was obtained using either the dye-binding methods of Bradford (1976) or the more sensitive method of the BCA reaction (Pierce Chemical Company, Technical bulletin #23220 1989). For both methods a standard curve was obtained using Bovine Serum Albumin (BSA) at known concentrations. The procedures for each method are listed below.

A) Dye-Binding assay Coomassie Brilliant Blue Reagent was made according to the increased sensitivity protocol of Löffler and Kunze (1989). Normally 100 μl of sample was added to 900 μl of Reagent; incubated at room temperature for 2 min and the resulting absorbance measured in microcuvettes (1 ml) at 595 nm. A typical calibration curve obtained is shown in figure 2.4.

B) Pierce BCA® assay Based on the reduction of Cu²⁺ to Cu⁺, in an alkaline solution. The reduced Cu⁺ reacts with bichinchoninic acid causing a colour change (light green to deep purple). The Working Reagent was prepared according to the enhanced sensitivity protocol (incubation temperature 70°C) of Smith et al. (1985) adapted for microtiter plates. Normally 10 μl of sample was added to 50 μl of BCA Working Reagent and incubated at 70°C for 30 min. The resulting absorbance was measured at 562 nm using an automatic microtiter plate reader. A typical calibration curve obtained is shown in figure 2.5.
Figure 2.4 A calibration curve obtained using Coomassie Brilliant Blue

![Absorbance at 595 nm vs. Amount of protein (μg BSA)](image)

Standard errors are contained within the area of each symbol.

Figure 2.5 A standard calibration curve obtained with the Pierce BCA kit

![Absorbance at 562 nm vs. Amount of protein (μg BSA)](image)

Standard errors are contained within the area of each symbol.

Protein concentrations were estimated in column fractions (optical absorbance at 280 nm), using an in-line real-time UV detector or manually using a spectrophotometer.

Materials and Methods
2.6 Enzyme Characteristics

2.6.1 Enzyme assays

2.6.1.1 Reaction Buffers

Buffer pH was determined at the temperature of intended use (pH_{(temp)}). This was necessary since the temperature compensation dial on pH meters only adjusts for the change in the H⁺/glass//Ag/AgCl electrode potential difference at varying temperatures (i.e. a thermodynamic difference). As a buffer solution is in chemical equilibrium a change in temperature causes a change in the equilibrium position and hence pH. This difference can be seen in table 2.11, using the values obtained from the standard enzyme buffer, 0.2 M MES/NaOH (pH_{(70^\circ C)} 5.70) and 1 mM Ca^{2+}.

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Temperature compensation dial at 20^\circ C</th>
<th>Temperature compensation dial at 70^\circ C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffer at 20^\circ C</td>
<td>6.09</td>
<td>6.21</td>
</tr>
<tr>
<td>Buffer at 70^\circ C</td>
<td>5.45</td>
<td>5.70</td>
</tr>
</tbody>
</table>
Figure 2.7 Typical PAHBAH reducing sugar calibration curve obtained

A) Total Range

B) Linear Range

C) Absorbance obtained using equal molar sugar concentration

Standard errors are contained within the area of each symbol

Materials and Methods
2.6.1.2 PAHBAH reducing sugar assay

On hydrolysis, polysaccharides release oligosaccharides which increase the number of reducing sugars present and this increase can be used as an indication of enzyme activity. Reducing sugar activity was determined by using PAHBAH (p-hydroxybenzoic acid hydrazide) reagent, formula listed below. The PAHBAH reagent reacts quantitatively with the reducing sugars present. It is oxidised, changing colour from clear to a deep yellow ($\lambda_{\text{max}} = 420$ nm) (Lever 1973).

**PAHBAH reagent:** $1.2 \text{ g NaOH} + 0.76 \text{ g PAHBAH} + 0.63 \text{ g Na}_2\text{SO}_3 + 0.22 \text{ g CaCl}_2 + 0.558 \text{ g Na}_3\text{ citrate} + \text{H}_2\text{O}$ to 100 ml.

Standard assay conditions were determined empirically to be 100 $\mu$l of sample (containing 200 $\mu$g of pullulan) and 900 $\mu$l of PAHBAH reagent mixed and boiled for 5 min. The increase in absorbance was determined at 420 nm against a reference sample containing enzyme extract and substrate, but not incubated. A standard calibration curve was determined for three saccharides (glucose, maltose and maltotriose), as shown in figure 2.7.

A linear response was obtained in the region of 0.2-2.0 mmol l$^{-1}$ reducing sugar, which equated to 2-35 $\mu$g ml$^{-1}$ for glucose, 5-70 $\mu$g ml$^{-1}$ for maltose and 5-100 $\mu$g ml$^{-1}$ for maltotriose.

For each experiment a new standard curve had to be determined due to batch variation within the PAHBAH reagent, the age of the PAHBAH reagent and the difference in reducing background of each enzyme sample.

This assay formed the basis of the standard enzyme procedures as it was more sensitive than the dye-linked substrate assays.

Assay procedure used: Pullulanase extract (2 $\mu$l), 200 mg pullulan (8 $\mu$l), and 90 $\mu$l of enzyme buffer (0.2 M, pH 5.7 with 1 mM Ca$^{2+}$). For each assay sub-samples were removed and assayed to determine the initial reaction rate (i.e when steady-state kinetics existed) regardless of the amount of enzyme activity present. The amount of substrate used was optimised empirically, so that the substrate concentration was in excess without causing a large reducing background. Enzyme activity was expressed as either $\mu$g maltotriose released per min per amount of enzyme, or Units ($\mu$mols of maltotriose released per min per amount of enzyme).

2.6.1.3 Starch-iodine complex

In potassium iodide solution, iodine forms a stable chromogenic complex with the linear $\alpha$-1,4-D-glucose chains of oligosaccharides and polysaccharides. Four properties of the starch-iodine complex (which are related to chain length) were determined; a) colour, b) Blue Value (B.V.), c) Peak Value (P.V.) and d) $\lambda_{\text{max}}$. Blue Value was defined as optical density at 680 nm under standard assay...
conditions. Peak Value refers to the absorbance at \( \lambda_{\text{max}} \) of the starch-iodine complex under standard assay conditions. Standard assay conditions were 10 \( \mu \)l of KI/I\( _2 \) solution (I\( _2 \) 2 mg ml\(^{-1} \), KI 20 mg ml\(^{-1} \)), 980 \( \mu \)l of 0.15 M acetic acid and 100 \( \mu \)l of saccharide sample. The absorbance spectrum from 450 - 700 nm was determined spectroscopically and values of B.V., P.V. and \( \lambda_{\text{max}} \) were determined (Bailey and Whelan 1961).

2.6.1.4 Dye-linked substrates

Another method for detecting enzyme activity was by dye-linked substrates. These chromogenic substrates are polysaccharides with dye groups covalently bound to sugar residues and in common with all polysaccharides they are insoluble in 50\% (v/v) ethanol. On hydrolysis, oligosaccharides and any dye groups attached to them are released. Enzyme activity was determined by the addition of ethanol to 50\% (v/v) which caused the undigested polysaccharides to precipitate, whereas the oligosaccharides released by enzyme activity remained soluble. The coloured supernatant was removed and its absorbance determined spectrometrically.

Chromogenic substrates used were Remazol Brilliant Blue R dye-linked pullulan, made according to the protocols of Kanno and Tomimura (1985) and Biely et al. (1988); Azurine-Blue cross-linked pullulan (MegaZyme); and Blue amylose (Phadebas\textsuperscript{®}, Pharmacia).

2.6.1.5 Substrate preparation

All polysaccharides were purified before use by separation of low molecular mass oligosaccharides to minimise the substrate reducing sugar background. Purification was based on the principle that solubility in ethanol decreases with increased saccharide chain length.

Polysaccharides were dissolved in H\(_2\)O (10 ml) and ethanol added slowly until polysaccharide precipitation occurred (25-50\% ethanol v/v). The oligosaccharides and saccharides remained in solution. The precipitate was collected by centrifugation, washed twice in 50\% ethanol, dried and weighed.

2.6.2 Protein gel electrophoresis

2.6.2.1 SDS-PAGE

Denaturing discontinuous gel electrophoresis (SDS-PAGE) was used to analyse protein samples according to the method of Laemmli (1970), which separates proteins based on molecular size as they move through a polyacrylamide gel matrix towards the anode. Proteins were made soluble by boiling in the presence of SDS (4\%) and disulfide bonds were reduced using either dithiothreitol (DTT) or Materials and Methods
2-Mercaptoethanol (2-ME). Polyacrylamide gels were prepared according to the method of Gallagher and Smith (1987), normally at final acrylamide concentration of 8% in the separating gel that resolved proteins of 25-150 kDa.

Samples were boiled for five minutes with an equal volume of 2x sample buffer (0.0625 M Tris- HCl, pH 6.8; 2.5% SDS; 5% 2-ME and 10% glycerol) centrifuged and loaded onto the gel. Gels were run at a constant voltage of 140 V, at room temperature, until the bromophenol blue dye was 1 cm from the bottom of the gel.

2.6.2.2 Non-denaturing PAGE

Non-denaturing discontinuous gel electrophoresis (Native-PAGE) separates proteins on the basis of a combination of molecular properties including size, shape and charge. Gels and buffers were prepared and run as in Section 2.6.2.1 except that SDS and 2-ME were omitted from all buffers and reagents and the samples were not boiled.

2.6.2.3 Coomassie Blue Stain

Polyacrylamide gels were stained with 0.1% Coomassie blue G-250, 50% (v/v) methanol, 10% (v/v) acetic acid for thirty minutes, then washed several times with 10% (v/v) methanol, 7% (v/v) acetic acid until adequate destaining was achieved. This allowed the detection of protein bands containing 0.5-5 µg of proteins.

2.6.2.4 Phastsystem

Small quantities of protein (20-50 ng) were analysed using the Pharmacia PhastSystem® that consists of an automatic electrophoretic separation unit and gel staining module. The samples were loaded automatically onto ultra-thin acrylamide gels (PhastGels, 0.45 mm x 50 mm x 43 mm). Both gradient and homogeneous acrylamide gels were used depending on the molecular mass of the proteins to be separated. All gels were run according to the manufacturer’s instructions and protein bands visualised using silver stain.

2.6.2.5 Periodic acid-Schiff’s reagent

The presence of glycosylated protein was determined using the periodic acid-Schiff’s reagent method (Jay et al. 1990). Protein samples were run on SDS-PAGE gels (see 2.6.2.1) with identical samples on each side. The gels were then halved and one side was stained by standard Coomassie Blue Stain (2.6.2.3) while the other was periodic acid-Schiff’s reagent stained to detect glycosylated proteins, which stained magenta. Normally, the purified enzyme sample was run on an SDS-PAGE gel together with two controls. The positive control was heavily glycosylated crystalline style proteins.
from bivalve molluscs, provided by W. Judd, Auckland University (Judd, 1987). The negative control was standard SDS-PAGE molecular mass protein markers (Pharmacia).

**Periodic acid-Schiff's reagent staining procedure:** Fix gels (50% v/v methanol, 60 min); rinse (H₂O, 20 min); oxidise carbohydrate with periodic acid (2%, 15 min); rinse (H₂O, 2 min); stain with Schiff's reagent (Sigma) until bands turn magenta; rinse (H₂O, 2 min); remove excess colour with sodium metabisulfite (2%, 6h).

**2.6.2.6 Activity gels**

Active pullulanase bands in acrylamide gels were detected by the Zymogram dye-linked substrate overlay method (Yang and Coleman 1987). Protein samples were run on SDS-PAGE gels (see 2.6.2.1) with identical samples on each side. The gels were then halved and one side was stained with standard Coomassie Blue Stain (2.6.2.3) while the other was overlaid with dye-linked substrate to detect pullulanase active protein bands.

**Zymogram procedure:** Remove SDS buffer (25% v/v isopropanol and 50 mM acetate buffer pH 5.0, 20 min); rinse (50 mM acetate buffer pH 5.0, 20 min); air dry; overlay with a precast dye-linked substrate replicate (2% agar in 0.2 M MES/NaOH (pH70°C) 5.70) and 2 mM Ca²⁺) and incubate at 70°C until a zone of clearing appears (normally 4-16 h).

**2.6.2.7 Agarose G.E.**

An agarose based matrix (FMC BioProducts, Prosieve™ gel system) was used to isolate protein samples after separation by gel electrophoresis. Gels were cast and run according to the manufacturers instructions. SDS-denatured protein samples were run with identical samples on each side. The gels were then halved and one side was stained with standard Coomassie Blue Stain (2.6.2.3) while the protein sample was extracted from the other. The location of the desired band was determined by comparison to the stained gel.

**Protein recovery procedure:** Excise gel portion containing the protein; weigh; dilute with added extraction buffer (20 volumes of 100 mM Tris/HCl and 0.1% v/v Triton X-100); melt at 70°C; mix; freeze at -70°C to disrupt the agarose matrix; thaw and centrifuge (4°C, 13 000 g, 30 min) to separate the soluble protein from the agarose matrix.

**2.6.3 Paper chromatography**

The composition of the final hydrolysis products was determined by descending paper chromatography over 36 hr. The resolving solvent used was n-Butanol:pyridine:water (6:4:3 v/v), due
to its ability to resolve oligosaccharides of up to 9 glucose residues in length. Samples (1-10 mg) were deionised before loading and run on Whatman 3MMChr paper (55 cm in length). Sugars were visualised using the AgNO$_3$/NaOH oxidising method (Krebs et al. 1969).

Table 2.12 Oligosaccharide $R_g$ values

<table>
<thead>
<tr>
<th>Saccharide</th>
<th>$R_g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>1.000</td>
</tr>
<tr>
<td>Maltose</td>
<td>0.704</td>
</tr>
<tr>
<td>Maltotriose</td>
<td>0.502</td>
</tr>
<tr>
<td>Maltotetraose</td>
<td>0.333</td>
</tr>
<tr>
<td>Maltopentaose</td>
<td>0.210</td>
</tr>
<tr>
<td>Maltohexose</td>
<td>0.138</td>
</tr>
<tr>
<td>Maltoheptaose</td>
<td>0.093</td>
</tr>
<tr>
<td>Isomaltose</td>
<td>0.545</td>
</tr>
<tr>
<td>Isomaltotriose</td>
<td>0.368</td>
</tr>
</tbody>
</table>

§ migration relative to glucose in $n$-Butanol:pyridine:water (6:4:3 v/v).
2.7 Computer Analysis

Except for restriction fragment analysis (see section 2.7.5), which was carried out using an Apple Macintosh Classic computer, all analyses were performed on a Digital MicroVax II, using programs from the GCG molecular biology package supplement (Genetics Computer Group Devereux et al. 1984) with compatible programs from other sources.

2.7.1 Standard Programs

For general analysis the following GCG programs were used.

<table>
<thead>
<tr>
<th>Table 2.13 GCG programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program/s</td>
</tr>
<tr>
<td>MAP, MAPSORT, MAPPLOT</td>
</tr>
<tr>
<td>TOSTADEN, FROMSTADEN</td>
</tr>
<tr>
<td>TRANSLATE</td>
</tr>
<tr>
<td>REVERSE</td>
</tr>
<tr>
<td>REPEAT</td>
</tr>
</tbody>
</table>

2.7.2 Sequence analysis

Sequence data was analysed using the most appropriate program from either the GCG package or the SHOTGUN sequencing program package (Staden 1980). Sequence data was entered using the program SEQED (GCG) via a Graf/Bar Mark II digitizer, gels were aligned to each other using VTCOMP (SHOTGUN) and consensus sequences generated using VTUTIN (SHOTGUN).

2.7.3 Database searches

Up to date DNA and protein databases were included as part of the GCG package (listed below). The different DNA sequence databases (Genbank, EMBL and NBRF-Nucleic acid) were combined into one hybrid database called "GenEMBL"[72.0]. In a similar way the protein databases were combined into SWISSPROT [22.0].

Also included were a number of search programs designed to access these databases and find sequence homologies to the queried search sequence. Listed in table 2.15 are the programs most commonly used.
Table 2.14 Databases used from the GCG package

<table>
<thead>
<tr>
<th>Database</th>
<th>Sequence</th>
<th>Release Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleic Acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GenEMBL (72.0)</td>
<td>73 052 entries; 95 025 892 base pairs</td>
<td>June 1992</td>
</tr>
<tr>
<td>Protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWISSPROT (22.0)</td>
<td>25 044 entries; 8 375 696 amino acids</td>
<td>June 1992</td>
</tr>
</tbody>
</table>

Table 2.15 Database search programs

<table>
<thead>
<tr>
<th>Program</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRINGS</td>
<td>Searches database entry documentation with character pattern queries</td>
</tr>
<tr>
<td>FASTA</td>
<td>Searches for sequence similarity between the query sequence and the databases</td>
</tr>
<tr>
<td>TFASTA</td>
<td>Translates database DNA sequences into 6 reading frames before doing a FASTA search</td>
</tr>
<tr>
<td>CODONPREFERENCE</td>
<td>Determines potential genes from ORF’s based on codon usage</td>
</tr>
</tbody>
</table>

2.7.4 Homology comparisons

Homology comparison programs are designed to produce a “best” alignment between a group of sequences (normally two). All homology comparison programs used were from the GCG (GAP, BESTFIT, COMPARE and the multi-alignment program PILEUP).

2.7.5 Restriction fragment analysis

Restriction fragments were sized using the program ‘Macgel V1.6’ (Kessell and Gibbs, unpublished) in combination with a Bausch and Lomb digitizing pad. This program uses cubic-spline interpolation to fit data to molecular mass standard curves. It can make correction for differing migration rates between lanes when a band or bands of a known size are present within each lane.
Chapter Three

The Pullulanase expressed from *Caldocellum saccharolyticum*

3.1 Introduction

To determine the properties of the pullulanase when expressed in the native organism it was isolated from *Caldocellum saccharolyticum*. Expression studies involving different carbohydrate sources were then performed to determine whether pullulanase expression was substrate induced (see section 3.2). Using these results, enzyme expression was maximised to aid in further purification.

The *C. saccharolyticum* pullulanase was purified to homogeneity from the culture supernatant (see section 3.3). It was then possible to obtain sufficient pure pullulanase to perform enzyme characterisations (see section 3.4). In chapter 9 these results are compared to those obtained from the pullulanase expressed in *E. coli*.

3.2 Substrate Induction of Pullulanase Expression

For a number of pullulanases, it has been shown that the carbohydrate source used for growth affects the level of enzyme produced as well as the amount exported into the culture medium and the location of the enzyme within the cell (see section 1.3.3). Substrate induction tests were performed to ascertain the effect of different carbohydrate sources on the production of pullulanase from *C. saccharolyticum* (see sections 3.2.1-3.2.2).

For all experiments in this section, pullulanase activity was determined using the dye-linked substrate method (see section 2.6.1.4), rather than the standard PAHBAH reducing sugar assay (see section 2.6.1.2), due to the high reducing background from the growth media and the reducing sugar added for growth. Hence the errors are greater in this section due to the lower accuracy and reproducibility of the dye-linked substrate assay compared to the PAHBAH reducing sugar assay. Standard enzyme samples were used as controls to correlate the amount of dye released with a known amount of maltotriose released, as measured by the reducing sugar assay.
3.2.1 The effect of carbohydrate source on pullulanase synthesis

The enzyme expression levels of a number of other glucoytic enzymes from thermophiles have been shown to be substrate inducible (Takase and Horikoshi 1988, Hyun and Zeikus 1985c, Madi et al. 1987). These studies showed that enzyme expression was induced by maltose and repressed by glucose, in a similar fashion to the *E. coli* maltose regulon (see section 1.3.1). Substrate induction tests were performed using *C. saccharolyticum* to determine whether pullulanase synthesis was also substrate inducible. Maltose was tested to see if it acted as an inducer, whereas glucose was tested as a repressor and xylose as a control for the basal level of enzyme expression. The interaction between glucose and maltose was tested to determine whether repression was stronger than induction, and the level of induction was tested by using two polysaccharides (amylopectin and pullulan) which are substrates for the pullulanase.

The *C. saccharolyticum* cells were grown anaerobically overnight in 2/1 minimal media with xylose (0.2%) as the sole carbohydrate source (as in section 2.2.4). The 5-carbon sugar xylose was used to return cell concentrations of glucose and maltose (and hence enzyme expression) to their basal levels, as it is not involved in substrate induction or repression of the maltose induction pathway. From this culture portions were removed and grown in six different carbohydrate sources as listed in table 3.1.

<table>
<thead>
<tr>
<th>Carbohydrate source (w/v)</th>
<th>Proposed effect on maltose regulon</th>
<th>Expression level</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Xylose 0.2%</td>
<td>control</td>
</tr>
<tr>
<td>B</td>
<td>Maltose 0.2%</td>
<td>induction</td>
</tr>
<tr>
<td>C</td>
<td>Maltose 0.1%, glucose 0.1%</td>
<td>induction/repression</td>
</tr>
<tr>
<td>D</td>
<td>glucose 0.2%</td>
<td>repression</td>
</tr>
<tr>
<td>E</td>
<td>pullulan 0.2%</td>
<td>induction</td>
</tr>
<tr>
<td>F</td>
<td>amylopectin 0.2%</td>
<td>induction</td>
</tr>
</tbody>
</table>

As *C. saccharolyticum* cells are lysed during the stationary phase, cells were grown to late log phase only (personal communication K. Borges, Department of Cell Biology, Auckland University). Due to the different growth rates obtained from each carbohydrate source, pullulanase activity was determined at late log phase (defined as OD$_{600}$ = 0.6).
The amount of pullulanase activity was determined in both the culture medium and whole cell extracts using the dye-linked substrate assay. The ratio of cellular to extracellular pullulanase appeared to be approximately 1:1 (a more detailed study is shown in section 3.2.2). The results were summed and are shown in table 3.2.

Table 3.2 Effect of carbohydrate source on the amount of pullulanase produced by C. saccharolyticum

<table>
<thead>
<tr>
<th>Carbohydrate source in 2/1 media</th>
<th>Activity (μg maltotriose / min/ml of culture)</th>
<th>Units/litre (μmoles maltotriose / min/litre of culture)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Xylose 0.2%</td>
<td>0.77 ± 0.2</td>
<td>1.5</td>
</tr>
<tr>
<td>B Maltose 0.2%</td>
<td>4.1 ± 0.4</td>
<td>8.1</td>
</tr>
<tr>
<td>C Maltose 0.1%, glucose 0.1%</td>
<td>0.39 ± 0.1</td>
<td>0.77</td>
</tr>
<tr>
<td>D glucose 0.2%</td>
<td>0.19 ± 0.05</td>
<td>0.38</td>
</tr>
<tr>
<td>E pullulan 0.2%</td>
<td>3.8 ± 0.6</td>
<td>7.5</td>
</tr>
<tr>
<td>F amylopectin 0.2%</td>
<td>3.7 ± 0.5</td>
<td>7.3</td>
</tr>
</tbody>
</table>

These results show that enzyme expression levels were dependent on carbohydrate source. Maltose, pullulan and amylopectin were found to induce enzyme expression above the basal level, and glucose to repress enzyme expression. The repression effect of glucose was reduced by the induction effect of maltose. This result implies that pullulanase expression in C. saccharolyticum was both substrate inducible and repressible in a similar manner to the E. coli /Klebsiella maltose operon (see section 1.3.1) and a number of other thermophilic glucolytic enzymes (see section 1.3.3).

3.2.2 Cellular location of the C. saccharolyticum pullulanase

As shown previously in section 3.2.1, the C. saccharolyticum pullulanase was exported into the culture medium when expressed in the native organism. To determine whether the carbohydrate source had any effect on the location of the pullulanase, cellular fractions were produced (see section 2.5.1.4) for cells from each carbohydrate source. Although this fractionation protocol was developed for E. coli cells, it is a general fractionation protocol for bacteria. Lysozyme is used to degrade the C. saccharolyticum cell wall when extracting genomic DNA (see method 2.3.10), hence this protocol was likely to be an effective fractionation method for C. saccharolyticum. A schematic representation of possible pullulanase locations within the cell is shown in figure 2.3. Pullulanase activity was
determined for each fraction and the results are shown in figure 3.1.

Figure 3.1 Pullulanase activity in different fractions from different carbohydrate sources

The values obtained are the means of two enzyme determinations of each fraction. Since experimental error between duplicates can account for any difference between the means of each carbohydrate source, no significant variation in enzyme distribution was observed with carbon source used. The enzyme distribution was typical for an extracellular enzyme with most exported to the culture media (approximately 50%). The remainder was located predominantly in the soluble cytoplasmic fraction (approximately 30%), with some exported to the periplasmic space (approximately 5%) or outer membrane/cell wall associated fraction (approximately 10%), presumably in the process of being exported out of the cell. In contrast to the recombinant enzyme, very little activity was present in the insoluble cytoplasmic fractions (fractions 2 (2%) and 0 (<1%)).

3.2.3 Validation of strain identity

As different batches of *C. saccharolyticum* were used during the course of this thesis, it was important to show that the DNA sequence (see section 4.0) and the purified pullulanase (see section 3.0) were both obtained from the same organism, and that this organism was *C. saccharolyticum*.
Southern hybridisations were performed to show that all *C. saccharolyticum* cells and DNA used in this thesis were from the same organism. The probe used in southern hybridisations was the 1.83 kb fragment of the *C. saccharolyticum* pullulanase gene from pNZ1072. This fragment was isolated (see section 2.3.5.2), radioactively labelled (section 2.3.7) and used to probe other batches of *C. saccharolyticum* DNA. Genomic DNA from each source of *C. saccharolyticum* was isolated (see section 2.3.10), digested with restriction endonuclease *Hind* III, electrophoresed on an agarose gel (see section 2.3.5) and blotted onto nylon membrane (see section 2.3.8.1). Hybridisations between the 1.83 kb probe and the genomic DNA were performed at maximum stringency (see section 2.3.9). Figure 3.2 shows a typical hybridisation result.

**Figure 3.2 Hybridisation of the 1.83 kb fragment to the *C. saccharolyticum* genome.**

The results show that the 1.83 kb *Hind* III fragment from plasmid pNZ1072 hybridised to only an identically sized fragment produced from a *Hind* III restriction endonuclease digestion of the *C. saccharolyticum* genome.

Arbitrarily primed PCR has shown that all samples of *C. saccharolyticum* genomic DNA prepared in our laboratory gave an identical banding pattern (personal communication P. Bergquist, Department of Cell Biology, Auckland University) implying that they were all from the same organism.
3.3 Purification of the Pullulanase From *Caldocellum saccharolyticum*

The *C. saccharolyticum* pullulanase was purified to homogeneity from the culture supernatant using the purification procedure outlined in figure 3.3 and the purification table is shown in Table 3.3. The purification scheme is given in more detail in the following sections.

Figure 3.3 Summary of the procedures used in the purification of the pullulanase enzyme

Growth medium from *C. saccharolyticum* (4 litres)

Ammonium sulphate fractionation. 30%-70%

Ion exchange (DEAE Sephadex)

Affinity binding (Amylose powder)

Gel filtration (Sephacryl S-200)

Affinity binding (Amylose powder)

Pullulanase band extracted from Prosieve™ Agarose gel

The major purification steps are shown in bold with thick arrows
Table 3.3 Purification table of the C. saccharolyticum pullulanase

<table>
<thead>
<tr>
<th></th>
<th>Protein (µg)</th>
<th>Units</th>
<th>Specific Activity U/µg</th>
<th>Percentage recovered</th>
<th>Purification ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Culture supernatant</td>
<td>59400</td>
<td>53</td>
<td>0.000892</td>
<td>100</td>
<td>1.0</td>
</tr>
<tr>
<td>(NH₄)₂SO₄ Extract 1/5 continue</td>
<td>31500</td>
<td>43</td>
<td>0.00136</td>
<td>81.1</td>
<td>1.52</td>
</tr>
<tr>
<td>IEX DEAE Sephacel</td>
<td>6300</td>
<td>8.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gel filtration Sephacryl S-200</td>
<td>435</td>
<td>5.1</td>
<td>0.0117</td>
<td>48</td>
<td>13.1</td>
</tr>
<tr>
<td>Amylose binding</td>
<td>98</td>
<td>3.4</td>
<td>0.0326</td>
<td>32</td>
<td>36.5</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>1.8</td>
<td>0.0651</td>
<td>16.9</td>
<td>72.9</td>
</tr>
</tbody>
</table>

Note only 1/5 of the (NH₄)₂SO₄ extract (5 ml out of 25 ml) was applied to the ion exchange column.

Figure 3.4 The purification steps shown on SDS-PAGE

lane 1 Ammonium sulphate fraction
lane 2 Ion Exchange fraction
lane 3 Gel filtration Sepharose 6B fraction
lane 4 Gel filtration Sephacryl S-200 fraction
lane 5 Amylose binding fraction
lane 6 Pharmacia low molecular weight protein markers (sizes are in kDa)
Proteins were visualised with coomassie brilliant blue (4-15% phastgel SDS-PAGE)
3.3.1 Ammonium sulphate precipitation

Since *Caldocellum saccharolyticum* is an obligate anaerobe, it must be grown under anaerobic conditions. As the facilities for large scale anaerobic growth were not available at Auckland University, a 4 litre culture (2/1 media supplement with 0.2% maltose w/v) was supplied by the Thermophile and Microbial Biochemistry & Biotechnology Unit at Waikato University, New Zealand. The cells were harvested by centrifugation (4°C, 3000 g, 10 min) and the amount of pullulanase activity was determined in both the culture medium and whole cell extracts using the dye-linked substrate assay (as in section 3.2.1). A similar ratio of cellular to extracellular pullulanase (approximately 50% cellular extract, 50% growth medium) was obtained. As the protein concentration in the growth medium was significantly less than in the cellular extract, the pullulanase was purified from growth medium.

The initial purification step was an ammonium sulphate precipitation (see section 2.5.1.3). This fractionation procedure has the two-fold advantage of sample concentration and elimination of reducing sugars present in the growth media. Initial tests were performed with small sub-samples to determine the optimum degree of saturation, as shown graphically in figure 3.5.

Figure 3.5 The distribution of enzyme activity at different ammonium sulphate saturation values

As the ammonium sulphate percentage saturation increased, more of the pullulanase activity was present in the precipitated fraction. Using these results, an ammonium sulphate cut of 30-70% saturation was used to concentrate the growth medium from 4000 ml to 25 ml.

3.3.2 Ion exchange

Initial tests, using both cation and anion ion exchange media at varying pH, were performed to

The Pullulanase expressed from *Caldocellum saccharolyticum*
test the binding of the pullulanase. The results are summarised in Table 3.4.

<table>
<thead>
<tr>
<th>Media</th>
<th>Type</th>
<th>pH (°C)</th>
<th>Pullulanase Binds</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM-cellulose</td>
<td>Cation</td>
<td>5.5</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.0</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.0</td>
<td>No</td>
</tr>
<tr>
<td>DEAE-cellulose</td>
<td>Anion</td>
<td>8.5</td>
<td>Some</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.0</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.0</td>
<td>Yes, eluted with 300 mM NaCl</td>
</tr>
</tbody>
</table>

After initial tests it was determined that anion exchangers at pH 8.0 (°C) efficiently bound the pullulanase, which could then be eluted using a 0-1000 mM NaCl step gradient. For large scale preparation 1-2 ml of ammonium sulphate extract was deionised using a PD-10 column (see section 2.5.4.1), applied to a DEAE Sephacel column (column buffer 20 mM Tris-HCl pH 8.0 (°C) and 0.01% v/v Triton X-100) (see section 2.5.4.2), and eluted using a 0-500 mM NaCl gradient. Pullulanase activity was determined using the PAHBAH reducing sugar assay (see section 2.6.1.2) for every fraction. A typical chromatogram is shown in figure 3.6 together with the location of pullulanase active fractions.

Figure 3.6 A DEAE-Sephacel ion exchange chromatogram
In an attempt to further purify the pullulanase by ion exchange, active fractions were deionised as before, applied to an FPLC Mono Q chromatofocussing column (Pharmacia) and eluted using a NaCl gradient. This procedure did not result in any significant improvement in purity.

### 3.3.3 Gel filtration

Initially, due to the possibility that the enzyme was exported as either a multimeric form or a high molecular mass multi-enzyme complex, a gel filtration medium with a large molecular mass separation range was used (Sepharose 6B, 4000-10 kDa). Two samples were applied to the column, the original ammonium sulphate extract and the combined pullulanase active fractions from the DEAE Sephadex ion exchange column (section 3.3.2). Pullulanase activity was determined using the PAHBAH reducing sugar assay (see section 2.6.1.2) for every fraction. A calibration curve obtained with molecular mass markers (thyroglobulin, 669 000; apoferritin, 443 000; alcohol dehydrogenase, 150 000; albumin, 66 000; and cytochrome C, 12 400) was used to determine the relative molecular mass of the non-denatured, exported pullulanase. The values obtained were approximately 120 kDa for the ammonium sulphate extract, and approximately 135 kDa for the ion exchange fractions. The calibration curve is shown in figure 3.7.

![Figure 3.7 Molecular mass calibration curve for the Sepharose 6B column](image)

Standards used were Thyroglobulin, 669 000; Apoferritin, 443 000; Alcohol Dehydrogenase, 150 000; Albumin, 66 000; and Cytochrome C, 12 400.

The Pullulanase expressed from *Caldocellum saccharolyticum*
Figure 3.8 A Sepharose 6B gel filtration chromatogram

Ammonium sulphate extract

Total volume (ml)

Combined pullulanase active fraction from DEAE-Sepharose column

Total volume (ml)
The difference in the relative molecular mass values between the samples was likely to be due to either the higher protein concentration in the ammonium sulphate extract retarding the pullulanase (and hence reducing the apparent molecular mass) or experimental error associated with such a large molecular mass separation range, gel filtration media. The apparent molecular mass of non-denatured protein (as determined by gel filtration) was approximately the same as the denatured molecular mass (as determined by SDS-PAGE), thus implying that the enzyme was not exported in a multimeric form or as a high molecular mass multi-enzyme complex. A typical chromatogram is shown in figure 3.8 together with the location of pullulanase active fractions.

As the enzyme did not appear to be exported as a large molecular mass complex, it was more appropriate to use a different gel filtration medium that had a molecular mass fractionation range closer to the molecular mass of the proteins present. The medium used was Sephacryl S-200, an allyl dextran covalently cross-linked to acrylamide, with a molecular mass fractionation range of 5 - 250 kDa (see section 2.5.4.3).

**Figure 3.9 Molecular mass calibration curve for the Sephacryl S-200 column**

![Molecular mass calibration curve](image)

Standards used were β-amylase, 200 000; Alcohol dehydrogenase, 150 000; Albumin, 66 000; Carbonic anhydrase, 29 000; and Cytochrome C, 12 400.

The combined pullulanase fractions from the DEAE Sephacel ion exchange column (section 3.3.2) were applied to the column. Pullulanase activity was determined using the PAHBAH reducing

The Pullulanase expressed from *Caldosellum saccharolyticum*
sugar assay (see section 2.6.1.2) for every fraction. A calibration curve obtained with molecular mass markers (β-amylase, 200 000; alcohol dehydrogenase, 150 000; albumin, 66 000; carbonic anhydrase, 29 000; and cytochrome C, 12 400) was used to determine the relative molecular mass of the non-denatured, exported pullulanase. The value obtained was approximately 120 kDa. The calibration curve, together with the location of the pullulanase activity fraction, is shown in figure 3.9.

Again, the apparent molecular mass obtained for gel filtration was similar to that obtained by SDS-PAGE and the previous values obtained with the Sepharose 6B gel filtration column. A typical chromatogram is shown in figure 3.10, together with the location of pullulanase active fractions.

**Figure 3.10 A Sephacryl S-200 gel filtration chromatogram**

![Chromatogram](image)

### 3.3.4 Affinity binding

Affinity chromatography was used to further purify the pullulanase from *C. saccharolyticum*. As no commercial ligand was readily available, the principle of pullulanase enzyme binding to various substrates was tested. These substrates included amylose, amylopectin, pullulan, cross-linked amylose, cross-linked amylopectin and cross-linked pullulan. The procedure is detailed in section 2.5.4.4.
Affinity binding was found to be a very effective method on a small scale for binding very small amounts of protein. Attempts to utilise affinity binding on a large scale, by using columns, were not successful due to very low flow rates obtained with even a small amount of binding ligand (e.g. 100 mg of amylose powder in Pharmacia K9/15 column gave a maximum flow rate of only 0.5 ml/h). Pasteur pipettes (50 μl of adsorbent) with step-gradient elution (using gravity flow) were found to be an efficient small scale procedure.

Affinity binding was used with two different samples, the original ammonium sulphate extract and the combined pullulanase active fractions from the Sephacryl S-200 gel filtration column (section 3.3.3). Pullulanase activity was determined using the PAHBAH reducing sugar assay for every fraction (see section 2.6.1.2). Protein concentration was determined as in method 2.5.5. In the case of the ammonium sulphate extract, the pullulanase was purified 20 fold with high recovery (approximately 70%) but only on a very small scale (20 μg sample). A typical elution profile is shown in figure 3.11 together with the location of pullulanase active fractions.

Figure 3.11 A chromatogram of ammonium sulphate extract affinity binding an elution for amylose powder

The Pullulanase expressed from \textit{Caldocellum saccharolyticum}
While this was an effective method, it yielded only low levels of enzyme and was very time consuming, since it could not be automated. Hence affinity binding did not replace the previous two-step purification method of ion exchange (see Section 3.3.2) followed by gel filtration (see Section 3.3.3).

However, unlike the purification procedures previously used, which relied on the gross properties of the enzyme, affinity binding involves the active site of the enzyme. Hence it was used to further purify the semi-purified Sephacryl S-200 gel filtration fraction. The level of purification obtained was sufficient for N-terminal sequencing to be performed and enzymatic characterisation to be determined.

This sample still hydrolysed α-1,4 glucosidic bonds, possibly due to the presence of contaminating enzyme. SDS-denatured protein samples were run on Prosieve™ agarose gels and the protein band corresponding to the pullulanase band was extracted and tested for activity (see Section 2.6.2.7). This method separates proteins on the basis of their denatured molecular mass, allowing the removal of contaminating proteins with differing molecular mass, however the sample still hydrolysed α-1,4 glucosidic bonds following this procedure (see section 3.4.4).
3.4 Enzyme Characterisation

Using the purification method listed in section 3.3, it was possible to obtain sufficient pure pullulanase to perform enzyme characterisations. The enzyme extract used for all characterisations, except for the final substrate characterisation, was the purified amylose-binding fraction (see section 3.3). For the final substrate characterisation, the purest enzyme extract was used, i.e. the affinity binding purified sample, that was further purified by running and isolating from Prosieve agarose (see section 3.3.4).

The characterisations performed are outlined in sections 3.4.1-3.4.4.

3.4.1 Glycosylation and activity studies

Periodic acid-Schiff's reagent staining was used to determine whether the pullulanase from *C. saccharolyticum* was glycosylated (see section 2.6.2.5), the results are shown in figure 3.12.

**Figure 3.12 Periodic acid-Schiff's reagent staining of the *C. saccharolyticum* pullulanase**

<table>
<thead>
<tr>
<th>Lane</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Glycosylated proteins used as positive controls</td>
</tr>
<tr>
<td>2</td>
<td><em>C. saccharolyticum</em> pullulanase</td>
</tr>
<tr>
<td>3</td>
<td>Pharmacia high molecular weight markers (size in kDa)</td>
</tr>
</tbody>
</table>

Identical samples were loaded onto each side and electrophoresed, one half was stained with coomassie brilliant blue (A) and the other with Periodic acid-Schiff's reagent (B) (SDS-PAGE 7.5%).

The Pullulanase expressed from *Caldocellum saccharolyticum*
The pullulanase stained magenta with periodic acid-Schiff's reagent, implying that it was glycosylated. Although this method is not quantitative, the degree of staining was lower than for the highly glycosylated proteins used as a positive control (40-50% glycosylation, personal communication W. Judd, Department of Cell Biology, Auckland University). Attempts to remove the sugar residues from the pullulanase by digestion with endoglycosidase F and H (Boehringer Mannheim) were unsuccessful.

Attempts to determine the amino-terminal sequence of the enzyme yielded no sequence, implying that the terminal amino acid was blocked (personal communication D. Christie, Department of Biochemistry, Auckland University).

Activity gels were performed with the purified enzyme extract (see section 2.6.2.6) and examples are shown in figure 3.13.

**Figure 3.13 Activity gels showing the location of the pullulanase**

![](image)

<table>
<thead>
<tr>
<th>Lane</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C. saccharolyticum pullulanase</td>
</tr>
<tr>
<td>2</td>
<td>Pharmacia low molecular weight standards (sizes are in kDa)</td>
</tr>
</tbody>
</table>

Identical samples were loaded onto each side of a 12% SDS-PAGE gel and electrophoresed. One half was overlaid with Pullulan Red (A) and the other stained with coomassie brilliant blue.

These gels show that the major band present was exclusively responsible for the pullulanase activity, implying that this protein was the pullulanase.
3.4.2 The effect of metal ions and inhibitors on pullulanase activity

3.4.2.1 Initial studies

The effects of a number of different metal ions on the activity of the pullulanase were determined using two different samples, the purified enzyme (see section 3.3) and the same sample dialysed (16h 4°C) with 10 mM EDTA to complex out any metal ions present. As expected, the dialysed sample had very low activity, implying that the removal of metal ions reduced enzyme activity. Specific metal ions were then added to each sample and percentage activity measured, relative to the purified enzyme.

Table 3.5 Effect of different metal ions on pullulanase activity

<table>
<thead>
<tr>
<th>Metal ion (1 mM) added</th>
<th>non-dialysed</th>
<th>EDTA dialysed</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>100</td>
<td>40</td>
</tr>
<tr>
<td>Ca^{2+}</td>
<td>410</td>
<td>390</td>
</tr>
<tr>
<td>Zn^{2+}</td>
<td>245</td>
<td>96</td>
</tr>
<tr>
<td>Ni^{2+}</td>
<td>284</td>
<td>126</td>
</tr>
<tr>
<td>Cu^{2+}</td>
<td>205</td>
<td>80</td>
</tr>
<tr>
<td>Fe^{2+}</td>
<td>194</td>
<td>81</td>
</tr>
<tr>
<td>Mg^{2+}</td>
<td>209</td>
<td>78</td>
</tr>
<tr>
<td>Mn^{2+}</td>
<td>207</td>
<td>79</td>
</tr>
<tr>
<td>Co^{2+}</td>
<td>296</td>
<td>75</td>
</tr>
<tr>
<td>Fe^{3+}</td>
<td>155</td>
<td>48</td>
</tr>
<tr>
<td>Ag^{+}</td>
<td>107</td>
<td>42</td>
</tr>
<tr>
<td>Na^{+}</td>
<td>96</td>
<td>35</td>
</tr>
<tr>
<td>K^{+}</td>
<td>92</td>
<td>46</td>
</tr>
<tr>
<td>Li^{+}</td>
<td>85</td>
<td>43</td>
</tr>
</tbody>
</table>

The values represent the average of 3 measurements, the variation between measurements was approximately 10%. Activity of the undialysed purified enzyme without the addition of metal ions was defined as 100% (0.001 U pullulanase). Assay procedure used: 0.001 U pullulanase (2 μl), 200 μg pullulan (8 μl), metal ion 1 mM (2 μl) and 88μl of enzyme buffer (Mes/NaOH 0.2 M, pH 5.70°C). Reducing sugar released was assayed by the PAHBAH reducing sugar assay (see section 2.6.1.2).

The addition of monovalent cations had little or no effect on enzyme activity in either sample. For the divalent cations the addition of metal ions to the dialysed sample increased enzyme activity significantly, and in the case of Ca^{2+} this reversal was almost complete with activity in the dialysed

The Pullulanase expressed from *Caldocellum saccharolyticum*
increased enzyme activity, and as some residual Ca\(^{2+}\)/EDTA complex was still present in the enzyme sample the addition of other cations released some of the Ca\(^{2+}\), allowing them to bind to the metal ion binding site.

The reduced activity on addition of other divalent cations to the non-dialysed samples was possibly due to the replacement of the Ca\(^{2+}\) at the metal ion binding site with other divalent cations, forming an active enzyme but with lower activity than addition of Ca\(^{2+}\) only.

### 3.4.2.2 Calcium ion studies

In common with other pullulanases (see section 1.4.2), the Ca\(^{2+}\) had the greatest effect on pullulanase activity. For this reason further experiments were performed at different concentrations of Ca\(^{2+}\) and EDTA. Their interaction is shown in figure 3.14.

---

**Figure 3.14** The interaction of varying concentrations of Ca\(^{2+}\) and EDTA on pullulanase activity

Undialysed purified enzyme (0.001 U pullulanase) was used. To test the effect of Ca\(^{2+}\), varying concentrations were added and assayed as in table 3.5. To test the effect of EDTA the pullulanase was dialysed in varying concentrations of EDTA (16h, 4\(^{\circ}\)C) and assayed as in table 3.5.
The results show that enzyme activity increased proportionally with increased concentration of Ca\(^{2+}\) (until values of 1 mM) and decreased proportionally with increased concentration of EDTA. A similar result was obtained for the unpurified enzyme extract and hence extra Ca\(^{2+}\) (1 mM) was added in all enzyme assays.

### 3.4.2.3 Monovalent cations

Monovalent cations in low concentrations (1 - 10 mM) were shown to have no effect on pullulanase activity. However, as high concentrations of NaCl were used during the purification procedure (up to 1500 mM) and some samples required further concentrating before assaying, the effect of high concentrations of NaCl on pullulanase activity was determined (see figure 3.15).

![Figure 3.15: The effect of high concentrations of NaCl on pullulanase activity](image)

The values represent the average of 3 measurements, the standard deviation contained within the area of the symbol. 100% activity was defined as maximum activity, which was in fact slightly higher than when no NaCl was added.

While extremely high concentrations of NaCl (>1500 mM) decreased enzyme activity, increasing the NaCl concentration (up to 1000 mM) caused a small increase in enzyme activity. Since NaCl concentrations of up to 1500 mM were used in the purification procedure, and enzyme activity was not significantly lower at these levels, pullulanase activity could be assayed directly from samples eluted by salt gradient, without the need for dialysis.

The Pullulanase expressed from *Caldocellum saccharolyticum*
3.4.3 Thermostability and temperature and pH optima of the pullulanase

To determine the temperature and pH optima of the pullulanase, standard enzyme assays (see section 2.6.1.1) were performed using the PAHBAH reducing sugar assay (see section 2.6.1.2). The temperature optimum of the pullulanase was determined by measuring the enzyme activity at different temperatures. The pH optimum was determined in a similar manner, with the enzyme reaction buffer at different pH values. The results of both temperature and pH optimal studies are shown in figure 3.16.

Figure 3.16 Temperature and pH optima of the C. saccharolyticum pullulanase

The values represent the average of 3 measurements, the standard deviation contained within the area of the symbol. 100% activity was defined as maximum activity.

These results show that the maximum activity of the purified enzyme extract was at 70°C and pH 5.7. These values are the same as those obtained for the unpurified enzyme extract and hence correspond to the standard enzyme assay conditions used.
To determine the thermostability of the pullulanase, a known amount of enzyme was preincubated at different temperatures for different lengths of time, without substrate. The remaining activity in each sample was determined by standard assay procedure using the PAHBAH reducing sugar assay. The results are shown in figure 3.17.

**Figure 3.17 Thermostability of the pullulanase from C. saccharolyticum**

The values represent the average of 3 measurements, the standard deviation contained within the area of the symbol. 100% activity was defined as maximum activity.

Note: Due to the sensitivity of the assay procedure, thermostability was determined using low ionic strength buffer at very low protein concentrations without the presence of substrate.

The enzyme appeared to be thermostable at temperatures up to 75°C, but was irreversibly denatured at higher temperatures. Longer incubation times were performed at both 70 and 75°C and the enzyme half-life at these temperatures was estimated to be approximately 100 h and 24 h respectively. These results correlate well with the known growth temperature of the organism.

### 3.4.4 Substrate characterisation

For the final substrate characterisation the purest enzyme extract was used, i.e. the affinity

The Pullulanase expressed from *Caldocellum saccharolyticum*
binding purified extract, that was further purified by running and isolating from Prosieve agarose (see section 3.3.4).

The resulting activity was determined as the number of bonds hydrolysed (i.e. increase in the amount of reducing sugar) and was measured by the PAHBAH reducing sugar assay. The hydrolysis products formed were determined by paper chromatography (see section 2.6.3). For the starch-based substrates (amylose, amylopectin and β-limit dextrin) activity was also measured by the starch-iodine binding assay (see section 2.6.1.3). The results of these analyses are shown in figure 3.18.

**Figure 3.18 Enzymatic activity measured as amylopectin-iodine binding**

![Graph showing enzymatic activity measured as amylopectin-iodine binding for C. saccharolyticum and Klebsiella](image)

**Summary of changes in starch-iodine values before and after incubation**

<table>
<thead>
<tr>
<th>Enzyme (Pullulanase)</th>
<th>Iodine colour</th>
<th>λ max (nm)</th>
<th>Peak value</th>
<th>Blue value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. saccharolyticum</em></td>
<td>red → clear</td>
<td>550 → 550</td>
<td>0.47 → 0.11</td>
<td>0.25 → 0.06</td>
</tr>
<tr>
<td><em>Klebsiella</em></td>
<td>red → purple</td>
<td>550 → 575</td>
<td>0.47 → 0.73</td>
<td>0.25 → 0.41</td>
</tr>
</tbody>
</table>

When other starch-based substrates (e.g. amylose and β-limit dextrin) were used the results obtained were essentially the same.

**Table 3.6 Relative activity of the *C. saccharolyticum* pullulanase on different substrates.**

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Pullulan</th>
<th>β- limit dextrin</th>
<th>Amylopectin</th>
<th>Amylose (Rabbit liver)</th>
<th>Glycogen (Oyster)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td>100</td>
<td>34.7</td>
<td>41.6</td>
<td>4.7</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Activities are expressed as a percentage of activity on pullulan and were measured by PAHBAH reducing sugar assay.
Enzymes that hydrolyse pullulan are classified according to their final hydrolysis end products (see section 1.4 and figure 1.4.2). As the *C. saccharolyticum* pullulanase only produced maltotriose as its final end product, it is classified as a pullulanase rather than an isopullulanase, neopullulanase or glucoamylase. Since the pullulanase also showed α-1,4 activity when β-limit dextrin, amylopectin and amylose were used as substrates, it is further classified as a pullulanase with α-1,4 activity (see section 1.4.2).

To reduce this α-1,4 activity, three α-amylase inhibitor proteins (Type I, II, III; Sigma) from wheat were added to various pullulanase samples. These proteins bind reversibly to some α-amylases, reducing their activity (O'Donnell and McGeeney 1976). Even though all the inhibitors were used in massive excess no detected change in activity or starch-iodine binding was detected.

During the purification procedure a number of initial tests of substrate characterisation were performed using semi-purified extracts. The results obtained were essentially identical to those shown above.

The Pullulanase expressed from *Caldocellum saccharolyticum*
3.5 Summary

Substrate induction studies showed that expression of the pullulanase from *C. saccharolyticum* was inducible with the addition of maltose (as well as pullulan and amylopectin), and that expression was repressed in the presence of glucose.

The enzyme distribution was typical for an extracellular enzyme, with most exported to the culture media (approximately 50%). The remainder was located predominantly in the soluble cytoplasmic fraction (approximately 30%) with some exported to the periplasmic space (approximately 5%) or outer membrane/cell wall associated fraction (approximately 10%) presumably in the process of being exported out of the cell. Southern hybridisations showed that all *C. saccharolyticum* cells and DNA used were from the same organism.

The enzyme was purified from the culture using a protocol that involved ammonium sulphate precipitation, DEAE ion exchange chromatography, gel filtration chromatography and amylose powder affinity binding chromatography. The non-denatured molecular mass of the pullulanase was determined to be 120-135 kDa by gel filtration, and the denatured molecular mass determined to be 130 kDa by SDS-PAGE.

The enzyme characterisation (section 3.4) of the pullulanase expressed by *C. saccharolyticum* showed that the enzyme was glycosylated and its N-terminal amino acid residue was blocked, preventing N-terminal amino acid sequencing. Enzyme activity was shown to increase in the presence of Ca$^{2+}$ and decrease in the presence of EDTA. The temperature optimum was determined to be 70°C and at this temperature the pH optimum was determined to be 5.7. Enzyme stability, as measured by half life, was 100 h at 70°C, 87 min at 80°C and 9 min at 90°C.

Substrate characterisation showed that the pullulanase from *C. saccharolyticum* had high activity on pullulan, producing only maltooltriose as a final end product. The enzyme also produced a mixture of smaller oligomaltodextrins from high activity on β-limit dextrin, amylopectin and amylose.
Chapter Four

DNA Sequence of the Pullulanase Gene from *Caldocellum saccharolyticum*

4.1 Introduction

The information presented in this chapter covers the sequencing and analysis of the genetic structure of the pullulanase gene from *Caldocellum saccharolyticum*. A genomic fragment (genomic λ phage, see section 1.6.1) that was phenotypically pullulanase positive had been isolated previously. Genetic manipulations were then performed in *E. coli*. As the recombinants were also phenotypically pullulanase positive, the approximate location of the pullulanase gene was able to be determined by enzyme assay. Plasmids used in this chapter are shown diagrammatically in Figure 4.1.

4.2 Restriction endonuclease map of pNZ1452 and pNZ1038

Before sequencing commenced a restriction endonuclease map was constructed from pNZ1038 and the smallest pullulanase positive pCGN recombinant (pNZ1452). For completeness, digests were performed (as in method 2.1.3.1) using every unique restriction endonuclease available in the department (58 in total). Table 4.1 summarises the results of the digestions of each recombinant.

The positions of potentially "useful" restriction endonuclease sites (defined as sites that can be used for directional cloning) were located using a combination of multi-enzyme digestion techniques and digestion of isolated internal fragments. Their relative positions are shown in figure 4.2.
DNA Sequence of the Pullulanase Gene puA
Figure 4.2 Restriction enzyme map of pNZ1452 and pNZ1038

The noticeable lack of "useful" restriction endonuclease sites correlates with two facts. The nucleotide base percentage composition of the recognition sites of most commonly used restriction enzymes is normally more than 50% guanine and cytosine (high G/C ratio), and the genomic G/C ratio of *C. saccharolyticum* DNA was only 34.5%, as determined by thermal denaturation (which correlates to the figure of 38.41% determined for analysis of the DNA sequence of other genes from *C. saccharolyticum*, see Appendix A4).

An example of this rarity of restriction endonuclease sites is shown in the comparison (table 4.1) of the first 375 bp of pBR322 (1st column) with the 1830 bp of *C. saccharolyticum* from pNZ1072 (4th column). In the first 375 bp of pBR322 there were 16 different restriction endonuclease enzymes with 1 or more sites while in the 1.83 kb of *C. saccharolyticum* DNA from pNZ1072 there were only 12 such enzyme sites, giving a ratio of 16/375, or 1 per 23 bp, for pBR322 but only 12/1830, or 1 per 152 bp, for *C. saccharolyticum*.

On completion of the sequence of pNZ1038 it became possible to generate a restriction endonuclease site map from the DNA sequence (see Appendix A2). This map was compared to that generated from the physical restriction endonuclease site map with the only noticeable improvements being an increase in the number of sites of some frequently cutting enzymes (e.g. *Alu* I sites increased from 5+ to 21 sites) and the exact location of all restriction endonuclease sites.
Table 4.1 Summary of restriction endonuclease sites detected in pullulanase recombinants

<table>
<thead>
<tr>
<th>Restriction Enzyme</th>
<th>Recognition site G/C ratio</th>
<th>Number of sites pBR322 bp1-375</th>
<th><em>Caldicellum</em> saccharolyticum</th>
<th>DNA from</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aau I</td>
<td>67</td>
<td>0</td>
<td>5+</td>
<td>5+</td>
</tr>
<tr>
<td>Alu I</td>
<td>50</td>
<td>0</td>
<td>5+</td>
<td>5+</td>
</tr>
<tr>
<td>Alw NI</td>
<td>67</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Apa I</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Aus I</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ava I</td>
<td>80</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Bsl I</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bam HI</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bcl I</td>
<td>33</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Bgl I</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bgl II</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bsp 1286</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bst III</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cla I</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dra I</td>
<td>0</td>
<td>0</td>
<td>5+</td>
<td>5+</td>
</tr>
<tr>
<td>Eco RI</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Eco RII</td>
<td>80</td>
<td>1</td>
<td>5+</td>
<td>5+</td>
</tr>
<tr>
<td>Eco RV</td>
<td>33</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hae II</td>
<td>80</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Hae III</td>
<td>100</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Hae I</td>
<td>80</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Hef Al</td>
<td>67</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hha I</td>
<td>100</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Hind III</td>
<td>33</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Hpa I</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Kpn I</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mlu I</td>
<td>67</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Msp I</td>
<td>100</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Nae I</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nar I</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nci I</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nco I</td>
<td>67</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Nde I</td>
<td>33</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Nhe I</td>
<td>67</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Not I</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nru I</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pst I</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pvu II</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rsa I</td>
<td>50</td>
<td>1</td>
<td>5+</td>
<td>5+</td>
</tr>
<tr>
<td>Sac II</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sal I</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sau 3AI</td>
<td>50</td>
<td>1</td>
<td>5+</td>
<td>3</td>
</tr>
<tr>
<td>Sau 96I</td>
<td>90</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Sca I</td>
<td>33</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Sfi I</td>
<td>90</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sma I</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sna BI</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Spe I</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sph I</td>
<td>67</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Sst I</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Suv I</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Taq I</td>
<td>50</td>
<td>2</td>
<td>5+</td>
<td>5+</td>
</tr>
<tr>
<td>Xba I</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Xho I</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Xma III</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Xma IV</td>
<td>33</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: Restriction endonuclease map of pBR322 (1-375) was determined by analysis of the published sequence. The restriction endonuclease listed for the other plasmids fragments is based on the number of fragments determined by restriction endonuclease digestion minus the number in pBR322 fragment as listed in the table above.

DNA Sequence of the Pullulanase Gene *pullA*
4.3 Sequencing of pNZ1452

4.3.1 Sequencing of previously constructed recombinants

The initial sequencing strategy was based on plasmid double-stranded dideoxy-sequencing (see 2.4.3.1), of nested exonuclease deletions ligated into pCGN565/566 prepared by previous workers (see section 1.6). Recombinant plasmids of pNZ(1452-1470) were sequenced, but all except pNZ1452 were found to contain fragments of plasmid pBR322 DNA rather than C. saccharolyticum DNA. These results and subsequent computer analysis are shown in Appendix 1.

4.3.2 Pseudo random library

Figure 4.3 Construction of pseudo-random libraries from the Eco RI / Sal I fragment of pNZ1452

![Diagram showing construction of pseudo-random libraries from the Eco RI / Sal I fragment of pNZ1452]

Alu Library

Hae Library

Rsa Library
Due to the problem outlined above, a new strategy was devised based on sequencing the recombinant plasmid pNZ1452. This plasmid was chosen as it was the smallest recombinant that expressed pullulanase activity. The initial 301 bp of sequence obtained from plasmid sequencing showed no homology to pBR322 and the G/C nucleotide percentage ratio of 36% was similar to the C. saccharolyticum genomic ratio; thus implying that the pNZ1452 contained the C. saccharolyticum pullulanase gene.

Pseudo-random libraries were constructed from the 3974bp fragment (Eco RI and Sal I) of pNZ1452 by digestion with 4 base pair recognition site restriction endonucleases (see Section 2.4.5). Three separate libraries were constructed using the enzymes Alu I, Hae III and Rsa I, and the resulting fragments were ligated into the Sma I site of M13mp10 for subsequent sequencing. Their construction is shown diagrammatically in figure 4.3.

From each of these pseudo-random libraries 48 recombinant phages (as determined by lacZ induction with BCIG/IPTG plates) were selected and single-stranded DNA was extracted. The resulting single-stranded DNA templates were T-tracked to determine their uniqueness, insert size and quality of sequence. The results are shown in table 4.2.

Table 4.2 Results of T-tracking recombinant phages from pseudo-random libraries

<table>
<thead>
<tr>
<th>Library</th>
<th>sequenced</th>
<th>producing readable sequence</th>
<th>with no inserts</th>
<th>with non-unique sequences</th>
<th>to be sequenced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alu</td>
<td>44</td>
<td>30</td>
<td>3</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>Hae</td>
<td>48</td>
<td>26</td>
<td>2</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>Rsa</td>
<td>44</td>
<td>26</td>
<td>4</td>
<td>2</td>
<td>20</td>
</tr>
</tbody>
</table>

Those recombinants that fulfilled the conditions outlined above, and T-tracked “acceptably” were resequenced by standard single-stranded dideoxy-sequencing. The sequences obtained were aligned to themselves and to the restriction enzyme map by using both the GCG (Genetics Computer Group, Devereux et al. 1984) and Staden (SHOTGUN/VTCOMP, Staden 1980) computer sequence analysis packages (see section 2.7).

DNA Sequence of the Pullulanase Gene purA
4.3.3 Random shear library

Random libraries were constructed by sonication of the Eco RI / Sal I fragment of pNZ1452 to produce randomly sheared fragments that were ligated into the Sma I site of M13mp10 for subsequent sequencing. Their construction is shown diagrammatically in figure 4.4.

Figure 4.4 Construction of a random shear library from the Eco RI / Sal I fragment of pNZ1452

This was repeated giving two random libraries, from which recombinant phages (as determined by lac Z induction with BCIG/IPTG plates) were selected and single-stranded DNA extracted. The resulting single-stranded DNA's were T-tracked to determine their uniqueness, insert size and quality of sequence. Those recombinants that fulfilled the conditions outlined above were sequenced by standard single-stranded dideoxy-sequencing sequencing. The results are shown in table 4.3.
Table 4.3 Results of T-tracking recombinant phages from random libraries

<table>
<thead>
<tr>
<th>Library</th>
<th>sequenced</th>
<th>producing readable sequence</th>
<th>with no inserts</th>
<th>with non-unique sequences</th>
<th>to be sequenced</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>84</td>
<td>80</td>
<td>15</td>
<td>-</td>
<td>65</td>
</tr>
<tr>
<td>second</td>
<td></td>
<td>T-TRACKING NOT PERFORMED</td>
<td></td>
<td></td>
<td>40</td>
</tr>
</tbody>
</table>

The sequenced fragments were aligned to a previously constructed restriction enzyme map, and to themselves, with the aid of both the GCG and the Staden computer sequence analysis packages.

A summary of the results obtained for sequencing the three pseudo-random and random sheared libraries by standard single-stranded dideoxy-sequencing is given in table 4.4. Overall, these libraries produced an estimated 86% of the sequence in one direction and 77% in the other with 2 gaps of unsequenced DNA of around 50 bp each.

Table 4.4 Summary of the results obtained from sequencing the three pseudo-random and two random sheared libraries

<table>
<thead>
<tr>
<th>library</th>
<th>sequenced</th>
<th>producing readable sequence</th>
<th>unknown</th>
<th>containing only M13§</th>
<th>containing only pBR322§</th>
<th>ENTERED INTO DATA BASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALU</td>
<td>22</td>
<td>17</td>
<td>-</td>
<td>2</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>HAE</td>
<td>17</td>
<td>15</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>RSA</td>
<td>20</td>
<td>17</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>1st Random</td>
<td>65</td>
<td>58</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>46</td>
</tr>
<tr>
<td>2nd Random</td>
<td>30</td>
<td>18</td>
<td>5</td>
<td>2</td>
<td>-</td>
<td>11</td>
</tr>
</tbody>
</table>

§ These recombinants were shown to be fragments of either M13 or pBR322 vector sequences rather than C. saccharolyticum DNA and were therefore not included in the sequence.
Figure 4.11 Amino acid homology lineup of amino acids upstream of ATG

Identical amino acids with respect to the *C. saccharolyticum* pullulanase are shown in black reverse font, where as similar amino acids are in grey background font

Cspula *Caldocellum saccharolyticum* pullulanase
Klpula Klebsiella sp. (both *pneumonia* and *aerogenes*) pullulanase
Bstpula *Bacillus stearothermophilus* pullulanase

### 4.4.3 Sequencing the remainder of pNZ1038

As stated above, it appeared that the plasmid pNZ1452 did not contain the total pullulanase gene even though phenotypically it was pullulanase-positive. This missing portion of sequence was obtained by either oligonucleotide primer sequencing or directional cloning of the remaining 1180 bp of pNZ1038. The location and direction of each primer and cloned fragment are shown diagrammatically in figure 4.12. To further extend ORF3, pNZ1037 was sequenced using oligonucleotide primer PRIM15 yielding a further 188 bp of DNA sequence beyond the *Sph I* site of pNZ1038. The entire sequence of pNZ1038 together with the translation of the ORF's is shown in Appendix 2.

Shown in figure 4.12A is the complete sequence of the pullulanase gene (*pulA*). Together with the deduced amino acid sequence of the gene product.
statistic shown and is based on the concept of synonymous codons (i.e. amino acids are coded by different codons at different frequencies). Analysis of all three statistics allows the differentiation of gene coding sequences from non-coding ORF's. In the case of pNZ1452 these statistics show that the sequence upstream of the putative start codon (shown in orange) appears to be part of a coding sequence i.e. a lack of rare codons, a consistently high codon bias and third position AT bias as found in the coding sequence of other genes from *C. saccharolyticum*.

**Figure 4.10 Codon preference of pNZ1452**

Further evidence that the sequence upstream of the putative ATG appeared to be coding sequence was obtained by homology comparisons (at the amino acid level) with other pullulanase sequences. This information showed that a translation of the DNA sequence upstream of the putative start codon was homologous to a conserved region found in other pullulanase genes. The lineup of the sequence is shown in figure 4.11.
4.4.2 Incomplete gene on pNZ1452

Analysis of the complete sequence of pNZ1452 (the smallest Pul+ clone) showed that the second ORF (pullulanase gene) did not have a stop codon upstream of its ATG, assigned as its putative start codon, implying that the N-terminal portion of the gene might be missing from pNZ1452. This result is shown graphically in figure 4.9, produced again by the program FRAMES but showing ORF’s only.

Figure 4.9 Potential ORF’s from pNZ1452

Analysis of the sequence upstream of the putative ATG using codon usage patterns further implied that the N-terminal portion of the gene was missing. Figure 4.10 (produced by CODONPREFERENCE) shows the codon usage of the three reading frames of pNZ1452 in the reverse direction. Three codon usage/protein coding sequence statistics are shown. These statistics are based on the assumption that reading frames that code for proteins use codons at the same frequency throughout the entire genome. The first of these is the frequency of usage of rare Caldocellum saccharolyticum codons (data determined by summation of previous sequences as shown in appendix A4). Within the coding region, some codons are used at a significantly lower frequency than expected by random chance and are referred to as rare codons. The absence of these rare codons can be used as an indication of ORF’s. The second statistic shows the third position nucleotide bias of codons in each reading frame (AT bias shown). The codon preference statistic (Gribskov et al. 1984) is the third...
Figure 4.7 Deduced translational start

**ORF1** Pep X

```
TTGAAATTTTCAATATTATGGGCTATAATAGAATAAACAGAAATTAAAAAGAGTTAATAATCCAGCAATG
-35  -10
```

**ORF2** PulA (pNZ1452)

```
TTTAGACTCTGGGCTCACAATCTGCTTATAAAAGATTCAGATTTGATGAGCATGAAAATTTCAATTTTTCAAGAAATG
-35  -10
```

Database searches of the GCG DNA database (see section 2.7) were performed using the programs FASTA and TFASTA (see appendix 2). No significant homology was detected for ORF1, while ORF2 showed slight homology to *Klebsiella* pullulanase. Further database searches were performed using the translation (deduced amino acid peptide sequence) of both of these ORF's. Again, ORF1 showed no significant homology at the peptide level, whereas ORF2 showed significant homology to *Klebsiella* pullulanase at this level. ORF2 was assigned as the pullulanase gene by combining this information with that already known about the location of the pullulanase gene (i.e. exonuclease deletions). This is shown in figure 4.8.

Figure 4.8 Diagrammatic summary of ORF's detected on sequencing pNZ1452

**pNZ1452**

A C. *saccharolyticum* DNA segment containing two ORFs is shown. ORF1 is 1131 bp long and ORF2 is 1735 bp long. ORF1 encodes a 44 kDa peptide and ORF2 encodes a 65 kDa peptide. Homologous regions to *Klebsiella* are indicated, as well as the position of key enzyme regions.
4.4 Analysis of the Sequence of pNZ1452

4.4.1 Determination of open reading frames (ORF)

On completion of the sequence of pNZ1452, the DNA sequence was searched for ORF’s using the GCG program FRAMES (Section 2.7). The output from this program is shown below. It plots all the start codons and stop codons present in all six potential reading frames. The program’s data file has been modified to show three potential start codons (ATG, TTG, GTG) as marks above the line representing the reading frame. The stop codons (TAA, TAG, TGA) are represented as marks below the line representing the reading frame (figure 4.6). Potential ORF’s are shown as regions of the line representing the reading frame devoid of stop codons.

Figure 4.6 Potential ORF’s from pNZ1452 showing all start and stop codons

Visual inspection of the figure above showed two ORF’s which were potential gene coding regions (highlighted in orange), each of which possessed a number of potential translational start sites. By searching upstream of each potential start codon for sequences indicative of transcription and translation, the most likely start codon was determined. These are shown in figure 4.7.
At this stage it was decided that no further reasonable sequence fragment overlaps were likely to be found (also all “good” T-Tracking M13 recombinants had been sequenced). Consequently a new strategy based on oligonucleotide primers and direct cloning of internal fragments was used to finish the sequencing of the Eco RI-Sal I fragment of pNZ1452.

4.3.4 Recombinant plasmids

The remaining regions of the missing sequence were resolved by forced directional cloning of restriction enzyme fragments into pGEM7 (Promega Corporation) and the DNA from the resulting recombinants was sequenced as either single stranded phage or double stranded plasmid. Oligonucleotide primers were constructed where suitable restriction enzyme sites were not present. The position of these internal fragments and the location of the oligonucleotide primers are shown in figure 4.5.

Figure 4.5 Location and position of primers and restriction enzyme fragments used to complete sequence of pNZ1452.

<table>
<thead>
<tr>
<th>Clones</th>
<th>Location</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNZ1072 Hind III-Hind III</td>
<td>[Diagram showing position]</td>
<td></td>
</tr>
<tr>
<td>pNZ1501 Dra I-Mlu I</td>
<td>[Diagram showing position]</td>
<td></td>
</tr>
<tr>
<td>pNZ1502 Mlu I-Dra I</td>
<td>[Diagram showing position]</td>
<td></td>
</tr>
</tbody>
</table>

Oligonucleotide primers location and direction

1  2  3  4  5  6  7

The exact location and sequence of the oligonucleotide primers are shown in table 2.11. The sequence generated from this primer is shown in figure 4.13.
Figure 4.12A Nucleotide sequence of the prlA gene

ACAAAAAGGTATTAAGAATTTAAAAGGCAAGATCCTTTCACCTGCTTTATAGATTTGCTGTTTATGCCTTTAGGT

Chapter 4
Figure 4.13 Sequencing strategy of pNZ1038

pNZ1038

Key:  pseudo random libraries ; Alu = A, Hae = H, Rsa = R,
random sheared library = S
oligonucleotide primers = PRIM
cloned fragments named after their R.E. generation sites.

DRML1F = Dra I MuI fragment 1 forward primer
DRML1R = Dra I MuI fragment 1 reverse primer
DRML2F = Dra I MuI fragment 2 forward primer
DRML2R = Dra I MuI fragment 2 reverse primer
SCS1F = Sca I Sca I fragment forward primer
SCS1R = Sca I Sca I fragment reverse primer
HFRW = Hind III / Hind III fragment forward primer
HINDR = Hind III / Hind III fragment reverse primer
SFI1F = Sph I MuI fragment forward primer
SFI1R = Sph I MuI fragment reverse primer

DNA Sequence of the Pullulanase Gene pulA
Figure 4.12 Location and position of primers and restriction enzyme fragments used to complete sequence of pNZ1038

pNZ1038

Recombinant plasmids
pNZ1503 Mlu I-Sph I
pNZ1504 Mlu I-Pst I

Oligonucleotide primers
location and direction

1KB

The exact location and sequence of the oligonucleotide primers are shown in table 2.11. The sequence generated from this primer is shown in figure 4.13.

A summary of the overall sequencing strategy showing the direction, location and size of every sequencing fragment obtained from all the pseudorandom and random libraries plus the restriction enzyme fragments and oligonucleotide primers is shown in figure 4.13.
Figure 4.13 Sequencing strategy of pNZ1038

Key: pseudo random libraries; Alu = A, Hae = H, Rsa = R, random sheer library = S
oligonucleotide primers = PRIM
cloned fragments named after their R.E. generation sites.

DRML1F = Dra I Mlu I fragment 1 forward primer
DRML1R = Dra I Mlu I fragment 1 reverse primer
DRML2F = Dra I Mlu I fragment 2 forward primer
DRML2R = Dra I Mlu I fragment 2 reverse primer
SCS1 = Sca I Sca I fragment forward primer
SCSCR = Sca I Sca I fragment reverse primer
HFRW = Hind III / Hind III fragment forward primer
HINDR = Hind III / Hind III fragment reverse primer
SPIMLU = Spa I Mlu I fragment forward primer
PSTMLU = Pat I Mlu I fragment forward primer

DNA Sequence of the Pullulanase Gene pulA
4.5 Genetic Structural elements of pNZ1038

4.5.1 ORF 1, 2, 3 and 4 gene assignment

On completion of the sequence of the 4.9 kb *Caldocellum saccharolyticum* fragment in pNZ1038 the analysis performed on pNZ1452 was repeated. Analysis of the six potential reading frames was performed using the program FRAMES (see section 4.3.1) and the output is shown in figure 4.14. Three potential ORF's were detected, these are shown in orange.

![Potential ORF's of pNZ1038 with all start and stop codons shown](image)

The start codon of the pullulanase gene (*pUL*A) was determined as before (see section 4.4.1) and is shown in figure 4.15. Sequences that were homologous to standard prokaryotic translation and transcription initiation sequences are also shown (see appendix A4).
The sequence of the 4.9 kb *Caldocellum saccharolyticum* fragment of the plasmid pNZ1038 was analysed again for gene coding regions (as in section 4.4.2) using the program CODONPREFERENCE. Based on the results of this program the three major ORF's previously identified were shown to be gene coding regions (see figures 4.16 and 4.17).

Besides the two ORF's detected before (or, more correctly, one complete and one partial ORF) another partial ORF was detected at the N-terminal end of the pullulanase gene. There were also a number of smaller potential ORF's detected at the C-terminal end of the gene (see figure 4.18).

**Figure 4.15** Putative start codon of *pulA*

**PuLA (pNZ1038)**

```
TTAAATAATCAATATAAGATGAAATTAGATATATCTATTGTATATCTAAATCATAAGAATGATAGACCTGAGTTT
-35  -10  S.D  start
```

DNA Sequence of the Pullulanase Gene *pulA*
Database searches of the GCG DNA and protein database (see section 2.7) were performed using the programs FASTA and TFASTA. No significant homology was detected for ORF1 and ORF3, while ORF2 showed homology to Klebsiella pullulanase at the amino acid level. The results from these searches are shown in appendix A2.

Sequences homologous to the prokaryotic RNA polymerase Rho independent termination sites were detected downstream from each ORF (for a full discussion of terminators found and comparison to other Caldocellum saccharolyticum terminators see section 4.5.3 and appendix A4). Their locations are shown diagrammatically in Figure 4.18.

In the maltose operon of Klebsiella sp. and E. coli, transcription is activated by the binding of the MalT\textsubscript{\tau} protein to a consensus DNA region, called a maltose box (5'-GGGGA\textsubscript{\tau}GAGG-3'). This region is located 34-35 base pairs upstream of the transcriptional start point. All MalT\textsubscript{\tau} binding sites determined so far are 70-100% homologous to this consensus (see section 1.3). Searches upstream of each ORF detected in pNZ1038 for sequences homologous to the maltose box revealed three sites but none in the correct location, implying that these sequences are unlikely to be inducible in E. coli.
Figure 4.18 Diagrammatic summary of ORF's detected on sequencing pNZ1038

4.5.2 Repeat sequences

pNZ1038 was searched for the presence of repeat sequences using the program REPEAT (see section 2.7). The majority of repeats identified were either located within genes or were too short to be of any significance. The notable exception was a set of three repeats (15 bp in length) located upstream of the pulA gene. The repeats formed two groups, the most homologous located 290 bp upstream of the pulA start codon and spaced 22 bp apart. The second group had weaker homology to each other and to the first set and were located 60 bp upstream of the pulA start codon with a spacing of 110 bp. The apparent function of repeats within DNA sequences is in translational or transcriptional...
regulation mediated by the attachment of DNA binding proteins to the repeat element (see Figure 4.19).

Figure 4.19 Direct repeat sequences identified upstream of the start of the pulA

<table>
<thead>
<tr>
<th>repeat</th>
<th>YHCYTGCTTTTWHRADMWCWTDHH</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,659:</td>
<td>ACCTT cccctgagttttgaataaagttaa CCCAC mis=5</td>
</tr>
<tr>
<td>3,768:</td>
<td>AAAGT gtctttcttttaaaaaaaatcaat CTTTC mis=5</td>
</tr>
<tr>
<td>3,867:</td>
<td>GAGCT ttctcactttttcaaatcttat CCCAC mis=3</td>
</tr>
<tr>
<td>4,061:</td>
<td>CATCT ttattttttatatccctaaa GTTTT mis=5</td>
</tr>
</tbody>
</table>

repeat /Rev DDHAWWGWKHTYDWAAAAAGCARGDR

| 4,060: | TCATC ttattgatatttttaatccctaaa AGTTT mis=5 |
| 4,082: | ACCTT aaaaatttttattgaaaaaagaaaa GGGGA mis=5 |
| 4,417: | TAAAC ataaattctactgtaaaaagtgttgt ATCCT mis=5 |

The location and homology of three highly conserved direct repeats are numbered 1-3 on figures 4.18 and 4.19. These were located upstream of the pul A gene but before PepY. A consensus sequence was used to search for the presence of other repeats on either the forward or reverse strands. A number were found near the start of the pulA gene with varying degrees of mismatch homology, and these are also shown in figure 4.19. Although the degree of homology was low for these other repeats the space between them was regular and consistent with the three major repeats. It is possible that these repeats are the binding site of either a substrate inducer or repressor protein analogous to the maltose box sequence of Klebsiella sp. or the CRP-binding site.

Searching the GenEMBL DNA database revealed that this consensus repeat was not present until the degree of mismatch was two. At this level of mismatch, approximately 1000 entries possessed at least one copy of this repeat in either orientation and, of these, only eight had two or more repeats closely spaced. Examination of these sequences revealed that these repeats were located inside the coding regions. Although these repeats were similar to the consensus sequence that contained a number of variable positions, there was no direct homology between them and the pulA repeats, implying that they were unrelated.

4.5.3 Terminator sequences

The sequences in figure 4.20 have been identified as putative terminator sequences (factor independent) by the GCG program TERMINATOR (Brendel and Trifonov 1984, Brendel et al. 1986). This program is based on searching for sequences similar to a consensus E. coli factor independent terminator sequence (GC dyad symmetry stem loop followed by a run of 4-8 T's). In bold are those in
expected locations, including those located downstream of ORF’s and those located upstream of PepX. For comparison, the highest scoring sequences located within each ORF are shown as a guide to the specificity of this method, since a major problem with this algorithm is the weighting placed on long runs of T’s. These long runs of T’s, although rare in E. coli, are common in C. saccharolyticum, hence a number of strong putative terminator sequences were detected within genes of C. saccharolyticum. No information is available on whether these sites function as transcriptional terminators in C. saccharolyticum.

Figure 4.20 Putative terminators detected in pNZ1038

TERMINATOR search on: pNZ1038 : 1 to: 4986

Primary structure threshold: 3.50  Secondary structure threshold: 0

| US of ORF1 (PepX) | 259-> | ATTTTTAAACAACAAAAATTTTGGCAATTTTCAATTTTGAAATTTTCAATTATT | 4.10 | 54 |
| DS of ORF1 (PepX) | 342-> | GAGGTAATAAAATCTGCATGAAAACATTTGGTTTTTTATTTATGGCTGTTCA | 4.04 | 17 |
| DS of ORF1 (PepX) | 539-> | CCATGGGTAAAGTCTGAGTTTGAAATCTCTTCAICTTTTTTCATGGAAAAAT | 3.91 | 12 |
| DS of ORF3 (PepX) | 1586-> | TCAAAGAGTGCGGGAACACACATGCATCATTGCTTTTACATACGAGCTA | 3.79 | 32 |
| DS of ORF3 (PepX) | 2463-> | AGTAGATAATAGTGCAAGGATGAAATCTCACCAGATTGTTTCTCTGTTG | 4.43 | 57 |
| DS of ORF3 (PepX) | 3454-> | TTGTATTTTTAATTGGCCCAAAGCTTTTGTTGTAATTTTACCCAT | 3.87 | 52 |

TERMINATOR search on reverse of: pNZ1038 from: 1 to: 4986

Primary structure threshold: 3.50  Secondary structure threshold: 0

| DS of ORF3 (PepX) | <= 4272 | TGGCATCAGCAACAGGGCTTGGCAATTTTGTACTATTTGAGCTTGTGG | 3.57 | 8 |
| DS of ORF3 (PepX) | <= 4139 | GGGACTTTCTCAAAGGGCAAGTGGAATCTGCTCTTTTCTTTTTTACAA | 3.82 | 17 |
| DS of ORF3 (PepX) | <= 3504 | ATGCGCAAATTGAGACTGACCTGCTTGGGAAATTTTAAATAACAA | 4.29 | 52 |
| DS of PulA | <= 1487 | ATTTTCAACTCTTTTGGCCAAAATATCTCAGATTCTCATATAAT | 4.31 | 0 |
| DS of PulA | <= 1389 | CCAGAGAGTGAAATTACACTGAAATCTCCATCTGTTGAGCAAAG | 3.73 | 13 |
| DS of PulA | <= 1292 | CATTAGCTAAAAGCTTCAAAGATGATTTACTGAAATTTTAAATTGCGCA | 4.33 | 33 |
| DS of PulA | <= 1091 | TAATTCCAAATTCATGATTTATTATTAAAACCTCTTTCTCTCATACT | 4.25 | 87 |
| DS of PulA | <= 309 | TAATATGAAAATCTACACTAGAATTTTGCTAAAAATTGTTTAATTAAAT | 4.27 | 53 |

DNA Sequence of the Pullulanase Gene pulA
4.5.4 Leader sequence analysis

Proteins to be exported are made as larger precursors containing an amino-terminal signal peptide of 20-40 amino acid residues which determines the final cellular location of the protein and is cleaved during export. The signal peptide consists of three regions with varying amino acid sequence, the N-terminal region consisting of one or more basic amino acids (e.g. K, R, H), a middle region of mainly hydrophobic amino acids (e.g. A, V, L, I, P, F, W, M, and the C-terminal region of polar amino acids (e.g. G, S, T, Y, C, Q, N), where helix-breaking amino acid residues determine the cleavage site (Verner and Schatz 1988).

The leader sequence of PulA showed similarity to a typical prokaryotic leader sequence with one important exception, the presence of negatively charged amino acids in the middle section. Searches of the database showed two other enzymes with a similar pattern of negatively charged amino acids; a xylanase, XynA, from C. saccharolyticum (Lüthi et al. 1990a) and a ligninase, LigA, from Pseudomonas paucimobilis (Noda et al. 1990). These three leader sequences are shown in figure 4.21.

![Figure 4.21 Proposed leader sequence of PulA](chart)

The optimal position of leader sequence cleavage was determined using the consensus signal sequence (Watson 1984), the statistical model of von Heijne (1986) and the experimental results of Shen et al. (1991). Since one of these enzymes was also from C. saccharolyticum, it is possible that the presence of negatively charged amino acids could be part of a specialised export mechanism in C. saccharolyticum. As other enzymes from C. saccharolyticum do not show this pattern of charge
distribution within their leader sequence (e.g. CelA, CelB and ManA, see section 1.5.2) it would appear to be a non-universal mechanism.

4.5.5 Codon preference

While all amino acids (excluding methionine and tryptophan) are coded for by two or more codons the frequency distribution of codon usage shows a specific pattern. A number of theories have been proposed to explain this observation, referred to as codon preference. From a biotechnological point of view codon preference is important since the codon usage frequency varies between organisms. Transferring a gene from one organism to another with widely different codon usage could affect the efficiency of expression.

Table 4.5 Codon usage patterns of the pulA gene

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Codon</th>
<th>Number</th>
<th>/1000</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td>GGG</td>
<td>11.00</td>
<td>13.33</td>
<td>0.17</td>
</tr>
<tr>
<td>Gly</td>
<td>GGA</td>
<td>24.00</td>
<td>29.09</td>
<td>0.38</td>
</tr>
<tr>
<td>Gly</td>
<td>GGT</td>
<td>16.00</td>
<td>19.39</td>
<td>0.25</td>
</tr>
<tr>
<td>Gly</td>
<td>GGC</td>
<td>12.00</td>
<td>14.55</td>
<td>0.19</td>
</tr>
<tr>
<td>Glu</td>
<td>GAG</td>
<td>22.00</td>
<td>26.67</td>
<td>0.37</td>
</tr>
<tr>
<td>Glu</td>
<td>GAA</td>
<td>38.00</td>
<td>46.06</td>
<td>0.63</td>
</tr>
<tr>
<td>Asp</td>
<td>GAT</td>
<td>37.00</td>
<td>44.85</td>
<td>0.61</td>
</tr>
<tr>
<td>Asp</td>
<td>GAC</td>
<td>24.00</td>
<td>29.09</td>
<td>0.39</td>
</tr>
<tr>
<td>Val</td>
<td>GTG</td>
<td>14.00</td>
<td>16.97</td>
<td>0.27</td>
</tr>
<tr>
<td>Val</td>
<td>GTA</td>
<td>15.00</td>
<td>18.18</td>
<td>0.29</td>
</tr>
<tr>
<td>Val</td>
<td>GTT</td>
<td>21.00</td>
<td>25.45</td>
<td>0.40</td>
</tr>
<tr>
<td>Val</td>
<td>GTC</td>
<td>2.00</td>
<td>2.42</td>
<td>0.04</td>
</tr>
<tr>
<td>Ala</td>
<td>GCG</td>
<td>3.00</td>
<td>3.64</td>
<td>0.07</td>
</tr>
<tr>
<td>Ala</td>
<td>GCA</td>
<td>22.00</td>
<td>26.67</td>
<td>0.54</td>
</tr>
<tr>
<td>Ala</td>
<td>GCT</td>
<td>13.00</td>
<td>15.76</td>
<td>0.32</td>
</tr>
<tr>
<td>Ala</td>
<td>GCC</td>
<td>3.00</td>
<td>3.64</td>
<td>0.07</td>
</tr>
<tr>
<td>Arg</td>
<td>AGG</td>
<td>17.00</td>
<td>20.61</td>
<td>0.46</td>
</tr>
<tr>
<td>Arg</td>
<td>AGA</td>
<td>16.00</td>
<td>19.39</td>
<td>0.43</td>
</tr>
<tr>
<td>Ser</td>
<td>AGT</td>
<td>5.00</td>
<td>6.06</td>
<td>0.13</td>
</tr>
<tr>
<td>Ser</td>
<td>AGC</td>
<td>8.00</td>
<td>9.70</td>
<td>0.21</td>
</tr>
<tr>
<td>Lys</td>
<td>AAG</td>
<td>28.00</td>
<td>33.94</td>
<td>0.40</td>
</tr>
<tr>
<td>Lys</td>
<td>AAA</td>
<td>42.00</td>
<td>50.91</td>
<td>0.60</td>
</tr>
<tr>
<td>Asn</td>
<td>AAT</td>
<td>21.00</td>
<td>25.45</td>
<td>0.55</td>
</tr>
<tr>
<td>Asn</td>
<td>AAC</td>
<td>17.00</td>
<td>20.61</td>
<td>0.46</td>
</tr>
<tr>
<td>Met</td>
<td>ATG</td>
<td>15.00</td>
<td>18.18</td>
<td>1.00</td>
</tr>
<tr>
<td>Ile</td>
<td>ATA</td>
<td>21.00</td>
<td>25.45</td>
<td>0.34</td>
</tr>
<tr>
<td>Ile</td>
<td>ATT</td>
<td>32.00</td>
<td>38.79</td>
<td>0.52</td>
</tr>
<tr>
<td>Ile</td>
<td>ATC</td>
<td>9.00</td>
<td>10.91</td>
<td>0.15</td>
</tr>
<tr>
<td>Thr</td>
<td>ACG</td>
<td>2.00</td>
<td>2.42</td>
<td>0.06</td>
</tr>
<tr>
<td>Thr</td>
<td>ACA</td>
<td>21.00</td>
<td>25.45</td>
<td>0.60</td>
</tr>
<tr>
<td>Thr</td>
<td>ACT</td>
<td>7.00</td>
<td>8.48</td>
<td>0.20</td>
</tr>
<tr>
<td>Thr</td>
<td>ACC</td>
<td>5.00</td>
<td>6.06</td>
<td>0.14</td>
</tr>
</tbody>
</table>

DNA Sequence of the Pullulanase Gene pulA
The codon preference of the *C. saccharolyticum* pullulanase gene is listed in table 4.5. These values were compared to the codon preference of other *C. saccharolyticum* genes, as listed in appendix 4. This comparison is shown graphically in figure 4.22.

**Figure 4.22 Comparison of codon usage patterns**

For mathematical reasons the three stop codons (TGA, TAG, TAA) were assigned a value of 1.

As expected the codon usage pattern of the pullulanase gene was similar to that found in other *C. saccharolyticum* genes. In keeping with a high genomic A/T ratio the third codon position showed a marked A/T bias. Comparisons were also made to three other codon preference data sets; yeast genes and *E. coli* genes expressed at high and low frequencies (these data sets were obtained from the GCG package). These comparisons are shown graphically in figure 4.23.

The comparison between the codon usage of the pullulanase and highly expressed *E. coli* genes showed marked differences. Those codons that were common in the *E. coli* genes were rare in *pulA*, while those codons that were rare in the *E. coli* genes were common in *pulA*. Therefore when the *pulA*
gene was expressed in *E. coli* the tRNA population was non-optimal for high expression. In particular, four codons that were common in the *pulA* gene were rare in highly expressed *E. coli* genes. This could effect the overall level of expression in *E. coli* as the tRNA codon population would not match the optimal population for gene expression.

**Figure 4.23 Comparison of codon usage of the pullulanase with other data sets**

![Comparison graph of codon usage](image)

For mathematical reasons the three stop codons (TGA, TAG, TAA) were assigned a value of 1. Codons not present are also assigned a small non-zero value (0.0001).

Similar comparisons were performed using *E. coli* genes expressed at low frequencies. These are also shown in figure 4.23. In these cases the codon frequency usage was more similar than with the highly expressed *E. coli* genes. This could partially account for the low levels of pullulanase expression in *E. coli* as found in chapters 5 and 6.

As the expression levels of the *pulA* gene in *Saccharomyces cerevisiae* were low (see section 7.2), the codon frequency of the *pulA* gene was compared to the standard *S. cerevisiae* data set. While the frequency of codon usage was different there appeared to be no major discrepancy between the codon frequency of *S. cerevisiae* and the *pulA* gene. Hence it is unlikely that codon preference mismatch was an explanation for the low levels of enzyme expression in yeast.

**DNA Sequence of the Pullulanase Gene pulA**
4.6 Summary

The initial sequencing strategy was based on plasmid sequencing of exonuclease deletions ligated into pCGN565/6 recombinants that had been constructed previously (see section 1.6.2 and appendix 1). On sequencing, all but one of these deletions were determined to contain pBR322 vector rather than C. saccharolyticum DNA. The one recombinant shown to contain C. saccharolyticum DNA (see section 4.3.1) was also phenotypically pullulanase positive (pNZ1452, 3618 bp) and hence was thought to contain the pullulanase gene. The sequence of this plasmid was determined using a combination of M13 sequencing libraries, nested enzyme deletions and oligonucleotide primers (see section 4.3). Analysis of the DNA sequenced showed that the N-terminal portion of the pullulanase gene was not present (381 bp or 15.4%). The remainder of pNZ1038 (1180 bp) was sequenced to complete this gene (see section 4.4).

Analysis of the sequence of pNZ1038 identified two major Open Reading Frames (ORF), ORF1 (1131 bp, 44 142 kDa peptide) and ORF2 (2478 bp, 95 732 kDa peptide). A partial ORF (ORF3) was also identified (at least 607 bp, 30 682 kDa peptide) (see section 4.5.1). The pullulanase gene was identified as ORF2 based on sequence homology (at amino acid level) to other pullulanase sequences (see section 4.4). No significant homology to any sequence in the database, at either the DNA or protein level, could be detected for the other ORF’s.
141

Chapter Five
Expression of the Pullulanase in Escherichia coli

5.1 Summary

The pullulanase gene was transcribed from its own promoters by the *E. coli* transcription/translation system, in common with a number of other *Caldocellum saccharolyticum* genes whose promoters are recognised by *E. coli*. A number of plasmids were constructed that contained the pullulanase gene and these are shown in figure 5.1. The plasmid pNZ1452 is included since it was phenotypically pullulanase positive, although the pullulanase gene was truncated with the first 381 bp of pullulanase gene sequence not present.

5.2 Enzyme Expression and Cell Growth

For all the plasmids listed above, i.e. those that were phenotypically pullulanase positive and transcribed from their own promoters, the following parameters of their expression and growth were determined: the amount of enzyme produced per cell (see section 5.2.1); whether enzyme synthesis was substrate-induced or repressed (see section 5.2.2); whether pullulanase synthesis was dependent on cellular growth phase (see section 5.2.3); the consequences of pullulanase expression on the *E. coli* cells (see section 5.2.4); the cellular location of the enzyme (see section 5.2.5); and that the pullulanase gene was truly being transcribed from its own promoters (see section 5.2.6). To minimise sample variation in the following experiments, all plasmids were transformed into the same background *E. coli* strain, JM83 (PB2946).

5.2.1 Expression levels

The amount of pullulanase produced from each recombinant plasmid mentioned above was determined using both the PAHBAH reducing sugar assay (see section 2.6.1.2) and the dye-linked pullulan assay (see section 2.6.1.4). Cultures were grown to stationary phase (OD<sub>600</sub> = 4.0). The cells were washed in 56/2 and disrupted by sonication. The pullulanase activity of each total cell extract and the relative plasmid copy numbers were determined (see section 2.2.3.3), and these data are listed in table 5.1.
Figure 5.1 Recombinants expressing the pullulanase gene

Expression of the Pullulanase in *E. coli*
Table 5.1 Enzyme expression levels from different plasmid constructions

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Activity (μg maltotriose / min/ ml of culture)</th>
<th>Units/l (μmoles maltotriose / min / litre of culture)</th>
<th>Relative Copy Number (ng DNA /kb of plasmid)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNZ1037</td>
<td>0.79 ± 0.08</td>
<td>1.6</td>
<td>8.2</td>
</tr>
<tr>
<td>pNZ1038</td>
<td>0.84 ± 0.07</td>
<td>1.7</td>
<td>7.7</td>
</tr>
<tr>
<td>pNZ1451</td>
<td>2.5 ± 0.2</td>
<td>5.0</td>
<td>34.4</td>
</tr>
<tr>
<td>pNZ1452</td>
<td>1.4 ± 0.1</td>
<td>2.8</td>
<td>37.2</td>
</tr>
<tr>
<td>pNZ1503</td>
<td>10.2 ± 0.7</td>
<td>20</td>
<td>103.5</td>
</tr>
<tr>
<td>pNZ1504</td>
<td>9.8 ± 0.7</td>
<td>19</td>
<td>98.7</td>
</tr>
</tbody>
</table>

The values obtained show, as expected, a direct correlation between plasmid copy number and pullulanase activity.

Initial kinetic characterisations were also performed to prove that the activity determined was enzymatic in nature. Thus the amount of pullulanase activity measured was shown to be dependent on: the amount of cellular extract, the amount of substrate, and the time of incubation.

5.2.2 Substrate induction of enzyme synthesis

Pullulanase synthesis from *Caldocellum saccharolyticum* was substrate inducible (see section 3.2), in common with a number of other thermophilic, glucolytic enzymes. The transcription of the genes from the maltose regulon of *E. coli* and *Klebsiella* sp. have been shown to be both substrate-induced and repressed. It has also been shown that the *E. coli* maltose regulon can complement genes from the *Klebsiella* maltose regulon (see section 1.3).

Substrate induction experiments were performed to determine whether genes of the *E. coli* maltose regulon could complement those from *Caldocellum saccharolyticum* and whether any of the other open reading frames identified were involved in substrate induction in *E. coli*. The carbon sources used were maltose as an inducer, glucose as a repressor and glycerol as a control. The cells
containing each recombinant plasmid were grown overnight on minimal media (56/2 with vitamin B1 and appropriate antibiotic) with glycerol (0.5%) as the sole carbon source. Glycerol was used to return cell concentrations of glucose and maltose (and hence enzyme expression) to their basal levels, as it is not involved in substrate induction or repression of the maltose induction pathway. From each overnight culture portions were removed and grown in six different carbon sources as listed in table 5.2.

<table>
<thead>
<tr>
<th>Table 5.2 Carbon sources</th>
<th>Expected effect on expression if substrate-inducible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon source</td>
<td>Effect on maltose regulon</td>
</tr>
<tr>
<td>1</td>
<td>Glycerol 1.2%</td>
</tr>
<tr>
<td>2</td>
<td>Maltose 0.2%, glycerol 1%</td>
</tr>
<tr>
<td>3</td>
<td>Maltose 1.2%</td>
</tr>
<tr>
<td>4</td>
<td>Maltose 1.2%, glucose 0.2%</td>
</tr>
<tr>
<td>5</td>
<td>Maltose 1.2%, glucose 1.2%</td>
</tr>
<tr>
<td>6</td>
<td>Glucose 1.2%</td>
</tr>
</tbody>
</table>

Due to the different growth rates obtained from each carbon source, pullulanase activity was determined at stationary phase (defined as OD₆₀₀ = 3.0). Pullulanase activity was determined by the PAHBAH reducing sugar method (see section 2.6.1.2) using whole cell extracts. These results showed that enzyme expression levels were identical, within experimental error, regardless of the carbon source. This result implies that enzyme expression was not substrate-inducible or repressible in *E. coli* for all recombinants tested.

### 5.2.3 Dependence of cellular growth phase on enzyme expression

A time course was performed using each recombinant plasmid, to test whether the growth phase affected the level of pullulanase expression. Cells containing each recombinant plasmid were grown in L-broth with appropriate antibiotic. Samples were removed at different growth phases (OD₆₀₀ = 0.1-4.0) as determined by optical density and assayed for pullulanase activity. The results are shown in figure 5.2.
Figure 5.2 The relationship between cell growth phase and enzyme expression

![Graph showing the relationship between cell growth phase and enzyme expression.]

For clarity, data are shown from only two plasmids, since the data from all other plasmids falls between these values. Each value is the mean of two repeats with the standard deviation contained within the area of the symbol.

Cultures were also left overnight (16-24 h) and assayed for pullulanase activity. Although these cultures normally had a higher optical density (as they contained both dead and living cells) they had reduced pullulanase activity. These results showed that the level of enzyme expression per absorbance unit (i.e. cell mass) did not change as the culture growth phase altered (i.e. expression levels were unaffected by growth phase).

5.2.4 Consequences of pullulanase expression on the E. coli cells

The presence of plasmids in E. coli impose a metabolic burden on the cells (Bentley et al. 1990). This effect can be seen in the different growth rates of cells containing plasmids compared to non-plasmid containing cells (Marr 1991). If the plasmid also expresses a protein deleterious to the cells, growth rates will be further reduced, with a possible change in cell structure. To ascertain the effect of pullulanase expression on the E. coli cell metabolism, the following experiments were performed: growth rate studies (see section 5.2.4.1); microscopic examination (see section 5.2.4.2); and plasmid stability tests (see section 5.2.4.3).

5.2.4.1 Growth rate studies

To ascertain the effect of pullulanase expression on cell growth rate, time course experiments were performed for E. coli containing each plasmid construct. Growth rate was measured as change in...
absorbance (OD_{600}) over time, with the maximum growth rate occurring during log phase. By plotting log (OD_{600}) versus time, the slope of the graph during log phase (maximum slope) defined the maximum growth rate. The maximum growth rates obtained for each plasmid are shown graphically in figure 5.3.

**Figure 5.3 Maximum growth rate of each recombinant**

![Graph showing maximum growth rate of each recombinant plasmid.]

Where 100% activity = \ln [\text{ABS OD}_{600}] / \text{min} \times 1.06/100. The background E. coli strain for all studies was JM83.

Overall these results show a reduction in cell growth rate for cells that contain plasmid. This reduction is proportional to the copy number of the plasmid (see section 5.2.1). A greater reduction occurred that was proportional to the expression level of the pullulanase.

5.2.4.2 Microscopic examination

Phase contrast light microscopy was used to examine the E. coli cells containing the recombinant plasmids at different phases of growth. Photographs were taken using a Leitz Diaplan microscope and a selection is shown in figure 5.4.

Expression of the Pullulanase in E. coli
Figure 5.4 Photomicrographs showing the effect of pullulanase expression on *E. coli* cells

Linear phase *E. coli* containing pGEM5 without the pullulanase gene.
Magnification X1500 Phase contrast

Linear phase *E. coli* containing pNZ1503 (pGEM5 + pulA gene) and expressing the pullulanase gene. Magnification X1500 Phase contrast

Linear phase *E. coli* containing pNZ1503 (pGEM5 + pulA gene) and expressing the pullulanase gene. Magnification X8750 Phase contrast

Chapter 5
The results indicate that as the level of pullulanase increases within the cells, cell division is inhibited causing the E. coli to remain as chains of 3-5 cells.

The distribution of each cell type was determined for each plasmid. Cells were grown to late linear phase (OD_{600} = 3.0), diluted and counted using a phase contrast light microscope. For counting purposes, duplicate slides were made for each plasmid and 100 cells counted (multimeric E. coli were counted as only one cell) and scored according to their phenotype (see figure 5.5).

**Figure 5.5 Distribution of cell type in different plasmids**

![Distribution of cell type in different plasmids](image)

where # mer = # full length cells joined

The cell distribution of the JM83 control was identical, regardless of which plasmid vector was present (pBR322, pCGN565 or pGEM5zf(+)).

As shown in figure 5.5, there was a correlation between enzyme expression level and the formation of multimeric elongated cells. This result implies that pullulanase expression was responsible for the inhibition of cell division. A similar result was obtained on induction with the heat inducible expression plasmid pJLA602 (see section 6.2.1) and hence this phenomenon was not related to the cloning vector used.

**Expression of the Pullulanase in E. coli**
5.2.4.3 Plasmid stability tests

Stability tests were performed using the method of serial dilution and viable count (see section 2.2.3.2). *E. coli* cells containing the plasmid (1 μl inoculum) were grown for 16 hours at 37°C in L-broth (10 ml) with no antibiotic selection. From this culture 1 μl was removed, innoculated into L-broth and grown for a further 8 hours at 37°C. Viable cell counts were determined by serial dilution onto L-agar plates with and without antibiotic selection. The cells were grown twice without antibiotic selection to magnify the result of plasmid instability as determined by differential growth rate. The fraction of cells still containing the plasmid was determined as the ratio of cells grown on antibiotic selection to cells grown without antibiotic selection. The results are shown in figure 5.6.

Figure 5.6 Plasmid stability without antibiotic selection

These results showed that the plasmids were generally unstable without antibiotic selection. This instability was greater with plasmids that also expressed the pullulanase gene.

Recombinants have also been detected that are antibiotic resistant although phenotypically pullulanase negative. Plasmid DNA from these recombinants was isolated and restriction enzyme mapped, revealing that deletions had taken place within the pullulanase gene.
5.2.5 Cellular location of the enzyme

The translational products of foreign genes expressed in *E. coli* are not normally exported out of the cytoplasm even though *E. coli* exports a large number of its own proteins to the periplasmic space, inner and outer membranes, and into the growth medium (Sjöström *et al.* 1987). Proteins exported include those of the maltose regulon (see section 1.3.1), and the pullulanase export proteins of *Klebsiella* sp. (see section 1.3.2). Proteins to be exported are made as larger precursors containing an amino-terminal signal peptide of 20-40 amino acid residues which determines the final cellular location of the protein and is cleaved during export (for discussion on the proposed PulA leader sequence, see section 4.5.4).

![Figure 5.7 Enzyme location within different cellular fractions]

**Figure 5.7 Enzyme location within different cellular fractions**

<table>
<thead>
<tr>
<th>Enzyme location / Fraction number</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3) Periplasm</td>
</tr>
<tr>
<td>(2) Cytoplasm insoluble but extractable with NaCl / Triton X-100</td>
</tr>
<tr>
<td>(1) Cytoplasm soluble</td>
</tr>
<tr>
<td>(0) Cytoplasm insoluble but non-extractable with NaCl / Triton X-100</td>
</tr>
</tbody>
</table>

Expression of the Pullulanase in *E. coli*
As shown previously, the *Caldocellum saccharolyticum* pullulanase gene was exported into the culture medium when expressed in the native organism. Cellular fractions were produced (see section 2.5.1.4) for each recombinant plasmid, to determine the location of the pullulanase when expressed in *E. coli*. A schematic representation of possible pullulanase locations within the *E. coli* cell is shown in figure 2.2.

As enzyme distribution can vary depending on the growth phase of the cells, cells were harvested at two different growth phases, mid-log phase (OD<sub>600</sub> = 0.5) and stationary phase (OD<sub>600</sub> = 4.0). The results are shown in figure 5.7.

This diagram shows that the location of the pullulanase was predominantly cytoplasmic (approximately 30% soluble, 50% insoluble) with some exported to the periplasmic space (approximately 20%). No enzyme was detected (<1%) in either the outer membrane/cell wall-associated fraction or the culture growth media. The pullulanase was exported out of the cells in the native organism but remained predominantly cytoplasmic when expressed in *E. coli*. However there appeared to be some recognition of the *Caldocellum saccharolyticum* PulA leader sequence by the *E. coli* protein export pathway since some pullulanase reached the periplasmic space.

Overall, two trends were evident. Between the mid-log and stationary growth phases, the enzyme distribution changed, increasing within the periplasm and decreasing in the soluble cytoplasmic fraction. There was also an increase in the insoluble cytoplasmic fraction corresponding to higher expression levels.

The difference in the percentage of pullulanase in the periplasmic space was likely to be an effect of the different antibiotic resistance enzyme coded for on each plasmid, β-lactamase (pNZ1037-8 and pNZ1503-4) or chloramphenicol acetyl-transferase (pNZ1451-2). It has previously been shown that a high level of β-lactamase expression, and hence export from the cells, increased the permeability of the *E. coli* membranes (Georgiou et al. 1988), whereas the chloramphenicol acetyltransferase was active in the cytoplasm and not exported, and therefore did not alter membrane permeability. Further evidence for this proposed increased permeability was found in an increase in the amount of periplasmic proteins (as determined by soluble protein assay method, section 2.5.5) for those cells that expressed β-lactamase compared to those that expressed chloramphenicol acetyl-transferase.

### 5.2.6 Proof that the pullulanase gene in all constructs was expressed from a *C. saccharolyticum* promoter

Until the nucleotide sequence of the *C. saccharolyticum* insert in pNZ1038 was determined (see section 4.4) it was impossible to say unequivocally that each construct did not express the pullulanase gene as the result of a gene fusion or using an *E. coli* vector sequence for expression.
Expression studies with the pCGN565-6 exonuclease deletions showed that the expression level was independent of the gene orientation (see section 1.6). This result implied that expression was solely dependent on *C. saccharolyticum* DNA. Although the *C. saccharolyticum* DNA containing the *pulA* gene was ligated into the middle of an already transcribed plasmid vector gene, it was always in the opposite orientation to the direction of transcription. Also, analysis of the sequence upstream of the start codon of each construct revealed a number of stop codons in each reading frame. This information is shown in figure 5.8.

**Figure 5.8 Stop codons between *pulA* and *PepY***

For clarity only stop codons and amino acids from *PepY* and *PulA* are shown. The stop codons present in all three reading frames between *PulA* and *PepY* are shown as *. The nucleotides are numbered from the methionine start codon of *PulA*.

The number of translational stop codons in each reading frame between the end of *PepY* and the start of *PulA* (4, 8, 4 respectively), would suggest that gene fusion was unlikely as a mechanism of enzyme expression.

Expression of the Pullulanase in *E. coli*
5.3 Purification of the *C. saccharolyticum* Pullulanase Expressed in *E. coli*

This section details the purification of the pullulanase expressed in *E. coli*. Initially all constructs were used. A strategy for the purification of the pullulanase expressed in *E. coli* by pNZ1503 is shown in figure 5.9. A purification table is shown in table 5.3.

---

**Figure 5.9 Strategy for purification of the pullulanase expressed in *E. coli***

1. Lysis of *E. coli* cells expressing the pullulanase enzyme
2. Heat treatment to denature *E. coli* proteins (70°C, 3 h)
3. Ammonium sulphate fractionation. 50%-70%
4. Gel filtration (Sepharose 6B)
5. Ion exchange (DEAE Sephacel)
6. Ultra Filtration (centricon™ 100)
7. Pullulanase band extracted from agarose gel

The major purification steps are shown in bold with thick arrows.
<table>
<thead>
<tr>
<th></th>
<th>Protein (µg)</th>
<th>Units</th>
<th>Specific activity U/µg</th>
<th>Percentage recovered</th>
<th>Purification ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sonicated</td>
<td>1 350 000</td>
<td>97.4</td>
<td>0.000072</td>
<td>100</td>
<td>1.0</td>
</tr>
<tr>
<td>extract (from</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 l culture)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heat treated</td>
<td>45 000</td>
<td>46.5</td>
<td>0.00103</td>
<td>47.7</td>
<td>14.3</td>
</tr>
<tr>
<td>extract</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(NH₄)₂SO₄</td>
<td>8 100</td>
<td>21.3</td>
<td>0.00263</td>
<td>21.9</td>
<td>36.5</td>
</tr>
<tr>
<td>Extract</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IEX DEAE</td>
<td>1 540</td>
<td>9.7</td>
<td>0.00629</td>
<td>6.3</td>
<td>87.4</td>
</tr>
<tr>
<td>Sephacel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ultra</td>
<td>200</td>
<td>3.2</td>
<td>0.016</td>
<td>3.28</td>
<td>222.2</td>
</tr>
<tr>
<td>filtration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As the purification strategy was to produce sufficient pullulanase for N-terminal sequencing and substrate characterisation, higher specific activity was favoured over total recovery.

Figure 5.10 Protein fractions obtained during purification

Lane 1 Ammonium sulphate fraction
Lane 2 Ion exchange fraction
Lane 3 Ultra filtration fraction
Lane 4 Pharmacia low molecular weight markers (sizes in kDa)

Samples were electrophoresed on a 12.5% SDS-PAGE gel and visualised by staining with coomassie brilliant blue.
5.3.1 Initial purification studies

Initial attempts to purify the pullulanase used a number of different plasmid constructs expressed in *E. coli*. The basis of this early purification was the heat purification procedure (see section 2.5.1.2). This method was based on the difference in thermostability between the foreign thermophilic protein and the heat-labile *E. coli* proteins. The thermophilic protein was determined by analysis of SDS-PAGE gels (see section 2.6.2) that had been loaded with the heat purified soluble cellular extracts in tandem with control samples not expressing the pullulanase gene. Analysis of these gels revealed that a number of thermostable proteins were present and different from the control samples, but no major band could be identified as the pullulanase, unlike other *C. saccharolyticum* enzymes (Love et al. 1988, Lüthi et al. 1990b). Therefore, further purification was required.

5.3.2 Heat purification and ammonium sulphate precipitation

The initial purification procedure was similar to that used for the pullulanase from *C. saccharolyticum* (see section 3.3) i.e. ammonium sulphate precipitation and ion exchange column chromatography. Initial attempts to purify the pullulanase used all the recombinant plasmids listed in section 5.2, but it soon became apparent that only those strains that had the high copy number plasmids (pNZ1503, pNZ1504) produced sufficient pullulanase for large scale purification. In the case of *E. coli* strains containing the other plasmids, the active pullulanase fractions were detected in similar locations to those determined for the large scale purification of pNZ1503, implying that each strain expressed the enzyme in the same form.

For the large scale purification of the pullulanase, *E. coli* strain PB4974 (containing pNZ1503) was grown in L Broth (100 μg / ml ampicillin) to stationary phase (OD_{600} = 3.0) and harvested by centrifugation (4°C, 3000 g, 10 min). The cells were disrupted as in section 2.5.1.1 and the soluble cellular extract was isolated by centrifugation (4°C, 13000 g, 30 min). Protease-free DNase (final concentration 10 μg/ml, 15 min at 37°C) was added to the extract to remove DNA fragments, which had previously been shown to complex with the pullulanase and interfere with subsequent purification steps. The extract was then heat-purified (as in section 2.5.1.2) followed by an ammonium sulphate precipitation (see section 2.5.1.3). This fractionation procedure had the two-fold advantage of sample concentration and elimination of reducing sugars present in the growth media. Initial tests were performed with small sub-samples to determine the optimum degree of saturation, shown graphically in figure 5.11.
Figure 5.11 The distribution of enzyme activity at different ammonium sulphate saturation values

Figure 5.11 shows that most of the pullulanase activity was present in the 30-80% saturated fraction. As the object of the purification procedure was to produce pure enzyme for N-terminal amino acid sequencing and substrate characterisation, only the 50-70% fraction was used for further purification work.

5.3.3 Ion Exchange

Initial ion exchange studies, using both cation and anion media at varying pH, were performed to test the binding of the pullulanase, as for the pullulanase from the native organism (see section 3.3.2). These tests showed a similar result to that obtained for the pullulanase from C. saccharolyticum. It was determined that anion exchangers at pH 8.0 (_{wC}) efficiently bound the pullulanase, which could then be eluted using a 0-500 mM NaCl step gradient. Extracts that had not been DNAse treated also bound at pH 8.0 (_{wC}), but only some activity could be eluted even at high salt concentration (900-1500 mM NaCl). For purification, 1-2 ml of ammonium sulphate extract was deionised using a PD-10 column (see section 2.5.4.1), applied to a DEAE Sephadex column (column buffer 20 mM Tris-HCl pH 8.0 (_{wC}) and 0.01% v/v Triton X-100) (see section 2.5.4.2) and eluted using a 0-500 mM NaCl gradient. Pullulanase activity was determined using the PAHBAH reducing sugar assay (see section 2.6.1.2) for every fraction. A typical chromatogram is shown in figure 5.12 together with the location of pullulanase active fractions.

Expression of the Pullulanase in E. coli
Figure 5.12 A DEAE-Sepacel ion exchange chromatogram of the pullulanase expressed in E. coli

Analysis of the fractions by SDS-PAGE showed that a protein band of the correct molecular mass was present in the active fractions, and that the density of the protein band increased with increasing activity, thus implying that it was the pullulanase (see figure 5.13).

Figure 5.13 SDS-PAGE of pullulanase active fractions from ion exchange column

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
</table>

lane 8 and 9 Pharmacia low molecular weight markers (sizes shown in kDa)
lane 1-7 and 10-16 column fractions with pullulanase activity (maximum activity, fraction 7 and 10)
Samples were electrophoresed on a (12.5 %) SDS-PAGE and silver stained
5.3.4 Gel filtration

Initially, due to the possibility that the enzyme was in the form of a high molecular mass aggregate, a gel filtration medium with a large molecular mass separation range was used (Sepharose 6B, 4000-10 kDa). The original ammonium sulphate extract was deionised and applied to the column as in section 3.3.3. Pullulanase activity was determined using the PAHBAH reducing sugar assay (see section 2.6.1.2) for every fraction. The relative molecular mass of the non-denatured, pullulanase was determined using a calibration curve previously obtained with molecular mass markers, thyroglobulin, 669 000, apoferritin, 443 000, alcohol dehydrogenase, 150 000, albumin, 66 000, and cytochrome C, 12 400 (see section 3.3.3). The value obtained was approximately 100 kDa. Earlier work using non-DNAse treated samples detected the presence of high molecular mass aggregates with nominal molecular mass greater than the exclusion value of the column (4000 kDa).

The apparent molecular mass of non-denatured protein (as determined by gel filtration) was approximately the same as the denatured molecular mass (as determined by SDS-PAGE), thus implying that the enzyme was not in the form of a high molecular mass multi-enzyme complex. A typical chromatogram is shown in figure 5.14 together with the location of active pullulanase fractions.

Expression of the Pullulanase in E. coli
As the enzyme did not appear to be a large molecular mass complex, it was more appropriate to use a different gel filtration medium that had a molecular mass fractionation range closer to the molecular mass of the proteins present. The medium used was Sephacryl S-200 (see section 2.5.4.3), an allyl dextran covalently cross-linked to acrylamide, with a molecular mass fractionation range of 5-250 kDa.

The combined pullulanase active fractions from the DEAE Sephacel ion exchange column were applied to the column. Pullulanase activity was determined using the PAHBAH reducing sugar assay (see section 2.6.1.2) for every fraction. A calibration curve previously obtained with molecular mass markers, β-amylase, 200 000, alcohol dehydrogenase, 150 000, albumin, 66 000, carbonic anhydrase, 29 000, and cytochrome C, 12 400, was used to determine the relative molecular mass of the non-denatured, pullulanase (see section 3.3.3). The value obtained was approximately 95 kDa. The location of the pullulanase active fraction is shown in figure 5.15.

Figure 5.15 A Sephacryl S-200 gel filtration chromatogram

Again, the apparent molecular mass obtained for gel filtration was similar to that obtained by SDS-PAGE and the previous values obtained with the Sepharose 6B gel filtration column.
5.3.5 Purification using ultrafiltration

As gel filtration could not separate the high molecular mass pullulanase from the other proteins, ultrafiltration was used. The pullulanase-positive ion exchange fractions were further purified by the use of a Centricon™ 100 concentrator (nominal M, exclusion of 100 kDa, i.e. retains proteins with M, > 100 kDa and also a significant number of the proteins with M, slightly < 100 kDa). As the pullulanase band was only slightly less than 100 kDa, most of the enzyme was retained and concentrated. The resulting fraction was sufficiently pure for N-terminal amino acid sequencing and enzymatic characterisation.

SDS-denatured protein samples were run on Prosieve™ agarose gels and the protein band corresponding to the pullulanase was extracted and tested for activity (see Section 2.6.2.7). Since this method separates proteins on the basis of their denatured molecular mass, it allowed the removal of contaminating proteins with differing molecular mass. Enzyme from this single protein band was used for substrate characterisation of the pullulanase expressed in *E. coli*.
5.4 Enzyme Characterisation

Using the purification method listed in section 5.3, it was possible to obtain pullulanase of sufficient purity to perform enzyme characterisations. The enzyme extract used for all characterisations, except for the final substrate characterisation, was the ultrafiltration (Centricon™ 100) purified extract (see section 5.3). For the final substrate characterisation, the purest enzyme extract (run and isolated from Prosieve agarose) was used (see section 5.3.5). The characterisations performed are outlined in sections 5.4.1-5.4.4.

5.4.1 Activity gels and N-terminal sequencing

Unlike the native enzyme, that appeared to be blocked (see section 3.4.1), the amino-terminal sequence of the enzyme expressed in E. coli could be sequenced. A total of four amino acid residues were determined and these residues corresponded exactly with the first four putative amino acids of the PulA gene product (see figure 5.16).

Figure 5.16 The N-terminal sequence of the purified PulA gene product

| Deduced amino acid sequence of PulA: | M I V K A Y I |
| N-terminal sequence of pullulanase: | M I V K |

For N-terminal sequencing the pullulanase fraction was electrophoresed on an SDS-PAGE and blotted on membrane.

Activity gels were performed with the purified enzyme extract (see section 2.6.2.6) and examples are shown in figure 5.17.

These gels show that besides the full length PulA gene product, a number of pullulanase-positive proteins of lower molecular mass were present. These proteins could be the result of either proteolytic degradation of the full-length PulA gene product, or translation of shortened mRNA transcripts. As there were a number of sequences in the middle of the gene homologous to expression start sites and RNA polymerase termination sites, it is possible that shorter mRNA transcripts (and hence proteins) could have arisen. Also the plasmid pNZ1452 expressed pullulanase activity even though it did not contain the full-length gene, hence this gene product must have arisen from an internal start site. Isolation of the pullulanase in the presence of protease inhibitors (e.g. PMSF and EDTA) did not alter either enzyme activity or the population distribution pattern of proteins present. This result implies that proteolytic degradation was either a minor cause or had taken place during cell lysis, before the protease inhibitors could inactivate the proteases.
Figure 5.17 Activity gels showing the location of the pullulanase

![Activity Gels Image]

Lane 1: Ion exchange fraction containing the pullulanase enzyme
Lane 2: Pharmacia low molecular weight standards (sizes are in kDa)

Identical samples were loaded onto each side of a 12% SDS-PAGE and electrophoresed. One half was overlaid with Pullulan Red (A) and the other stained with coomassie brilliant blue.

5.4.2 The effect of metal ions and inhibitors on pullulanase activity

5.4.2.1 Initial studies

The effects of a number of different metal ions on the activity of the pullulanase expressed in *E. coli* were determined using a similar method to section 3.4.2. Two different samples were used, the purified enzyme (see section 5.3) and the same sample dialysed (16h 4°C) with 10 mM EDTA to complex out any metal ions present. As expected, the dialysed sample had very low activity, implying that the removal of metal ions reduced enzyme activity. Specific metal ions were then added to each sample and percentage activity measured, relative to the purified enzyme.

This result is similar to that obtained in section 3.4.2. The addition of monovalent cations had little or no effect on enzyme activity in either sample. For the divalent cations the addition of metal ions to the dialysed sample increased enzyme activity significantly, and in the case of Ca²⁺ the reversal was almost complete with activity in the dialysed sample approximately equal to 400% activity achieved by adding calcium ions to the purified sample. This reversal was the result of two effects, the binding of the metal ion at the metal ion binding site increased enzyme activity, and as some residual...
Ca\textsuperscript{2+}/EDTA complex was still present in the enzyme sample the addition of other cations released some of the Ca\textsuperscript{2+}, allowing them to bind to the metal ion binding site.

### Table 5.4 Effect of different metal ions on pullulanase activity

<table>
<thead>
<tr>
<th>Metal ion (1 mM) added</th>
<th>non-dialysed</th>
<th>EDTA dialysed</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>100</td>
<td>36</td>
</tr>
<tr>
<td>Ca\textsuperscript{2+}</td>
<td>392</td>
<td>405</td>
</tr>
<tr>
<td>Zn\textsuperscript{2+}</td>
<td>256</td>
<td>91</td>
</tr>
<tr>
<td>Ni\textsuperscript{2+}</td>
<td>263</td>
<td>106</td>
</tr>
<tr>
<td>Cu\textsuperscript{2+}</td>
<td>217</td>
<td>87</td>
</tr>
<tr>
<td>Fe\textsuperscript{2+}</td>
<td>176</td>
<td>72</td>
</tr>
<tr>
<td>Mg\textsuperscript{2+}</td>
<td>202</td>
<td>83</td>
</tr>
<tr>
<td>Mn\textsuperscript{2+}</td>
<td>194</td>
<td>88</td>
</tr>
<tr>
<td>Co\textsuperscript{2+}</td>
<td>285</td>
<td>72</td>
</tr>
<tr>
<td>Fe\textsuperscript{3+}</td>
<td>137</td>
<td>53</td>
</tr>
<tr>
<td>Ag\textsuperscript{+}</td>
<td>86</td>
<td>39</td>
</tr>
<tr>
<td>Na\textsuperscript{+}</td>
<td>102</td>
<td>35</td>
</tr>
<tr>
<td>K\textsuperscript{+}</td>
<td>89</td>
<td>32</td>
</tr>
<tr>
<td>Li\textsuperscript{+}</td>
<td>91</td>
<td>43</td>
</tr>
</tbody>
</table>

The values represent the average of 3 measurements, the variation between measurements was approximately 10%. Activity of the undialysed purified enzyme without the addition of metal ions was defined as 100% (0.001 U pullulanase). Assay procedure used: 0.001 U pullulanase (2 µl), 200 µg pullulan (8 µl), metal ion 1 mM (2 µl) and 88 µl of enzyme buffer (Mes/NaOH 0.2 M, pH 5.7\textsubscript{20°C}). Reducing sugar released was assayed by the PAHBAH reducing sugar assay (see section 2.6.1.2).

The reduced activity on addition of other divalent cations to the non-dialysed samples was possibly due to the replacement of the Ca\textsuperscript{2+} at the metal ion binding site with other divalent cations, forming an active enzyme but with lower activity than addition of Ca\textsuperscript{2+} only.

#### 5.4.2.2 Calcium ion studies

In common with the pullulanase expressed in \textit{C. saccharolyticum} (see section 3.4.2.2), the Ca\textsuperscript{2+} had the greatest effect on pullulanase activity. For this reason further experiments were performed at different concentrations of Ca\textsuperscript{2+} and EDTA. Their interaction is shown in figure 5.18.
The interaction of varying concentrations of Ca\(^{2+}\) and EDTA on pullulanase activity

![Graph showing the interaction of Ca\(^{2+}\) and EDTA on pullulanase activity](image)

The results show that enzyme activity decreased proportionally with increased concentration of EDTA and increased proportionally with increased concentration of Ca\(^{2+}\) (until values of 1 mM). A similar result was obtained for the unpurified enzyme extract and hence extra Ca\(^{2+}\) (1 mM) was added in all enzyme assays.

### 5.4.2.3 Monovalent cations

Monovalent cations in low concentrations (1 - 10 mM) have been shown to not affect pullulanase activity. However, as high concentrations of NaCl were used during the purification procedure (up to 1500 mM) and some samples required further concentrating before assaying, the effect of high concentrations of NaCl on pullulanase activity was determined (see figure 5.19).
Figure 5.19 The effect of high concentrations of NaCl on pullulanase activity

![Graph showing the effect of NaCl concentration on pullulanase activity. The values represent the average of 3 measurements, with standard deviations contained within the area of the symbol. 100% activity was defined as maximum activity, which was in fact higher than when no NaCl was added.]

While increasing the NaCl concentration (up to 1000 mM) caused a small increase in enzyme activity, extremely high concentrations of NaCl (>2500 mM) decreased enzyme activity. Since NaCl concentrations of up to 1500 mM were used in the purification procedure, and enzyme activity was not significantly lower at these levels, pullulanase activity could be assayed directly from samples eluted by salt gradient, without the need for dialysis.

5.4.3 Thermostability and temperature and pH optima of the pullulanase

To determine the temperature and pH optima of the pullulanase expressed in E. coli, standard enzyme assays (see section 2.6.1.1) were performed using the PAHBAH reducing sugar assay (see section 2.6.1.2). The experimental regime was identical to that used in section 3.4.3. The temperature optimum was determined by measuring the enzyme activity at different temperatures and the pH optimum was determined in a similar manner, with the enzyme reaction buffer at different pH values. The results of both temperature and pH optimal studies are shown in figure 5.20.
These results show that the maximum activity of the purified enzyme extract was at 70°C and pH 5.7. These values are the same as those obtained for the unpurified enzyme extract, as well as from the other pullulanase expressing plasmids, and hence correspond to the standard enzyme assay conditions used.

To determine the thermostability of the pullulanase expressed in *E. coli*, a known amount of enzyme was pre-incubated, without substrate, at different temperatures for different lengths of time. The remaining activity in each sample was determined by standard assay procedure using the PAHBAH reducing sugar assay. The results are shown in figure 5.21.

The enzyme appeared to be thermostable at temperatures up to 75°C, but was irreversibly denatured at higher temperatures. Longer incubation times were performed at both 70°C and 75°C and the enzyme half-life at these temperatures was estimated to be approximately 75 h and 18 h respectively. These results correlated well with the known growth temperature of the *C.*

Expression of the Pullulan in *E. coli*
saccharolyticum organism, hence showing that the peptide sequence of the pullulanase conferred intrinsic thermostability.

Figure 5.21  Thermostability of the full length pullulanase expressed in E. coli

Although the pH and temperature optima of the shortened pullulanase expressed from pNZ1452 were identical to those of the full-length form, a reduction in thermostability did occur. The thermostability of this shortened pullulanase was determined using semi-purified pullulanase extract, as shown in figure 5.22.

The enzyme appeared to be thermostable at temperatures up to 70°C, but was irreversibly denatured at higher temperatures. Longer incubation times were performed at both 70°C and 75°C. The enzyme half-life at these temperatures was estimated to be approximately 45 h and 11 h respectively.
5.4.4 Substrate characterisation

The enzyme was incubated with various substrates to determine the substrate activity profile of the pullulanase expressed in *E. coli*. For the final substrate characterisation the purest enzyme extract was used, i.e. the affinity binding purified extract, that was further purified by running and isolating from Prosieve agarose (see section 5.3.5).

The resulting activity was determined as the number of bonds hydrolysed (i.e. increase in the amount of reducing sugar) and was measured by the PAHBAH reducing sugar assay. The hydrolysis products formed were determined by paper chromatography (see section 2.6.3). For the starch-based

Expression of the Pullulanase in *E. coli*
substrates (amylose, amylopectin and β-limit dextrin) activity was also measured by the starch-iodine binding assay (see section 2.6.1.3). The results of these analyses are shown in figure 5.23.

Figure 5.23  Enzymatic activity measured as amylopectin-iodine binding

Summary of changes in starch-iodine values before and after incubation

<table>
<thead>
<tr>
<th>Enzyme (Pullulanase)</th>
<th>Iodine colour</th>
<th>λ max (nm)</th>
<th>Peak value</th>
<th>Blue value</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNZ1503</td>
<td>red → red</td>
<td>550 → 560</td>
<td>0.47 → 0.52</td>
<td>0.25 → 0.26</td>
</tr>
<tr>
<td>Klebsiella</td>
<td>red → purple</td>
<td>550 → 575</td>
<td>0.47 → 0.73</td>
<td>0.25 → 0.41</td>
</tr>
</tbody>
</table>

When other starch-based substrates (e.g. β-limit dextrin) were used the results obtained were essentially the same.

Table 5.5 Relative activity of the pullulanase expressed in E. coli (pNZ1503) on different substrates.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Pullulan</th>
<th>β- limit dextrin</th>
<th>Amylopectin</th>
<th>Amylose</th>
<th>Glycogen (Rabbit liver)</th>
<th>Glycogen (oyster)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1.43</td>
<td>0.35</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Activities are expressed as a percentage of activity on pullulan and were measured by PAHBAH reducing sugar assay.
Figure 5.24 Paper chromatogram of the hydrolysis products formed.

Lane 1-7 Glucose, Maltose, maltotriose, maltotetraose, maltopentaose, maltohexose and maltoheptaose respectively.
Lane 8 Panose
Lane 9-12 Hydrolysis products formed from digestion with pullulan, amylopectin, beta limit dextrin and amylose respectively.

During the purification procedure a number of initial tests of substrate characterisation were performed using semi-purified extracts. The results obtained were essentially identical to those shown above. Similar results were also obtained using semi-purified extracts of the pullulanase expressed from pNZ1452. Denaturing refolding experiments (8 M urea followed by dialysis) were performed in an attempt to improve the pullulanase activity on β-limit dextrin and amylopectin. These experiments had no apparent effect.

Enzymes that hydrolyse pullulan are classified according to their final hydrolysis end products (see section 1.4 and figure 1.4.2). As the pullulanase expressed in E. coli only produced maltotriose as its final end product, it was classified as a pullulanase rather than an isopullulanase, neopullulanase or glucoamylase. Since the pullulanase showed no α-1,4 activity when β-limit dextrin, amylopectin and amylose were used as substrates, it was further classified as a pullulanase without α-1,4 activity (see section 1.4.2). However since it also showed only low α-1,6 activity on β-limit dextrin and amylopectin this was probably not a true indication of its substrate hydrolysis profile.

Expression of the Pullulanase in E. coli
5.5 Summary

Expression studies using each of these recombinants have shown that pullulanase expression is directly proportional to plasmid copy number (see section 5.2.1). For all recombinants, the level of enzyme expression is unaffected by either cell growth phase (see section 5.2.3) or carbon source (see section 5.2.2). Plasmid stability tests have shown that recombinants expressing the pullulanase gene are unstable in *E. coli* without antibiotic selection and that expression is deleterious to the cells, reducing growth rate and interfering with cell division causing the formation of elongated chains of cells (see section 5.2.4).

The location of the enzyme was determined to be predominantly cytoplasmic (approximately 30% soluble, 50% insoluble) with some exported to the periplasmic space (approximately 20%). No enzyme was detected (<1%) in either the outer membrane/cell wall associated fraction or the culture growth media (see section 5.2.5).

The enzyme was purified from the soluble/cytoplasmic fraction using a protocol that involved cell lysis, heat treatment, ammonium sulphate precipitation, DEAE ion exchange, and ultrafiltration chromatography. The non-denatured molecular mass of the pullulanase was determined to be 90 - 100 kDa by gel filtration, and the denatured molecular mass determined to be 95 kDa by SDS-PAGE.

The N-terminal amino acid sequence of the pullulanase expressed in *E. coli* was shown to be identical to the deduced amino acid sequence of the pullulan gene (section 5.4.1). Enzyme activity was shown to increase in the presence of Ca²⁺ and decrease in the presence of EDTA (section 5.4.2). The temperature optimum was determined to be 70°C and, at this temperature, the pH optimum was determined to be 5.7. Enzyme stability, as measured by half life, was 38 min at 80°C and 7 min at 90°C (section 5.4.3). Substrate characterisation showed that the enzyme had high activity on pullulan, producing only maltotriose, but only low activity on β-limit dextrin and amylopectin. No activity could be detected when amylose was used as a substrate.

Comparison of the pullulanase isolated from the native organism to the pullulanases expressed from *pulA* gene in *E. coli* showed some major differences (see section 9.3 for a detailed discussion).
Chapter Six

Inducible Enzyme Expression in *E. coli*

6.1 Introduction

In this chapter experiments relating to the overexpression of the pullulanase from *C. saccharolyticum* in *E. coli* are detailed. While the pullulanase gene was expressed in *E. coli* from its own promoter without substrate regulation, the level of enzymatic activity was low and hence the use of overexpression plasmids was investigated.

6.2 Optimisation of Expression Levels

6.2.1 Expression plasmid construction

6.2.1.1 Expression plasmid choice

The pullulanase gene, *pulA*, was ligated into an inducible overexpression vector to increase the level of expression. Initially, attempts were made to ligate the *pulA* gene into a number of different inducible overexpression vectors. Due to the different multi-cloning sites of each vector, slightly different portions of the *pulA* gene were used. Inducible overexpression vectors used were pT7-7 (Tabor 1989), pNH18a (Stratagene) and pJLA602 (Schauder *et al.* 1987).

Due to the lack of suitable restriction enzyme sites, the *pulA* gene was first ligated into M13mp10, then removed by restriction enzyme digests (Eco RI/Bam HI) and ligated into the Bam HI/Eco RI site of pJLA602 (see figure 6.1). After many attempts, a pullulanase-positive recombinant with an insert of the correct size was obtained using pJLA602.

For the other vectors all the ligation and transformation controls were positive, but although a number of transformants were detected, none were pullulanase-positive and produced the correct restriction endonuclease digestion pattern. A possible explanation for this phenomenon was a "leaky" expression system (i.e gene expression was not completely repressed). It has previously been shown that expression of the pullulanase gene was deleterious to the health of the *E. coli* cell, with this effect increasing with the level of expression (see section 5.2.4). It was later shown by other members of our group that the pT7-7 vector was not fully repressible and expression of deleterious enzymes affected the growth of the *E. coli* cell (personal communication, G. Munro, Cell Biology Department, Auckland University).
Figure 6.1 Construction of the pullulanase gene in pJLA602

Inducible Enzyme Expression in *E. coli*
6.2.1.2 Structure of pJLA602

The pJLA602 inducible overexpression vector was derived from the vector pBR322 (Bolivar et al. 1977) with the addition of \( \lambda \) sequences (major leftward and rightward promoters and clts857 repressor protein). Other sequences present included translational stop codons from the fd bacteriophage, a translation initiation region (atpE leader sequence) and a synthetic multi-cloning site. The location of these sequences is shown in table 6.1.

<table>
<thead>
<tr>
<th>Sequence location pJLA602</th>
<th>DNA Source</th>
<th>function</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-1147</td>
<td>( \lambda (36967-38107) )</td>
<td>( \lambda ) clts857 and promoter ( p_R )</td>
</tr>
<tr>
<td>1148-1222</td>
<td>E. coli ori C</td>
<td></td>
</tr>
<tr>
<td>1223-1471</td>
<td>( \lambda (rev\ 35715-35467) )</td>
<td>promoter ( p_L )</td>
</tr>
<tr>
<td>1480-1527</td>
<td>atpE leader sequence (968-1020)</td>
<td>translation initiation region</td>
</tr>
<tr>
<td>1528-1586 multi cloning</td>
<td>synthetic</td>
<td>multi cloning site</td>
</tr>
<tr>
<td>1587-1921</td>
<td>pBR322 (651-974)</td>
<td></td>
</tr>
<tr>
<td>1922-1996</td>
<td>M13mp7 (1573-1649)</td>
<td></td>
</tr>
<tr>
<td>1997-2046</td>
<td>fd (1524-1573)</td>
<td>fd transcription terminator</td>
</tr>
<tr>
<td>2047-2144</td>
<td>M13mp7 (1425-1524)</td>
<td></td>
</tr>
<tr>
<td>2145-2597</td>
<td>pBR322 (974-1429)</td>
<td>Origin of replication</td>
</tr>
<tr>
<td>2599-4891</td>
<td>pBR322 (2068-4361)</td>
<td>amp resistance</td>
</tr>
</tbody>
</table>

The major features of the expression vector pJLA602 were the presence of the two major \( \lambda \) promoters \( \lambda p_R \lambda p_L \) in tandem and a strong RNA polymerase terminator (fd). Expression was regulated by the \( \lambda \) repressor protein (clts857) which binds to the \( \lambda o_R \lambda o_L \) operators near the \( \lambda p_R \lambda p_L \) promoter, blocking transcription. Induction was achieved by increasing the temperature to 42\(^\circ\)C, denaturing the temperature-sensitive clts857 repressor protein. Translation was initiated from the atpE leader Shine-Dalgarno sequence which was adjacent to a synthetic polylinker sequence containing three restriction enzyme sites (Nco I, Nde I, Sph I), each of whose recognition sequences contained an ATG start codon.
6.2.2 Optimisation of induction

The standard induction procedure was growth at 28°C to produce sufficient cell mass, followed by heat induction at 42°C. Induction tests were performed to maximise the level of induction and hence enzyme expression levels. Heat induction at 42°C was performed at three different cell growth phases (early log phase, mid log phase and late log phase) and induction at late log phase was also performed at 37°C and a combined induction at 37°C with an initial 42°C heat step. For all induction experiments, cells were grown at 28°C pre-induction.

6.2.2.1 28 → 42 heat shift induction

Cells were grown in L-broth (with 50 μg/ml ampicillin) at 28°C to three different growth phases then induced at 42°C. The growth phases used were early log phase (OD₆₀₀ = 0.08), mid-log phase (OD₆₀₀ = 0.4) and late log phase (OD₆₀₀ = 1.4). The method used for each growth phase was identical. At various time points (0, 15, 30, 60, 120 and 240 minutes) 5 ml portions were removed and analysed with respect to cell growth and pullulanase activity. The number of cells surviving at each time point was determined by serial dilution (see section 2.2.3.2) followed by plating on to L-plates (50 μg/ml ampicillin). Cell growth was also measured as ABS OD₆₀₀ (see section 2.2.3.1) and cells were examined under a phase contrast microscope to determine their morphology. The results are shown graphically in figure 6.2.

Pullulanase activity was measured by the standard reducing sugar assay, as in section 2.6.1.2, for each time point. Activity was measured at two levels of purification, total cell extract (sonicated whole cell extract) and extractable soluble protein (heat treated supernatant). To determine whether the increased activity at later time points was due to increased cell mass rather than increased induction levels, activity per absorbance was also calculated. These results are also shown graphically in figure 6.2.

Cell growth, as measured by OD₆₀₀, increased with time of induction, while cell viability decreased. The level of pullulanase expression increased with increasing cell mass and induction time, except in late log phase where a maximum was reached between 60 and 120 min, and after this time activity decreased. The increased level of expression in late log phase cannot be accounted for by a greater cell mass alone since the ratio of activity per cell absorbance was greater in late log phase than mid log phase which in turn was higher than early log phase. During the 4 hour induction, cell viability decreased by approximately one order of magnitude as measured by CFU on L-plates (with 50 μg/ml ampicillin). These results show that maximum induction was obtained by inducing late log phase cells, hence for future experiments only late log phase cells were used.
Figure 6.2 Cell growth and pullulanase activity for 28 → 42 heat shift induction

Early log phase induction

Mid log phase induction

Late log phase induction
6.2.2.2 28 → 37 heat shift Induction

At growth temperatures above 37°C E. coli expresses a group of proteins known as "heat shock" proteins (including a number of proteases). To reduce the synthesis of "heat shock" proteins, expression studies were performed at the lower temperature of 37°C since the clts857 repressor protein is denatured at 37°C. A similar method to 6.2.2.1 was used with an OD₆₀₀ = 0.8 and an induction temperature of 37°C. The results are shown in figure 6.3.

Figure 6.3 Cell growth and pullulanase activity for 28 → 37 heat shift induction

The overall trend was similar to that obtained in section 6.2.2.1, but enzyme production levels were lower, implying that clts857 was only partially denatured at 37°C.

6.2.2.3 28 → 42 → 37 heat shift induction

As stated in 6.2.2.2 E. coli grown at high temperatures induces "heat shock" protein synthesis. Although the E. coli cells grew faster and were more viable at 37°C than at 42°C, the clts857 repressor protein appeared to be only partially denatured at 37°C. Hence, a combined method was used consisting of 42°C heat shock (to completely denature the clts857 repressor protein) followed by growth at 37°C.

A similar procedure was used except that cells were first induced at 42°C (for 0, 5, 15, 30 or 60 minutes) followed by continued induction at 37°C until total induction time was 2 hours. The results are shown in figure 6.4.

Inducible Enzyme Expression in E. coli
Figure 6.4 Cell growth and pullulanase activity for $28 \rightarrow 42 \rightarrow 37$ heat shift induction

The overall trend was again similar, with enzyme production levels lower than $28 \rightarrow 42$ heat shift induction (section 6.2.2.1) but higher than $28 \rightarrow 37$ heat shift induction.

6.2.2.4 Alkaline pH shift induction

Due to the problems experienced with the low yields obtained from the heat induced system a modified alkaline pH shift induction method was used (Poindexter and Gayle 1991). Cells were grown at $28^\circ C$ to stationary phase ($OD_{600} = 4.0$) and NaOH added to shift the pH from 6.5 to 9.0. The effect of the pH change was twofold, neutralising the organic acids inhibiting cell growth and denaturing the clts857 gene product. As the cells continued to grow the pH decreased, hence more NaOH was added to maintain the pH at 9.0.

Poindexter and Gayle found this alkaline shift produced 15 times more protein than heat induction. Although repeated twice, the maximum level of induction achieved was only 1% of that obtainable from heat induction, although the cell density did continue to increase with the addition of NaOH. Possible reasons for this unexpected low level of induction were the use of a different plasmid, and the high cellular pH affecting the stability or folding of the enzyme. In addition, an automated fermenter was unavailable, requiring the hand monitoring of pH and addition of NaOH.

6.2.2.5 Optimal induction procedure

Analysis of the results obtained in sections 6.2.2.1-4 showed that the optimal induction was obtained using late log phase cells induced at $42^\circ C$ for 1-2 hours. In contrast to the results obtained in
section 5.2, the majority of the enzyme activity remained in the cellular pellet and hence could not be purified by the standard heat purification procedure.

6.2.3 Enzyme location within the cell

The location of pullulanase produced from heat inducible expression in *E. coli* was determined as in section 2.5.1.4. As enzyme distribution could vary depending on the time of induction, 6 different time points were used (0, 15, 30, 60, 120 and 240 minutes) 5 ml portions were removed and fractionated. The results are shown in figure 6.5.

**Figure 6.5** The effect of induction time on the distribution of enzyme activity within different cellular fractions

![Diagram showing enzyme distribution](image)

The fraction positions are as obtained in section 2.5.1.4 and are shown graphically in figure 2.2

- Fraction / Position 3 = Enzyme secreted into periplasm
- Fraction / Position 2 = Enzyme insoluble or membrane associated but extractable with NaCl/ Triton X-100
- Fraction / Position 1 = Enzyme soluble in the cytoplasm
- Fraction / Position 0 = Enzyme insoluble or membrane associated but non-extractable with NaCl/ Triton X-100

This diagram shows that the location of the pullulanase was predominantly cytoplasmic (approximately 15% soluble, 22.5% insoluble and 55% non-extractable) with some exported to the periplasmic space (approximately 7.5%). No enzyme was detected (<1%) in either the outer membrane/cell wall associated fraction or the culture growth media. The pullulanase was exported out of the cells in the native organism but remained cytoplasmic when expressed in *E. coli*. The overall trend was an increase in the insoluble cytoplasmic fractions (0, 2) and a decrease in the soluble cytoplasmic fraction (1) corresponding to higher expression levels.
6.2.4 Microscopic examination of heat induced cultures

Phase contrast light microscopy was used to examine the *E. coli* cells containing the recombinant plasmid at different times of heat induction. Photographs were taken using a Leitz Diaplan microscope. A selection of photographs are shown over the page. These photographs indicate that as pullulanase expression increased during heat induction, cell division was inhibited causing the *E. coli* to remain as chains of 3-6+ cells.

Figure 6.6 Photomicrographs showing the effect of pullulanase expression on *E. coli* cells

Non induced linear phase *E. coli* containing pJLA602 without the pullulanase gene. Magnification X1600 Phase contrast

Induced (2hr at 42°C) linear phase *E. coli* containing pNZ1507 (pJLA602 + pull) expressing the pullulanase gene. Magnification X1600 Phase contrast

Induced (2hr at 42°C) linear phase *E. coli* containing pNZ1507 (pJLA602 + pull/A) expressing the pullulanase gene. Magnification X8750 Phase contrast
The distribution of each cell type was determined at various heat induction time points. Cells were grown to late log phase (OD₆₀₀ = 1.0), induced as per section 6.2.2.5 for various time points, diluted and counted using a phase contrast light microscope. For counting purposes, duplicate slides were made for each plasmid and 100 cells counted (multimeric *E. coli* were counted as only one cell) and scored according to their phenotype (see figure 6.7).

**Figure 6.7** The effect of induction time on the distribution of cell multiplicity

![Figure 6.7: The effect of induction time on the distribution of cell multiplicity](image)

where \( \# \text{ mer} = \# \text{ full length cells joined} \)

As shown in figure 6.7 there was a correlation between enzyme expression level and the formation of multimeric elongated cells. This result implies that pullulanase expression was responsible for this phenomenon. A similar result was obtained using the other plasmids expressing pullulanase off their own promoters (see section 5.2.4).

Small refractile inclusion bodies were also detected in induced cells under the phase contrast microscope. These inclusion bodies probably consisted of insoluble pullulanase protein which had been precipitated in the cell forming aggregates.

*Inducible Enzyme Expression in E. coli*
6.3 Attempted Purification of Overexpressed Enzyme

Initial attempts to purify the pullulanase produced from the overexpression plasmid, used a similar strategy to that used to purify the pullulanase expressed in *E. coli* (see section 5.3).

Cell lysis was followed by heat treatment to denature the thermolabile *E. coli* proteins (see section 2.5.1.2), which then precipitated and were removed by centrifugation (4°C, 13 000 g, 30 min). The heat stable soluble proteins, including the pullulanase, remained in the supernatant.

The conditions for maximum pullulanase expression for pNZ1507 were determined to be a two hour induction at 42°C using cells that had previously been grown to late-log phase at 28°C (see section 6.2.2.5). These conditions produced a maximum enzymatic activity of 264 Units per litre (μmoles maltotriose / min / litre of culture) for a sample volume of 5 ml. After heat purification only 89 Units per litre remained in the supernatant (see figure 6.2). For purification a larger culture volume of 500 ml was used and induced as before. After cell lysis the enzymatic activity was determined to be 79 Units per litre and after heat purification 14 Units per litre remained in the supernatant. This reduction could be due to a number of conditions affecting the rapid heating from 28°C to 42°C and vigorous agitation required for maximum induction. As the volume of culture increased, the rate of heating (using a water bath) decreased, and this could have affected the level of induction. Also, for larger volumes, the rate of agitation was decreased.

After heat purification, the soluble pullulanase supernatant was further purified using ammonium sulphate precipitation (see section 2.5.1.3). Maximum pullulanase activity was recovered in the same fraction as found in section 5.3.2 (50 - 70% saturation). Following the method described in section 5.3.3 this ammonium sulphate fraction was deionised, applied to a DEAE Sephacel column (see section 2.5.4.2) and eluted using a 0-1000 mM NaCl gradient. However, only approximately 5% of the total pullulanase activity could be eluted (at approximately 200 mM) from the column, even at higher salt concentrations (up to 2000 mM). The protein that bound to the column effectively blocked the column, increasing back pressure and requiring the column to be repacked.

The following modifications were performed to the pullulanase purification scheme in an attempt to recover more enzyme. A different enzyme induction method (constant 37°C) was used with no resulting improvement. The pullulanase was also extracted using high concentrations of urea (4-8 M) (Chaffotte *et al.* 1992). While solubilising the enzyme, the urea reduced the effectiveness of the heat purification step by also solubilising the *E. coli* proteins. On removal of the urea the problems of pullulanase retention to the column returned.

As the amount of enzyme recovered was substantially less than that obtained for the enzyme expressed from its own promoter in *E. coli* (see section 5.3), and as the enzymatic characterisation showed that both enzymes were essentially the same, no further purification work was performed using the pullulanase overexpressed in *E. coli*. 

Chapter 6
6.4 Enzyme Characterisation

Although the pullulanase was not purified to homogeneity some enzyme characterisations were performed on fractions obtained during the purification procedure. These results were nearly identical to those obtained for the pullulanase expressed in *E. coli* from its own promoter. The results are summarised in Table 6.2.

Table 6.2 Summary of enzyme characterisations of the pullulanase obtained from overexpression in *E. coli*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH optimum</td>
<td>5.7</td>
</tr>
<tr>
<td>Temperature optimum</td>
<td>70°C</td>
</tr>
<tr>
<td>Ca²⁺ increased enzyme activity</td>
<td></td>
</tr>
<tr>
<td>EDTA decreased enzyme activity</td>
<td></td>
</tr>
<tr>
<td>Thermostability was determined as half-life of 27 min at 80°C and 3 min at 90°C</td>
<td></td>
</tr>
<tr>
<td>Complete hydrolysis of pullulan yielding maltotriose only</td>
<td></td>
</tr>
<tr>
<td>Low activity on β-limit dextrin and amylopectin</td>
<td></td>
</tr>
<tr>
<td>No activity on amylose</td>
<td></td>
</tr>
</tbody>
</table>

These results show that the enzymatic properties of the pullulanase expressed from the heat inducible plasmid pNZ1507 were essentially the same as the properties of the enzyme expressed from its own promoter in *E. coli* (see section 5.4). Also, the only major difference between this enzyme and that produced in *C. saccharolyticum* related to substrate characterisation, which is discussed in section 5.4.
6.5 Summary

The pullulanase gene from *C. saccharolyticum* was ligated into the heat inducible over expression vector pJLA602 (*clts857* repressor protein and *λpR* *λpL* expression promoters in tandem) (see section 6.2.1). Many different conditions of induction were used to determine the optimal conditions for maximum enzyme expression. Variables examined included time of induction, temperature, and the effect of pH on induction (section 6.2.2). The conditions for maximum pullulanase expression for pNZ1507 were determined to be a two hour induction at 42°C using cells that had previously been grown to late-log phase at 28°C (see section 6.2.2.5). These conditions produced a maximum enzymatic activity of 264 Units per litre (µmoles maltotriose / min / litre of culture) for a sample volume of 5 ml.

Cellular fractionation studies determined the location of the pullulanase to be predominantly cytoplasmic (approximately 15% soluble, 22.5% insoluble and 55% non-extractable) with some exported to the periplasmic space (approximately 7.5%). No enzyme was detected (<1%) in either the outer membrane/cell wall associated fraction or the culture growth media (see section 6.2.3). Examination of the cells under a light microscope showed that as pullulanase expression increased during heat induction, cell division was inhibited causing the *E. coli* to remain as chains of 3-6+ cells. These cells also contained small refractile inclusion bodies (see section 6.2.4).

While the concentration of the pullulanase expressed from the pJLA promoter was higher than that expressed from its own promoter in *E. coli*, subsequent purification problems resulted in enzyme loss. Since this reversed the situation no further purification work was performed using the pullulanase overexpressed in *E. coli*.

The enzymatic characterisation (section 6.4) of the pullulanase expressed from the heat inducible plasmid pNZ1507 was essentially the same as the characterisation of the enzyme expressed from its own promoter in *E. coli*. Substrate characterisation showed that the enzyme had high activity on pullulan producing only maltotriose, but only low activity on β-limit dextrin and amylopectin. No activity could be detected when amyllose was used as a substrate.

While the pJLA heat inducible overexpression plasmid pNZ1507 was an efficient method for the production of large quantities of pullulanase, the purification problems associated with the enzyme from this construct showed that higher levels of enzyme activity could be recovered from high copy number plasmids that expressed the pullulanase from its own promoter in *E. coli*. 

Chapter 6
Chapter Seven

Pullulanase Expression Studies using Yeast

7.1 Plasmid Construction

The *Caldocellum saccharolyticum* pullulanase was expressed in *Saccharomyces cerevisiae* as a comparison to expression in *E. coli* and the native organism. As the major potential application of the pullulanase is in the food industry (see section 1.2), and *S. cerevisiae* is classified as a food grade organism, this could also be an advantageous expression system for the commercial exploitation of the pullulanase. Other advantages of a yeast based expression system over an *E. coli* expression system include; enzyme glycosylation (Kumagai et al. 1990), enzyme export into the media and lower growth temperatures (this can be important for those thermophilic enzymes that are deleterious to the cell, as enzymatic activity is proportionately lower) (Botstein and Fink 1988, Ratner et al. 1989).

Of the yeast-*E. coli* shuttle vectors available in our laboratory, only two had appropriate restriction endonuclease sites, these were pMV2Adel (Wong et al. 1988) and pYELC5 (Macreadie 1990). The construction of pNZ1508 for expression of the pullulanase gene in yeast is shown in figure 7.1. Ligations of each construct were performed (see section 2.3.6.1), transformed into *E. coli* (see section 2.3.6.2), and plasmid DNA isolated (see section 2.3.3). Only the pYELC5 vector produced transformants that contained a plasmid of the correct size and restriction endonuclease digestion pattern (see section 2.3.4). All the ligation and transformation controls were positive for the pMV2Adel vector, but although a number of transformants were detected none contained plasmids with the correct restriction endonuclease digestion pattern.

The pYELC5 plasmid contained a Cu²⁺ inducible expression system (*CUP1* gene, promoter and two upstream activation sequences) which produced a gene product (metallothionein) that conferred Cu²⁺ resistance. However, when foreign genes were ligated into the middle of this gene, the resulting gene fusion no longer mediated copper resistance, although it remained Cu²⁺ inducible. Hence expression in a Cu²⁺ resistant yeast host was required. Of the yeast strains available in our laboratory only one, PB5321, was Cu²⁺ resistant (his3-11, 15; leu 2-3, leu 2-112; *CUP1*, Can1). This host also contained a mutation which stopped the synthesis of the amino acid leucine, but was complemented by the plasmid LEU2 gene present in pNZ1508. Thus only strains containing pNZ1508 did not require the addition of leucine for growth.

Plasmid pNZ1508 was transformed into yeast strain PB5321 and plated onto SD plates supplemented with histidine (see section 2.3.12.1). Again, growth without the addition of leucine was used as a selective marker to indicate recombinants that contained the plasmid pNZ1508. Three recombinants, which appeared to be identical (by restriction endonuclease digestion pattern), were thus selected for expression studies (producing Strain PB4977).
Figure 7.1 Construction of pNZ1508

M13mp10
7250 bp

Bam HI 6253
Sma I 6247
Pvu II 6054

Sca I

pul A gene
2.6 kb fragment
from pNZ1038

Sca I

Sca I / Sca I fragment ligated into the Sma I site of the M13mp10 giving pNZ1505

Bam HI 8921

pNZ1505
9928 bp

Pvu II 6054

Fragment isolated by Pvu II / Bam HI digestion and ligated into Bam HI / Pvu II site of pYELC5 forming pNZ1508

pYELC5
6.20 Kb

CUP18

on 2μ

LEU2

Amp

pNZ1508
8.80 Kb

CUP18

on 2μ

LEU2

Amp

Pvu II

Sal I/Pst I

Pullulanase Expression Studies using Yeast
7.2 Enzyme Induction and Expression Studies

The standard induction procedure for the expression of the \textit{CUP1B} promoter of plasmid pYELC5 is as follows:

Grow in rich media until late log phase, induce expression by the addition of 1 mM Cu\textsuperscript{2+} and out grow for a further 4 hours (Macreadie \textit{et al.} 1989).

Originally rich media was used but as no selective markers were present the yeast cells lost the plasmid during growth. Therefore, in later experiments, minimal media (SD with histidine) was used.

To test the effect of different Cu\textsuperscript{2+} concentrations on induction, and hence expression of the pullulanase, a time course (ranging from 30 - 240 min) was performed using different concentrations (ranging from 0.001-10 mM). The culture medium and the cellular extract (see section 2.5.3) were both tested for pullulanase activity by the dye-linked substrates method (see section 2.6.1.4). Activity was only detected in the cellular extract as shown in figure 7.2.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure72.png}
\caption{The effect of Cu\textsuperscript{2+} concentration on pullulanase activity}
\end{figure}

Where 100\% activity = 0.05 ± 0.02 µg maltotriose / min / ml of culture

or 0.1 Units/ l (µmoles maltotriose / min / litre of culture)
The conditions of maximum induction were determined to be similar to the standard conditions (i.e. 1 mM Cu²⁺ and 4 h induction). Using induction times greater than 4 hours decreased enzyme activity slightly. No more characterisations were performed due to the very low levels of enzyme expression obtained and that the amounts were insufficient for purification.

7.3 Plasmid Stability Tests

As the enzyme expression levels were lower than expected, and the plasmid construction appeared to be unstable, the following tests were performed to show that the plasmid was still present in the cells following induction. The standard method was plasmid isolation (see section 2.3.12.2), retransformation into E. coli (see section 2.3.6.2), re-isolation (see section 2.3.2) and restriction endonuclease digestion pattern analysis. The major problems with this method were the time taken and the indirect method of analysis. A PCR-based direct analysis method was also performed (see section 2.3.11). Both of these tests showed that the plasmid was still present after induction, although the concentration was impossible to determine.

7.4 Summary

Although the pNZ1508 expressed the pullulanase and was inducible with Cu²⁺, the level of expression was very low. Also, as the enzyme was not exported out of the cell it was an inappropriate expression system. Other bacterial starch hydrolysis enzymes have been expressed and exported efficiently, including a Bacillus α-amylase gene (Pretorius et al. 1988). Therefore the low level of expression could be due to the expression plasmid used rather than an intrinsic problem associated with the expression of the pulA gene in yeast.

After this work was completed other expression vectors became available in our laboratory including the pFlagU2 expression vector. Expression from this vector is controlled by ADH promoter which is repressed in the presence of glucose. On induction a gene fusion is formed between the protein, the α-factor leader sequence (for export out of the cell) and the FLAG protein (for protein purification). Work is currently underway using this vector for pullulanase expression in S. cerevisiae.
Chapter Eight

Analysis of the Deduced Peptide Sequence of the *C. saccharolyticum* Pullulanase

8.1 Introduction

Once the nucleotide sequence of the *C. saccharolyticum* pullulanase was determined it was possible to deduce the amino acid sequence of the enzyme. It then became possible to analyse the peptide sequence, and perform comparisons between the peptide sequence of the *C. saccharolyticum* pullulanase and other related sequences. The pullulanase sequence was also analysed using the protein structure prediction program available with the GCG package. Attributes predicted included hydrophilicity, flexibility, glycosylation sites and secondary structural elements (α-helices, β-sheets, and turns). The graphical output from the program PEPTIDESTRUCTURE is shown in figure 8.1

---

**Figure 8.1 Analysis of the peptide structure of the *C. saccharolyticum* pullulanase**

Note that the accuracy of this prediction method varies considerably, with hydrophilicity being the most accurate attribute and glycosylation site being the least accurate. Prediction of the secondary structural elements has been the most studied but remains very inaccurate.
8.2 Comparison of the Putative Peptide Sequence of the *C. saccharolyticum* Pullulanase to Other Related Enzymes

The deduced peptide sequence of the *pulA* gene was compared to the sequences of a number of related enzymes, listed in table 8.1.

### Table 8.1 Pullulanases and other related enzymes used

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pullulanase</td>
<td><em>Caldocellum saccharolyticum</em></td>
</tr>
<tr>
<td>pullulanase</td>
<td><em>Bacillus stearothermophilus</em> TRS128</td>
</tr>
<tr>
<td>pullulanase</td>
<td><em>Thermus</em> sp. AMD33</td>
</tr>
<tr>
<td>pullulanase</td>
<td><em>Klebsiella oxytoca</em> UNF5023</td>
</tr>
<tr>
<td>isoamylase</td>
<td><em>Pseudomonas</em> sp. SMP1</td>
</tr>
<tr>
<td>neopullulanase</td>
<td><em>Bacillus stearothermophilus</em> TRS40</td>
</tr>
<tr>
<td>α-amylase-pullulanase</td>
<td><em>Clostridium thermohydrosulfuricum</em></td>
</tr>
<tr>
<td>α-amylase-pullulanase</td>
<td><em>Clostridium thermosulfurogenes</em></td>
</tr>
<tr>
<td>α-1,6-Branching enzyme</td>
<td><em>E. coli</em></td>
</tr>
<tr>
<td>oligo-1,6-glucosidase</td>
<td><em>Bacillus thermoglucosidasius</em> KP1006</td>
</tr>
<tr>
<td>cyclodextrin glycosyltransferase</td>
<td><em>Bacillus circulans</em></td>
</tr>
<tr>
<td>α-amyrase</td>
<td>Pig pancreatic α-amyrase</td>
</tr>
</tbody>
</table>

The enzymes selected were every known example of enzyme that hydrolysed pullulan or debranched amylopectin (4 pullulanases, 2 α-amylase-pullulanases, 1 neopullulanase, 1 isoamylase) and four enzymes that were representative examples of related enzymes (α-1,6-branching enzyme, oligo-1,6-glucosidase, cyclodextrin glycosyltransferase and α-amyrase). Other sequences were also compared but showed no significant homology (e.g. β-amylase and glucoamylase) and were therefore not included. Where possible the representative sequences of the related enzymes were thermophilic eubacteria, except for the α-amylase sequence where the pig pancreatic sequence was used as this enzyme had been extensively characterised with respect to its 3-dimensional structure and amino acids involved in catalysis, substrate binding and metal ion binding.

The individual comparisons are shown in appendix 3 (figures A3.1 - A3.11). For each enzyme the comparison is shown in two forms, an amino acid sequence alignment produced by the GCG program BESTFIT (see section 2.7.4) and a graphical representation of homology produced by the GCG program COMPARE (see section 2.7.3), which compares each amino acid in the *C. saccharolyticum* pullulanase to the entire length of the other sequences. Amino acid comparisons
The percentage similarities and percentage identities obtained are listed in Table 8.2.

Table 8.2 Percentage homology between amino acid sequences of pullulanases and related enzymes

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>BSTPULA</th>
<th>THIPUL</th>
<th>KLEBPULA</th>
<th>PSEIAMA</th>
<th>BSTNPL</th>
<th>CTHAPUX</th>
<th>CTSPULA</th>
<th>ECRA</th>
<th>O16GLU</th>
<th>CGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSTPULA</td>
<td>59.4 (38.6)</td>
<td>55.6 (31.0)</td>
<td>51.0 (28.7)</td>
<td>47.9 (24.1)</td>
<td>48.6 (22.6)</td>
<td>45.7 (20.1)</td>
<td>46.5 (20.2)</td>
<td>49.5 (21.4)</td>
<td>45.1 (17.3)</td>
<td>41.4 (17.4)</td>
</tr>
<tr>
<td>THIPUL</td>
<td>55.6 (31.0)</td>
<td>70.7 (56.3)</td>
<td>52.4 (32.1)</td>
<td>44.9 (23.9)</td>
<td>44.6 (19.6)</td>
<td>44.8 (20.7)</td>
<td>45.2 (18.0)</td>
<td>47.2 (18.1)</td>
<td>45.6 (18.3)</td>
<td>35.8 (16.4)</td>
</tr>
<tr>
<td>KLEBPULA</td>
<td>51.0 (28.7)</td>
<td>52.4 (32.1)</td>
<td>51.4 (28.1)</td>
<td>46.9 (21.6)</td>
<td>45.6 (18.3)</td>
<td>44.1 (18.9)</td>
<td>45.6 (19.0)</td>
<td>45.2 (19.6)</td>
<td>44.4 (17.2)</td>
<td>40.6 (16.4)</td>
</tr>
<tr>
<td>PSEIAMA</td>
<td>47.9 (24.1)</td>
<td>44.9 (23.9)</td>
<td>46.9 (24.9)</td>
<td>46.9 (24.9)</td>
<td>47.7 (18.8)</td>
<td>42.4 (17.2)</td>
<td>39.7 (15.6)</td>
<td>49.8 (23.9)</td>
<td>44.4 (17.8)</td>
<td>44.9 (20.1)</td>
</tr>
<tr>
<td>BSTNPL</td>
<td>48.6 (22.6)</td>
<td>44.6 (19.6)</td>
<td>47.7 (22.4)</td>
<td>48.4 (22.4)</td>
<td>44.1 (18.9)</td>
<td>42.3 (17.2)</td>
<td>46.7 (21.4)</td>
<td>43.4 (20.1)</td>
<td>42.4 (17.8)</td>
<td>40.4 (16.4)</td>
</tr>
<tr>
<td>CTHAPUX</td>
<td>45.7 (20.1)</td>
<td>44.1 (18.9)</td>
<td>45.1 (20.2)</td>
<td>60.0 (35.3)</td>
<td>44.4 (17.8)</td>
<td>42.3 (17.2)</td>
<td>55.4 (32.5)</td>
<td>52.1 (22.7)</td>
<td>52.6 (29.2)</td>
<td>51.2 (23.1)</td>
</tr>
<tr>
<td>CTSPULA</td>
<td>46.5 (20.2)</td>
<td>45.2 (18.0)</td>
<td>46.7 (21.4)</td>
<td>82.9 (71.6)</td>
<td>49.8 (23.9)</td>
<td>43.4 (20.1)</td>
<td>55.4 (32.5)</td>
<td>52.1 (22.7)</td>
<td>52.6 (29.2)</td>
<td>51.2 (23.1)</td>
</tr>
<tr>
<td>ECRA</td>
<td>49.5 (21.4)</td>
<td>42.3 (17.2)</td>
<td>52.6 (29.2)</td>
<td>52.4 (21.5)</td>
<td>51.2 (23.1)</td>
<td>45.6 (22.7)</td>
<td>55.4 (32.5)</td>
<td>52.1 (22.7)</td>
<td>52.6 (29.2)</td>
<td>52.4 (21.5)</td>
</tr>
<tr>
<td>O16GLU</td>
<td>45.1 (17.3)</td>
<td>46.8 (23.9)</td>
<td>42.3 (17.2)</td>
<td>52.6 (29.2)</td>
<td>47.5 (23.7)</td>
<td>48.1 (26.0)</td>
<td>45.6 (22.7)</td>
<td>52.1 (22.7)</td>
<td>52.6 (29.2)</td>
<td>52.4 (21.5)</td>
</tr>
<tr>
<td>CGT</td>
<td>41.4 (17.4)</td>
<td>44.9 (20.1)</td>
<td>47.5 (23.7)</td>
<td>46.5 (24.5)</td>
<td>46.5 (24.5)</td>
<td>47.5 (23.7)</td>
<td>43.5 (21.5)</td>
<td>46.5 (24.5)</td>
<td>46.5 (24.5)</td>
<td>46.5 (24.5)</td>
</tr>
<tr>
<td>α-AMY</td>
<td>45.3 (21.0)</td>
<td>44.7 (18.2)</td>
<td>43.4 (19.3)</td>
<td>44.1 (18.9)</td>
<td>46.5 (24.5)</td>
<td>46.2 (24.5)</td>
<td>46.5 (24.5)</td>
<td>46.5 (24.5)</td>
<td>46.5 (24.5)</td>
<td>46.2 (24.5)</td>
</tr>
</tbody>
</table>

Values are similarity percentages obtained from the GCG sequence alignment program BESTFIT. Two values are shown for each enzyme comparison. The first value is the percentage similarity, which takes into account identical and similar amino acids, the value in brackets is the percentage identity and is obtained when only identical amino acids are counted.

Abbreviations for amino acid sequences are as follows:

- **CSPULA**: *Caldocellum saccharolyticum* pullulanase
- **BSTPULA**: *Bacillus stearothermophilus* pullulanase
- **THIPUL**: *Thermus* sp. AMD33 pullulanase
- **KLEBPULA**: *Klebsiella oxytoca* UNF5023
- **PSEIAMA**: *Pseudomonas* sp. SMP1 isoamylase
- **BSTNPL**: *Bacillus stearothermophilus* neopullulanase
- **CTHAPUX**: *Clostridium thermohydrodsulfuricum* α-amylose-pullulanase
- **CTSPULA**: *Clostridium thermosulfurogenes* α-amylose-pullulanase
- **ECRA**: *E. coli* α-1,6-Branching enzyme
- **O16GLU**: *Bacillus thermoglucosidasius* KP1006 oligo-1,6-glucosidase
- **CGT**: *Bacillus circulans* cyclodextrin glycosyltransferase
- **α-AMY**: Pig pancreatic α-amylose
Further comparisons were performed between each of the related enzymes using the GCG multi-alignment program PILEUP (see section 2.7.4). The output from this program is shown in figure 8.2 as a dendrogram. Note that this alignment is based on homology over the entire sequence length.

**Figure 8.2 A dendrogram of the alignment of the related sequences**

![Dendrogram of related sequences](image)

Abbreviations for amino acid sequences are as follows:
- CGT: *Caldocellum saccharolyticum* pullulanase
- BSTPULA: *Bacillus stearothermophilus* pullulanase
- THPULA: *Thermus* sp. AMD33 pullulanase
- KLEBPLA: *Klebsiella* sp. (both pneumonia and aerogenes) pullulanase
- PSEIAMA: *Pseudomonas* sp. SMPl isoamylase
- BSTNPL: *Bacillus stearothermophilus* neopullulanase
- CTAPUX: *Clostridium thermohydrodsulfuricum* α-amylase-pullulanase
- CTSPULA: *Clostridium thermosulfurogenes* α-amylase-pullulanase
- ECBRA: *E. coli* α-1,6-branching enzyme
- O16GLU: *Bacillus thermoglucosidasi* sp. KP1006 oligo-1,6-glucosidase
- COT: *Bacillus circularis* cyclodextrin glycosyltransferase
- α-AMY: Pig pancreatic α-amylase

Analysis of this dendrogram showed that the peptide sequences could be divided into two major groups. These groups correlated with enzyme hydrolysis pattern (predominantly α-1,4 or α-1,6 bonds). Within the α-1,6 grouping the three thermophilic pullulanases grouped together with their nearest “neighbour” being the mesophilic pullulanase from *Klebsiella* sp. The isoamylase sequence from *Pseudomonas* sp. was the next nearest “neighbour”. For the α-1,4 grouping the two *Clostridium* α-amylase pullulanases grouped together with the neopullulanase from *B. stearothermophilus*. The only exception to the α-1,4 / α-1,6 division was the oligo-1,6-glucosidase from *Bacillus thermoglucosidasi* sp. Note that for ease of comparison only one example of each of the related peptide sequences is shown, when other examples are added they group with their single representative sequence.

Analysis of the Deduced Peptide Sequence of the *C. saccharolyticum* Pullulanase
8.3 The Presence of Three Conserved Regions in Pullulanases

As shown in appendix 3 (figure A3.1-A3.11) all 12 of the related enzymes listed in table 8.1 showed sequences homologous to the three highly conserved α-amylase regions (100, 200 and 300), in the same order and spacing. These three regions were homologous to the three conserved regions found in α-amylases and have been shown to be involved in catalytic activity, substrate binding and metal ion binding (see appendix 5). A schematic representation of the alignment of each of these enzymes is shown in fig 8.3.

Figure 8.3 Schematic representation of the three highly conserved regions

<table>
<thead>
<tr>
<th>KLEBPULA</th>
<th>612</th>
<th>679</th>
<th>835</th>
<th>1090</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSPULA</td>
<td>456</td>
<td>520</td>
<td>630</td>
<td>825</td>
</tr>
<tr>
<td>THPULA</td>
<td>343</td>
<td>408</td>
<td>524</td>
<td>716</td>
</tr>
<tr>
<td>BSTPULA</td>
<td>286</td>
<td>360</td>
<td>473</td>
<td>669</td>
</tr>
<tr>
<td>PSEIAMLA</td>
<td>321</td>
<td>394</td>
<td>526</td>
<td>771</td>
</tr>
<tr>
<td>BSTNPL</td>
<td>248</td>
<td>323</td>
<td>418</td>
<td>588</td>
</tr>
<tr>
<td>CLOAPUX</td>
<td>523</td>
<td>562</td>
<td>700</td>
<td>1475</td>
</tr>
<tr>
<td>CLOPULA</td>
<td>525</td>
<td>564</td>
<td>702</td>
<td>1203</td>
</tr>
<tr>
<td>ECBRA</td>
<td>340</td>
<td>403</td>
<td>525</td>
<td>721</td>
</tr>
<tr>
<td>CGT</td>
<td>140</td>
<td>224</td>
<td>322</td>
<td>684</td>
</tr>
<tr>
<td>O16GLU</td>
<td>101</td>
<td>193</td>
<td>325</td>
<td>562</td>
</tr>
<tr>
<td>α-AMY</td>
<td>80-100</td>
<td>180-220</td>
<td>280-320</td>
<td>450-520</td>
</tr>
</tbody>
</table>

Boxed areas are conserved α-amylase regions (at approximate amino acid position of 100, 200, 300) implicated in enzymatic activity

KLEBPULA: Klebsiella sp. (both pneumonia and aerogenes) pullulanase
CSPULA: Caldocellum saccharolyticum pullulanase
THPULA: Thermus sp. AMD33 pullulanase
BSTPULA: Bacillus stearothermophilus pullulanase
PSEIAMLA: Pseudomonas sp. isoamylase
BSTNPL: Bacillus stearothermophilus neopullulanase
CTHAPUX: Clostridium thermohydrosulfuricum α-amylase-pullulanase
CTSPULA: Clostridium thermosulfurogenes α-amylase-pullulanase
CTG: Bacillus circulans cyclodextrin glycosyltransferase
O16GLU: Bacillus thermosaccharolyticus KP1006 oligo-1,6-glucosidase
ECBRA: E. coli α-1,6-branching enzyme
α-AMY: Highly conserved regions in both prokaryotic and eukaryotic α-amylases
Each peptide sequence possessed the three regions in the same order and spacing, with the location of the three regions within each peptide sequence being the only difference. As shown in section 8.5, these three regions were the basis of the major secondary structural element (an (α/β)8 barrel) which also contained the residues implicated in catalytic activity, substrate binding and metal ion binding. The amino acid sequences of these conserved regions are shown in figure 8.4.

**Figure 8.4** Amino acids found in the three highly conserved regions within pullulanases and other related enzymes

<table>
<thead>
<tr>
<th>BSTPULA</th>
<th>LKRAIHTLQNSNGIRUJMDVUVNHUYIRD 291</th>
</tr>
</thead>
<tbody>
<tr>
<td>THPULA</td>
<td>LKQAIHTLHENGRAVMDAQVNUYVSEPDXE 349</td>
</tr>
<tr>
<td>CSPULA</td>
<td>RATIKKMLHENG16DVUUVNTHYHTK 461</td>
</tr>
<tr>
<td>KLEBPULA</td>
<td>RTMIA1KDQDLMQVUVYVYTHNNAAG 617</td>
</tr>
<tr>
<td>PSEIAMA</td>
<td>QAMUIQAFHNAIKUYVUVUNHTAEGG 326</td>
</tr>
<tr>
<td>CLOAPUX</td>
<td>FEKLMEDAHAKGIKIDLGVENHTSDSI 529</td>
</tr>
<tr>
<td>CLOPULSA</td>
<td>FEKLMMDAHAKGIKIDLGVENHTSDSI 531</td>
</tr>
<tr>
<td>BSNPL</td>
<td>LKTLLIDRDLCKIRULDAVDNHCGYE 251</td>
</tr>
<tr>
<td>BAC16G</td>
<td>WDELLHEMHERNKLMLDUNVTSDH 108</td>
</tr>
<tr>
<td>ECBRA</td>
<td>FYFDIAAHRGLNUILOUWPAGFPTDD 345</td>
</tr>
<tr>
<td>CGT</td>
<td>FQNLTTAHAKGIKIDFAPNHTSPA 144</td>
</tr>
<tr>
<td>α-AMY</td>
<td>FRDMUTACNNVGVAVYVDANHMCNSG 106</td>
</tr>
<tr>
<td>BSTPULA</td>
<td>IDSURFWVEEHUNGRFRDLMGILDVETMAV 365</td>
</tr>
<tr>
<td>THPULA</td>
<td>VDSUFWAKEYIDGFRLMGUHDIETMAV 423</td>
</tr>
<tr>
<td>CSPULA</td>
<td>LQAIWTEFIDGFRLMGLLTLVQGQD 538</td>
</tr>
<tr>
<td>KLEBPULA</td>
<td>ADSLAVTTYKIDGFRLMGYHPKQISI 697</td>
</tr>
<tr>
<td>PSEIAMA</td>
<td>VDSLAYWANTMDGFRLASULGNSCLNGA 414</td>
</tr>
<tr>
<td>CLOAPUX</td>
<td>WLNDPGDK.NUGADGWRLVDNEVA ............... 635</td>
</tr>
<tr>
<td>CLOPULSA</td>
<td>WLNPDPGL.NDGADGWRLVDNEVA ............... 634</td>
</tr>
<tr>
<td>BSNPL</td>
<td>WI.R.EFIDGWLVDNEVAID .............. 334</td>
</tr>
<tr>
<td>ECBRA</td>
<td>VGNALYIEFQDIALVDUAVAS ... IY RDY 415</td>
</tr>
<tr>
<td>BAC16G</td>
<td>WL.EKFGGRDMLVUNIFSKEGLPTU 213</td>
</tr>
<tr>
<td>CGT</td>
<td>FKDRKILWLDGCMGDGIDUDAVHKMPL ........... 236</td>
</tr>
<tr>
<td>α-AMY</td>
<td>IADYNKLDIGUVRGFRDLASKHM .............. 202</td>
</tr>
</tbody>
</table>

Analysis of the Deduced Peptide Sequence of the *C. saccharolyticum* Pullulanase
<table>
<thead>
<tr>
<th>Abbreviations for amino acid sequences as follows:</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSPULA: Caldocellum saccharolyticum pullulanase</td>
</tr>
<tr>
<td>BSTPULA: Bacillus stearothermophilus pullulanase</td>
</tr>
<tr>
<td>THPULA: Thermus sp. AMD33 pullulanase</td>
</tr>
<tr>
<td>KLEBPULA: Klebsiella oxytoca UNIFS023</td>
</tr>
<tr>
<td>KLEBPULA: Pseudomonas sp. SMPI isoamylase</td>
</tr>
<tr>
<td>BSTNPL: Bacillus stearothermophilus neopolullanase</td>
</tr>
<tr>
<td>CLOAPUX: Clostridium thermosulfurogenes α-amylase-pullulanase</td>
</tr>
<tr>
<td>CLOAPULSA: Clostridium thermosulfurogenes α-amylase-pullulanase</td>
</tr>
<tr>
<td>ECBRA: E. coli α-1.6-branching enzyme</td>
</tr>
<tr>
<td>O16GLU: Bacillus thermoglucosidasis KP1006 oligo-1,6-glucosidase</td>
</tr>
<tr>
<td>CGT: Bacillus circulans cyclodextrin glycosyltransferase</td>
</tr>
<tr>
<td>α-AMY: Pig pancreatic α-amylase</td>
</tr>
</tbody>
</table>

These sequences show that within each of the three regions some amino acids were totally conserved, particularly those implicated in catalytic activity, substrate binding and metal ion binding. The residues that were not involved in catalytic activity, substrate binding or metal ion binding were not conserved and showed no obvious pattern. Mutagenesis studies on these conserved residues in the Bacillus stearothermophilus neopolullanase have shown them to be involved in enzyme activity (Kuriki et al. 1991). The overall pattern (some amino acid positions totally conserved and others variable) was similar to that obtained for the α-amylases, as shown in figure A5.1. Highly conserved regions implicated in enzymatic activity have also been found in other enzyme groups (Gilkes et al. 1991).

The homology lineups from figure 8.4 suggest that aspartic acid residues were at the catalytic site for all of these related enzymes. Ionisation of the side chain carboxyl group for free amino acids in solution has a pKa of 4.4. This value correlates well with the experimentally determined pH optimum of 5.7 for the pullulanase (see sections 3.4.3 and 5.4.3). Note that as the pH optima of all these related enzymes were also slightly acidic (i.e. around 5-6), aspartic acid residues were also likely to be found at the catalytic site.
Further comparisons were performed on each of the conserved regions individually, using the GCG multi-alignment program PILEUP. The output from these comparisons is shown in figure 8.5 as three dendrograms.

Figure 8.5 Dendrograms showing the alignment of the three conserved regions in pullulanases and related sequences

A. Alignment based on the 100 region

B. Alignment based on the 200 region

Analysis of the Deduced Peptide Sequence of the C. saccharolyticum Pullulanase
C. Alignment based on the 300 region

Abbreviations for amino acid sequences are as follows:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>KLEBPULA</td>
<td>Klebsiella (both pneumonia and aerogenes) pullulanase</td>
</tr>
<tr>
<td>CSPULA</td>
<td>Caldocellum saccharolyticum pullulanase</td>
</tr>
<tr>
<td>THPULA</td>
<td>Thermus sp. AMD33 pullulanase</td>
</tr>
<tr>
<td>BTPULA</td>
<td>Bacillus stearothermophilus pullulanase</td>
</tr>
<tr>
<td>PSEIAMA</td>
<td>Pseudomonas sp. isoamylase</td>
</tr>
<tr>
<td>CTSPULUX</td>
<td>Clostridium thermohydrosulfuricum α-amylase-pullulanase</td>
</tr>
<tr>
<td>BSTPNL</td>
<td>Clostridium thermosulfurogenes α-amylase-pullulanase</td>
</tr>
<tr>
<td>α-AMY</td>
<td>Highly conserved regions in both prokaryotic and eukaryotic α-amylases</td>
</tr>
<tr>
<td>CGT</td>
<td>Bacillus circulans cyclodextrin glycosyltransferase</td>
</tr>
<tr>
<td>O16GLU</td>
<td>Bacillus thermopsis KP1006 oligo-1,6-glucosidoase</td>
</tr>
<tr>
<td>ECBRA</td>
<td>E. coli α-1,6-branching enzyme</td>
</tr>
</tbody>
</table>

Analysis of these dendrograms shows that the peptide sequences formed three distinct groups, correlating with enzyme hydrolysis pattern. Within the pullulanase cluster the four pullulanases grouped together with their order depending on which conserved region was used. The isoamylase sequence from Pseudomonas sp. was the next nearest "neighbour" to the pullulanase sequences for two out of three of the conserved regions. The two Clostridium α-amylase-pullulanases grouped together with the neopullulanase from B. stearothermophilus in all cases. The third group consisted of the related enzymes (α-1,6-branching enzyme, oligo-1,6-glucosidase, cyclodextrin glycosyltransferase and α-amylase) whose grouping followed no distinct pattern and changed depending on the conserved region used. Note that for ease of comparison only one example of each of the related peptides is shown. When other examples are added they group with their single representative sequence.
8.4 Comparison of the Peptide Sequence of the C. saccharolyticum Pullulanase to other Pullulanases

An extensive alignment of the four pullulanase sequences to each other was performed using the GCG multi-alignment program PILEUP. The results are shown in figure 8.6.

This alignment shows significant homology between the amino acid sequences of the four pullulanases, especially around the three highly conserved regions, 100, 200 and 300. These regions were homologous to the α-amylase regions that are involved in catalytic activity, substrate binding and metal ion binding. Other highly conserved regions were also detected. They are labelled from the N-terminal end as A-F.

Previous workers have analysed the amino acid sequences of similar enzymes to discover whether a pattern existed that could explain thermostability (see section 1.1.4). Argos, Zuber and co-workers (see section 1.1.4.3) determined which amino acid residues were different over the entire protein of a number of highly conserved enzymes from both mesophilic and thermophilic sources. They then formulated some general rules regarding observed amino acid replacement for enhanced protein thermostability. Using regression analysis, Suzuki and co-workers suggested that for certain enzyme groups, including pullulanases, the proline content increases with increasing thermostability (Suzuki et al. 1991).

Comparisons were performed between the three thermophilic pullulanase amino acid sequences and the mesophilic Klebsiella sp to investigate whether this pattern of amino acid replacement existed for the three thermophilic pullulanases. These comparisons showed that some of the amino acid substitutions between the thermophilic and mesophilic pullulanase sequences followed Argos' preferred substitution rules. However, the proline content of the C. saccharolyticum pullulanase (3.5%) appeared to be unrelated to its thermostability. In fact, the proline content percentage was lower than that found for the pullulanase from Klebsiella sp (4.7%).
<table>
<thead>
<tr>
<th>Figure 8.6 An alignment of the four pullulanase amino acid sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bs</strong></td>
</tr>
<tr>
<td>MTRYTRNALV</td>
</tr>
<tr>
<td>HKVEKLMPH</td>
</tr>
<tr>
<td>IAGNSAVYDS</td>
</tr>
<tr>
<td>MAPFTLVAS</td>
</tr>
<tr>
<td>LAPKSLFSGD</td>
</tr>
<tr>
<td>MAPKVSLV</td>
</tr>
<tr>
<td>DPKTEKAEBYI</td>
</tr>
<tr>
<td>CSFFASSWFS</td>
</tr>
<tr>
<td>TAYKVIQIF</td>
</tr>
<tr>
<td>TAQQDVYVVY</td>
</tr>
<tr>
<td>IDLDA</td>
</tr>
<tr>
<td>VVRDDRLGEN</td>
</tr>
<tr>
<td>YFDPADGLI</td>
</tr>
<tr>
<td>QVVLNDSL</td>
</tr>
<tr>
<td>HPEGSG</td>
</tr>
<tr>
<td>HPSG</td>
</tr>
<tr>
<td>DQSSG</td>
</tr>
<tr>
<td>RLDS</td>
</tr>
</tbody>
</table>

--- B ---
<table>
<thead>
<tr>
<th>Bs</th>
<th>AsGVThVhEl</th>
<th>MFDsFAGVDE</th>
<th>E</th>
<th>227</th>
</tr>
</thead>
<tbody>
<tr>
<td>Th</td>
<td>ElGVThVQlL</td>
<td>MFMDsFAGVDE</td>
<td>E</td>
<td>285</td>
</tr>
<tr>
<td>Cs</td>
<td>ElGVThVhLl</td>
<td>MFDsFAGVDE</td>
<td>E</td>
<td>398</td>
</tr>
<tr>
<td>Kl</td>
<td>AsGVThVhEl</td>
<td>MFDsFAGVDE</td>
<td>E</td>
<td>398</td>
</tr>
<tr>
<td>Bs</td>
<td>YNWNPLHYNa</td>
<td>PEGSYATDFPN</td>
<td>DPYARIRELk</td>
<td>265</td>
</tr>
<tr>
<td>Th</td>
<td>YNWNPLHLYA</td>
<td>PEGSYATDFPN</td>
<td>DPYARIRELk</td>
<td>323</td>
</tr>
<tr>
<td>Cs</td>
<td>YNWNPLHLYA</td>
<td>PEGSYATDFPN</td>
<td>DPYARIRELk</td>
<td>404</td>
</tr>
<tr>
<td>Kl</td>
<td>TlVdQdSdYn</td>
<td>PEGSYATDFPN</td>
<td>DPYARIRELk</td>
<td>592</td>
</tr>
<tr>
<td>Bs</td>
<td>RAIHTLQSN</td>
<td>EIVmDvLdVe</td>
<td>NHyYirdQ</td>
<td>233</td>
</tr>
<tr>
<td>Th</td>
<td>QAHTLhEn</td>
<td>GLRVVMDAV</td>
<td>NHyYiDrQ</td>
<td>358</td>
</tr>
<tr>
<td>Cs</td>
<td>TmIKLKLhN</td>
<td>GIVmDvLdVe</td>
<td>NHTYHTKg</td>
<td>473</td>
</tr>
<tr>
<td>Kl</td>
<td>TmLlqAxqKdl</td>
<td>GINvMvLdVe</td>
<td>NHTyNAGGDT</td>
<td>632</td>
</tr>
<tr>
<td>Bs</td>
<td>VEEHhNFGf</td>
<td>FDMgILdVe</td>
<td>TmKAaVREMId</td>
<td>300</td>
</tr>
<tr>
<td>Th</td>
<td>AKEyHIDgFf</td>
<td>FDMgILdVe</td>
<td>TmKAaVREMId</td>
<td>358</td>
</tr>
<tr>
<td>Cs</td>
<td>TecHtHIDGf</td>
<td>FDMgILdVe</td>
<td>TmKAaVREMId</td>
<td>473</td>
</tr>
<tr>
<td>Kl</td>
<td>TTHhIDGf</td>
<td>FDMgILdVe</td>
<td>TmKAaVREMId</td>
<td>627</td>
</tr>
<tr>
<td>Bs</td>
<td>EYWdL..PTp</td>
<td>LSSEoQKATMQ</td>
<td>NAEKLyPFgY</td>
<td>418</td>
</tr>
<tr>
<td>Th</td>
<td>EYWdL..PTL</td>
<td>FHRNIRRWRs</td>
<td>NAaKoLPfPAY</td>
<td>476</td>
</tr>
<tr>
<td>Cs</td>
<td>EYWdMGNStC</td>
<td>LVEEMaTIlS</td>
<td>SCHqGYSiGL</td>
<td>593</td>
</tr>
<tr>
<td>Kl</td>
<td>EYWdSNQSDR</td>
<td>FEAISQIlnK</td>
<td>GTG..IGT</td>
<td>748</td>
</tr>
<tr>
<td>Bs</td>
<td>GtStFqLPdRg</td>
<td>FALgDSDDRE</td>
<td>K</td>
<td>439</td>
</tr>
<tr>
<td>Th</td>
<td>GtStFqLPdRg</td>
<td>FALgDNPRGg</td>
<td>Q</td>
<td>407</td>
</tr>
<tr>
<td>Cs</td>
<td>GtDLDGYktG</td>
<td>YVHqGSLDvG</td>
<td>R</td>
<td>613</td>
</tr>
<tr>
<td>Kl</td>
<td>GgGFDsGDa</td>
<td>LRQNQGIGsG</td>
<td>AgvLPnelAs</td>
<td>788</td>
</tr>
<tr>
<td>Bs</td>
<td>VlVlAISGS</td>
<td>I</td>
<td>.GKKnellFL</td>
<td>456</td>
</tr>
<tr>
<td>Th</td>
<td>VlVlLlAUGs</td>
<td>I</td>
<td>.RALGlLFC</td>
<td>514</td>
</tr>
<tr>
<td>Cs</td>
<td>LlQGIKAA</td>
<td>I</td>
<td>.DDFAK</td>
<td>627</td>
</tr>
<tr>
<td>Kl</td>
<td>DlTRLGmAGN</td>
<td>LADFvMDKd</td>
<td>GAAKKGsE</td>
<td>828</td>
</tr>
<tr>
<td>Bs</td>
<td>HFTQTINYVE</td>
<td>SHDNHtFdWk</td>
<td>MEIAngHsE</td>
<td>495</td>
</tr>
<tr>
<td>Th</td>
<td>hPROSINyVe</td>
<td>CHDNHtFdWk</td>
<td>MEANHDEPE</td>
<td>553</td>
</tr>
<tr>
<td>Cs</td>
<td>EPDECVNYvS</td>
<td>CHDNLtLdWk</td>
<td>LQKTVGdE</td>
<td>666</td>
</tr>
<tr>
<td>Kl</td>
<td>DpTEVvNYvS</td>
<td>CHDNLtLdWk</td>
<td>IYSvKASQeAd</td>
<td>868</td>
</tr>
<tr>
<td>Bs</td>
<td>AtAVlLsSQG</td>
<td>IFELHsCSOcE</td>
<td>YRTKQGIDNS</td>
<td>535</td>
</tr>
<tr>
<td>Th</td>
<td>AtAVlLsSQG</td>
<td>IFELHsCSOcE</td>
<td>YRTKQGIDNS</td>
<td>593</td>
</tr>
<tr>
<td>Cs</td>
<td>AtnALVTSQG</td>
<td>VAFELHGCVE</td>
<td>nRStKgGHPNT</td>
<td>706</td>
</tr>
<tr>
<td>Kl</td>
<td>SLAVMLGOC</td>
<td>IAfDQCGsG</td>
<td>LRSsKFSFTrs</td>
<td>908</td>
</tr>
<tr>
<td>Bs</td>
<td>DWhKESts</td>
<td>I</td>
<td>.E</td>
<td>54</td>
</tr>
<tr>
<td>Bs</td>
<td>IE</td>
<td>D</td>
<td>VRY</td>
<td>VQG</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Th</td>
<td>YE</td>
<td>B</td>
<td>D</td>
<td>S</td>
</tr>
<tr>
<td>Cs</td>
<td>FF</td>
<td>RT</td>
<td>FK</td>
<td>F</td>
</tr>
<tr>
<td>Kl</td>
<td>LK</td>
<td>Q</td>
<td>MT</td>
<td>A</td>
</tr>
</tbody>
</table>

| Bs  | ...  | A  | ...  | Y  | ...  | H  | ...  | L  | ...  | D  | ...  | V  | QQ  | X  | GP  | ...  | ...  | WR | D  | I  | IV | ...  | I  | HH  | N  | QQ  | G  | GK  | QS  | 614 |
| Th  | ...  | P  | ...  | V  | ...  | T  | ...  | A  | ...  | Y  | ...  | R  | L  | D  | ...  | A  | AV | Y  | GP  | ...  | ...  | WE | D  | I  | IV | ...  | V  | HH  | N  | EE  | K  | ETA  | 671 |
| Cs  | ...  | D  | ...  | G  | V  | V  | ...  | A  | FI  | I  | ...  | T  | P  | BE  | ...  | ...  | ...  | ...  | KK | I  | IV | ...  | A  | YN  | PF | FKE | KK | I  | 782 |
| Kl  | D  | QQ  | A  | G  | GL  | VMT | V  | D  | GM  | KAC  | AS  | LDS  | RL  | D  | GL  | LV  | A  | I  | NA  | A  | AP  | ES  | RT  | 1028 |

| Bs  | VF  | LP  | D  | GG  | GH  | VIC  | DG  | IR  | S  | GV  | TP  | VF  | S  | ...  | .  | VH  | KE  | IN  | LD  | GI  | 649 |
| Th  | IA  | LF  | D  | RE  | WA  | VVC  | DG  | QR  | CG  | GT  | TP  | FG  | Q  | ...  | .  | AR  | GL  | ML  | RL  | LD  | GI  | 706 |
| Cs  | LO  | LP  | HE  | G  | SWY  | V  | K  | AND  | G  | IV  | V  | FP  | D  | S  | FE  | KE  | ...  | ...  | ...  | ...  | 817 |
| Kl  | LN  | EF  | AG  | E  | GL  | T  | Q  | LS  | AI  | Q  | QT  | A  | GE  | NS  | LAN  | GV  | Q  | IA  | AD  | GT  | VT  | IP  | AW  | 1068 |

| Bs  | GT  | WV  | LV  | V  | K  | Q  | S  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | 658 |
| Th  | GT  | WV  | LV  | H  | P  | AG  | *  | .  | 716 |
| Cs  | SL  | F  | AT  | Y  | Q  | K  | *  | .  | .  | 825 |
| Kl  | SV  | AV  | IE  | LP  | Q  | G  | EA  | Q  | GAG  | LP  | VS  | SK  | 1090 |

Abbreviations for amino acid sequences are as follows:

- **BSTPULA**: *Bacillus stearothermophilus* pullulanase
- **THPULA**: *Thermus* sp. AMD33 pullulanase
- **CSPULA**: *Caldocellum saccharolyticum* pullulanase
- **KLPULA**: *Klebsiella oxytoca* UNF5023
8.5 Pullulanase Secondary Structure Predictions

Homology lineups and secondary structure prediction algorithms (Chou and Fasman 1978, Garnier et al. 1978) predicated that the related enzymes (table 8.1) possessed a similar secondary structure, including an \((\alpha/\beta)_8\) barrel structure and a multi-\(\beta\)-sheet structure.

The three-dimensional structures (X-ray crystallography) of only two \(\alpha\)-amylases are known (see appendix 5). The peptide sequence of each enzyme shows little homology except for the three highly conserved regions (5-8 amino acids) at amino acid positions approximately 100, 200 and 300 (see figure A5.3). They have similar structural elements, \(\alpha\) helices and \(\beta\) sheets in similar arrangement and spatial position with an N-terminal \((\alpha/\beta)\)\(_8\) barrel structure. In both cases the three conserved regions form part of the active site cleft and/or \(\text{Ca}^{2+}\) -binding site (see figure A5.3). The 3-dimensional structure of a related enzyme, cyclodextrin glycosyltransferase (CGTase) from *Bacillus circulans*, has also been determined by X-ray crystallography (Klein and Schulz 1991). It possesses a similar domain structure to the two \(\alpha\)-amylase sequences.

As stated in section 8.3, homology comparisons between the peptide sequences of pullulanases, \(\alpha\)-amylases and other related enzymes showed the same three conserved regions, in the same order and spacing. Due to their similarity these conserved regions have been proposed to contain the active site.

The secondary structures of \(\alpha\)-glucosidase, CGTase and oligo-1,6-glucosidase have been proposed using a mixture of homology lineups and secondary structure prediction algorithms (MacGregor and Svensson 1989, Watanabe et al. 1991). The resulting model structures were very similar to the three dimensional \(\alpha\)-amylase structures determined by X-ray crystallography (see appendix 5).

Using a combination of the structural information obtained from the \(\alpha\)-amylases (X-ray structure, predicted secondary structure and homology lineups) and the proposed secondary structure of other related enzymes (\(\alpha\)-glucosidase, CGTase and oligo-1,6-glucosidase), together with the peptide sequence homology lineup, it was possible to produce an overall proposed secondary structure for pullulanases. An alignment of the proposed two-dimensional structural elements of the pullulanase was obtained and is shown in figure 8.7.

Analysis of the Deduced Peptide Sequence of the *C. saccharolyticum* Pullulanase
Figure 8.7 An alignment of the proposed two-dimensional structure of pullulanases

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Position</th>
<th>Sequence</th>
<th>Structure</th>
<th>Enzyme</th>
<th>Position</th>
<th>Sequence</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSPULA</td>
<td>GENLR</td>
<td>GKFGLF CQ.EGYKEGIS TGLLHKLKEG</td>
<td>V6T HVHL L. 388</td>
<td>KLPULA</td>
<td>PAELR</td>
<td>GKYKLAL TAGDSNMQHLK T....LSASG</td>
<td>V6T HVELE L. 484</td>
</tr>
<tr>
<td>BSTPULA</td>
<td>PFDDFAAGVDE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>227</td>
<td></td>
</tr>
<tr>
<td>THPULA</td>
<td>PPMDFAAGVDE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>285</td>
<td></td>
</tr>
<tr>
<td>CSPULA</td>
<td>PISDFGSVDD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>398</td>
<td></td>
</tr>
<tr>
<td>KLPULA</td>
<td>PVFDLATXKVDSFKAVEIQQPFLQPVNFQKIPGHPVYR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>537</td>
<td></td>
</tr>
<tr>
<td>BSTPULA</td>
<td>. ..... RNPLQQ.YNWGNPNLHYNAPEGSYATDPNDPYAIR</td>
<td>262</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THPULA</td>
<td>. ..... RDQQA.YNWGNPNLHYNAPEGSYATDPNDPYAIR</td>
<td>320</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSPULA</td>
<td>. ..... KNEDKR.YNWGYDVLQYCEPYYWSTKGGIEA.LK</td>
<td>432</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KLPULA</td>
<td>. ..... LKQSDQDNPQVQLNITLAQVTDSSNYWGDYPHFTVPEGSYATDPEGT.TRIK</td>
<td>589</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α2</td>
<td>ELKRAIHTLQS N.GI RVIMDV VYNHVVYIRDQ...SSFEKIVPGYFRR</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α3</td>
<td>YDAYGNPSNGTGTVGNDIAERKM VRKWIIDSVRFWVUE YHVNY</td>
<td>352</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β3</td>
<td>B20 α-amylase region</td>
<td>637</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α4</td>
<td>B20 α-amylase region</td>
<td>684</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β5</td>
<td>LMGILD VETHKAVRE MLTDLSIPV LSFHNL.RPT LSSQK</td>
<td>394</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α5</td>
<td>B20 α-amylase region</td>
<td>452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSTPULA</td>
<td>ATMQNAEKLVP RIGY KNDY FRDSVK GTEFQLPDRFALGDSDDREK</td>
<td>439</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THPULA</td>
<td>RRWSNAKLQP RFAY FDNR FRDAVK GSTFHLPDGRFALGNPQGREQ</td>
<td>497</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSPULA</td>
<td>TILSSCHQGY SIGL FNDR IRDAIR G.DLDGKTYGVHNLSVDVR</td>
<td>613</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KLPULA</td>
<td>INLKG .... GIGT FSDR LDASVR GGGPFSDGALRQNGQIGSGA</td>
<td>769</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSTPULA</td>
<td>. ..... VKVAISGSI</td>
<td>451</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THPULA</td>
<td>. ..... VKLAIAGSIL</td>
<td>509</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSPULA</td>
<td>. ..... LKQGIKAII</td>
<td>622</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KLPULA</td>
<td>. ..... GVLPNELASDSDQVRHLADTLRLGMAGNLADFVMIDKDGAACKGSEIDYNGAP</td>
<td>823</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chapter 8
### 300 α-amyrase region

<table>
<thead>
<tr>
<th>α&lt;sub&gt;6&lt;/sub&gt;</th>
<th>β&lt;sub&gt;7&lt;/sub&gt;</th>
<th>α&lt;sub&gt;7&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BSTPULA</strong></td>
<td><strong>NGLET</strong></td>
<td><strong>HPTQT</strong></td>
</tr>
<tr>
<td><strong>THPULA</strong></td>
<td><strong>GGLFC</strong></td>
<td><strong>HPQSI</strong></td>
</tr>
<tr>
<td><strong>CSPULA</strong></td>
<td><strong>DDFAK</strong></td>
<td><strong>EPDEC</strong></td>
</tr>
<tr>
<td><strong>KLPULA</strong></td>
<td><strong>GGYAA</strong></td>
<td><strong>DPTEV</strong></td>
</tr>
</tbody>
</table>

### 2-dimensional structure

<table>
<thead>
<tr>
<th>α&lt;sub&gt;8&lt;/sub&gt;</th>
<th>β&lt;sub&gt;12&lt;/sub&gt;</th>
<th>β-sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BSTPULA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>THPULA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CSPULA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>KLPULA</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>α&lt;sub&gt;8&lt;/sub&gt;</th>
<th>β&lt;sub&gt;12&lt;/sub&gt;</th>
<th>β-sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BSTPULA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>THPULA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CSPULA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>KLPULA</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations for amino acid sequences are as follows:
- **BSTPULA**: *Bacillus stearothermophilus* pullulanase
- **THPULA**: *Thermus* sp. AMD33 pullulanase
- **CSPULA**: *Caldocellum saccharolyticum* pullulanase
- **KLPULA**: *Klebsiella oxytoca* UNF5023

If this 2-dimensional structure is correct the pullulanase is likely to be folded to form an (α/β)<sub>8</sub> barrel structure followed by a structure composed of multi-β-sheets. This is shown diagrammatically in figure 8.8. To date no actual secondary structural information is known for the pullulanases.

Analysis of the Deduced Peptide Sequence of the *C. saccharolyticum* Pullulanase
Figure 8.8 Pseudo-two-dimensional structure of the pullulanase

A 2-dimensional representation of the proposed secondary structural elements of pullulanase based on the x-ray crystal structure of α-amylase and homology lineups to the α-amylases and other related enzymes. The numbers in the circles and squares refer to the position of the α-helices and β-sheets in the (α/β)8 barrel domain. The internal numbers (100, 200, 230 and 300) refer to the position of the highly conserved α-amylase regions. It is these regions that were used as marker points to align the position of the (α/β)8 barrel within the consensus pullulanase peptide sequence.
8.6 Summary

Analysis of the deduced peptide sequence of the *Caldocellum saccharolyticum* pulA gene product showed that the proposed sequence had high homology to the other three known pullulanase peptide sequences (see section 8.4). It also showed homology to other similar enzymes (e.g. isoamylase and α-amylase-pullulanases) and other related enzymes including α-amylase (see section 8.2). All of these peptide sequences contained the three highly conserved α-amylase sequences in the same order and spacing. These highly conserved regions have been implicated in catalytic activity, substrate binding and metal ion binding (see section 8.3).

Using a combination of the structural information obtained from the α-amylases and the proposed secondary structure of other related enzymes, together with the peptide sequence homology lineup, it was possible to produce a proposed secondary structure for pullulanases. This model predicted a central (α/β)$_8$ barrel structure followed by a C-terminal structure composed of a multi-β-sheet structure (see section 8.5).
Chapter Nine
Discussion

9.1 Comparison of the DNA sequence of the pNZ1038 *Caldocellum saccharolyticum* pullulanase to other known *Caldocellum saccharolyticum* genes and the DNA sequences of other related enzymes

The DNA sequence of pNZ1038 was searched for sequences indicative of substrate regulation, since expression of the pullulanase gene was substrate-inducible in the native organism. While sequences homologous to the "maltose box" sequence from *E. coli* and *Klebsiella* sp. were detected, they were not in the expected positions and did not appear to be active in substrate induction (i.e. were not recognised by the *E. coli* host cell). A number of amylolytic enzymes that are substrate inducible in their native organisms are non-inducible when expressed in *E. coli* from their own promoters (see section 1.3.3), implying that substrate induction is a more complex and species-variable phenomenon than gene expression. Analysis of the promoter regions of substrate-inducible DNA sequences of foreign genes expressed in *E. coli* supported this assumption.

The only sequences detected in pNZ1038 that possibly could be involved in substrate induction were three direct repeats located upstream of the start of the pullulanase gene. These sequences appeared to be unique, as searches of other *C. saccharolyticum* sequences and the GenEMBL DNA database revealed no homologous sequences. Thus these unique sequences could be involved in substrate inducible pullulanase expression in *C. saccharolyticum*.

The proposed ribosome (Shine-Dalgarno) and RNA polymerase binding sites (-10, -35) of the pullulanase gene (*pulA*) were homologous to standard prokaryotic translation and transcription initiation sequences (Dreyfus 1988), as well as the consensus *C. saccharolyticum* sequence. As all of these sequences were determined by homology to standard prokaryotic sequences, and since no experiments have been performed on gene expression in *C. saccharolyticum*, there is no way of knowing whether these sequences are used in the native organism or in *E. coli*. As the *C. saccharolyticum* genes were expressed in *E. coli* from their own DNA sequences recognised by the *E. coli* transcriptional and translational cellular machinery, it is logical to assume that these sequences were homologous to the standard *E. coli* consensus initiation sequences. Proof of this assumption can be seen in the N-terminal amino acid sequence of the pullulanase as the peptide start codon was identical to the proposed start codon. The sequence identified as the Shine-Dalgarno region was
located 10 base pairs from the ATG (which is close to the optimal distance of 9 bp): hence it is likely to have been used in *E. coli*. Another method of determining an optimal Shine-Dalgarno sequence is by sequencing the 16S ribosomal RNA gene since the last 4-9 bases at the C-terminal end are involved in ribosome binding to the mRNA and hence the complement of this sequence is the theoretical ideal sequence (Gualerzi and Pon 1990). Unfortunately whilst the *C. saccharolyticum* 16S ribosomal RNA sequence had been determined for taxonomic purposes, the last bases had not been sequenced as they coincided with one of the amplification primers. Hence this method could not be used to determine the theoretical optimum Shine-Dalgarno region.

Studies on mRNA transcription of the *celA* gene from a related thermophile (*Clostridium thermocellum*) have shown that two different sets of -35 and -10 sequences were present. When the gene was expressed in *E. coli*, sequences homologous to the standard $\sigma^{+0}$ consensus were used, whereas when expressed in *Clostridium thermocellum*, sequences homologous to a *B subtilis $\sigma^{+8}$* consensus were used (Béguin et al. 1986). While it was possible that the *pulA* gene also used different sequences for expression, searches of the DNA sequence upstream of the start codon revealed no other potential -35 and -10 sequences. Later work on mRNA initiation of two other genes from *Clostridium thermocellum*, *celF* and *celD*, showed that transcription of *celD* used the standard $\sigma^{+0}$ consensus, while for the *celF* gene, the RNA polymerase bound at a -10 and -35 sequence that showed no homology to any known mRNA promoter (Mishra et al. 1991).

Alignment and analysis of RNA polymerase promoters showed that the degree of conservation of each base was not uniform (Hawley and McClure 1983). While the genomic G/C ratio has not been shown to affect the RNA polymerase promoter or ribosome binding site, studies on genes from *Micrococcus luteus* (G/C 74%) showed that sequences similar to the consensus hexamers were present in the right location. The only differences involved the substitution of A or T for G or C in the less conserved bases, resulting in the promoter sequence having a higher G/C ratio than that found in *E. coli* (Nakayama et al. 1989). The reverse situation existed for the *pulA* gene and other *C. saccharolyticum* genes where the A/T ratios of the promoter sequences (both RNA polymerase and ribosomal) were greater than that found in *E. coli*. Further work is necessary with other species of varying G/C ratios to determine whether the overall genomic G/C ratio influences the G/C ratio of expression promoters.

Putative factor-independent terminator sequences have been identified downstream of the *pulA* genes (see section 4.5.3) and all other *C. saccharolyticum* genes (see Appendix 4). While it has not been shown that any of these sequences were terminators, each gene possessed sequences homologous to the Shine-Dalgarno region and RNA polymerase binding site upstream of its start. The presence of these homologous sequences would indicate that these genes were produced as monocistronic units of transcription, and hence the need for an RNA polymerase terminator sequence at the 3' end. The program also detected a major terminator sequence 500 bp downstream from the start of the pullulanase gene. While this sequence may not be recognised in *C. saccharolyticum*, it could be in *E. coli*. Evidence that it is recognised in *E. coli* can be seen in the detection of shortened peptides (see Discussion.
section 5.4.1) that could have been produced from sequences acting as internal promoters for initiation and/or reinitiation of translation (see section 4.4).

It should be noted in this context that mutagenesis studies on strong *E. coli* transcriptional terminators, in which hybrid terminators were constructed, showed that some theoretically effective terminators were not active, implying that other sequences and/or other factors were also involved in Rho independent termination (Reynolds *et al.* 1992, Reynolds and Chamberlin 1992).

The identification and function of the other ORF’s detected on pNZ1038 (PepX and PepY) have not been determined. These ORF’s showed no homology to any other known sequence (see appendix 2), and pullulanase expression studies in *E. coli* using plasmids containing these genes showed no apparent effect on either the level of expression or the cellular location (see section 5.2). Detailed sequence comparisons were performed (at the DNA and peptide levels) between PepX, PepY and the *Klebsiella* genes (*pulB*-*pulN*) located adjacent to the *Klebsiella pulA* gene and no similarity was detected. Although these peptides did not show any homology to other known genes, they could still be involved in either substrate induction or export out of the cell. It is of interest that ORF’s of unknown function, and without homology to any other known sequences, have also been detected adjacent to the *Thermus* AMD33 pullulanase gene (Nakamura and Sashihara 1991).
9.2 Comparison of the peptide sequence of the \textit{C. saccharolyticum} pullulanase to the peptide sequences of other related enzymes

One of the problems in amino acid sequence comparison is determining which sequence differences account for differences between enzymes (e.g. activity, thermostability, pH and temperature optima). However, one advantage of working with sequences that have a high degree of variability is that any region of homology between them is likely to be significant.

\subsection*{9.2.1 Comparison of the amino acid sequences}

The four known pullulanase sequences were obtained from widely different eubacterial species with little DNA homology. Therefore any regions of common amino acid sequence were likely to be important in elucidating the essential amino acid sequences critical for the enzyme (i.e. those sequences that form the essential enzyme "signature"). As shown in figure 8.6, at least six regions (6 amino acids or greater) of identical amino acid homology were detected, besides the three \(\alpha\)-amylase regions. These regions were not evenly spaced over the entire peptide but concentrated in the middle section, implying that this portion was essential for enzyme activity.

The computer alignment of the four sequences showed that while the length of the four enzymes varied remarkably (659, 716, 825 and 1091 amino acids), they were nearly perfectly aligned at the C-terminal end (3 sequences finishing within 2 amino acids of each other and the fourth within 12 amino acids of the others). To accommodate this alignment at the C-terminal end, the N-terminal portion of each enzyme was offset depending on the peptide sequence length (see figures 8.3 and 8.6). The location of the homologous regions and the N-terminal sequence offset implied that the central and C-terminal portions contained the essential pullulanase motifs. Experimental evidence to support this assumption was seen by the fact that pNZ1452 was missing the first 100 amino acids but was still able to hydrolyse pullulan. Pugsley and co-workers have also shown that the mutation in the N-terminal portion of the \textit{Klebsiella} pullulanase affected enzyme export but not pullulanase activity (see section 1.3.2).

Having identified sequences that were common to all pullulanases, the next step was to determine whether amino acid sequence differences could explain the differences in thermostability and reported substrate characterisation of each enzyme.

\subsection*{9.2.2 Differences related to thermostability}

Comparison of the amino acid sequences of the three thermophilic enzymes to the one mesophilic enzyme showed that within the highly conserved central and C-terminal portions (480 amino acids),
the *Klebsiella* pullulanase had five regions of sequence not present in the other enzymes (51, 11, 19, 21 and 34 amino acids in length). These extra sequences could confer increased flexibility at mesophilic temperatures since they were located between the conserved regions of the *Klebsiella* pullulanase. Analysis of individual amino acid replacement patterns between the mesophilic and thermophilic enzymes, according to the generally observed rules of Argos and co-workers (see sections 1.1.4 and 8.4), revealed that only five preferred amino acid exchanges occurred. The location of these exchanges was close to the three highly conserved α-amylase regions previously implicated in enzymatic activity (see section 8.4). As these exchanges would theoretically increase the rigidity of the protein, their location would increase the rigidity of the active site (relative to the mesophilic enzyme). This increase in rigidity would be needed for activity at thermophilic temperatures.

Using regression analysis, Suzuki and co-workers have suggested that for certain enzyme groups, including pullulanases, the proline content increases with increasing thermostability (Suzuki et al. 1991). However the proline content of the *C. saccharolyticum* pullulanase appeared to be unrelated to its thermostability (see section 8.4). The proline content of 3.5% was lower than that found for the pullulanase from *Klebsiella* sp (4.7%) and in fact the theory would predict that the *C. saccharolyticum* pullulanase was cryophilic. Sequencing errors were discounted since to increase the proline content to the predicted level of approx 5.5% would require the addition of 16 proline residues. Closer inspection of the data used by Suzuki showed that some examples were inappropriate. These included using the DNA sequence of the *Klebsiella pulA* gene as determined by Murooka that is shown in appendix 6 to be incorrect (note that using the correct sequence only alters the percentage proline content of the *Klebsiella* sp. pullulanase slightly), and the use of data from a total amino acid analysis of purified extracellular *Klebsiella* sp. pullulanase that still included the LamB protein.

It may be possible to account for the differences in thermostability between the thermophilic and mesophilic pullulanases using differences in amino acid sequence. However, detailed analysis of amino acid differences among the three thermophilic pullulanases revealed no detectable pattern that could explain the apparent differences in their activity.

### 9.2.3 Differences related to substrate characterisation

The four pullulanases with known amino acid sequence can be divided into two groups based on their substrate characterisation. The first group consists of the Type I pullulanases (*Klebsiella* sp. and *Bacillus stearothermophilus*) and Type II pullulanases form the second group (*Thermus* AMD33 and *C. saccharolyticum*). Although the amino acid sequence comparisons of the four pullulanases could be expected to show the same division, this is not the case (as shown in chapter 8). For example, in the dendrograms shown in figures 8.2 and 8.5, the *B. stearothermophilus* pullulanase actually groups with the *Thermus* pullulanase three times and the *C. saccharolyticum* once, while the *Klebsiella* sp. pullulanase groups with the *C. saccharolyticum* pullulanase once and the *Thermus* pullulanase once but never with the *Bacillus stearothermophilus* pullulanase. A similar situation is seen for the type II
pullulanases, i.e. the \textit{C. saccharolyticum} pullulanase never groups with the \textit{Thermus} pullulanase. Analysis of the overall amino acid sequences (see figure 8.6) showed that there were no "extra" regions of homology within either group. In fact, as shown in table 8.2 the highest degree of homology was between the \textit{Thermus} and \textit{Bacillus stearothermophilus} pullulanases. Hence the four pullulanases grouped together and were indistinguishable by amino acid sequence analysis except for changes likely to affect thermostability.

Amino acid sequence comparisons between the \textit{Caldocellum saccharolyticum} pullulanase and other related enzymes (see appendix 3) showed that while the three highly conserved $\alpha$-amylase regions were present, none of the other pullulanase regions were present even for the more closely related enzymes (i.e. isoamylase, neopullulanase and $\alpha$-amylase-pullulanases). In the comparison between the \textit{C. saccharolyticum} pullulanase and the \textit{Pseudomonas} isoamylase, there were only two homologous regions of four amino acids each (besides the $\alpha$-amylase consensus regions). For the neopullulanase and the \textit{Clostridium} $\alpha$-amylase-pullulanase only the $\alpha$-amylase regions were present.

As shown in section 8.5, the predicted secondary structures of pullulanases, including the \textit{C. saccharolyticum} pullulanase, were similar. These structures were devised from a combination of homology lineups, X-ray crystal structure of related enzymes and proposed secondary structure of the pullulanase genes. These prediction methods have been shown to be relatively inaccurate for the prediction of the secondary structure of an individual peptide sequence. However the overall accuracy increased when analysis was performed on a multi-sequence alignment and in tandem with the known X-ray structure of related peptides (reviewed by Taylor 1987).
9.3 Enzymatic properties of the *Caldocellum saccharolyticum* pullulanase compared to the enzyme expressed in *E. coli*

Comparison of the enzymatic characterisation of the pullulanase isolated from the native organism to the pullulanase expressed from the *pulA* gene in *E. coli* showed some major differences (see table 9.1). The significance and possible reasons for these differences are discussed in this section.

<table>
<thead>
<tr>
<th>Enzyme from the native organism</th>
<th>Enzyme expressed from the <em>pulA</em> gene in <em>E. coli</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>pH optimum</td>
<td>pH optimum</td>
</tr>
<tr>
<td>5.7</td>
<td>5.7</td>
</tr>
<tr>
<td>temperature optimum</td>
<td>temperature optimum</td>
</tr>
<tr>
<td>70°C</td>
<td>70°C</td>
</tr>
<tr>
<td>Enzyme glycosylated</td>
<td>Enzyme non-glycosylated</td>
</tr>
<tr>
<td>Ca(^2)+ increased enzyme activity</td>
<td>Ca(^2)+ increased enzyme activity</td>
</tr>
<tr>
<td>EDTA decreased enzyme activity</td>
<td>EDTA decreased enzyme activity</td>
</tr>
<tr>
<td>Molecular mass 120 kDa</td>
<td>Molecular mass 96 kDa</td>
</tr>
<tr>
<td>Thermostability was determined as half life of 87 min at 80°C and 9 min at 90°C</td>
<td>Thermostability was determined as half life of 38 min at 80°C and 7 min at 90°C</td>
</tr>
<tr>
<td>Complete hydrolysis of pullulan yielding maltotriose only</td>
<td>Complete hydrolysis of pullulan yielding maltotriose only</td>
</tr>
<tr>
<td>Complete hydrolysis of β-limit dextrin, amylopectin and amylose yielding oligomaltosaccharides</td>
<td>Low activity on β-limit dextrin and amylopectin No activity on amylose</td>
</tr>
</tbody>
</table>

Theoretically, it is possible that different parent organisms were used in the isolation of the pullulanase gene (*pulA*) expressed in *E. coli* and the enzyme from the native organism. As stated in section 3.2.3 the DNA from the plasmid that expressed the pullulanase in *E. coli* hybridised (at very high stringency) to an identical sized fragment from a restriction endonuclease digest of the genomic DNA from which the native enzyme was isolated. The G-C percentage ratio, codon preference and doublet preference of the *pulA* gene appeared similar to those previously obtained for the other *C.*
saccharolyticum genes (see chapter 4 and appendix 4). Hence all the evidence stated above and elsewhere in this thesis indicates that the same organism was used to isolate both the pulA gene and the enzyme expressed from the native organism. While it is still possible that they were from different organisms the only definitive proof would be to amino acid sequence the pullulanase from the native organism.

It has recently been shown that C. saccharolyticum contains three different xylanases side by side on the same DNA fragment (personal communication V. Te'o, Department of Cell Biology, Auckland University). Hence it is possible that there are also multiple copies of other genes in the C. saccharolyticum genome. The differences between the two enzymes could be accounted for if C. saccharolyticum contained two different pullulanase genes, with one enzyme exported and the other remaining in the cell (i.e. the intracellular enzyme was expressed in E. coli and the one exported was isolated from the native organism). However, only one gene was detected with southern hybridisation. The gene was expressed in E. coli and showed pullulanase activity on pullulan (a large polymer only present externally) and as only small oligosaccharides can enter the cell, it would be expected that the intracellular enzyme would show greater homology to an oligo-1,6-glucosidase, which it does not.

Finally, the deduced protein sequence of the pullulanase was homologous to the other pullulanase sequences and hence its enzymatic activity should be closer to a pullulanase than an oligo-1,6-glucosidase. Possible methods for searching for other pullulanase genes are discussed in section 9.7. Without identifying another gene, amino acid sequencing of the pullulanase expressed in the native organism would be the only definitive method to show that both enzymes have the same peptide sequence.

Assuming that two genes were not involved, the following explanations could also account for the differences.

It is possible that the presence of post-expression modification of the enzyme could explain the differences in the enzymatic characterisation of the pullulanase. These differences could modify the peptide sequence to produce two slightly different enzymes from the same gene sequence. While enzyme expression from the native organism was substrate-regulated and the enzyme was glycosylated and exported out of the cell, enzyme expression in E. coli was not substrate-regulated and the enzyme was non-glycosylated (as expected) and not exported from the cell. Glycosylation of the enzyme expressed in the native organism could account for the difference in molecular mass between the enzyme expressed in the native organism and that expressed in E. coli.

Since most enzymes that are exported from the cell are modified during export (including peptide leader sequence cleavage) it is likely that a similar situation occurred for the enzymes exported from C. saccharolyticum, whereas sequencing of the N-terminus of the pulA gene product showed that no leader sequence cleavage had taken place, implying that the enzyme had not been modified. While proteins acting as chaperones for the pullulanase are likely to be present in C. saccharolyticum, The

Discussion
same proteins specific to the pullulanase are unlikely to be present in E. coli. These chaperones assist the enzyme to form the correctly folded conformation and can also aid in export (reviewed by Ellis and van der Vies 1991). Since the growth temperatures of C. saccharolyticum and E. coli are different, the cellular environment in which the peptide was folded was also different and could also affect folding, peptide sequence modification and export (reviewed by Meyer 1988).

Non-export of extracellular enzymes expressed in E. coli is a common phenomenon and these enzymes are sometimes incorrectly folded (reviewed by Mitraki and King 1989). Studies on the Klebsiella pullulanase expression system showed that a 19.2 kb fragment of genomic DNA was needed for maltose inducible pullulanase expression, modification and cell-surface exposition in E. coli (see section 1.3.2). Site directed mutagenesis studies on the twelve glycosylation sites of yeast acid phosphatase showed that the absence of glycosylation severely affects protein folding and enzyme export (Riederer and Hinnen 1991). As the pullulanase was not glycosylated when expressed in E. coli, this could account for the difference in export from the cells and enzymatic activity differences based on folding.

While the pH and temperature optima of the two enzymes were similar, there was a major difference in the thermostability of each enzyme. These differences could be due to a lack of glycosylation and/or incorrect folding of the pullulanase expressed in E. coli. Olsen and Thomsen (1991) expressed a β-glucanase gene from B. macerans (normally glycosylated) in both E. coli and Saccharomyces cerevisiae. Thermostability studies at 70°C showed the half life of the glycosylated enzyme produced in S. cerevisiae to be 26 min while that of the non-glycosylated enzyme produced in E. coli was 10 min.

The greatest difference between the enzymes was seen in substrate characterisation. The enzyme expressed from the native organism possessed both α-1,4 and α-1,6 activities while the enzyme expressed in E. coli had only α-1,6 activity. There was also a difference in the ratio of enzymatic activity on a number of substrates, for example the enzyme expressed in E. coli had little activity on β-limit dextrin and amylopectin. To date this issue of differences in the substrate characterisation has not been fully resolved despite a large amount of effort. Possible reasons for these differences include incorrect folding of the enzyme, glycosylation and contaminating enzyme activity.

As stated above, the environment for folding the pullulanase in E. coli was different. This different environment could have resulted in a sub-optimal enzyme conformation that, while active on the soluble and easily accessible substrate pullulan, had lower activity on the less-soluble and less-accessible substrates (e.g. β-limit dextrin and amylopectin). Evidence for this suggestion is seen in the higher enzyme activity on β-limit dextrin than on amylopectin.

Glycosylation has been shown to affect the binding of enzymes (via hydrogen bonding) to the surface of a number of starch polymers (Evans et al. 1990). Affinity binding studies have shown that
the native enzyme bound with high affinity to both amylose and amylopectin. The enzyme expressed in *E. coli* showed only low affinity with amylose and amylopectin, which could explain the different enzymatic substrate characterisations.

Although the pullulanase expressed from the native organism was purified to apparent homogeneity, it is possible that contaminating enzymes (e.g. α-amylase) were still present. Since the specific activity of α-amylases is higher than for pullulanases a minor contamination would produce a substantial effect on substrate characterisation. The results obtained with α-amylase inhibitors do not eliminate the possibility that contaminating enzymes were still present, as not all α-amylases were affected to the same extent or even inhibited at all. In one study of over 66 α-amylases from different sources only 33 were inhibited, to varying degrees (Silano *et al.* 1975).

Hence it is possible that the mechanism of expression and the cellular environment in which the enzyme was expressed were the cause of the enzymatic differences between the enzymes rather than any difference in the gene sequence. Thus the same gene could produce the same peptide that, due to the expression environment, could produce a slightly different enzyme.
9.4 Enzymatic properties of the *Caldocellum saccharolyticum* pullulanase compared to the enzymatic properties of other related enzymes

Before discussing the enzymatic properties of the *C. saccharolyticum* pullulanase, it is important to examine the true substrate activity of the enzyme, since the recombinant enzyme and the gene expressed in the native organism have different activities on starch, but not on pullulan. As discussed previously, the pullulanase expressed in the native organism would be classified as a pullulanase with both α-1,4 and α-1,6 activities (see sections 3.4 and 9.3), while the pullulanase expressed in *E. coli* would be classified as a pullulanase with only α-1,6 activity. Since the pullulanase expressed in *E. coli* has only low activity on the α-1,6 linkage of amylopectin and β-limit dextrin, it is probable that the enzyme was affected by expression in *E. coli*, hence this sample cannot be used as an indication of the enzyme’s true activity. This apparent contradiction is similar to that found for *Thermus* AMD 33 (Nakamura and Sashihara 1991).

Analysis of the substrate characterisation data of purified thermophilic pullulanases showed that the majority had both α-1,4 and α-1,6 activities (see section 1.4.2). The only thermophilic pullulanases to be reported possessing only α-1,6 activity are the pullulanases from *Bacillus acidopullulyticus* and *Bacillus stearothermophilus*, both of which are only moderate thermophiles with temperature optimums of 60°C and 65°C respectively. Thermophiles in general, and enzymes from *C. saccharolyticum* in particular, have been shown to have dual enzyme activities (see sections 1.1.2 and 1.5). Hence it is plausible to suggest that the pullulanase did have intrinsic α-1,4 activity (i.e. no contaminating enzyme was responsible for this activity).

In a recent review, Antranikian (1990) divided pullulanases into two groups depending on whether the enzyme showed exclusively α-1,6 activity (pullulanase Type I) or mixed α-1,4 and α-1,6 activity (pullulanase Type II). Note that in this thesis the Type II pullulanases have been further subdivided into two groups (pullulanases with α-1,4 activity, see section 1.4.2.2, and pullulanases with α-amylase activity, see section 1.4.2.3). While this division may appear subtle in terms of enzyme activity, there are major differences in the molecular mass of the enzymes and their DNA sequences (see section 8.3).

A number of pullulanases have been purified (at least 16), but only four enzymes have been sequenced and expressed in a different organism (Klebsiella sp., *Bacillus stearothermophilus*, *C. saccharolyticum* and *Thermus* AMD33). Analysis of the deduced protein sequences of these four pullulanases has failed to detect any sequence differences that correlate with activity, i.e. between the
two enzymes with only \(\alpha-1,6\) activity (type I) and the two with combined \(\alpha-1,4\) and \(\alpha-1,6\) activity (type II). This analysis is discussed in section 9.2. Hence, while the enzyme-substrate characterisation of the \textit{C. saccharolyticum} pullulanase groups the enzyme with other type II pullulanases with combined \(\alpha-1,4\) and \(\alpha-1,6\) activity, analysis of the protein sequence does not distinguish between the pullulanases. Thus it is theoretically possible that no type II pullulanase really has intrinsic \(\alpha-1,4\) activity (i.e. undetected contaminating enzymes are responsible for the \(\alpha-1,4\) activity found in type II pullulanases). In contrast, both the amino acid sequence and substrate activities of the \(\alpha\)-amylase-pullulanases group together i.e. they form a group separate from the other pullulanases (see sections 8.2 and 8.3). Hence the pullulanases can be arranged into three groups based on substrate characterisation, whereas analysis of amino acid sequence homology revealed only two groups, implying that there was a factor missing in either the analysis of the amino acid sequence homology or the analysis of the substrate characterisation. This factor could be a subtle difference in the amino acid sequence that affected the three-dimensional structure of the enzyme but was undetected in a simple amino acid one-dimensional lineup or a specific post-translational modification that affected only type II pullulanases. As more sequences become available it may be possible to determine the missing factor(s).
9.5 The potential biotechnological applications of the *Caldocellum saccharolyticum* pullulanase

The major biotechnological application of a thermophilic pullulanase is in the hydrolysis of the α-1,6-linkages of starch polymers to produce glucose and fructose (see section 1.2). Knowledge of the true substrate profile of the *Caldocellum saccharolyticum* pullulanase will determine its real biotechnological value. It could be used in the hydrolysis of amylopectin even if it was shown to have intrinsic α-1,4-activity. However, before commercial exploitation the level of enzyme expression would need to be increased to reduce the cost of enzyme production (see section 9.6).

The currently used commercial pullulanase is the enzyme from *Bacillus acidopullulyticus*, tradename Promozyme™ 200L, marketed by Novo Industri A/S. This enzyme, while having only α-1,6-activity, is only moderately thermophilic (temperature optimum 55-60°C), therefore the reaction vessel needs to be cooled before the enzyme is added. Since only mesophilic isomerases are used for the following step (the isomerisation of glucose to fructose) this cooling is acceptable, but with the isolation (and eventual commercial development) of thermophilic isomerases (Lehmacher and Bisswanger 1990, Dekker et al. 1991) the use of moderate thermophiles for the saccharification process will need to be reviewed. The reaction vessel would need to be cooled for the saccharification step then reheated for isomerisation, which would make the process costly in both energy and time.

A further potential biotechnological application involves using the pullulanase in an immobilised form, together with various amylases, to produce oligosaccharides of specific lengths, e.g. the co-immobilisation of amyloglucosidase and pullulanase to enhance starch hydrolysis (Chakrabarti and Storey 1990), and the use of a dual immobilised system of pullulanase and maltotetraose-forming amylase to produce maltotetraose (Kimura et al. 1989, 1990).

The synthesis of branched cyclomaltooligosaccharides using pullulanases is another potential application. These cyclodextrins show both hydrophobic and hydrophilic properties, and have potential applications in industry and medicine, but are limited by the high cost of production by chemical synthesis. Using the pullulanase “in reverse” to synthesise α-1,6-bonds (in low water conditions) can result in the production of branched cyclodextrins from various maltooligosaccharides cheaply and efficiently on a laboratory scale (Hizukuri et al. 1989).
9.6 Future work directed towards solving expression problems

The major problem to solve is to determine the true substrate activity of the *C. saccharolyticum* pullulanase and pullulanases in general. As discussed in sections 9.3 and 9.4, it appeared that the expression organism affected final enzyme activity. The enzyme was also only expressed at low levels, compared to other recombinant enzymes, and hence an overall improvement in the level of enzyme activity is needed.

9.6.1 Possible methods to improve pullulanase expression in *E. coli*-based expression systems

The overall level of enzyme production needs to be increased before successful biotechnological exploitation of the pullulanase can take place. The level of enzyme expression in *E. coli*, even from the pJLA602 overexpression plasmid, was still low, although higher than from the native organism. The pullulanase protein was estimated to be less than 1% of the total cell protein (between 0.1 and 0.5%). This amount is significantly lower than the maximum obtained for the *C. saccharolyticum* xylanase with 40% of the total cell protein (Lüthi et al. 1990b). Even if the level of enzyme activity was increased to these levels, the enzyme expressed in pJLA602 was insoluble and could not be extracted efficiently (see chapter 6.3). This section discusses a number of methods that have been applied successfully to increase the expression of other recombinant enzymes in *E. coli*.

(i) From the expression work in chapter 5 and sequence analysis in chapter 4, the *pulA* gene appeared to contain a strong internal RNA polymerase termination site that was recognised in *E. coli*. This site caused the RNA polymerase to prematurely terminate, reducing the amount of mRNA transcribed, and hence the amount of protein synthesised. Therefore, altering the DNA sequence at this termination site, without changing the amino acid sequence, could increase expression by reducing postulated premature termination.

(ii) Work on viral glycoprotein overexpression in *E. coli* has shown that hydrophobic amino acid domains (approx 15-20 amino acids long) are toxic to *E. coli*. Removal of these regions increased recombinant protein from undetectable levels to approximately 10-15% of total cell protein (Sisk et al. 1992). The *C. saccharolyticum* pullulanase contained three hydrophobic amino acid domains (see figure 8.1, 23 amino acids around position 370, 22 around 565 and 20 around 670). While it has not been shown that these sequences were toxic, expression of the pullulanase did affect the *E. coli* cell, as seen by decreased plasmid stability and the formation of multimeric cells (see section 5.2). To delete these domains completely would result in changes to the amino acid sequence and hence could alter the enzyme, but it may be possible to replace some of the amino acids near the centre of these hydrophobic domains without affecting the enzymatic activity.

Discussion
Differences in preferential codon usage between *E. coli* and *C. saccharolyticum* could have resulted in decreased expression due to a shortage of appropriate tRNAs. These differences in codon preference could be reduced by altering the rarest codons (from an *E. coli* “perspective”) in the pullulanase gene to conform to the optimal codon usage for the host. Hence the shortage of appropriate tRNAs would decrease and expression levels would increase. Although it would be impractical to alter every codon in the *C. saccharolyticum* pullulanase gene, altering three of the rarest codons could be sufficient.

Another possible limiting factor in expression of proteins in *E. coli* is a lack of oxygen. Recombinant cells, due to their higher metabolic burden compared to plasmid free cells, grow at a slower rate in oxygen-limiting conditions. The addition of a bacterial oxygen binding and transport protein from *Vitreoscilla* has been shown to reduce this problem of oxygen limitation. Expression studies using recombinants that expressed this gene together with the *Bacillus stearothermophilus* α-amylose, produced 3.3 times more α-amylase activity compared to recombinants that only expressed the α-amylase (Khosravi et al. 1990). The major advantage of this method is its simplicity once the appropriate strains have been constructed.

If these techniques were applied to the expression of the *C. saccharolyticum* pullulanase in *E. coli*, the level of enzyme expression is likely to increase with a possible change in the form of the enzyme. Some of the problems associated with expression in *E. coli* are also likely to be resolved.

### 9.6.2 Possible methods to improve pullulanase expression in expression systems other than *E. coli*

While it may be possible to increase the level of enzyme expression in *E. coli* by using some of the above methods, the enzyme produced is unlikely to be glycosylated, exported into the growth media or processed correctly. Theoretically, the simplest method to overcome some of these problems of expression in *E. coli* would be to express the *pulA* gene in a different organism that glycosylates and/or exports proteins to the growth media. Several organisms have been successfully used to overexpress and export genes.

(*Bacillus subtilis*) has been used as an alternative prokaryotic expression system. While this organism has been extensively studied and used biotechnologically to export enzymes to the growth media, it has a number of undesirable host properties including non-glycosylation and proteolytic degradation of exported proteins (reviewed by Errington and Mountain 1990). Therefore, although this expression system is potentially superior to *E. coli* it would still produce less than optimal results.
(ii) Yeast-based expression systems have also been extensively studied, due to their appropriate post-translational modification and export of heterologous proteins into the growth media. The yeast commonly used is *Saccharomyces cerevisiae* due to its long history in biotechnology and the availability of a number of stable inducible expression plasmids. This system results in an enzyme being glycosylated and exported out of the cell. However, problems associated with *S. cerevisiae* systems include low product yields, plasmid instability, hyperglycosylation and only partial export of secreted proteins to the periplasmic space. To overcome some of these problems other non-*Saccharomyces* yeasts could be used (reviewed by Buckholz and Gleeson 1991).

(iii) Filamentous fungi, including *Aspergillus* sp., are another group of organisms already used in biotechnology. They are also potentially useful for expression of heterologous proteins due to their high expression levels and appropriate glycosylation. The only major disadvantage of this system is the availability of appropriate expression systems for non-fungal proteins (reviewed by Turner 1990). Until appropriate expression systems are available it is impossible to determine their usefulness for pullulanase expression.

(iv) The most appropriate expression system for the pullulanase gene may be expression in a thermophilic host using a plasmid-based expression system, since the higher growth temperatures could replicate folding at 70°C as in the native organism. Koyama (1992) has recently constructed a *Thermus* expression plasmid that allows expression of non-thermus genes in *T. thermophilus*. This plasmid could be used to express the pullulanase at thermophilic temperatures. Expression at these temperatures could elucidate the effect of temperature on protein folding and hence enzyme activity.

(v) It is possible to express genes in a cell free expression system (*in vitro* expression). While expensive to set up and run, these expression systems produce large quantities of the enzyme in controlled environments without the problems associated with expression in *E. coli* (Spirin et al. 1988). However, it is unlikely that this technique would be cost-effective.
9.7 Future work on the study and modification of enzyme activity

A number of other thermophiles have been shown to contain pullulanases. Previously, prior to gene isolation, expression libraries needed to be constructed, requiring the genes to be expressed from their own promoters. However, these other genes can now be isolated directly using sequence alignments and PCR (from the information shown in section 8). The alignment of the pullulanase sequences showed that the most highly conserved regions were not around the three α-amylase homology regions but in the conserved pullulanase “signature” regions. The amino acid sequence of the conserved regions could be back-translated to a DNA sequence using the most appropriate translation table for the organism. From this DNA sequence, sets of oligonucleotide primers could be synthesised to locate and generate fragments of the pullulanase gene from the genomes of other thermophilic bacteria using the PCR reaction. The entire pullulanase gene could then be recovered by southern hybridisation of the PCR fragment to a genomic restriction endonuclease digest and isolation of the band detected.

The following methods could be used to try and locate other pullulanase genes in *C. saccharolyticum*. Genomic walking could be used to isolate the DNA fragments either side of the pulA gene (Shyamala and Ames 1989, Parker *et al*. 1991). These fragments could be ligated into an *E. coli* expression vector and tested for pullulanase activity. If other genes are present but not adjacent they could be isolated using the PCR reaction (at varying stringency) with the consensus pullulanase primers. This could result in the generation of internal fragments of each pullulanase gene. The remainder of each gene could be isolated by DNA hybridisation of the PCR-generated fragment to the *C. saccharolyticum* genome. Expression libraries of the entire *C. saccharolyticum* genome could also be constructed and tested for pullulanase activity. Those recombinants that express pullulanase could be differentiated by further testing involving either DNA hybridisation, restriction endonuclease digestion or enzyme assay at different temperatures or pH.

The isolation of new pullulanase genes would allow for enzymes to be available with slightly different enzyme characteristics, especially those with higher temperature optima (e.g. the pullulanase from *Pyrococcus furiosus*, temperature optimum of 105°C, Brown *et al*. 1990) or with different pH optima (e.g. *Bacillus flavocaldarius* KP 1228 at pH 6.3, Suzuki *et al*. 1991). However, from a commercial perspective, enzymes that have just α-1,6 activity are the most useful (e.g. the pullulanase from *Bacillus acidopullulyticus*, Jensen and Norman 1984). As Novo Industri A/S have a patent on their pullulanase from *Bacillus acidopullulyticus* it is unlikely that this enzyme could be marketed directly (even if the gene was reisolated), but the sequence information obtained from the gene could be used to alter other pullulanases. Analysis of any differences in the amino acid sequence of this enzyme compared to the amino acid sequence of the *C. saccharolyticum* pullulanase (or any other pullulanase sequence) could reveal those sequence differences that distinguish whether a pullulanase
has α-1,4 and α-1,6 activities, or just α-1,6 activity. A hybrid pullulanase could be constructed from the *C. saccharolyticum* pullulanase by incorporating these changes into the amino acid sequence to produce an enzyme that had exclusively α-1,6 activity. In this way the substrate preference of a thermophilic xylose isomerase from *Clostridium thermosulfurogenes* was altered from D-xylose to D-glucose using site-directed mutagenesis to change the substrate binding pocket (Meng *et al.* 1991).

While changes in the DNA sequence, and hence the peptide sequence, can increase an enzyme’s thermostability (see section 1.1.4), normally the increase is only minor. As the thermostability of the *C. saccharolyticum* pullulanase is similar to that of the other two sequenced thermophiles, there is insufficient DNA sequence information available for site-directed mutagenesis to increase thermostability.

Using molecular modelling computer programs (reviewed by Bowie *et al.* 1991) a better model of protein secondary structure could be obtained. This improved model would be important in determining which amino acid to change to alter substrate specificity. Increased protein synthesis would make more protein available, hence the X-ray and three-dimensional structures of the pullulanase gene could be determined directly.

Thus, using the pullulanase amino acid lineup and the PCR reaction it will be possible to isolate and sequence a large number of enzymes. As more pullulanase sequences become available the alignment shown in section 8.4 will become more reliable in predicting amino acid changes needed to produce different characteristics.
Appendices

Appendix One
Results of Sequencing pCGN Recombinants

The initial sequencing strategy was based on plasmid sequencing of two sets of exonuclease deletions produced by previous workers, using the original plasmid, pNZ1038, and the procedure detailed in section 1.6.2. After sequencing one set of deletions it was determined that their inserts were pBR322 sequence rather than Caldocellum saccharolyticum DNA. This is shown in table A1.1. It appeared that the pBR322 inserts, which formed a set of nested deletions, were produced from Bal 31 exonuclease deletions from the original plasmid pNZ1038 as shown by the alignment of the inserts relative to pBR322 (see figure A1.1).

Figure A1.1 Alignment of pCGN565 deletion series to pBR322 vector

![ Alignment of pCGN565 deletion series to pBR322 vector ]

Vector pBR322 4363 bp

Appendices
Table A1.1 Summary of sequence data obtained from pCGN recombinants, showing homology to pBR322

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>sequence length</th>
<th>pBR322 % homology</th>
<th>location within pBR322</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNZ1452 (Eco2)</td>
<td>292</td>
<td>none detected</td>
<td>C. saccharolyticum DNA</td>
</tr>
<tr>
<td>pNZ1453 (Eco3)</td>
<td>278</td>
<td>86</td>
<td>680-958</td>
</tr>
<tr>
<td>pNZ1454 (Eco4)</td>
<td>211</td>
<td>97</td>
<td>914-1125</td>
</tr>
<tr>
<td>pNZ1455 (Eco5)</td>
<td>355</td>
<td>96</td>
<td>1043-1398</td>
</tr>
<tr>
<td>pNZ1456 (Eco6)</td>
<td>295</td>
<td>92</td>
<td>1242-1537</td>
</tr>
<tr>
<td>pNZ1457 (Eco7)</td>
<td>363</td>
<td>80</td>
<td>1513-1876</td>
</tr>
<tr>
<td>pNZ1458 (Eco8)</td>
<td>48</td>
<td>none detected</td>
<td>poor sequence</td>
</tr>
<tr>
<td>pNZ1459 (Eco9)</td>
<td>369</td>
<td>87</td>
<td>1850-2179</td>
</tr>
<tr>
<td>pNZ1460 (Eco10)</td>
<td>221</td>
<td>none detected</td>
<td>poor sequence</td>
</tr>
<tr>
<td>pNZ1461 (Eco11)</td>
<td>312</td>
<td>75</td>
<td>2135-2447</td>
</tr>
<tr>
<td>pNZ1462 (Eco12)</td>
<td>sequence unreadable</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pNZ1463 (Eco13)</td>
<td>sequence unreadable</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pNZ1464 (Eco14)</td>
<td>sequence unreadable</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pNZ1465 (Eco15)</td>
<td>sequence unreadable</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pNZ1466 (Eco16)</td>
<td>sequence unreadable</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pNZ1467 (Eco17)</td>
<td>sequence unreadable</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pNZ1468 (Eco18)</td>
<td>sequence unreadable</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pNZ1469 (SalI-21)</td>
<td>not sequenced</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

A proposed explanation for the occurrence of these pBR322 nested deletions is shown schematically in figure A1.2. The insert (thermophilic DNA) and the vector (pBR322) were approximately the same size for the plasmid pNZ1038 and Bal 31 exonuclease digested from both ends at approximately equal rates. Thus for each time point the deleted pBR322 and thermophilic DNA fragments were approximately the same size. Hence a competitive ligation took place between the pBR322 fragment and the Caldocellum saccharolyticum fragment into Eco RI/Sma I cut pCGN565. Previous experiments have shown that E. coli vector DNA ligates and transforms into E. coli vectors at a higher efficiency than thermophilic DNA. Therefore most inserts contained pBR322 vector rather than Caldocellum saccharolyticum DNA.

Appendices
Figure A1.2 Proposed mechanism for deletions from pBR322

Bal31 Exonuclease
Time point deletion

End repaired and cut with Eco RI

Igated into the SmaI, Eco RI site of pCGN 565

Sal I Sma I Eco RI

pCGN 565
2.5 Kd
Appendix Two

Analysis of pNZ1038

This chapter contains data relating to the computer analysis of the sequence of pNZ1038. The sequence of pNZ1038 is shown in Figure A2.1 together with the deduced amino acid translation of the three detected ORF's. The location and number of restriction endonuclease sites detected in pNZ1038 are shown graphically in figure A2.3. Figure A2.2 shows the results of searches of both the GenEMBL DNA database and the Swissprot amino acid database with each of the detected ORF's (ORF1, PepX; ORF2, pulA; ORF3, PepY). These searches identified ORF2 as the pullulanase gene but failed to identify any other homologous sequence.

Figure A2.1 The nucleotide sequence of pNZ1038
As the genes were expressed from different strands the first 1500 base pairs show the top strand (peptide sequence of PepX) and the direction of transcription is left to right, top to bottom. The remainder (1500-4986) shows the bottom or reverse compliment (peptide sequence of PepY and PulA) and the direction of transcription is right to left, bottom to top.

## Database Searches

As the size of the database increases the probability of finding matches caused by random variation increases.

A TFASTA database search of the GenEMBL DNA database with the amino acid sequence of PepY identified a region of high homology to a translation of one reading frame of the *Chlamydia pneumoniae* GroES and GroEL data sequence. However, this region was also located in the middle of another *Chlamydia pneumoniae* gene (in a different reading frame), together with a number of stop codons. Hence, while this translation was homologous to PepY, it is likely to be the result of random chance, as shown below.

\[
\text{# bp in database} \times (6\text{ frames} / 3) \times \text{length} = \# \text{ of comparisons}
\]

\[
95\,025\,892 \times 2 \times 333 = 6.33 \times 10^{10} \text{ comparisons}
\]

The probability of matching 8 amino acids in a row is

\[
1 / 20^8 = 1 / 256 \times 10^8 \text{ i.e. } 2.56 \times 10^{10} \text{ comparisons would be needed to find a region of homology of 8 amino acids by random variation.}
\]
The probability of matching 8 amino acids in a row is

\[
\frac{1}{20^8} = \frac{1}{256} \times 10^8 \text{ i.e. } 2.56 \times 10^{10} \text{ comparisons would be needed to find a region of homology of 8 amino acids by random variation.}
\]

Figure A2.2 Data base searches using ORF1-3 of pNZ1038

Results of a TFASTA search of GenEMBL DNA database with Cspula.pep

TO: GenEMBL:*  Sequences: 73 052  Symbols: 95 025 892  Word Size: 2

The best scores are:

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Frame</th>
<th>initl</th>
<th>initn</th>
<th>opt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>gb_ba:knpulaa</td>
<td>M16187 K.aerogenes pullulanase (pulA) gen...</td>
<td>(3)</td>
<td>195</td>
<td>476</td>
<td>395</td>
</tr>
<tr>
<td>gb_ba:knnulb</td>
<td>X52181 Klebsiella pneumoniae pulA gene for...</td>
<td>(2)</td>
<td>195</td>
<td>472</td>
<td>398</td>
</tr>
<tr>
<td>gb_ba:ecoglg</td>
<td>J01616 E.coli ADP-glucose synthetase (glgC)...</td>
<td>(3)</td>
<td>169</td>
<td>264</td>
<td>355</td>
</tr>
<tr>
<td>gb_ba:psiaim</td>
<td>J03871 P.amylofermoosa amyloamylase gene, com...</td>
<td>(2)</td>
<td>115</td>
<td>215</td>
<td>275</td>
</tr>
<tr>
<td>gb_ba:psiam</td>
<td>M25247 Pseudomonas sp. amyloamylase gene, co...</td>
<td>(1)</td>
<td>115</td>
<td>215</td>
<td>281</td>
</tr>
<tr>
<td>gb_ba:bacglgb</td>
<td>M35089 B.stearothermophilus branching enz...</td>
<td>(3)</td>
<td>99</td>
<td>192</td>
<td>166</td>
</tr>
<tr>
<td>gb_un:m28138</td>
<td>M28138B.stearothermophilus. Neopullulanase ...</td>
<td>(2)</td>
<td>97</td>
<td>130</td>
<td>167</td>
</tr>
<tr>
<td>gb_ba:ecoglgba</td>
<td>M13751 Escherichia coli glgB gene encodin...</td>
<td>(1)</td>
<td>97</td>
<td>123</td>
<td>112</td>
</tr>
<tr>
<td>gb_pl:psgb</td>
<td>M76739 Saccharomyces cerevisiae 1,4-gluca...</td>
<td>(1)</td>
<td>86</td>
<td>121</td>
<td>157</td>
</tr>
<tr>
<td>gb_ba:clopulsa</td>
<td>M57692 Clostridium thermosulfurogenes pul...</td>
<td>(3)</td>
<td>54</td>
<td>114</td>
<td>55</td>
</tr>
<tr>
<td>gb_ba:cloapux</td>
<td>M28471 C.thermosulfurogenes alpha-amyla...</td>
<td>(1)</td>
<td>53</td>
<td>109</td>
<td>60</td>
</tr>
</tbody>
</table>

Results of a FASTA search of Swissprot protein database with Cspula.pep

TO: SwissProt:*  Sequences: 25 044  Symbols: 8 375 696  Word Size: 2

The best scores are:

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>initl</th>
<th>initn</th>
<th>opt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>sw:pula_kleae</td>
<td>P07911 klebsiella aerogenes. pullulanase ...</td>
<td>195</td>
<td>475</td>
<td>395</td>
</tr>
<tr>
<td>sw:pula_klep</td>
<td>P07206 klebsiella pneumoniae. pullulanase ...</td>
<td>195</td>
<td>472</td>
<td>398</td>
</tr>
<tr>
<td>sw:glx_ecoli</td>
<td>P15067 escherichia coli. glx protein. 11/91</td>
<td>169</td>
<td>277</td>
<td>355</td>
</tr>
<tr>
<td>sw:isoa_psea</td>
<td>P10342 pseudomonas amylofermoosa. isoamylase...</td>
<td>115</td>
<td>215</td>
<td>275</td>
</tr>
<tr>
<td>sw:glgb_ecoli</td>
<td>P07762 escherichia coli. 1,4-alpha-glucan ...</td>
<td>97</td>
<td>150</td>
<td>112</td>
</tr>
<tr>
<td>sw:apu_clotf</td>
<td>P16950 clostridium thermosulfurogenes. a...</td>
<td>53</td>
<td>109</td>
<td>60</td>
</tr>
<tr>
<td>sw:amy2_dicth</td>
<td>P14898 dictyoglomus thermophilum. alpha-am...</td>
<td>68</td>
<td>109</td>
<td>121</td>
</tr>
<tr>
<td>sw:amy_bacst</td>
<td>P06279 bacillus stearothermophilus. alpha-a...</td>
<td>44</td>
<td>83</td>
<td>54</td>
</tr>
</tbody>
</table>

Appendices
Results of a TFASTA search of GenEMBL DNA database with PepX

TO: GenEMBL: Sequences: 73 052 Symbols: 95 025 892 Word Size: 2
The best scores are:

<table>
<thead>
<tr>
<th>LOCUS</th>
<th>YSkk1P</th>
<th>8874 bp da-DNA</th>
<th>PIN</th>
<th>15-DEC-1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFINITION</td>
<td>Kluyveromyces lactis killer plasmid k1 DNA. ACCESSION</td>
<td>X07127</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEYWORDS</td>
<td>killer DNA; killer plasmid; overlapping genes; plasmid; toxin;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCORES</td>
<td>Frame: (5) Initl: 65 Initn: 141 Opt: 79 20.4% identity in 113 aa overlap</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PepX

<table>
<thead>
<tr>
<th>yskklp</th>
<th>LLPXGSKKQQESEATIFPLXPSTLFRASFIv{tYl,sC-LIFSILNFISLGISLPLIEYI</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>130 140 150 160</td>
</tr>
</tbody>
</table>

PepX

<table>
<thead>
<tr>
<th>ysXKLp</th>
<th>EiYFEEVKSSEQEIHIVERIYNGKLIPF-WKSEF----EISSFFMENSKNYVVGDK</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>50 60 70 80</td>
</tr>
</tbody>
</table>

PepX

<table>
<thead>
<tr>
<th>yskklp</th>
<th>SIKSRVIIIEAEAXPFSRGSPFGSTYNNFPLPNVSFSFLTLATSTKXLEPSLysiFNSL</th>
</tr>
</thead>
<tbody>
<tr>
<td>170</td>
<td>180 190 200 210 220</td>
</tr>
</tbody>
</table>

PepX

<table>
<thead>
<tr>
<th>yskklp</th>
<th>LRYISIFFFLPPQIVRRHFKVATKRGFYKDKIYTVGLDLGLATADRCYVYNDTLTVY</th>
</tr>
</thead>
</table>

Results of a FASTA search of Swissprot protein database with PepX

TO: SwissProt: Sequences: 25 044 Symbols: 8 375 696 Word Size: 2
The best scores are:

<table>
<thead>
<tr>
<th>ID</th>
<th>COXI_LEITA STANDARD: PRT; 549 AA.AC P14544;</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCORES</td>
<td>Initl: 54 Initn: 95 Opt: 59 34.6% identity in 26 aa overlap</td>
</tr>
</tbody>
</table>

PepX

<table>
<thead>
<tr>
<th>cxG1_1</th>
<th>HYVLSLGAVGFGTGFHFLAKWPIELYLFWMFYISTFLFIGSNMLFFPHSLGMYAFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>380</td>
<td>390 400 410 420 430</td>
</tr>
</tbody>
</table>

ID       | NF1_HUMAN STANDARD: PRT; 2485 AA.AC P21359;              |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SCORES</td>
<td>Initl: 40 Initn: 94 Opt: 47 33.3% identity in 30 aa overlap</td>
</tr>
</tbody>
</table>

PepX

<table>
<thead>
<tr>
<th>nFl_hu</th>
<th>GLCIPANNTLPIVSIKTALANEPHILFLLEECIGSFKSSIELKHHCLLEYMTFWLNSL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1550</td>
<td>1560 1570 1580 1590 1600</td>
</tr>
</tbody>
</table>

Appendices
Results of a TFASTA search of GenEMBL DNA database with PepY

TRANSLATE of: therm.rev check: 1477 from: 1 to: 1000
generated symbols 1 to: 333.
REVERSE-COMPLEMENT of: therm.seqqq check: 4233 from: 1 to: 4986
FROMSTADEN of: Therm.Datt check: 6267 from: 1 to: 4985

TO: GenEMBL:* Sequences: 73 052 Symbols: 95 025 892 Word Size: 2
The best scores are: frame initl initn opt..

gb:ba:chtgro M69217 Chlamydia pneumoniae GroES and GroEL...(2) 139 209 227
gb:or:riccposxx X15901 Rice complete chloroplast genome ... (4) 48 121 50

gb:or:riccposxx M19359 Rat gamma-cystallin gene cluster, ... (5) 62 109 77

gb:or:riccposxx M62622 S. cerevisiae mitochondrion DNA, com... (6) 44 107 48

gb:or:riccposxx X04465 Marchantia polymorpha (liverwort) c... (5) 53 107 54

DEFINITION Chlamydia pneumoniae GroES and GroEL genes, complete cds.ACCESSION M69217
SCORES Frame: (2) Initl: 139 Initn: 209 Opt: 227 45.6% identity in 90 aa overlap

PepY YAEFEKEILKSRESEPAETLCLNIGWNLKKTYGDNEVDKLATILIWARPHYMNFIIVY
chtgro ISHRKQXLPRRYGUVTSDSLSALEWARPHYMNFIIVY

PepY KYATSMAAANEVVKNT--EKGTNTYKI-EFLKAGSSDPINVLKAGVDTMSTKPVNDLL
chtgro QYATGIIAALSFAEKILTQEPGSLNFIXNFLKSGRDFPLNIKLKSGDMTTSASIRXSL

PepY TYFGQLVDEMEKILKKQGKTXFQRWDRWXMRKGDYSGKVNPLEVFSLKLLGYYKINK
chtgro CIHYEKDRPTFLAFRRLSFHLKLSSAKIIAQKKTLCYRDCRNSKVSNVKHKEHINVX

Results of a FASTA search of Swissprot protein database with PepY

TO: Swissprot:* Sequences: 25 044 Symbols: 8 375 696 Word Size: 2
The best scores are: initl initn opt..

sw:apb_human P04114 homo sapiens (human). apolipoprotein...
sn:apb_human P04114 homo sapiens (human). apolipoprotein...
sn:naca:canfa P23685 canis familiaris (dog). sodium/calc...
sn:ace_rabbit P13282 oyster glycogen (rabbit). angi...
sn:ace_rabbit P13282 oyster glycogen (rabbit). angi...
sn:ace_rabbit P13282 oyster glycogen (rabbit). angi...
sn:ace_rabbit P13282 oyster glycogen (rabbit). angi...
sn:ace_rabbit P13282 oyster glycogen (rabbit). angi...
sn:ace_rabbit P13282 oyster glycogen (rabbit). angi...
sn:acy_rabbit P13282 oyster glycogen (rabbit). angi...
sn:acy_rabbit P13282 oyster glycogen (rabbit). angi...
sn:acy_rabbit P13282 oyster glycogen (rabbit). angi...
sn:acy_rabbit P13282 oyster glycogen (rabbit). angi...

ID APB HUMAN STANDARD: PRT: 4563 AA. AC P04114;
DE APOLIPOPROTEIN B-100 PRECURSOR (APO B-100/APO B-48). . .
SCORES Initl: 57 Initn: 89 Opt: 59 20.8% identity in 24 aa overlap

PepY GTVYTQVMYAEFEKEILKSRESEPAETLCLNIGWNLKKTYGDNEVDKLATILIWARPI...
apb_hu GTVYTQVMYAEFEKEILKSRESEPAETLCLNIGWNLKKTYGDNEVDKLATILIWARPI

Appendices
Figure A2.3 Location of restriction endonuclease sites detected from the DNA sequence of pNZ1038
Appendices

Appendix Three
Peptide Comparisons

The comparisons of the deduced peptide sequence of the *C. saccharolyticum* pullulanase gene to a number of other related sequences are shown in this section. Sequences compared are shown in table 3.1.

Table 3.1 Pullulanases and other sequences used in peptide comparisons

<table>
<thead>
<tr>
<th>Source</th>
<th>Enzyme</th>
<th>Computer Code</th>
<th>Figure</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caldodcellum saccharolyticum</td>
<td>pullulanase</td>
<td>CSPULA</td>
<td>-</td>
<td>chapter 4 and 8</td>
</tr>
<tr>
<td>Bacillus stearothermophilus TRS128</td>
<td>pullulanase</td>
<td>BSTPULA</td>
<td>A3.1</td>
<td>Kuriki <em>et al.</em> 1990</td>
</tr>
<tr>
<td>Thermus sp. AMD33</td>
<td>pullulanase</td>
<td>THPULA</td>
<td>A3.2</td>
<td>Nakamura and Sashihara 1991</td>
</tr>
<tr>
<td>Klebsiella oxytoca UNFS023</td>
<td>pullulanase</td>
<td>KLPULA</td>
<td>A3.3</td>
<td>Kornacker and Pugsley 1990</td>
</tr>
<tr>
<td>Pseudomonas sp SMP1</td>
<td>isoamylase</td>
<td>PSEIAMA</td>
<td>A3.4</td>
<td>Tognoni <em>et al.</em> 1989</td>
</tr>
<tr>
<td>Bacillus stearothermophilus TRS40</td>
<td>neopullulanase</td>
<td>BSTNPL</td>
<td>A3.5</td>
<td>Kuriki and Imanaka 1989</td>
</tr>
<tr>
<td>Clostridium thermohydrosulfuricum</td>
<td>α-amylase-pullulanase</td>
<td>CTHAPUX</td>
<td>A3.6</td>
<td>Melasniemi <em>et al.</em> 1990</td>
</tr>
<tr>
<td>Clostridium thermosulfurogenes</td>
<td>α-amylase-pullulanase</td>
<td>CTSPULA</td>
<td>A3.7</td>
<td>Burchhardt, <em>et al.</em> 1990</td>
</tr>
<tr>
<td>E. coli</td>
<td>α-1,6-Branching enzyme</td>
<td>ECBRA</td>
<td>A3.8</td>
<td>Romeo <em>et al.</em> 1988</td>
</tr>
<tr>
<td>Bacillus thermoglucosidasius KP1006</td>
<td>oligo-1,6-glucosidase</td>
<td>O16BRA</td>
<td>A3.9</td>
<td>Watanabe <em>et al.</em> 1991</td>
</tr>
<tr>
<td>Bacillus circulans</td>
<td>cyclodextrin glycosyltransferase</td>
<td>CGT</td>
<td>A3.10</td>
<td>Klein and Schulz 1991</td>
</tr>
<tr>
<td>Pig pancreatic α-amylase</td>
<td>α-amylase</td>
<td>α-AMY</td>
<td>A3.11</td>
<td>Buisson <em>et al.</em> 1987</td>
</tr>
</tbody>
</table>

The enzymes selected were every known example of enzyme that hydrolysed pullulan or debranched amylpectin (4 pullulanases, 2 α-amylase-pullulanases, 1 neo-pullulanase, 1 isoamylase) and four enzymes that were representative examples of related enzymes (α-1,6-branching enzyme, oligo-1,6-glucosidase, cyclodextrin glycosyltransferase and α-amylase). Other sequences were also compared but showed no significant homology (e.g. β-amylase and glucoamylase) and were therefore not included. Where possible the representative sequences of the related enzymes were thermophilic eubacteria, except for the α-amylase sequence where the pig pancreatic sequence was used as this enzyme has been extensively characterised with respect to its 3-dimensional structure and amino acids involved in catalysis, substrate binding and metal ion binding.
These comparisons are shown in two ways on the following pages (figures A3.1-A3.11). The first comparison is an alignment of the amino acid sequences using the GCG program BESTFIT (see section 2.7.4). Identical amino acids in each sequence are linked by a vertical line (|), closely homologous amino acids by a colon (:) and slightly homologous amino acids by a dot (.). The second is a graphical representation of homology using the GCG program COMPARE. This program compares each amino acid in C. saccharolyticum pullulanase to the entire length of the other sequences. The stringency was set higher than the default value to reduce the background and show only highly conserved regions.

### Table A3.2 Percentage homology between amino acid sequences

<table>
<thead>
<tr>
<th>BSTPULA</th>
<th>59.4 (38.6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>THPUL</td>
<td>55.6 (31.0)</td>
</tr>
<tr>
<td>KLEBPULA</td>
<td>51.0 (28.7)</td>
</tr>
<tr>
<td>PSEIAMA</td>
<td>47.9 (24.1)</td>
</tr>
<tr>
<td>BSTNPL</td>
<td>48.6 (22.6)</td>
</tr>
<tr>
<td>CTHAPUX</td>
<td>45.7 (20.1)</td>
</tr>
<tr>
<td>CTSPULA</td>
<td>46.5 (20.2)</td>
</tr>
<tr>
<td>ECBRA</td>
<td>49.5 (21.4)</td>
</tr>
<tr>
<td>O16GLU</td>
<td>45.1 (17.3)</td>
</tr>
<tr>
<td>CGT</td>
<td>41.4 (17.4)</td>
</tr>
<tr>
<td>α-AMY</td>
<td>45.3 (21.0)</td>
</tr>
<tr>
<td>CSPULA</td>
<td>54.3 (18.2)</td>
</tr>
<tr>
<td>BSTPULA</td>
<td>59.4 (38.6)</td>
</tr>
<tr>
<td>THPUL</td>
<td>55.6 (31.0)</td>
</tr>
<tr>
<td>KLEBPULA</td>
<td>51.0 (28.7)</td>
</tr>
<tr>
<td>PSEIAMA</td>
<td>47.9 (24.1)</td>
</tr>
<tr>
<td>BSTNPL</td>
<td>48.6 (22.6)</td>
</tr>
<tr>
<td>CTHAPUX</td>
<td>45.7 (20.1)</td>
</tr>
<tr>
<td>CTSPULA</td>
<td>46.5 (20.2)</td>
</tr>
<tr>
<td>ECBRA</td>
<td>49.5 (21.4)</td>
</tr>
<tr>
<td>O16GLU</td>
<td>45.1 (17.3)</td>
</tr>
<tr>
<td>CGT</td>
<td>41.4 (17.4)</td>
</tr>
<tr>
<td>α-AMY</td>
<td>45.3 (21.0)</td>
</tr>
<tr>
<td>CSPULA</td>
<td>54.3 (18.2)</td>
</tr>
</tbody>
</table>

Values are similarity percentages obtained from the GCG sequence alignment program BESTFIT. Two values are shown for each enzyme comparison: The first value is the percentage similarity, which takes into account both identical and similar amino acids, the value in brackets is the percentage identity and is obtained when only identical amino acids are counted.

**Abbreviations for amino acid sequences are as follows:**

- **CSPLUA**: Caldocellum saccharolyticum pullulanase
- **BSTPULA**: Bacillus stearothermophilus pullulanase
- **THPUL**: Thermus sp. AMD33 pullulanase
- **KLEBPULA**: Klebsiella oxytoca UNF5023
- **PSEIAMA**: Pseudomonas sp. SMP1 isoamylase
- **BSTNPL**: Bacillus stearothermophilus neopullulanase
- **CTHAPUX**: Clostridium thermohydrosulfuricum α-amylase-pullulanase
- **CTSPULA**: Clostridium thermosulfurogenes α-amylase-pullulanase
- **ECBRA**: E. coli α-1,6-Branching enzyme
- **O16GLU**: Bacillus thermoglucosidasius KP1006 oligo-1,6-glucosidase
- **CGT**: Bacillus circulans cyclodextrin glycosyltransferase
- **α-AMY**: Pig pancreatic α-amylase

**Appendices**
Figure A3.1 Amino acid sequence comparison between the C. saccharolyticum pullulanase and the Bacillus stearothermophilus TRS128 pullulanase

<table>
<thead>
<tr>
<th>C. saccharolyticum pullulanase</th>
<th>Bacillus stearothermophilus TRS128 pullulanase</th>
</tr>
</thead>
<tbody>
<tr>
<td>KYECIDILKQIGKQFAPGSRKLYWYFCGYN</td>
<td>FFDDREAEATISNYKTTTTQVLMANLKYVKNKNEKAPKELTVNAYFAGLARYE</td>
</tr>
<tr>
<td>EYCTRMTWTPYTVKQISDHEHNFKQGKNGNMAETDMDLYPGGLNKHYESIHYNYEDDIQIVYRVDPFYSASSSNQSGESTFIPDAGLID</td>
<td></td>
</tr>
<tr>
<td>156</td>
<td>144</td>
</tr>
<tr>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>315</td>
<td>315</td>
</tr>
<tr>
<td>214</td>
<td>214</td>
</tr>
<tr>
<td>145</td>
<td>145</td>
</tr>
<tr>
<td>308</td>
<td>308</td>
</tr>
<tr>
<td>144</td>
<td>144</td>
</tr>
<tr>
<td>340</td>
<td>340</td>
</tr>
<tr>
<td>437</td>
<td>437</td>
</tr>
<tr>
<td>612</td>
<td>612</td>
</tr>
<tr>
<td>709</td>
<td>709</td>
</tr>
<tr>
<td>538</td>
<td>538</td>
</tr>
<tr>
<td>800</td>
<td>800</td>
</tr>
<tr>
<td>638</td>
<td>638</td>
</tr>
</tbody>
</table>

Appendices
Figure A3.3 Amino acid sequence comparison between the C. saccharolyticum pullulanase and the Klebsiella oxytoca UNF5023 pullulanase

**CSPULA**

3 VRAYDFDEFNVIYTVAILQVGVKVEKEDFQVFVVFVIQEDIGIEKVERKLMPHSENPEAE...FGTECQDKQKIRFVIAED... HDHRAK HDHRAK HDHRAK HDHRAK

**KLPUL**

30 GPIDEPTIQNVVVRTPQV. AVPIEVAATVVENQDVKHVLGDSGAGTSSAAYSSKKNLJHNNETCDASAPVMDAMNVTSTPSGDGPQCYTVWVNILKES 127

**CSPUL**

86 TVRFVFGIDGMENVWLSEPEWELTYSLKAR....YELIKLDKDIQCKFBEAGSK RKKNYPVDGQIVDVIDS ... DREAFTFM 166

**KLPUL**

128 GCINVWIVODGDKL. IDSDLILAVIFDPDTRTVGSAIVNASDRAFADCRAFCAVOLAEAIWHDKNLIPQCQDPFIVSRYSHSSKAVGDDQFEDK 226

**CSPUL**

167 VRITTKRLMNALTKYVTYKETRYARPLHEILNREFFEYFADOL... GTEQSTCTTFLAPWTAQYKYK 231

**KLPUL**

227 YLKLPTTPVSQVSMRFFHLSSYAFKLDPMVVLQGETVRAIALEDGILSATQYKTGDLVDIRAALEALELYGQALGAQVFFMDWAPTAQVQ 326

**CSPUL**

232 IGIFDEHENKFKGERMRS...AENGTWDVYIGDGLKNHPYLVETHWYEDDGCYTVYPDPSKASSSSNGSKYIFWCDAGOL...DWQADSFDVNIKEQ 328

**KLPUL**

327 VYVYXADKXYGSHYFMRDEEAGSKNQGGDGSDKCAGRYXKMYRNYN... FGKXEYVQTVDPTAILNLSRETSQTVLVDKALSGDPPQGLQEIHPQDRTK 425

**CSPUL**

329 DE...ALYEMFPQDFTIDQSGGCGGNNLCRGFLGQECQEQYGQDGIISTGLHNL...GTVYHLLPLSIDFVSVDD... 398

**KLPUL**

426 ADLAMKTHEHDBLS.AWQTVPTEALGKTYILAL...TADSMSMVQKLELTSAGSYTHVELFIVDIAVNEFSDKVADIQPFSSKLCEVNSAVK 518

**CSPUL**

399 KNPDO...KRNQNYGVPYLQCEFWYSTRSCHCEALKERTMLKXNLG.IGUVMDVAVHVTY...H 459

**KLPUL**

519 SEFACGDGSGTVFEINJLQKDQPSQMAPQVTAIQTVGDYDYSQYMBNGFQDFFHTVPSGSAIDPKEESTERIKFRTYQALSKQGQGGVWNYSHNL 518

**CSPUL**

460 TKGKAFSJDKIFKVFTTVFDVYDYSQYMBNGFQDFFHTVPSGSAIDPKEESTERIKFRTYQALSKQGQGGVWNYSHNL 518

**KLPUL**

619 GPTORTSLDKVMFYQRNETGGSVASEATCCDSSAPFMKFLKLIASLVWTDTTYKIDGRDFMDTHHPAQILSAWRIKALAPDITFGEQGWSDR 718

**CSPUL**

560 NSTICKSMATILSSQNGQSICLGDNRDIIRAI... DLDGYKQYHN... GNLQGVRLEQK 616

**KLPUL**

719 QRSFF...EIAGTX.LKTCGQFQDD0HLSAVGGDPDSQOLAQQGQGSGGSGAVNPIEASDSDQVRHLADTLRSLMGANLGVDFVMDGDKGAEKK 814

**CSPUL**

617 DTLKAIPOFASISKPVSSQICDGIVNPMT Interestingly...IAQKTVGFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQFIVQFVQ 712

**KLPUL**

815 SKIDYNAGGQDYAUPETQEEVYSDKQHNLQICTYYSKASEAQLATLVWKRQVLDTVQVLVIAIQGCOIAFDSELHLRAKSGFLXQSVFORDYVDSIQD 914

**CSPUL**

713 KQDF... Interestingly... KFRCYWDNLXIRHARFMRSSGEGKIRYKAPFPCDFPAVPTTPYDEWKK 769

**KLPUL**

915 NHTYVMPREDSGSYEVTIRKHVPATCAGREELKURKSFSLPFLGDDGSSAVQKVFSLRTNGDDCOAGLVMVTDQDMGACASLISR 1014

**CSPUL**

770 . AILYEMFPQDFTIDQSGGCGGNNLCRGFLGQECQEQYGQDGIISTGLHNL...GTVYHLLPLSIDFVSVDD... 1015

**KLPUL**

GUVAIIAPESTRKIFQSTGIFLQCLAQTGATGSLQVIAAOQTVLFLPAHNSAVI 1075

Appendices
Figure A3.4 Amino acid sequence comparison between the *C. saccharolyticum* pullulanase and the *Pseudomonas* sp SMPI isoamylase

| 12 | EIVTVLAQWHVYVKKDFKVLDFSQEDIEKVEKLMPSRENPFEEATAGYICIEQKKGKIVFLKDEIFDFHKSTVRQVYFVGONNVKTSPFEMELTSR 111 |
| 2  | KCPI1ALLGCAVLAGPMRPMNQ..SM1CAGI3DIQQFFRNYSSQATIVILYSAGYQGGAT...YLLSPA 78 |
| 112 | LAGRYELIKDLKDIQSKFKEAGSRKMLWYYPGGNYEDFYDEAARNTMVKTITTKRLNANLKLKVNYEKTARFRELTRNYYFFAGELG 231 |
| 79 | GQSWAVTVPWSSK...AAGITGAVTYGTRA...WDPNYPYASNMKOGQAG...FYSVVIDA 132 |
| 212 | RYEPYTGYWPAWATYKVQQIQDEHENKFKEKMSRMGAEINWVYPDILKHFEYLFYEEVKVYDEDDGTVYVFDDTSKASSNNGKSTIFDPAD 311 |
| 133 | NCDRPNKKLLDPPYAEQGVDPLN...PHQNG...NHTGSAASYRTDGSYGIPKGV 185 |
| 312 | LIDGWQADS TDVNI KQGOAIYZEMVRQPFTIDQSSGTCEMLKFKFLGCQGEGYKGEISTGLHLKELTVHHLPI...SDFGSVDXKPKNYNV 497 |
| 186 | VLPVSTQSTCTKTRAGQVYIEYHNKQFT...EQDISFAQXQXGAIQKASY...LASIGVTVHLEFGQXTQDNDVWNKDANQNYV 274 |
| 408 | GYPDVLQCPQEPYNYSTKSCGIPQDLKELRTMIKLHENGIOVYNVVFNNTH...TKGKRBPIDKIVPEYTYVDYDGYSDHACNELEIT 498 |
| 275 | GYMTENSYSPRASYNAAAGDPTEAEFCMQVQAFAINQIKVYNYVNYHTAEGTWTSSDPITATTIYSWRGDLNNTTEYLTSTQFYNLNGNTIGNFNTY 374 |
| 499 | KPMVRFQGITEDTVMPFQZDPDLNGIL...TVQGSRQAVNEARPKVALYQGGWQNCTCGLVEEMAITLSG 574 |
| 375 | NTVAGLQPVSLAPLHAPMVGFDGRFSDVLGILSCNGAYTASSAFNCPNGYFDANDIYANLIRRVFVFYRAACGGSSLDFPAEWAIONQYQGQS 474 |
| 575 | CHQYSI..GLFNDKHKDQGDLDEYKTGYVGNLVLQKFLIKAAIDAFFAKIDDECVTVSCHUNILTFD 646 |
| 474 | FPQCSNGWILDGFDRLQA..QNELAS..MTLTYTQDANS..GSSLFQSSGSRPWSNQINFHDQMDLQYSCNGAGAASNQRFFWPCDGTSTNY 568 |
| 647 | KLPKTVMEGDIPW..AATLNASVLTVQGQVAFLHGYEFVSNEGIKQHPNYDTITIVCDINLRKVIAPFFRSSSSEGCEK 744 |
| 569 | SWQCGMADAETGQAOQRAARTCGMEGLASATQMQDSLQCGQATLQQGNSAMKLTVKINDEXQSGMQTYTAPQLNAPRLAFFNSWSVQYDQO 668 |
| 745 | YLFIPAPDO..............VAVFITTYP..DYEEKRVAYTNVPKEKKQLPPGGSN...........YVKHHDNVRPFD..PEKAIASGFEAPV 817 |
| 669 | LIVWKPYGAVADENNYWNTSNTAYATAYNGPGLDENSVYAVGNNSSSTVTPLPAPSGQYRVDCTCWDQGASTFVAPGETLIGGATTTYGCCQ 768 |

**Appendices**
Figure A3.5 Amino acid sequence comparison between the C. saccharolyticum pullulanase and the Bacillus stearothermophilus TRS40 neopullulanase

73 VLKEGHSDFIRKSTVARPVFGMDMNWTSPEWELTYSKLGRYELIKLKDGIKIQKFKEAGDACR.LIWYPFGYNIVIEDFDEEAATINVKT171
1 HRKAEITH...RPA....DFASYHGETLMAATLTKDKIDK.ELRHGDFPDQWNGAWQYQYMKTEGDFEALDQYFRA......V.K.PP 78
172 TKRLNANLKKYVWYKYCKPFLRILTREYFAEGLARYEYPGTYGFRFRLWAPTYAKVIIQIFDEHENFKGKREMSRAENGTWDYVGDLKKNHYLYE 271
272 IWHNYEDGYIYHVPPDFSKASSNSGKSTIFDFAPDGLIDQWDSSFDND1EKQDDALYYEHMDFTDIQSGSIGEINL6RKFLODFCQEGTYREGIS 371
140 IPEAEPANPSIIPSEGRPM\_CEEPFPT2HYCGDLQIDIMLDO......183
372 TGLRLKELGTVTHILLPSDFGPVDXKDNRWYGPDLYCPEWYSTSKGS1ALKELKMKKLEENGIVQPMGVFNNTYHR......KGG 463
184 ...YLVQDLLTIGYIPLF...FRPSNLYKHTDADYFVEQDFPF,...GDKETLK...TLDCHKCIKIVMDLFVHNgCIEYEFAPFAQQVWNKG 262
464 KPSIFDK...IPEYFYPVDDYDYDNSATGCGN...ELATKPWVRAFILTQYTWEEFHDGFPRFDLMGDLTVQSRQVANEKFKNPLLYYGEMWNG 559
263 ESQYKWIFHIHEPPPLTERPYNDRFYPQGPKINTARVEKRYLLDVATWIREPDIGDIDLDOVANE1DHEWFKRESQEVKALKPSYLGEINHDA 362
560 NSTC1LEMAI1LSSCQOY6TSGFD3IAID1GRGDDVYCTVHGNLLGVDRLKLQCIKAIADDFKKEPE. CUNYVSCHKDNLJFOGKTQMTVEDIF 658
363 MPWQLGQFQDAMNY...PFTQGVLEFFAEKEISAR..............QFAAGQHMLHVLYKNNVNEAAGFLESHTGISL...TVQGDD 436
659 WIDRATIANALVISSQGVAFILIIGVEFKNASGGISHNPTYNAGD..NINRDKMLKKEFFDD7TPFKYCODLINIAKHIAEFBMRSSGEIKYKLRFIAPGDOVGA 757
437 ...RVRVLLTFQIJFTGCPYTCIGDIAGTDOODPCECKCMWNPDQOQKEEASLHQHQLVLLAI3KQYQRSLAEGISL.LMADDN.WLYIYKRDGDSTT 531
758 FIITFYDEWKKIIIVAYNPKEKKILQLPESNIVANKQDGGVFPDSFEEAIASGEISAP 816
532 LVIZN3SQQAKDFPFLDA.R6STKLYNLL7GER...KANRSTICTELPPPYGIVNIAE1 587

Appendices
Figure A3.6 Amino acid sequence comparison between the *C. saccharolyticum* pullulanase and the *Clostridium thermohydrosulfuricum* α-amylase-pullulanase

<table>
<thead>
<tr>
<th>Position</th>
<th>C. saccharolyticum</th>
<th>C. thermohydrosulfuricum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10</td>
<td>MIVKAV1DPN7VFYTV4AQWHSVRRE</td>
<td>RIVKAV1DPN7VFYTV4AQWHSVRRE</td>
</tr>
<tr>
<td>21-80</td>
<td>VFRKAAFLGFALFLFLFLYFTAVGSMMPFMFAKADTDAHALAIYNYGNQPLQKLSIDKWNSLDKSIDKTEDYKFGYTEEFPVTVPALPAGDYEVKADNSMMGGY</td>
<td>RFRKAAFLGFALFLFLFLYFTAVGSMMPFMFAKADTDAHALAIYNYGNQPLQKLSIDKWNSLDKSIDKTEDYKFGYTEEFPVTVPALPAGDYEVKADNSMMGGY</td>
</tr>
<tr>
<td>91-100</td>
<td>PTQSG1LPDFSLDSUPVTYYNTSSLATTSTRTYTFIPEDKL</td>
<td>PTQSG1LPDFSLDSUPVTYYNTSSLATTSTRTYTFIPEDKL</td>
</tr>
<tr>
<td>111-170</td>
<td>NDIVIDEYDREAAFTMKVIVVAKNKLKYYVNKNKRAPELRTKRENYYFAYGELGAYEYFYGTYFR.LNAP..AYKVKIQTIDFEHNFKPGK</td>
<td>NDIVIDEYDREAAFTMKVIVVAKNKLKYYVNKNKRAPELRTKRENYYFAYGELGAYEYFYGTYFR.LNAP..AYKVKIQTIDFEHNFKPGK</td>
</tr>
<tr>
<td>181-290</td>
<td>149</td>
<td>149</td>
</tr>
<tr>
<td>301-400</td>
<td>248</td>
<td>248</td>
</tr>
<tr>
<td>411-510</td>
<td>334</td>
<td>334</td>
</tr>
<tr>
<td>521-620</td>
<td>404</td>
<td>404</td>
</tr>
<tr>
<td>631-730</td>
<td>481</td>
<td>481</td>
</tr>
<tr>
<td>741-840</td>
<td>547</td>
<td>547</td>
</tr>
<tr>
<td>851-950</td>
<td>618</td>
<td>618</td>
</tr>
<tr>
<td>961-1060</td>
<td>686</td>
<td>686</td>
</tr>
<tr>
<td>1071-1170</td>
<td>755</td>
<td>755</td>
</tr>
<tr>
<td>1181-1290</td>
<td>825</td>
<td>825</td>
</tr>
<tr>
<td>1301-1400</td>
<td>893</td>
<td>893</td>
</tr>
<tr>
<td>1411-1510</td>
<td>961</td>
<td>961</td>
</tr>
</tbody>
</table>

Appendices
Amino acid sequence comparison between the C. saccharolyticum pullulanase and the Clostridium thermosulfurogenes α-amylase-pullulanase
Figure A3.8 Amino acid sequence comparison between the C. saccharolyticum pullulanase and the E. coli α-1,6-Branching enzyme

250

Appendices
Figure A3.9  Amino acid sequence comparison between the *C. saccharolyticum* pullulanase and the *Bacillus thermoglucosidasius* KP1006 oligo-1,6-glucosidase

```
326  EKQDDAIVEMAVPDFTDIQQSGIGENLGKGLGFCQGQYYKREGISGGLKIHKLOVTVHIILPLSI
dGQGVDIDMNIDPRYKWGDPVLYGQPEYWYSTK
3  MQKREAVYVQYPRSYDSNDQGDGD..............LQGVQKLDKIRIADVWLCP...FDSPQDDN........
425  GSGIALKELGTMKKLHNCCGVMQVFFTYNTHTK...GDKFSIFDKLVEPFYVRVDQGDYDNYGATGNELAT....
79  DMFQ......LIDEVRKROMKIMLDLRVNNHSDRASAAERKSSKDNFYTCRYCVPDKRAGGSEPNWGNQAI
171  EKPMVLRKIDTTIYTWTEEFHDIGFRLMGGLTVKQSRQVGANEVRKRPKALHYGEGQWVMNGSTLCVLLEMATILSCHQGYSI
172  ENEAVREYDLFTMFWDR.GVDQGRMDVIGIS........KFVDPFDYETDDRFTYV.......VGYSNSN.\GPRHEFQEMRNAEVS
245  598  GYKTDYV.HGNLSDVRKQKCIKAAIDF\AKEPDECVN.........................YVCHONLIFDKLQKTMVG
246  4RDCMTYVGEAGGOSQVEAKTYDPSREHNKMTTFFEHMDITKQHSPNGKQMDFDPIALKTTMRWQTALMVGWNLYENHOQPRVSI
345  655  EDIFWIDRAFLANAYLTVQQVAPFIRGQVEFIRKKGQNHPTYNAGDNIKIDWSL..KERRYTDPKFYCDILNLRKHKFAFRMS655QI
346  ..ELRKQRSQISNSSA.RHCRNPFGYQGIECMTSNEPLEMYYDDLEIKNAYRELVIENKTMTEEDFRXAVAKKGRDHRARPMQMDG....KYAGTF
440  753  DCWAPFIYTPYDHEKIIWAYNPKFKEKIIQLPEGSWY........KANDQGVPFD...SFEKEA.IGSF..EIAPV\LFIA 822
441  ...........EAKLAVNPFRYQIKXSSGKNLADSDIFSYYQKQGLRQKQMVVYGVRLLLEDPFRATYIREYGERKL\VPE 509
```
Figure A3.10  Amino acid sequence comparison between the \textit{C. saccharolyticum} pullulanase and the \textit{Bacillus circulans} cyclodextrin glycosyltransferase.

Appendices
Figure A3.11 Amino acid sequence comparison between the *C. saccharolyticum* pullulanase and the pig pancreatic α-amylase
Appendix Four

Genetic Analysis of *Caldocellum saccharolyticum* Genes

Another 21.375 kb of *Caldocellum saccharolyticum* sequence had been previously sequenced by other members of the thermophile genetic group at Auckland University (Love and Streiff 1987, Saul et al. 1989, Lüthi et al. 1990a, Te'o 1992 and Gibbs et al. 1992; see Section 1.5). These sequences comprised three major fragments, a lambda genomic recombinant 13.606 kb ($\lambda_2A$) and two recombinant plasmids pNZ1001 (1.702 kb) and pNZ1400 (6.067 kb). A schematic representation detailing the location of the genes found is shown in Figure A4.1.

**Figure A4.1** Location of other *C. saccharolyticum* genes previously sequenced.

![Diagram showing gene locations and sequences](attachment:image.png)

An analysis of these DNA sequences was performed to aid in the characterisation of the sequence of pNZ1038. Due to the all encompassing analysis performed and the larger data set available there is some variation in the location of features compared with previously published results.

**Transcriptional and Translational Initiation Sequences**

As the *C. saccharolyticum* genes were expressed in *E. coli* from their own DNA sequences recognised by the *E. coli* transcriptional and translational cellular machinery, it was logical to assume that these sequences were homologous to the standard *E. coli* consensus initiation sequences. Using these initiation sequences, searches for ribosome binding sites and RNA polymerase initiation sequences were performed upstream of the start codon of each *C. saccharolyticum* gene (see tables A4.1 and A4.2).

Appendices
Table A4.1 Putative *C. saccharolyticum* ribosomal binding site

<table>
<thead>
<tr>
<th>Gene</th>
<th>Ribosomal binding site.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> consensus</td>
<td>AGT GGAGG AAT</td>
</tr>
<tr>
<td><em>B. subtilis</em> consensus</td>
<td>AAA GGAGG TGA</td>
</tr>
<tr>
<td><em>bgl A</em></td>
<td>AAA GGAGG TTT</td>
</tr>
<tr>
<td><em>xyn A</em></td>
<td>GAA TGAGG TGT</td>
</tr>
<tr>
<td><em>xyn B</em></td>
<td>AGA GGAGG AAA</td>
</tr>
<tr>
<td><em>xyn C</em></td>
<td>AAA GGGGT CTT</td>
</tr>
<tr>
<td><em>cel A</em></td>
<td>TGA GGGGG TTA</td>
</tr>
<tr>
<td><em>man A</em></td>
<td>AGA GGGGG TTA</td>
</tr>
<tr>
<td><em>cel B</em></td>
<td>AGA GGGGG TTA</td>
</tr>
<tr>
<td><em>cel C</em></td>
<td>AGA GGGGG TTA</td>
</tr>
<tr>
<td><em>pep X</em></td>
<td>AAA AGAGG TAA</td>
</tr>
<tr>
<td><em>pul A</em></td>
<td>TGA ATAGG TGA</td>
</tr>
<tr>
<td>Consensus</td>
<td>AG/A GG/GAGG TTA/TT</td>
</tr>
</tbody>
</table>

Table A4.2 Putative *C. saccharolyticum* RNA polymerase gene sequences

<table>
<thead>
<tr>
<th>Gene</th>
<th>-35 Region</th>
<th>-10 Region</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> consensus σ70</td>
<td>TTGATT</td>
<td>TAATAT</td>
</tr>
<tr>
<td><em>B. subtilis</em> consensus σ43</td>
<td>TTGATT</td>
<td>TAATAT</td>
</tr>
<tr>
<td><em>bgl B</em></td>
<td>not sequenced</td>
<td>TATTAT</td>
</tr>
<tr>
<td><em>xyn A</em></td>
<td>TACTTA</td>
<td>TAAAAT</td>
</tr>
<tr>
<td><em>xyn B ORF 5</em></td>
<td>TTATTC</td>
<td>TAGACT</td>
</tr>
<tr>
<td><em>xyn C ORF 2</em></td>
<td>AAAAGA</td>
<td>TAAGAT</td>
</tr>
<tr>
<td><em>cel A</em></td>
<td>not sequenced</td>
<td>TAATAT</td>
</tr>
<tr>
<td><em>man A</em></td>
<td>AATAGT</td>
<td>TAAAAT</td>
</tr>
<tr>
<td><em>cel B</em></td>
<td>CCAATT</td>
<td>TAAAAT</td>
</tr>
<tr>
<td><em>cel C</em></td>
<td>AAGTTT</td>
<td>TACGAT</td>
</tr>
<tr>
<td><em>pep X</em></td>
<td>TTGAAT</td>
<td>TATAAT</td>
</tr>
<tr>
<td><em>pul A</em></td>
<td>TTAAAAA</td>
<td>TAGAAT</td>
</tr>
<tr>
<td>Consensus Sequence</td>
<td>T/AAA/ATT</td>
<td>TAA/ATT</td>
</tr>
</tbody>
</table>
Direct Repeats

Repeat sequences have been detected in other *C. saccharolyticum* genes with the most significant being the 13 bp sequence located 7-8 bp upstream of the ATG start codons of the 4 genes of λ2A. This sequence is not present elsewhere in other *C. saccharolyticum* sequences or in any other sequence in the GenEMBL DNA database. The location of the sequence near the translational start implied that it could be involved in the regulation of enzyme expression (by the action of a DNA binding protein). As the proposed binding site was within the Shine-Dalgarno region it was most likely to be a repressor protein binding site (Collado-Vides *et al.* 1991). The position of the major repeat upstream of the start codon of the genes of λ2A is shown in figure A4.2.

---

**Figure A4.2** Position of direct repeat in the four genes of λ2A

<table>
<thead>
<tr>
<th>Gene</th>
<th>Repeat</th>
<th>Start Codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>celA</td>
<td>TTATAATAAATCTGA</td>
<td>ATGAGGGGGTTAG AGATTGT ATG</td>
</tr>
<tr>
<td>manA</td>
<td>AAAAAATGCATAAAAA</td>
<td>AAGAGGGGGTTAG GTTTTAAG ATG</td>
</tr>
<tr>
<td>celB</td>
<td>AATTTAAAAATAGA</td>
<td>AAGAGGGGGTTAG TGGATAAA ATG</td>
</tr>
<tr>
<td>celC</td>
<td>ACTATCAAAATCAAA</td>
<td>AAGAGGGGGTTAG GTTAAAA ATG</td>
</tr>
</tbody>
</table>

Note: The presence of this sequence upstream of the proposed start codon of the celA gene is strong evidence that the whole gene is contained on λ2A.

---

Terminators

Sequences homologous to RNA polymerase rho-independent terminator sequences have been located downstream of all *C. saccharolyticum* genes. These sequences are shown in figure A4.3 and have been identified as putative terminator sequences (factor independent) by the GCG program TERMINATOR (see section 4.5). In bold are those sequences in expected locations, this includes those located downstream of *C. saccharolyticum* genes and those located upstream but near the end of the sequences. For comparison, sequences located within genes are shown as a guide to the specificity of this method, since a major problem with this algorithm is the weighting placed on long runs of T's. These, although rare in *E. coli*, are common in *C. saccharolyticum*.

Appendices
Figure A4.3 Putative terminators located in other *C. saccharolyticum* DNA sequences

**TERMINATOR search on:** pNZ1001 from: 1 to: 1720

<table>
<thead>
<tr>
<th>Terminator</th>
<th>Position</th>
<th>Score</th>
<th>p Value</th>
<th>s Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS <em>bgIA</em></td>
<td>1389</td>
<td>4.03</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

**TERMINATOR search on reverse of:** pNZ1001 from: 1 to: 1720

<table>
<thead>
<tr>
<th>Terminator</th>
<th>Position</th>
<th>Score</th>
<th>p Value</th>
<th>s Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS <em>clmA</em></td>
<td>1441</td>
<td>4.97</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>DS <em>clmA</em></td>
<td>1430</td>
<td>4.78</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

**TERMINATOR search on:** λ2a : 1 to: 13606

<table>
<thead>
<tr>
<th>Terminator</th>
<th>Position</th>
<th>Score</th>
<th>p Value</th>
<th>s Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>US <em>celA</em></td>
<td>1942</td>
<td>4.20</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>DS <em>clmA</em></td>
<td>5499</td>
<td>4.45</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>DS <em>manA</em></td>
<td>9945</td>
<td>4.42</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>DS <em>manA</em></td>
<td>11004</td>
<td>4.55</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>DS <em>celB</em></td>
<td>13425</td>
<td>3.61</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>DS <em>celB</em></td>
<td>13425</td>
<td>3.57</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

**TERMINATOR search on:** pNZ1400. from: 1 to: 6067

<table>
<thead>
<tr>
<th>Terminator</th>
<th>Position</th>
<th>Score</th>
<th>p Value</th>
<th>s Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>US <em>xynA</em></td>
<td>33</td>
<td>3.64</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>DS <em>xynA</em></td>
<td>1287</td>
<td>3.58</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>DS <em>xynC</em></td>
<td>2142</td>
<td>3.61</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>DS <em>xynC</em></td>
<td>2168</td>
<td>3.87</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>DS <em>ORF3/4</em></td>
<td>4915</td>
<td>3.97</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>DS <em>xynB</em></td>
<td>5777</td>
<td>4.20</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

**US** = upstream, **DS** = downstream
Downstream of each ORF were sequences homologous to RNA polymerase rho-independent terminator sequences, implying that each gene was transcribed as a separate transcript rather than in an operon-like structure. Note that very little mRNA work has been performed to verify this conclusion.

**Codon frequency**

Codon frequency was determined using the GCG program CODONFREQUENCY for each gene and is listed below in table A4.3.

In keeping with a high genomic A/T ratio the third codon position showed a marked A/T bias but there was also an added preference for G over C. This can be seen where four similar codons code for one amino acid (glycine, valine, alanine, threonine, serine, arginine, leucine and proline). Codons ending in guanine were used at a higher frequency than those ending in cytosine (up to 400% higher).
259

Table A4.3 Codon frequency for all C. sacclurolyticum gene sequences
Amino ecid

codon

cclA

celB

cclC

men A

Glv
Glv
Glv
Glv
Glu
Glu
Asp

ccc

29

to

0

l7

.,

C,GA

55

20

I

4l

GCT

46

33

l5

29
6

2

CrCC

0

6

6
6
0

GAC

5l

17

0

GAA

23

a'l

')

27

t6

GAI'

6Z

42

28
28
49

l8

l3

IE
IE

GAC
GTG

3E

zl

22

5

6

5

23

6

4

4

3

52

3l
.,,

22
42

6
6

AsD

Val
Val
VEI

GTA
CTT

26

0
2

Xvn A

XvnC

ORF4

I

meth

XynB
7

4

4

l3

6

2

3

l0

tz

I

0

l9

L

toul

4

7'l

J

154

.'

l6l

z

42

4

",

n

5

145

l0

29

5

2t2

zo
6

4
4

20

l0

0

248
t17

4
6

9

I

8

I4

6
4

80
167

II

6

3

I

l9

0

85

136

2E

E

8

GTC
C'cG
GCA

3

')

0

4

0

43

l5

0

0

0

4

z

79

23

4

l0

5

9

t3

5

l3

.,

209

Ale
Ala

c'cT

l8

l6

.,

IE
46
IE

0
2
4

1

t0

4

5

3

93

J

0

3

Art
Arr

AGG

6

I
a

o
0

75

Val
AIa
Ala

Ser
Ser

AGA
AGT
ACrC

J
t1

')

9

5

8

l9

l5
3l

0

1'l

t'l

0

35
30
12

9

'l

l0

4l

23

l4

25

n

I

60
34
20
60

l3

l8

2L

a

6

9
4

28

l4

37
48

AAC

74

Lvs

3l

Am

AAA
AAT

Asn

AAC

30

M€t

AlG

le

ATA
ATT

Thr

ATC
ACG
ACA

Itrr

AgT

lht
Tro

ACC
TGC

End

24
44
45
24

0

25

15

J

45

24

34
33

7
')

5

J

32
94
IE

t6

0

33

54

2
0

8l

17

?

68
0

39

TGA

Cvr

I'Gt-

4

/!l

1'GC

End

TAG

4
0

End

TAA
TAT

Tht

6

26

Lys

Ile
Ile

4

52

I

l4

")

6
6

l0
0

I

0
I

0

0

3
,,

5

l3

5

8

0

)1J

6

4

5

3

3

4

0
0

64
47

0
0

I

I

t

2

50

l0

l4

t5

5

3

25

5

I

7

Ser

l0

6

0

6

Ser
Ser

TCA
TCT

30

l0

J

?5

0
4

IL

il

0

Ser

TCC

a

5

0

Arp

CGG
0

Ars

CGA
CGT

tus

CGC

I

0

0
0
0
0

0

Ar!

0
4

Cln

CAG

42

t

26

cln

CAA
CAT

20
IE

l8
t2
t2

l2

l0

5

4

J

t0
l5

t2

l3

5

6

IJ

2l

2

8

'l

l5
t2

0

0
4

I

I

6

0
0
0

0
2

J

0

o
o

I

I

I

2

I

0

o

o

4

4

4

ll

I

3

J

8

0

7

I

t

J

4

z
0

2

9

3

E

J

CTT

l5

I

2

la

6

o

0

0

0

50

20

0

28

2

0

29

l
'l

1750

0

7

1040

66

t3n

!

J

7

I

1

0

l0

ccc

9

l0

0

2

?5

5

0

LI

0
0

4

0

I

6

t2

I6

I

l0
z

6

34

zl

0

5

t3

n

33

I
o

l4

CTG
CTA

CCT

0

2

4

cAc

ccA

I

0

l7

1

t6

)

5

.,

o
343

0

0
4

)

I

7

3

8
-l

l0

6
0

0

J

J

5

3

7
I

267

313

5

')

I
254
98

,,

I

108

t29
199

0
0
2

59

I
I

59

0

IE

9'l

l2
J

l2
4
J

ll6

2
o

77
70
28

8

?

0

t97

0
0

0

0
6

84

l3

o

0
0

Pro
Pm
Pro

Total

0

I

5

t72

3

0

tt

201

6

0

TTC
TCG

109

3

t9

0

l5

0
E

-t

E

IE

3

t0

lt

6

l4

tz

259
245
t21

3

I
I

28

8

20

2
I

l9

25

3

3

33

194

8

7

0

45

ccG

5

0

30

ho

4

'l

lTT

l-cu

9

2l

o

TTA

His
I.eu
Leu
[.eu

8

5

4

lt

t{is

ll9

0

33

J

I
l6
l4

lo

l0

9

Trc

j

5

E

40
0

2l

Phc

J

z
t2

8t

Phe

123

2
0

0

I

93

a

",

3l

[,cu
Leu

'l

t

6

.IAC

'l-w

I

I

6

,,

I

o
35

Tw

8

7l

ll

4

6

4
0

0

0

E

6

I
I

4'I

t

9E

J

6
0

106
53

tn
5

0

l4

0
6

l0

4

8

7

J

0
489

0
209

0

0
0
98

456

4l
E

94

t2
6358

Appendices


Appendix Five

Three-dimensional Structure of α-amylase

Protein sequence homology comparisons between all α-amylases, from eukaryotic to prokaryotic, show little homology except for three highly conserved regions (5-8 amino acids) at amino acid positions approximately 100, 200 and 300. There are also three weakly conserved regions at amino acid positions approximately 50, 70 and 230 (Svensson 1988). An homology line-up of 23 representative prokaryotic and eukaryotic α-amylases, showing the three highly conserved regions, is shown in figure A5.1. This also shows that sequences immediately adjacent to the 100, 200 and 300 regions have low universal homology or none at all.

The three dimensional structures (X-ray crystallography) of only two α-amylases are known (TAA, fungal α-amylase, Matsuura et al. 1984; PPA, pig pancreatic α-amylase, Buisson et al. 1987). They have similar structural elements, α-helices and β-sheets in similar arrangement and spatial position with an N-terminal (α/β)₈ barrel structure. In both cases the three conserved regions form part of the active site cleft and/or Ca²⁺-binding site (see figure A5.2).

The 100 and 200 regions are two parallel chains of adjacent amino acids running from the surface to the interior of the protein along one face of the active site cleft. These chains are held together by the binding of amino acids from the 100 region (ASN, N) and 200 region (HIS, H) to the essential Ca²⁺ ion. On the opposite side, half way down the surface of the active cleft, two amino acids from the 300 region protrude from the interior of the protein to the surface of the active cleft (see figure A5.3).

Substrate binding studies have shown that two aspartic acid (D) side chain residues (one each from the 200 and 300 regions) are involved in catalytic activity. Other amino acids on the surface of the active cleft are involved with substrate binding, notably the side chain carboxyl group of a glutamic acid (E) at position 230 in TAA (equivalent position 233 in PPA). This glutamic acid residue was proposed as an active site by Matsuura et al. (1984) whereas Buisson et al. (1987), using substrate binding studies, proposed that an aspartic residue 206 was involved in catalytic activity with the glutamine residue an essential substrate binding residue. Further evidence that the aspartic acid residues are involved in catalytic activity was found from studying reaction intermediates using ¹³C nuclear magnetic resonance. A β-carboxyl-acetal ester covalent enzyme-glycosyl intermediate was detected which was consistent with aspartic acid residues at the active site (Tao et al. 1989). Site-directed mutagenesis within the conserved regions has shown that these residues are essential for activity (Vihinen et al. 1990).
Figure A5.1 Amino acid sequence comparison around the three highly conserved regions shown by α-amylases.

<table>
<thead>
<tr>
<th>Sequence Comparison</th>
<th>AMYP$MOUSE</th>
<th>AMYP$RAT</th>
<th>AMYS$MOUSE</th>
<th>AMYC$HUMAN</th>
<th>AMYP$HUMAN</th>
<th>AMYS$HUMAN</th>
<th>AMYP$PIG</th>
<th>AMYAS$DROME</th>
<th>AMYS$STRHY</th>
<th>AMYS$STRLM</th>
<th>AMYS$BACSU</th>
<th>AMYS$ORYSA</th>
<th>AMY2$HORVU</th>
<th>AMY1$HORVU</th>
<th>AMYAS$VIGMU</th>
<th>AMYS$BACAM</th>
<th>AMYS$BACLI</th>
<th>AMYS$BACST</th>
<th>AMYS$SACFI</th>
<th>AMYAS$SPOR</th>
<th>AMYS$BACCI</th>
<th>AMYS$BACME</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMY$MOUSE</td>
<td>GNEDEFRDMM</td>
<td>TACNNUGVR</td>
<td>YUDAVINHM</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMY$RAT</td>
<td>GNEEFKDOM</td>
<td>TACNNUGVR</td>
<td>YUDAVINHM</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$MOUSE</td>
<td>GNEEFRDMM</td>
<td>NRCNNUGVR</td>
<td>YUDAVINHM</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYC$HUMAN</td>
<td>GNEEFREMM</td>
<td>TACNNUGVR</td>
<td>YUDAVINHM</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYP$HUMAN</td>
<td>GNEEFREMM</td>
<td>TACNNUGVR</td>
<td>YUDAVINHM</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$HUMAN</td>
<td>GNEEFREMM</td>
<td>TACNNUGVR</td>
<td>YUDAVINHM</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYP$PIG</td>
<td>GNEEFRDMM</td>
<td>TACNNUGVR</td>
<td>YUDAVINHM</td>
<td>103</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYAS$DROME</td>
<td>GNEEFREMM</td>
<td>KRCNNUGVR</td>
<td>YUDAVINHM</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$STRHY</td>
<td>GORDRASMM</td>
<td>SACHARGU</td>
<td>IADAVINHMA</td>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$STRLM</td>
<td>GORDRFSMM</td>
<td>DTCHRANGU</td>
<td>IADAVINHMA</td>
<td>123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$BACSU</td>
<td>GTEEEOKEM</td>
<td>AAAEEYDIK</td>
<td>YUDAVINH..</td>
<td>143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$ORYSA</td>
<td>GNEQRLLSI</td>
<td>EAFHGGCUGV</td>
<td>IADAVINHAT</td>
<td>119</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMY2$HORVU</td>
<td>GKNQDLSL</td>
<td>GALHGGCUGA</td>
<td>IADAVINHAT</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMY1$HORVU</td>
<td>GNAALCSSL</td>
<td>GALHGGCUGR</td>
<td>IADAVINHAC</td>
<td>119</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYAS$VIGMU</td>
<td>GSNQELLSI</td>
<td>AAFHEKGCIC</td>
<td>IADAVINHAT</td>
<td>116</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$BACAM</td>
<td>GTKSLQQAI</td>
<td>GSLHRSNQV</td>
<td>YGOULINHKA</td>
<td>136</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$BACLI</td>
<td>GTKGLQQAI</td>
<td>KSLHRSDINV</td>
<td>YGOULINHKG</td>
<td>136</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$BACST</td>
<td>GTKQYLOQAI</td>
<td>QAHHAGCQV</td>
<td>YARUDEDKGG</td>
<td>142</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$SACFI</td>
<td>GTAOLKSL</td>
<td>QELHRDQML</td>
<td>MODUTUNHYG</td>
<td>151</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$SCHOC</td>
<td>GTAOLKSL</td>
<td>QELHRDQML</td>
<td>MODULUNHYA</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYAS$SPOR</td>
<td>GTAOLKSL</td>
<td>QELHRDQML</td>
<td>MODULUNHMG</td>
<td>145</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$BACCI</td>
<td>GSMTDFANL</td>
<td>SSAHSRNFKV</td>
<td>YIDRPHNH..T</td>
<td>170</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMYS$BACME</td>
<td>GNLQPARKL</td>
<td>KEADKRUOK</td>
<td>MINDLVNHTS</td>
<td>143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Appendices
| AMYP$MOUSE | GEGWGLUPSD | RALUFVOND | NOG | GHHGAGGS | 322 |
| AMYP$RAT   | GEGGFUPTD | RALUFVOND | NOG | GHHGAGGA | 317 |
| AMYS$MOUSE | GEGGLMPSD | RALUFVOND | NOG | GHHGAGGA | 325 |
| AMYC$HUMAN | GEGGFMPSD | RALUFVOND | NOG | GHHGAGGA | 325 |
| AMYP$HUMAN | GEGGFUVD | RALUFVOND | NOG | GHHGAGGA | 325 |
| AMYP$PIG   | GEGGFMPSD | RALUFVOND | NOG | GHHGAGGA | 325 |
| AMYAS$DROME | GTANGFRASON | RSLUFVOND | NOG | GHHGAGGA | 316 |
| AMYS$STRHY | AOGKLWQRQA | RT... | FUONHD | TER...NG | 297 |
| AMYS$STLM  | GEDGYGMAEG | KSUFFOND | TER... | GG | 301 |
| AMYS$BACSU | ASD... | USAD | KLUTFUESHD | T...| YANDD | 316 |
| AMYS$ORYSA | APGM1GHWRPA | KATTFUONHD | TGST... | Q | 319 |
| AMYS$SHORVU | APGM1GHWRPA | KATTFUONHD | TGST... | Q | 318 |
| AMY15$SHORVU | APGM1GHWRPA | KATTFUONHD | TGST... | Q | 320 |
| AMYAS$VIGMU | PPGM1GUKPE | KATTFUONHD | TGST... | Q | 314 |
| AMY$BACAM  | TUVSKHP1E | KATTFUONHD | TQP...G | Q | 364 |
| AMYS$BACL1  | TUVSKHPE | KATTFUONHD | TQP...G | Q | 362 |
| AMYS$BACST  | TLMKDOQPTL | KATTFUONHD | TNPR...K | 370 |
| AMY15$SACFI | VASSCSDGTL | LT.NFFVONHD | NEAF...A | 329 |
| AMYS1$SCHO  | LQSSCSDVSL | LG.NFFVONHD | QUAF...P | 338 |
| AMYAS$ASPOR | VSCPDOSTL | LG.TFFVONHD | NPF...A | 323 |
| AMYS$BACCI | TADYYSUNDQV.TFFVONHD | MODRF...Q | 362 |
| AMYS$BACME | LFKSYNPNKIDGI.FFTTNHD | QNBU...M | 345 |

**Key**

- **Amyp$Mouse**: Mus musculus (mouse) pancreatic
- **Amyp$Rat**: Rattus norvegicus (rat) pancreatic
- **Amys$Mouse**: Mus musculus (mouse) salivary and hepatic
- **Amyc$Human**: Homo sapiens carcinoid
- **Amyp$Human**: Homo sapiens pancreatic
- **Amys$Human**: Homo sapiens salivary
- **Amyp$Pig**: Sus scrofa (pig) pancreatic
- **Amya$Drome**: Drosophila melanogaster (fruit fly)
- **Amys$Strhy**: Streptomyces hygroscopicus
- **Amys$Strlm**: Streptomyces limosus
- **Amys$Bacsu**: Bacillus subtilis
- **Amys$Orysa**: Oryza sativa (rice)
- **Amys2$Horvu**: Hordeum vulgare (barley)
- **Amys1$Horvu**: Hordeum vulgare (barley)
- **Amysa$Vigm1**: Vigna unguiculata (cowpea)
- **Amys$Bacam**: Bacillus amyloliquefaciens
- **Amys$Bacli**: Bacillus licheniformis
- **Amys$Bacst**: Bacillus stearothermophilus
- **Amys1$Sacfi**: Saccharomyces fibuliger (yeast)
- **Amys1$Scho**: Schwannomyces occidentalis (yeast)
- **Amys$a$Aspor**: Aspergillus oryzae
- **Amys$Bacci**: Bacillus circulans
- **Amys$Bacme**: Bacillus megaterium

- ¥ indicates residues implicated in catalytic activity
- † indicates residues implicated in substrate binding
- □ indicates residues implicated in Ca^{2+} binding
- ◊ indicates residues implicated in Cl⁻ binding

Data obtained from SWISSPROT protein database

Identical amino acid with respect to the consensus sequence are shown in black reverse font; where as those amino acids similar are in grey background font.

Appendices
Figure A5.2. Structure of TAA α-amylase
Figure A5.3 Structure of the active site of TAA α-amylase
The x-ray crystal structure of an acidic α-amylase from *Aspergillus niger* has also been determined. It shows the same overall homology to the other α-amylases with equivalent residues implicated in Ca²⁺-binding, substrate binding and catalytic residues (Boel et al. 1990).

The 3-dimensional structure of a related enzyme, cyclodextrin glycosyltransferase (CGTase) from *Bacillus circulans*, has also been determined by x-ray crystallography (Klein and Schulz 1991). Regions homologous to α-amylase show a similar overall structural pattern, with respect to the spatial positions of α helices, β sheets and the (α/β)₈ barrel structure. Amino acids homologous to the three highly conserved α-amylase regions are also present and are located at the active site.

Although the amino acids between the homologous regions of TAA and PPA amylase and CGTase are different, the tertiary and secondary structures in the form of α-helixes and β-sheets are identical. An overall alignment of the structural elements of α-amylase and CGTase is shown in figure A5.4.

**Figure A5.4 Alignment of conserved secondary structural elements**

<table>
<thead>
<tr>
<th>α1</th>
<th>β1</th>
<th>α2</th>
<th>β2</th>
<th>α3</th>
<th>β3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGT DPDTAVTNQSFSTD</td>
<td>VITQVF</td>
<td>FTDRFLDGNPSNNPSSNPTGAYDATCSNLKLYCGGD</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAA ...ATPADWRSQ...</td>
<td>SIYFL</td>
<td>.TDRFARTDGSTTATCNTADQK......</td>
<td>YCGGT</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>PPA ... .QYAPQTQSG...</td>
<td>TDIVHL</td>
<td>FEW.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α1</td>
<td>β1</td>
<td>α2</td>
<td>β2</td>
<td>α3</td>
<td>β3</td>
</tr>
<tr>
<td>CGT</td>
<td>...</td>
<td>...</td>
<td>QWGIKDKLDYIQ</td>
<td>GMGF.......</td>
<td>TAIWIP</td>
</tr>
<tr>
<td>TAA</td>
<td></td>
<td>...</td>
<td>...</td>
<td></td>
<td>VTAQLPQTAYGD....AYHGY..</td>
</tr>
<tr>
<td>PPA</td>
<td></td>
<td>...</td>
<td></td>
<td>...</td>
<td>SPPNENVVVTNP....SRPWER</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α1</td>
<td>β1</td>
<td>α2</td>
<td>β2</td>
<td>α3</td>
<td>β3</td>
</tr>
<tr>
<td>CGT</td>
<td>...</td>
<td>...</td>
<td>WARDFKKTNPYFGTM.</td>
<td>ADFOIrLITTAETI</td>
<td>GI</td>
</tr>
<tr>
<td>TAA</td>
<td>...</td>
<td>...</td>
<td>WQQDIYSLNENYGTAD</td>
<td>.DLKALSSALHER</td>
<td>GM</td>
</tr>
<tr>
<td>PPA</td>
<td>...</td>
<td>...</td>
<td>YQPVSYKLCRTSGNE.</td>
<td>NEFRDVMVTRCMNV</td>
<td>GV</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α1</td>
<td>β1</td>
<td>α2</td>
<td>β2</td>
<td>α3</td>
<td>β3</td>
</tr>
<tr>
<td>CGT</td>
<td>...</td>
<td>...</td>
<td>FAENGR L.--------</td>
<td>YD NG</td>
<td>TG</td>
</tr>
<tr>
<td>TAA</td>
<td>...</td>
<td>...</td>
<td>SVDYSV FKPFSSQDYFHPFCFIQ</td>
<td>NYEDQTQVYEC</td>
<td></td>
</tr>
<tr>
<td>PPA</td>
<td>...</td>
<td>...</td>
<td>TGTTCGS YCNPGRNE.FPAVP</td>
<td>YSA</td>
<td>WDFNDGK CKTASGGIES YNDPYVRDC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α1</td>
<td>β1</td>
<td>α2</td>
<td>β2</td>
<td>α3</td>
<td>β3</td>
</tr>
<tr>
<td>CGT</td>
<td>...</td>
<td>...</td>
<td>NLYDLADF. NHN...</td>
<td>MATIDXYZFKAIXLWLDM</td>
<td>GVD.</td>
</tr>
<tr>
<td>TAA</td>
<td>...</td>
<td>...</td>
<td>WLGDTVSLLDLDTTK</td>
<td>VKHENYDWVGSLV...</td>
<td>NYSID</td>
</tr>
<tr>
<td>PPA</td>
<td>...</td>
<td>...</td>
<td>QLVGLLDDLAEK</td>
<td>.DYVRSWIADYLKLID</td>
<td>.GVA.</td>
</tr>
</tbody>
</table>

Appendices
In CGT a further two domains are present.

**Domain D**

```
TIGHVG P VMG KPG NVTIDGR GFGSTKG TVYFG T TAV TGA
```

**Domain E**

```
TG DQVTVFVNN NATTALGE NLYLTGN VAELGNWSTGSTA IGPA FNQVIHQYPT
```

---

In CGT a further two domains are present.

**Domain D**

```
TIGHVG P VMG KPG NVTIDGR GFGSTKG TVYFG T TAV TGA
```

**Domain E**

```
TG DQVTVFVNN NATTALGE NLYLTGN VAELGNWSTGSTA IGPA FNQVIHQYPT
```
The alignment shown above is based on the secondary structural elements determined by x-ray crystallography. Even though the amino acid sequence homology for each enzyme is low there is a very high degree of secondary structural and 3-dimensional similarity. The only major exception is Domain B where there is no overall structural correlation.

Note the high correlation between the positions of the secondary structural elements of the (α/β)$_8$ barrel (β$_3$, β$_4$, β$_5$ and β$_7$) and the conserved residues involved in catalytic activity, substrate binding and metal ion binding.

Figure A5.5  Schematic representation of the secondary structure of α-amylases

A 2-dimensional representation of the secondary structure elements of α-amylase and cyclodextrin glycosyltransferase (CGTase), based on the x-ray crystal structure of PPA, TTA and B. circulans CGTase. Domains D and E are present only in CGTase.
Appendix Six

Possible Errors in Debranching Enzyme Sequences

When I started work on my thesis no debranching enzyme DNA sequences had been published. Since then three Pseudomonas isoamylase DNA sequences P. amylofera SB-15 (Amemura et al. 1988), P. amylofera JD210 (Chen et al. 1990) and Pseudomonas sp SMP1 (Tognoni et al. 1989), and two Klebsiella pullulanase DNA sequences, K. aerogenes W70 (Katsuragi et al. 1987) and K. pneumoniae UNF5023 (Kornacker and Pugsley 1990) have been published.

Homology comparisons at the amino acid level of the three Pseudomonas isoamylases show a high overall degree of similarity (see table A6.1), in fact two of the sequences, SMP1 and JD210, are nearly identical. All three genes share the same restriction enzyme pattern, and the gene products are enzymatically identical. When the same analysis is performed at the DNA level, the overall degree of homology is substantially higher. Normally homology comparisons between identical enzymes show a higher degree of homology at the amino acid level than the DNA level, due to the large degree of redundancy in the genetic code (61 DNA codons coding for only 20 amino acids) particularly at the third nucleotide position.

Table A6.1 Pseudomonas isoamylase homology comparisons (% identity) at the amino acid and DNA levels

<table>
<thead>
<tr>
<th>amino acid</th>
<th>JD210</th>
<th>SMP1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SB-15</td>
<td>95.052</td>
<td>94.922</td>
</tr>
<tr>
<td>JD210</td>
<td>-</td>
<td>99.871</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DNA</th>
<th>JD210</th>
<th>SMP1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SB-15</td>
<td>99.146</td>
<td>99.578</td>
</tr>
<tr>
<td>JD210</td>
<td>-</td>
<td>99.962</td>
</tr>
</tbody>
</table>

The situation is the same for the two pullulanase sequences. At the amino acid level percent homology is 86.098% whereas at the DNA level the corresponding value is 90.308%.

As these amino acid sequences are deduced from DNA sequences and not from peptide sequencing, and assuming the DNA sequence has been sequenced correctly, it is impossible for the degree of homology at the DNA level to be greater than that at the amino acid level for the coding
region. Due to the different weightings of amino acids compared to DNA a small variation is tolerated (see appendix 7 on computer algorithm). Outside this variability the only possibility is that the reading frame has been changed, generating a different set of codons and hence different amino acids translated from essentially the same DNA. As the regions of disputed homology at the DNA level are relatively small (1-6 nucleotides) it is unlikely to be caused by a natural deletion, so the most likely explanation is errors in reading the DNA sequence (insertion or deletion of a nucleotide base). These nucleotide base deletions or insertions can be found in the homology line-up of DNA sequences where gaps have been inserted to maximise DNA homology. The reading frame will change at these gaps unless the size of the gap is divisible by three.

If there are sequencing errors the next task is to decide which sequence is correct. Without resequencing each gene it is impossible to determine unequivocally which sequence is correct. By examining current and past literature it is possible to deduce which gene sequence is most likely to be correct.

With the *Pseudomonas* isoamylases it is relatively easy to assign the sequences from JD210 and SMP1 as correct. Firstly these two sequences are almost identical, and secondly an examination of the pertinent literature shows that the SB-15 paper was probably submitted with the knowledge that another group was nearing completion of a similar work.

With the *Klebsiella* pullulanase there are only two published sequences. Again, examination of previous literature from each group showed that the two were probably racing to be first to publish the sequence. When Kornacker and Pugsley finally published their sequence, three years after the other group, a more thorough sequencing strategy was apparent, i.e. no regions with sequencing of one strand only, and every nucleotide base sequenced at least four times.

As stated before it is impossible to determine definitively which sequence is the correct sequence without resequencing. The errors detected could also have been caused by a natural process, i.e. the gene sequences could really be different.

For comparison an alignment of the *ompA* genes from 18 enteric bacteria (18 100 bp in total) showed 60 insertion/deletion fragments (246 bp). Each of these fragments consisted of a multiple of 3 bp with 82 triplets involved in total. This maintained the reading frame and hence caused minimal peptide disruption (Lawrence *et al*. 1991).

Listed in the figures below are the homology line-ups of the sequences showing the frame shifts detected by the algorithm described in Appendix 7. For each enzyme group three alignments are shown, a DNA comparison (figures A6.1 and A6.4), a peptide comparison (figures A6.2 and A6.5), and the DNA sequence with a three reading frame amino acid translation (figures A6.3 and A6.6).
Figure A6.1  Lineup of *Klebsiella* pullulanase DNA sequences

```
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Start</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>W70</td>
<td>80</td>
<td>160</td>
</tr>
<tr>
<td>UNF</td>
<td>160</td>
<td>240</td>
</tr>
<tr>
<td>W70</td>
<td>240</td>
<td>320</td>
</tr>
<tr>
<td>UNF</td>
<td>320</td>
<td>400</td>
</tr>
<tr>
<td>W70</td>
<td>400</td>
<td>480</td>
</tr>
<tr>
<td>UNF</td>
<td>480</td>
<td>560</td>
</tr>
<tr>
<td>W70</td>
<td>560</td>
<td>640</td>
</tr>
<tr>
<td>UNF</td>
<td>640</td>
<td>720</td>
</tr>
<tr>
<td>W70</td>
<td>720</td>
<td>800</td>
</tr>
<tr>
<td>UNF</td>
<td>800</td>
<td>880</td>
</tr>
<tr>
<td>W70</td>
<td>880</td>
<td>960</td>
</tr>
<tr>
<td>UNF</td>
<td>960</td>
<td>1040</td>
</tr>
<tr>
<td>W70</td>
<td>1040</td>
<td>1200</td>
</tr>
<tr>
<td>UNF</td>
<td>1200</td>
<td>1280</td>
</tr>
</tbody>
</table>
```

Appendices
Where sequences differ the nucleotides are coded red, gaps have also been inserted to maximise the nucleotide lineup. The reading frame will change at these gaps unless the size of the gap is divisible by three.
Figure A6.2 Lineup of Klebsiella pullulanase peptide sequences.

1

UNF MLRTYKALYGLSSLVLSQCTGISSSS8AGG 80
W70 MLRTYKALYGLSSLVLSQCTGISSSS8AGG

81

UNF AGITSSAAAD8KNLYLWNNETCDLSAPVDWNVSTTPGSDKGYGYWPIPLKESGCINIVRDXT7KLIDSK8RV
W70 AGITSSAAAD8AFKNLYLWNNETCDLSAPVDWNVSTTPGSDKGYGYWPIPLKESGCINIVRDXT7KLIDSK8RV

161

UNF AFDFTDRTYTS1ADGSYYDRAFDRAFRAAQGVALAEAIWVQ5TLLWPGQ5KPTVRLYYSHSSKVAD3EG9PFX2X
W70 AFDFTDRTYTS1ADGSYYDRAFDRAFRAAQGVALAEAIWVQ5TLLWPGQ5KPTVRLYYSHSSKVAD3EG9PFX2X

241

UNF KLTTPTVLOIQSVSMPFHAGSAFKLDP1HNAVDELLQG6TIAYAAEGILS8MTQVAPTAVLDDGAEAEALSYGAGL
W70 KLTTPTVLOIQSVSMPFHAGSAFKLDP1HNAVDELLQG6TIAYAAEGILS8MTQVAPTAVLDDGAEAEALSYGAGL

321

UNF ADGGVTVRFRWFAPTAQQCVVVYSAUSDRKGI8SHMRDSASGA8SQQG6DLKGAFAFAYMTYHIPQS5RKVEQVEYTDYP4
W70 T6GGVTVRFRWFAPTAQQCVVVYSAUSDRKGI8SHMRDSASGA8SQQG6DLKGAFAFAYMTYHIPQS5RKVEQVEYTDYP4

401

UNF HSLFTINESYQSTQNDNSALATKGPQMDMTLMPHQAQKTDLAKLMTHTHESHRIDLSMDQTVPAELRKYLUALVATSMNQQV
W70 HSLFTINESYQSTQNDNSALATKGPQMDMTLMPHQAQKTDLAKLMTHTHESHRIDLSMDQTVPAELRKYLUALVATSMNQQV

481

UNF HLMKLGAVGTVHELFFVDFLAVFNEFSDKVAD1QCPFSRLCEVNSAVKSSSFYCDGASGSSTVEELIQLKQ5D8CNFQ4
W70 HLMKLGAVGTVHELFFVDFLAVFNEFSDKVAD1QCPFSRLCEVNSAVKSSSFYCDGASGSSTVEELIQLKQ5D8CNFQ4

561

UNF VQALNLVQTDTSNYMDYDPHYTFEGS8YATDE577R1KEKRMICAIKQDL85WMDMVYNNIFNASKPDTRTSVLD
W70 VQALNLVQTDTSNYMDYDPHYTFEGS8YATDE577R1KEKRMICAIKQDL85WMDMVYNNIFNASKPDTRTSVLD

641

UNF K1VFYQYRQINETFSSVSAATCCDSAPHERMFAKLASSMLASLVWTDYKIDGSFEDM6AHPKAIQ7LJWAMERIKLAPNDI
W70 K1VFYQYRQINETFSSVSAATCCDSAPHERMFAKLASSMLASLVWTDYKIDGSFEDM6AHPKAIQ7LJWAMERIKLAPNDI

721

UNF YPPFSWDMNSQSGSRFETASQINLKGITGTCPHDRLQD5VRGGPFTSDGALKQNGQGSGAGVLNP6LSLDDQVHLA
W70 YPPFSWDMNSQSGSRFETASQINLKGITGTCPHDRLQD5VRGGPFTSDGALKQNGQGSGAGVLNP6LSLDDQVHLA

801

UNF DLIQLMAGAGLADFVMIDGQACHGSEIDYNAGPYAYADTPETVNYVSHKDNYTLDI5IA5YQAEADLMTRVWNAQV
W70 DLIQLMAGAGLADFVMIDGQACHGSEIDYNAGPYAYADTPETVNYVSHKDNYTLDI5IA5YQAEADLMTRVWNAQV

881

UNF SLLAVMQGQAIAAQCDSEIESVPSSEIDSGYDNLQDVQEDC5GYSQ8NRSVTRVSKMVADTFPG4E
W70 SLLAVMQGQAIAAQCDSEIESVPSSEIDSGYDNLQDVQEDC5GYSQ8NRSVTRVSKMVADTFPG4E

961

UNF LKQMTAQYQELTS4LRKSSPLFPLGDGSAVKRVDFHFTG57DQAGGLLVMTVDDGM2AGASL2...9FLOTVLWAEINAPE
W70 LKQMTAQYQELTS4LRKSSPLFPLGDGSAVKRVDFHFTG57DQAGGLLVMTVDDGM2AGASL2...9FLOTVLWAEINAPE

1041

UNF ESOLSHFADAPQVSA...SQPPICRQVQAADG...NVTLPNEAVAVLELPQ3SAQVAWILPWS
W70 ESOLSHFADAPQVSA...SQPPICRQVQAADG...NVTLPNEAVAVLELPQ3SAQVAWILPWS

W70 (K. aerogenes W70) and UNF (K. pneumoniae UNF5023).
Where sequences differ the amino acids are coded red, whereas those coded blue are amino acids that are present in a different reading frame of K. aerogenes W70 DNA sequence. If frame shifts were added to the K. aerogenes W70 DNA sequence the amino acids would become identical to those of K. pneumoniae UNF5023.

Appendices
Figure A6.3 The *K. aerogenes* W70 pullulanase and a three reading frame peptide translation
Appendices
The stated reading frame is coded in black with the two nontranslated reading frames coded in yellow.
Where sequences differ the amino acids are coded red, whereas those coded blue are amino acids that are present in a different reading frame of *K. aerogenes* W70 DNA sequence. If frame shifts were added to the W70 DNA sequence the amino acids would become identical to those of *K. pneumoniae* UNF5023.

Appendices
Figure A6.4 Lineup of *Pseudomonas* isoamylase DNA sequences.

```
JD2I0  ATGAAGTGCACCAAGCTTCTGCTGCTGCGCCCTGGCCGGGCTCCGCGGCGGCGG
SMPI  ATGAAGTGCACCAAGCTTCTGCTGCTGCGCCCTGGCCGGGCTCCGCGGCGGCGG
SB-15 ATGAAGTGCACCAAGCTTCTGCTGCTGCGCCCTGGCCGGGCTCCGCGGCGGCGG

JD2I0  CATCACAAGCTGAGCTGGCGAAGACATACGACGACAGGCGAAGGCGAAGACATACGACGACAGG
SMPI  CATCACAAGCTGAGCTGGCGAAGACATACGACGACAGGCGAAGGCGAAGACATACGACGACAGG
SB-15 CATCACAAGCTGAGCTGGCGAAGACATACGACGACAGGCGAAGGCGAAGACATACGACGACAGG

JD2I0  CGCCACAGCTGAGCTGGCGAAGACATACGACGACAGGCGAAGGCGAAGACATACGACGACAGG
SMPI  CGCCACAGCTGAGCTGGCGAAGACATACGACGACAGGCGAAGGCGAAGACATACGACGACAGG
SB-15 CGCCACAGCTGAGCTGGCGAAGACATACGACGACAGGCGAAGGCGAAGACATACGACGACAGG

Appendices
JD210  ACCTGCCTGGCGCCACGCTCAAGCACGCTGTATTGTCGCTCAACGATACCTGCGAATGACCTGCTACCTGTT
SMP1  ACCTTCCTGGCGCCACGCTCAAGCACGCTGTATTGTCGCTCAACGATACCTGCGAATGACCTGCTACCTGTT
SB-15  ACCTTCCTGGCGCCACGCTCAAGCACGCTGTATTGTCGCTCAACGATACCTGCGAATGACCTGCTACCTGTT

JD210  TGGTCACCGGGCAGGAGCACATTGCCGCGGCGGCAACACCTAGGCAAACGCGTCGCGCGTCGCGG
SMP1  TGGTCACCGGGCAGGAGCACATTGCCGCGGCGGCAACACCTAGGCAAACGCGTCGCGCGG
SB-15  TGGTCACCGGGCAGGAGCACATTGCCGCGGCGGCAACACCTAGGCAAACGCGTCGCGCGG

JD210  TCTCCAAGTAG
SMP1  TCTCCAAGTAG
SB-15  TCTCCAAGTAG

JD210 (P. amylofera) SMP1 (P. sp. SMP1), SB-15 (P. amylofera) SB-15 using the GCG program "PILEUP". Where sequences differ the nucleotides are coded red, gaps have also been inserted to maximise the nucleotide lineup. The reading frame will change at these gaps unless the size of the gap is divisible by three.

Appendices
Figure A6.5 Lineup of *Pseudomonas* isoamylase peptide sequences

<table>
<thead>
<tr>
<th>Accession</th>
<th>Peptide Sequence</th>
<th>Identity with Reference Sequence</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>JD210</td>
<td>MKCPKILAL LGCAVLAVVP AMPAAAINR MSLGASYDAQ QANITFRVYS SQATRVLHL YSACHYQVES</td>
<td>70%</td>
<td>P. amyloferans strain SB-15 DNA</td>
</tr>
<tr>
<td>SMP1</td>
<td>MKCPKILAL LGCAVLAVVP AMPAAAINR MSLGASYDAQ QANITFRVYS SQATRVLHL YSACHYQVES</td>
<td>71%</td>
<td>P. amyloferans strain SB-15 DNA</td>
</tr>
<tr>
<td>SB-15</td>
<td>MKCPKILAL LGCAVLAVVP AMPAAAINR MSLGASYDAQ QANITFRVYS SQATRVLHL YSACHYQVES</td>
<td>71%</td>
<td>P. amyloferans strain SB-15 DNA</td>
</tr>
</tbody>
</table>

JD210 (P. amyloferans strain JD210), SMP1 (P. sp. SMP1), SB-15 (P. amyloferans strain SB-15). Where sequences differ the amino acids are coded red, whereas those coded blue are amino acids that are present in a different reading frame of *P. amyloferans* strain SB-15 DNA sequence. If frame shifts were added to the SB-15 DNA sequence the amino acids would become identical.

Appendices
Figure A6.6 The *Pseudomonas* SB-15 isoamylase and a three reading frame translation

1

281

---

Appendices
Where sequences differ the amino acids are coded red, whereas those coded blue are amino acids that are present in a different reading frame of *P. amylofera* strain SB-15 DNA sequence. If frame shifts were added to the SB-15 DNA sequence the amino acids would become identical.
An algorithm was developed for the detection of frame-shifts in DNA sequences resulting in different peptide sequences from similar DNA sequences.

This algorithm is based on the principle that amino acid sequence derived from DNA sequence is essentially the same data set as the original DNA sequence and should be “equivalent”. Therefore when comparisons between two related sequences at the amino acid level and the DNA level produce higher homology at the DNA level compared to the amino acid level, frame-shifts are expected. Due to the greater effect of changes in amino acids compared to DNA sequences a small variable is required. The variable corrects for this increased weighting and is %amino acid + 2(100 - %DNA). For example 300 nucleotides code for 100 amino acids, a 1 nucleotide difference in the DNA sequence can cause 1 amino acid to change, but the % homologies are 99.67% at the DNA level and 99% at the amino acid level. The variable in this case would be

\[
\text{%amino acid} + 2(100 - \%\text{DNA}) = \text{comparison \%} \\
99 + 2(100 - 99.67) = 99.67\%
\]

N.B. If comparison\% is < \%DNA a frame shift has taken place.

The algorithm (see figure A7.1) first determines the amino acid and DNA \% homology for the entire sequence length for two comparison sequences. A decision on the presence of frame-shifts is then made, using the above formula. If frame-shifts are likely to be present (groups 2 and 3 in figure A7.1) the procedure is repeated using a smaller window size to locate the position(s) of the frame-shifts.
Figure A7.1 Algorithm for detection of frame-shifts

1. Read DNA sequences, for $x_1 \rightarrow x_n$ name as (DNA) $x_1 \rightarrow x_n$
2. Translate DNA sequence from (DNA) $x_1 \rightarrow x_n$ to amino acid sequence (AA) $x_1 \rightarrow x_n$
3. Determine % homology between each DNA sequence and the corresponding amino acid sequence.
4. Compare amino acid homology % to DNA homology % for each set.
5. Repeat with a smaller window size.
An example involves the comparison between two *Pseudomonas* isoamylase sequences, JD210 and SB-15 (see appendix 6). A frame-shift was present in one sequence between amino acid positions 450 and 500. The results of the comparisons using the algorithm are shown in figure A7.2. As the comparisons were performed manually the window was not shifted incrementally but in a block jump equivalent to the window size.

A. DNA% = 99.146%, Amino acid% = 95.052%

B. \( \text{%amino acid} + 2(100 - \text{%DNA}) = \text{comparison %} \)
\[
95.052 + 2(100 - 99.146) = 96.76\%
\]
As DNA% > comparison % frame-shifts definitely present

C. Repeated with window size of 300 bp or 100 peptides. This is shown graphically in figure A7.2

---

**Figure A7.2**  % homology comparisons at the amino acid and DNA levels using the frame-shift detection algorithm

![Graphs showing % similarity at amino acid and DNA levels](image)

Window = 100 amino acids

Window = 300 nucleotides

---

Appendices
D. As the region between position 400 and 500 in the peptide sequence showed the greatest deviation this region was re-analysed with a window of 25 amino acids. This is shown graphically in figure A7.3.

Figure A7.3  The effect of decreasing comparison window size on locating frame-shift location

The results show that the algorithm detected a major frame-shift at sequence position 450 and 475 (amino acids). This is exactly the same position as seen in the manual sequence line-up (see appendix 6).
References


Burchhardt, G., Haeckel, K., Spreinat, A., Antranikian, G. and Bahl, H. Nucleotide sequence of the pullulanase gene from Clostridium thermosulfurogenes EM1 and processing of the enzyme. GENBANK DNA sequence database accession number M57692. 1990.


References


References


References


References


References


References


References


References


References


References


References


References


References


References


References


