Isolation, phylogeny and characterisation of proteases and *p*-hydroxyphenylacetic acid hydroxylase from thermophilic *Geobacillus* strains from Buranga Hot Springs in Uganda

by

JOSEPH HAWUMBA

Submitted in partial fulfilment of the requirements of the degree
Philosophiae Doctor
in the Faculty of Natural and Agricultural Sciences
University of Pretoria
Pretoria

August 2003

Isolation, phylogeny and characterisation of proteases and *p*-hydroxyphenylacetic acid hydroxylase from thermophilic *Geobacillus* strains from Buranga Hot Springs in Uganda

by

JOSEPH HAWUMBA

Submitted in partial fulfilment of the requirements of the degree
Philosophiae Doctor
in the Faculty of Natural and Agricultural Sciences
University of Pretoria
Pretoria

August 2003

ACKNOWLEDGEMENTS

I am most grateful to my supervisors: Prof. Volker S. Brözel and Prof. Jacques Theron for the guidance and advice that they have unreservedly provided me during the course of this study, and above all, the cordial working atmosphere, which has made my work enjoyable. The staff in the Department of Biochemistry has been so helpful, providing me with access to equipment and technical advice. Special recognition goes to Prof. A. Neitz, Sandra Van Wyngraardt and Ben Mans. Fellow research students are greatly acknowledged. Though I will not mention each and everyone by name, I will be cheating myself if I forgot to single out: Boet Weyers, who has always had time for me whenever I needed help from him, Sonya, for the encouragement she has always provided all of us on a daily basis, Julian Jaftha, for the help he has so willingly given whenever requested, Francois and Coenie, for their technical advice, and not forgetting Jacques and Antoinette for their friendliness.

I also wish to recognise the brotherly love I received from Claude Sabeta, John Muyonga, Yusufu Byaruhanga, Pius Chin-Molo, Dan Annor, Malebogo Legodi, Bennie Du Toit, Edison Mwanje and Fred Shaws. Working in the background, the following ladies have been an inspiration to me, throughout my research. My sisters - Ruth Nabwire and Naome Hasacha - have given me unsurpassable encouragement and financial support. Aida Bosch, Carol Inocencio and Laura Kirstein have been more than friends and have made my stay in South Africa meaningful through their encouragement and always having time for me.

I am greatly indebted to colleagues in the Academic Registrar's office and the Biochemistry Department (Makerere University), especially Paul Semwanga, Agara, Prossy Nammande and Jane Nabulya, who have worked hand-in-hand with the administration to see that my studies are not hampered by financial difficulties. In addition, they have unrelentingly supported me both morally and financially. Lastly, the whole family of the Hawumba's who has been with me throughout in their prayers. A sense of nearness to the family has made my stay have meaning.

Finally, I wish to acknowledge the following institutions: Makerere University (Staff Development Committee), International Foundation for Science (IFS) and the Council for the Development of Social Science Research in Africa (CODESRIA) for providing me with a staff development fellowship and research grant that have enabled me to further my studies. In the same spirit, I wish to acknowledge the generosity of Prof. Volker S Brözel who met my bench fees.

DECLARATION

I declare that the thesis, which I hereby submit for the degree, Philosophiae Doctor (Microbiology) at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at another University.

Signed: ..... Date: 05/08/2003

DEDICATION

To my parents; Pr. W.W. Hawumba and Suzan Hawumba, Charles Mukiibi and Kezia Mukiibi, Mr.and Mrs. Esau Kato, brothers and sisters, and, above all, my lord God.

SUMMARY

Isolation, phylogeny and characterisation of proteases and *p*-hydroxyphenylacetic acid hydroxylase from thermophilic *Geobacillus* strains from Buranga Hot Springs in Uganda

by

JOSEPH HAWUMBA

Supervisor: Dr Volker S. Brözel
Co-supervisor: Dr Jacques Theron
Department: Microbiology and Plant Pathology
University of Pretoria

for the degree Ph.D

Enzymatic processes that can be run at high temperatures are attractive, as the reaction rates and the substrate solution is often increased. Consequently, there is a continuous search for new thermostable enzymes with the required technological properties. In this study, two thermophilic bacterial isolates, *Geobacillus* PA-9 and PA-5, obtained from the Buranga hot springs in western Uganda, were characterised with the specific aim of isolating and characterising genes encoding novel enzymes.

Both bacterial isolates grew at an optimum temperature and pH of 60°C and 7.5-8.5, respectively, and zymogram analyses indicated that the isolates produced two (isolate PA-5) or more (isolate PA-9) extracellular protease enzymes. The optimum temperature and pH for casein-degrading activity were 70°C, pH 6.5 for isolate PA-9, but caseinolytic activity could also be observed at 2°C. Isolate PA-9 was thus selected for further characterisation. Although various strategies were used to isolate the protease-encoding genes, including enzyme purification and functional screening of a constructed genomic DNA library in *Bacillus megaterium* and in *Escherichia coli*, none resulted in the isolation of the desired

genes. The inability to purify the protease(s) may suggest that low amounts of the protease(s) are being produced or that the protease(s) may be distinct from other characterised proteases.

A clone containing the gene encoding the hydroxylase involved in the degradation of 4-hydroxylphenylacetic acid was, however, isolated from the *Geobacillus* sp. PA-9 genomic DNA library. Sequence analysis indicated the presence of three novel open reading frames (ORFs) of which *PheH* exhibited homology to several 4-hydroxyphenylacetate 3-hydroxylases (4-HPA hydroxylase), *PheH2* appeared to be unique and *PheC* exhibited homology to 2,3-dioxygenases. The 4-HPA hydroxylase has an optimum pH and temperature of 9.0 and 50°C, respectively. Purified PheH did not display hydroxylase activity, suggesting that the 4-HPA 3-hydroxylase from *Geobacillus* isolate PA-9 is composed of two proteins with PheH being the hydroxylase and PheH2 serving as a helper protein required for efficient substrate hydroxylation.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABC	ATP-binding cassette
Ala	alanine
amp^r	ampicillin resistance
Arg	arginine
ATP	adenosine trisphosphate
bp	base pair
ca.	approximately
CO	carbon monoxide
CO_2	carbon dioxide
Cys	cystein
°C	degrees Celsius
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside-5'-triphosphate
EARS	East African Rift valley Systems
e.g.	for example
EtOH	ethanol
4-HPA	4-hydroxyphenylacetic acid
FAD	flavin adenine dinucleotide
Fe	iron
Fig.	figure
Gly	glycine
h	hour
H_2	hydrogen
His	Histidine
H_2S	hydrogen sulphide
IPTG	isopropyl β-D-thiogalactoside
kb	kilobase pairs
kDa	kilodalton
l	litre
lacZ	β-galactosidase gene
LB-broth	Luria-Bertani broth
Lys	lysine
M	molar
mA	milliampere
min	minute

ml	millilitre
mM	millimolar
Mn	manganese
NAD^+	nicotinic adenine dinucleotide
nm	nanometer
nt	nucleotide
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
pmol	picomole
Pro	proline
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
rpm	revolutions per minute
s	second
SDS	sodium dodecyl sulphate
Ser	serine
3,4-DHPA	3,4-dihydroxyphenylacetic acid
TE	Tris-EDTA
tet^r	tetracycline resistance
Tyr	tyrosine
U	units
µg	microgram
µl	microlitre
V	volts
v/v	volume per volume
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

LIST OF FIGURES

FIGURE NUMBER

LIST OF TABLES

TABLE NUMBER

RESEARCH COMMUNICATIONS

Papers published:

1. **Hawumba, J.F.**, Theron, J. and Brözel, V.S. (2002). Thermophilic protease-producing *Geobacillus* from Buranga hot springs in western Uganda. Current Microbiology 45: 144-150

2. **Hawumba, J.F.**, Theron, J. and Brözel, V.S. (2001). Thermophilic protease-producing *Bacillus* from Buranga hot spring in western Uganda. Proceedings of the 8th NAPRECA Symposium, Nairobi, Kenya.

Papers submitted:

1. **Hawumba, J.F.**, Brözel, V.S. and Theron, J. (2003). Nucleotide sequence and biochemical characterization of a 4-hydroxyphenylacetic acid 3-hydroxylase from thermophilic *Geobacillus* sp. PA-9. Submitted for publication: *Applied and Environmental Microbiology*.

Conference contributions:

National conferences:

1. **J.F. Hawumba**, J. Theron and V.S. Brözel. Characterisation of three protease-secreting thermophilic bacteria from hot springs in western Uganda. BioY2K Combined Millennium Meeting, Grahamstown, South Africa, January 2000 (Poster).

2. **J.F. Hawumba**, J. Theron and V.S. Brözel. Characterisation of extracellular proteases produced by extremely thermophilic *Bacillus* isolates. BioY2K Combined Millennium Meeting, Grahamstown, South Africa, January 2000 (Poster).

International conferences:

1. **J.F. Hawumba**, J. Theron and V.S. Brözel. Isolation of thermophilic *Bacillus* producing small proteases active over a wide temperature range from a Ugandan hot spring. American Society for Microbiology Meeting, Orlando, USA, May 2001 (Poster).

2. **J.F. Hawumba**, J. Theron and V.S. Brözel. Thermophilic protease-producing *Bacillus* from Buranga hot spring in Western Uganda. NAPRECA Symposium, Nairobi, Kenya, 28 - 31 August 2001 (Paper presentation and poster).

CHAPTER ONE

INTRODUCTION

1.1. GENERAL INTRODUCTION

Many microorganisms thrive in diverse marine and terrestrial habitats at the extremes of temperatures. Based on the temperature at which they grow and carry out diverse metabolic processes, microorganisms are generally categorised into three major groups, namely:

a) Psychrophiles, which inhabit parts of oceans at temperatures of approximately 4°C or less (Feller *et al.*, 1997; Abe and Horikoshi, 2001). Although some psychrophiles grow well at temperatures as low as -10°C, e.g. in microscopic pockets of water at the ice-sea water interphase (Stanier *et al.*, 1995; Lengeler *et al.*, 1999), their optimal growth temperature is below 20°C and some are killed by a brief warming to room temperature (Lengeler *et al.*, 1999).

b) Mesophiles, which represent the majority of microorganisms, grow optimally between 20 and 42°C. They inhabit almost all terrestrial environments and many have been isolated from water bodies such as wells, springs, rivers and lakes (inclusive of salt lakes), marine environments, as well as soil and air (Stanier *et al.*, 1995; Lengeler *et al.*, 1999).

c) Thermophiles, which inhabit warm through extremely hot environments. Whereas thermophiles refer to microorganisms growing at temperatures up to 70°C, hyperthermophiles have growth temperature optima above 70°C (Stanier *et al.*, 1995; Lengeler *et al.*, 1999). The hyperthermophiles also include those organisms that had previously been grouped as caldo-active with a maximum growth temperature above 70°C (Williams, 1975), as well as barothermophiles (now referred to as piezophiles) that inhabit high-pressure deep-sea vents, and are capable of growing at 100°C and above (Clive, 1990; Abe and Horikoshi, 2001).

Many microorganisms, however, cannot be readily grouped into one of the above categories, as their growth temperatures overlap or span across these demarcations. Such organisms have, in many instances, been described as either psychrotolerant/thermotolerant or cold-tolerant (Lengeler *et al.*, 1999). While the former description has been applied mainly to mesophiles, the latter is applicable to thermophiles that defy the established lower limits of growth and grow at temperatures typical of mesophiles.

1.2. THERMOPHILIC ENVIRONMENTS

The environments typically inhabited by thermophiles include the upper layer of the soil surface that is heated in warm summer temperatures (Lengeler *et al.*, 1999), engineered systems operated at high temperatures (e.g. laundry heaters) (Clive, 1990), compost heaps (Maheshwani *et al.*, 2000), marine thermal vents (Lengeler *et al.*, 1999), the Earth's crust (Summit and Baross, 1998), ultra-deep mines and deep basalt aquifers (Baker *et al.*, 2003), and terrestrial hot springs. Hot springs are widely distributed in different parts of the world and are found in Yellowstone National Park (USA) (Nold and Ward, 1995; Brock, 1998; Ward, 1998), New Zealand (Saul *et al.*, 1999) and Uganda (Gislason *et al.*, 1994), as well as in the volcanic areas of Italy (Pienemann-Simon *et al.*, 1995), Iceland (Sonne-Hansen and Ahring 1999) and Japan (Koga *et al.*, 1998).

1.2.1. Hot springs

Hot springs are formed as a consequence of volcanic activity or due to the movement of the Earth's crust and tectonically active sites. The pressure created by these events leads to the upward mobilisation of heated water, which escapes in the form of geysers and fumaroles (Simoneit *et al.*, 2000). The heated water contains a range of dissolved minerals (Fe, Mn, etc.) and traces of gases (H_2S, CO_2, H_2, CO) (Clive, 1990; Cowan, 1992; Lengeler *et al.*, 1999; Reysenbach and Cady, 2001). The chemistry of terrestrial hot springs varies greatly and is mainly dependent on the properties of the parent or source rock. Also, the temperatures of hot springs range from barely tolerable to the human hand to that of boiling water. Due to temperature gradients being present in the out-flow channel of the hot springs, different microbial communities can establish themselves in these channels (Lengeler *et al.*, 1999).

The East African Rift Valley System (EARS) displays great geothermal activity (Gislason *et al.*, 1994; Darling, 1998; Simoneit *et al.*, 2000). It is divided into the "Eastern Rift", which stretches through Kenya and Ethiopia up to far-southern Tanzania, and the "Western Rift", which stretches through the Tanganyika trough southwards through western Uganda (Fig. 1.1). In Uganda, hot springs are mainly found in the western, south-western and north-western regions and surveyed areas include Kibiro (around Lake Albert), Katwe and Buranga (in Semuliki National Park) (Fig. 1.2) (Gislason *et al.*, 1994).

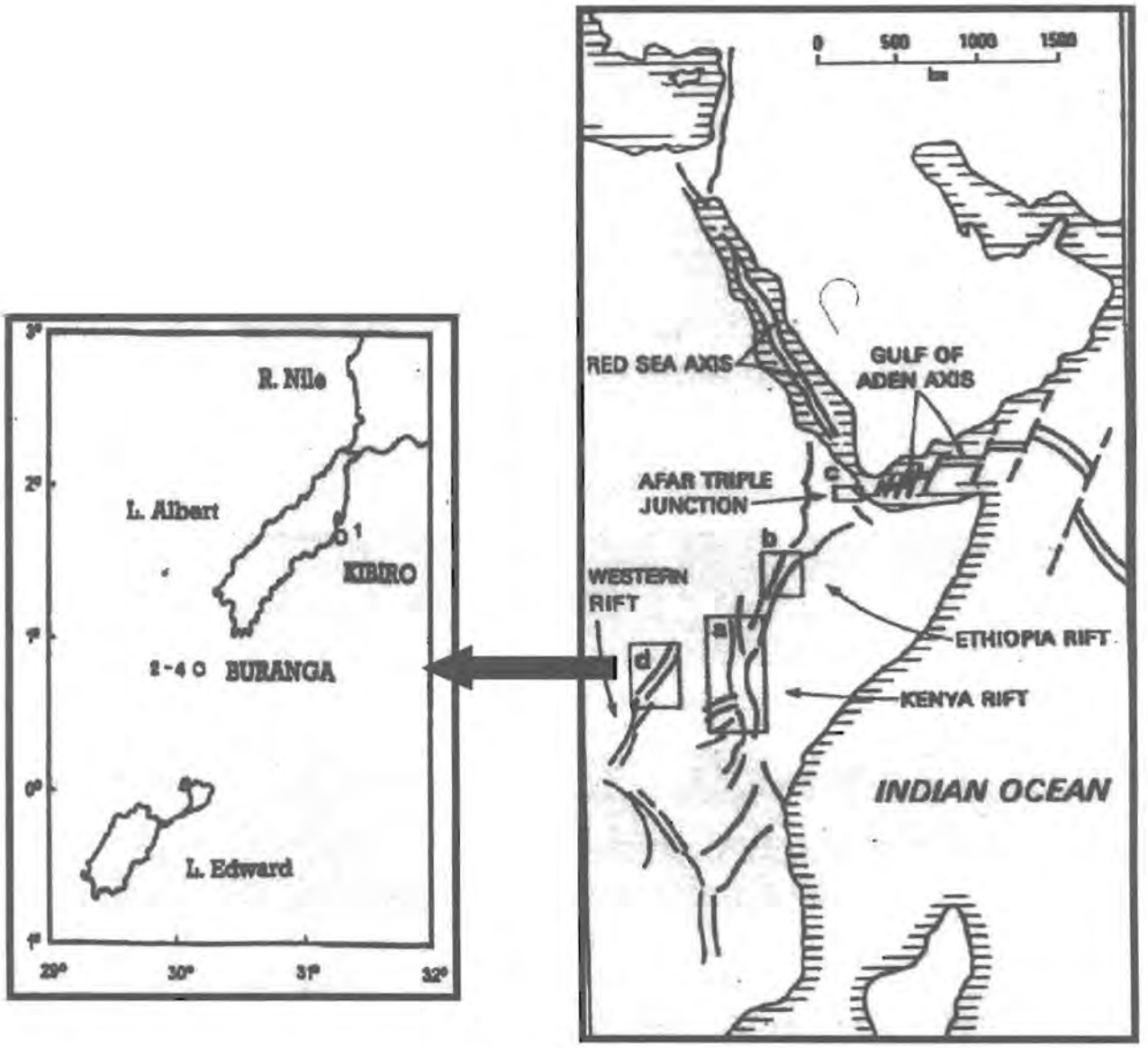


Fig. 1.1 Map of the East African Rift Valley System. The boxes (a-d) are indicative of areas displaying high geothermal activity. (a) Kenya, (b) Ethiopia, (c) Djibouti and (d) Uganda. An enlargement of the geothermal region of Uganda (d) is shown to the left. (*Adapted from Darling, 1998)*

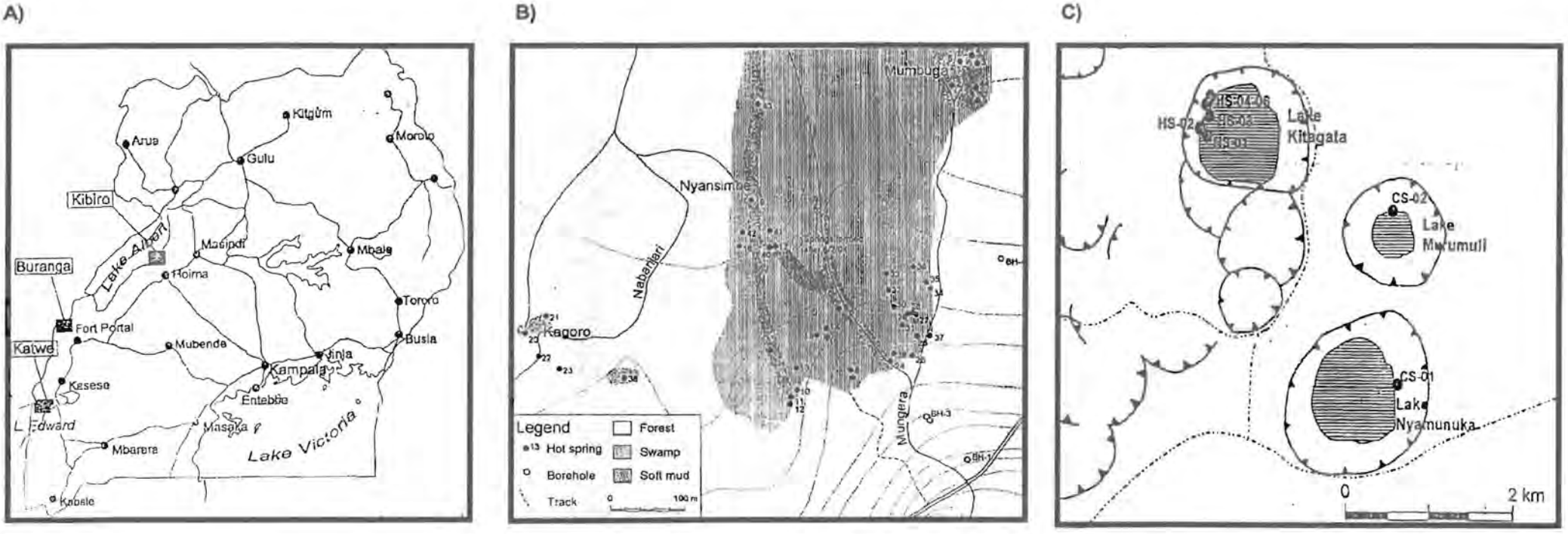


Fig. 1.2 A map of Uganda indicating the three major geothermal areas of Katwe, Buranga and Kibiro is shown in A, while survey maps indicating the distribution of hot springs in the Katwe and Buranga areas are shown in B and C, respectively.

1.3. MICROORGANISMS IN HOT ENVIRONMENTS

Hot environments have been successfully colonised predominantly by prokaryotes (Clive, 1990; Cowan, 1992; Stanier *et al.*, 1995; Ali *et al.*, 1998; Brown and Lupas, 1998; Summit and Baross, 1998; Lengeler *et al.*, 1999; Reysenbach and Cady 2001), and to a lesser extent by eukaryotes such as fungi (Maheshwani *et al.*, 2000) and algae (Sartoni and De Biasi, 1999). Thermophilic bacteria belong to two major domains; Bacteria, which include, among others, members of genera such as *Bacillus, Thermus, Clostridium, Thiobacillus, Thermotoga* and *Streptomyces* (Bergquist and Morgan, 1992; Kim *et al.*, 1998; Xu *et al.*, 1998; Saul *et al.*, 1999), and Archaea, which include, among others, members of genera such as *Methanococcus*, *Thermopterus*, *Sulfolobus*, *Pyrodictium*, *Thermoplasma* and *Pyrococcus* (Bergquist and Morgan, 1992; Adams and Kelly, 1998; Chen and Roberts, 1999).

Whereas thermophilic algae have been isolated from shallow hydrothermal vents and other warm marine environments (Sartoni and De Biasi, 1999), thermophilic fungi have been isolated mostly from heaped masses of plant material, piles of agricultural and forestry products, and other accumulations of organic matter. Notably, only 30 species from the approximately 50 000 recorded fungal species are thermophilic with optimal growth temperatures above 40 or 45°C (Maheshwani *et al.*, 2000). They comprise a heterogeneous physiological group and include *Mycelia sterilia*, various species of the *Phycomycetes* and *Ascomycetes* genera, as well as members of the Fungi imperfecti (Maheshwani *et al.*, 2000).

1.4. IMPORTANCE OF THERMOPHILIC MICROORGANISMS

Thermophilic organisms are promising sources of thermostable enzymes. Since these enzymes tend to remain stable in the presence of denaturing agents such as detergents and organic solvents, they are being exploited in a number of biotechnological bioprocesses (Sonnleitner, 1983; Jaenicke *et al.*, 1996). Thermostable enzymes have been used as constituents of laundry detergents (Nagami and Tanaka, 1986; Tran *et al.*, 1991; Mala *et al.*, 1998) and in the bioconversion of proteinaceous waste into biomass (Anwar and Saleemuddin, 1998), as well as in the degradation of xenobiotic compounds (Ali *et al.*, 1998). They are also being applied in different industries such as the leather tanning, waste management (Hameed *et al.*, 1996; Mehrotra *et al.*, 1999), dairy and food processing (Harwood 1992), as well as pharmaceutical industries (Cowan, 1992).

Due to the inherent stability of thermophilic proteins (Talbot and Sygusch, 1990), they are also used as models to elucidate the factors governing protein thermostability (Karshikoff and Landenstein, 1998; Kannan and Vishveshwara, 2000). The results from such investigations may in future aid in the manipulation of existing enzymes from mesophilic organisms to enhance their thermostability. Recombinant gene technology has already revolutionalized the enzyme and protein industry by enabling the successful isolation, cloning and expression of a number of genes from thermophilic origin in mesophilic host organisms (primarily *Bacillus subtilis* and *Escherichia coli)*. This, in turn, has allowed for the large-scale production of thermostable enzymes, which would be difficult to achieve in the original hosts, as few laboratory methods have been developed to grow Archaea economically.

1.5. AIM OF THIS STUDY

Thermophilic bacteria had been preveously isolated from soil samples collected from Buranga hot springs in western Uganda. The isolates (38) had previously been screened for production of protease enzymes and two isolates that displayed the highest proteolytic activity were selected for use in this study with the specific aim of isolating and characterising genes encoding novel enzymes. The characterisation of protease enzyme(s) and a 4-hydroxyphenylacetic acid hydroxylase is reported in this investigation.

1.6. REFERENCES

Abe, F. and Horikoshi, K. (2001). The biotechnological potential of piezophiles. *Trends in Biotechnology* 19: 102-108.

Adams, M. and Kelly, R. (1998). Finding and using hyperthermophilic enzymes. *Trends in Biotechnology* 16: 329-332.

Ali, S., Fernandez-Lafuente, R. and Cowan, D.A. (1998). Meta-pathway degradation of phenolics by themophilic *Bacilli*. *Enzyme and Microbial Technology* 23: 462-468.

Anwar, A. and Saleemuddin, M. (1998). Alkaline proteases: A Review. *Bioresource Technology* 64: 175-183.

Baker, B.J., Moser, D.P., MacGregor, B.J., Fishbain, S., Wagner, M., Fry, N.K., Jackson, B., Speolstra, N., Loos, S., Takai, K., Lollar, B.S., Fredrickson, J., Balkwill, D., Onstott, T.C., Wimpee, C.F. and Stahl, D.A. (2003). Related assemblages of sulphate-reducing bacteria associated with ultra-deep gold mines of South Africa and deep basalt aquifers of Washington State. Environmental Microbiology 5: 267-277.

Bergquist, P.L. and Morgan, H.W. (1992). The molecular genetics and biotechnological application of enzymes from extremely thermophilic eubacteria. In: *Molecular Biology of Extremophiles*. Herbert, R.A. and Sharp, R.J. (Eds.). Blackie, Chapman and Hall, New York, USA, pp. 44-72.

Brock, T.D. (1998). Early days in Yellowstone Microbiology. *ASM News* 64: 137-140.

Brown, R.J. and Lupas, N.A. (1998). What makes a thermophile? *Trends in Microbiology* 6: 349-351.

Chen, L. and Roberts, M.F. (1999). Characterization of a tetrameric inositol monophosphatase from the hyperthermophilic bacterium *Thermotoga maritima*. *Applied and Environmental Microbiology* 65: 4559-4567.

Clive, E. (1990). Thermophilic blue-green algae and the thermal environment. *Bacteriological Reviews* 33: 476-504.

Cowan, D.A. (1992). Enzymes from thermophilic archaebacteria: Current and future applications in biotechnology. *Biochemical Society Symposium* 58:149-169.

Darling, G.W. (1998). Hydrothermal hydrocarbon gases: Application in the East African Rift System. *Applied Geochemistry* 13: 825-840.

Feller, G., Arpigny, L.J., Narinx, E. and Gerday, C. (1997). Molecular adaptations of enzymes from psychrophilic organisms. *Comparative Biochemistry and Physiology* 118A: 495-499.

Gislason, G., Ngobi, G., Isabirye, E. and Tumwebaze, S. (1994). An inventory of three geothermal areas in west and southwest Uganda. Geothermal Exploration. UGA/92/002, pp. 6-32.

Hameed, A., Natt, M.A. and Evans, C.S. (1996). Production of alkaline protease by a new *Bacillus subtilis* isolate for use as a bating enzyme in leather treatment. *World Journal of Microbiology and Biotechnology* 12: 289-291.

Harwood, R.C. (1992). *Bacillus subtilis* and its relatives: Molecular biology and industrial workhorses. *Trends in Biotechnology* 10: 247-256.

Jaenicke, R., Schurig, H., Beaucamp, N. and Ostendorp, R. (1996). Structure and stability of hyperstable proteins: Glycolytic enzymes from hyperthermophilic bacterium *Thermotoga maritima*. *Advances in Protein Chemistry* 48:181-269.

Kannan, N. and Vishveshwara, S. (2000). Aromatic clusters: A determinant of thermal stability of thermophilic proteins. *Protein Engineering* 13: 753-761.

Karshikoff, A. and Landenstein, R. (1998). Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. *Protein Engineering* 11: 867-872.

Kim, B.S., Falconer, C., Williams, E. and Goodfellow, M. (1998). *Streptomyces thermocarboxydovorans* sp. nov. and *Streptomyces thermocarboxydus* sp. nov., two moderately thermophilic carboxydotrophic species from soil. *International Journal of Systematic Bacteriology* 48: 59-68.

Koga, Y., Morikawa, M., Haruki, M., Nakamura, H., Imanaka, T. and Kanaya, S. (1998). Thermostable glycerol kinase from a hyperthermophilic archaeon: Gene cloning and characterisation of the recombinant enzyme. *Protein Engineering* 11: 1219-1227.

Lengeler, J.W., Drews, G. and Schlegel, H.G. (Eds.). (1999). Biology of the Prokaryote. Blackwell Science Inc., New York, USA, pp. 652-915.

Maheshwani, R., Bharadwaj, G. and Bhati, M.K. (2000). Thermophilic fungi: Their physiology and enzymes. *Microbiology and Molecular Biology Reviews* 64: 461-488.

Mala, B.R., Aparn, M.T., Mohini, S.G. and Vasanti, V.D. (1998). Molecular and biotechnological aspects of microbial poteases. *Microbiology and Molecular Biology Reviews* 62: 597-635.

Mehrotra, S., Pandey, P.K., Gaur, R. and Darmwal, N.S. (1999). The production of alkaline protease by *Bacillus* species isolate. *Bioresource Technology* 67: 201-203.

Nagami, Y. and Tanaka, T. (1986). Molecular cloning and nucleotide sequencing of a DNA fragment from *Bacillus natto* that enhances production of extracellular proteases and levansucrase in *Bacillus subtilis*. *Journal of Bacteriology* 166: 20-28.

Nold, S.C. and Ward, D.M. (1995). Diverse *Thermus* species inhabit a single hot spring microbial mat. *Systematic and Applied Microbiology* 18: 274-278.

Pienemann-Simon, S., Ludwig, W., Vogt, B. and Gottschalk, G. (1995). Taxonomic analysis of the thermophilic bacterium strain SP83 producing H_2 from starch at 75°C. *Systematic and Applied Microbiology* 18: 231-236.

Reysenbsch, A. and Cady, S. (2001). Microbiology of ancient and modern hydrothermal systems. *Trends in Microbiology* 9: 79-85.

Sartoni, G. and De Biasi, A.M. (1999). A survey of the marine algae of Milos Island, Greece. *Cryptogamie Algol* 20: 271-283.

Saul, D.J., Reeves, R.A., Morgan, H.W. and Bergquist, P.L. (1999). *Thermus* diversity and strain loss during enrichment. *FEMS Microbial Ecology* 30: 157-162.

Simoneit, T.R.B., Aboul-Kassim, T.A.T. and Tiercelin, J.J. (2000). Hydrothermal petroleum from lacustrine sedimentary organic matter in the East African Rift. *Applied Geochemistry* 15: 355-368.

Sonne-Hansen, J. and Ahring, K.B. (1999). *Thermodesulfobacterium hveragerdense* sp. nov., and *Thermodesulfovibrio islandicus* sp. nov., two thermophilic sulphate reducing bacteria isolated from a Icelandic hot spring. *Systematic and Applied Microbiology* 22: 559-564.

Sonnleitner, B. (1983). Biotechnology of thermophilic bacteria: Growth, products and application. *Advances in Biochemical Engineering/Biotechnology* 28: 69-138.

Stanier, R.Y., Ingraham, J.L., Wheelis, M.L. and Painter, P.R. (1995). General Microbiology (5th Ed.). MacMillan Press Ltd., Prentice-Hall Inc., New Jersey, USA, pp. 204-210.

Summit, M. and Baross, A.J. (1998). Thermophilic sub-seafloor microorganisms from the 1996 North Gord Ridge eruption. *Deep-sea research II* 45: 2751-2766.

Talbot, C. and Sygusch, J. (1990). Purification and characterisation of thermostable β-mannanase and α-galactosidase from *Bacillus stearothermophilus*. *Applied and Environmental Microbiology* 56: 3505-3510.

Tran, L., Wu, X. and Wong, S. (1991). Cloning and expression of a novel protease gene encoding an extracellular neutral protease from *Bacillus subtilis*. *Journal of Bacteriology* 173: 6364-6372.

Ward, D.M. (1998). Microbiology in Yellowstone National Park. *ASM News* 64: 141-150.

Xu, L-H., Tiang, Y-Q., Zhang, Y-F., Zhao, L-X. and Jiang, C-X. (1998). *Streptomyces thermogriseus*, a new species of the genus *Streptomyces* from soil, lake and hot spring. *International Journal of Systematic Bacteriology* 48: 1089-1093

CHAPTER TWO

LITERATURE REVIEW

2.1. INTRODUCTION

Classic methods for the isolation and identification of microorganisms require their cultivation in pure culture. Often, only a minor fraction of bacteria present in an environmental sample can actually be grown under standard culturing conditions, and the selectivity of the cultivation techniques hardly reflects the true extent of the microbial community composition or diversity (Amann *et al.*, 1995). When it comes to isolating thermophilic bacteria, few laboratory methods have been developed whereby Archaea can be isolated or cultured. Many thermophilic microorganisms, as a result, have thus not been recovered from the environment (Adams and Kelly, 1998; Rondon *et al.*, 1999). Consequently, the true extent of microbial diversity in general and of thermophiles in particular, probably far exceeds the current estimates. With the advances being made in isolation and culturing techniques, culturing of some fastidious organisms is becoming increasingly more possible. Furthermore, advances in molecular biology methods (DNA- and RNA-based methods) and in the development of improved DNA extraction and amplification techniques applicable to soil and sediment samples (Tsai and Olson, 1991; Steffan *et al.*, 1998) have enabled the identification and phylogenetic placement of bacteria that have thus far eluded cultivation (Head *et al.*, 1998; Rondon *et al.*, 1999; Rondon *et al.*, 2000). These methods have opened new avenues for studying bacterial diversity and also for the isolation of potentially novel products from bacteria inhabiting many different and extreme environments.

2.2. PROKARYOTIC SYSTEMATICS

All life forms have been assigned to one of three taxonomic domains, namely: Eukarya, Bacteria and Archaea (Woese *et al.*, 1990). Although thermophilic members belonging to the domain Eukarya have been described (e.g. thermophilic fungi), most studies have been directed at the many genera and species of the two prokaryotic domains, i.e. Archaea and Bacteria (Peinemann-Simon *et al.*, 1995; Summit and Baross, 1998; Reysenbach and Cady, 2001; Kannan *et al.*, 2001; Hemmi *et al.*, 2002).

Members of the archaeal domain are regarded as ancient and primitive. Their membrane lipids characteristically consist of predominantly isoprenoid glycerol diethers or diglycerol tetraethers, and their ribosomes contain an archaeal type rRNA (Woese *et al.*, 1990). The

domain Archaea is subdivided into two phylogenetically distinct groups. The first group, Methanogens, comprises genera such as *Archaeoglobus* (extreme halophiles and sulphate-reducing species), *Thermoplasma* and the *Thermococcus-Pyrococcus*. The second group, Thermoacidophiles, comprises extremely thermophilic, sulphate-dependent archaebacteria (Yang *et al.*, 1985). The niches archaebacteria are entirely thermophilic and they constitute a relatively homogeneous physiological group.

Compared to Archaea, members of the domain Bacteria are regarded as advanced. Their definitive features include membrane lipids consisting of predominantly diacylglycerol diesters, and ribosomes containing an (eu)bacterial type of rRNA (Woese *et al.*, 1990). Within the domain Bacteria, various thermophilic and hyperthermophilic species have been isolated and described (Wery *et al.*, 2001). They include: (a) aerobic heterotrophs, many of which have been assigned to the genera *Thermus* and *Bacillus*, (b) microaerophilic autotrophs (*Aquifex* spp.), (c) anaerobic autotrophs (*Desulfurobacterium* spp.), and (d) anaerobic heterotrophs (*Clostridium* spp. and *Thermotoga* spp.) (Wery *et al.*, 2001).

2.2.1. Methodological approaches to prokaryotic systematics

2.2.1.1. Phenotypic traits used in prokaryotic systematics

Phenotypic methods form the foundation of numerical taxonomy and comprise all methods that are not directed toward characterisation of the DNA or RNA of the microorganism (Vandamme *et al.*, 1996). They include methods to characterise: (a) morphological features such as shape, endospore formation, flagella, inclusion bodies, Gram stain and colony colour, dimension and form (Chung *et al.*, 2000; Hattori *et al.*, 2000; Nazina *et al.*, 2001; Alain *et al.*, 2002; Seyfried *et al.*, 2002), (b) physiological features such as growth at different temperatures, pH, salt concentrations, atmospheric conditions (aerobiosis and anaerobiosis), ability to metabolise different carbon sources, resistance to antibiotics, and production of different metabolic products and enzymes (Fiala *et al.*, 1986; Huber *et al.*, 1992; Sokolova *et al.*, 2001; Seyfried *et al.*, 2002; Zavarzina *et al.*, 2002), and (c) biochemical features such as analysis of the cell wall composition (Kim *et al.*, 1999), polar lipids and fatty acid composition (Nunes *et al.*, 1992a; 1992b; Chung *et al.*, 2000; Nazina *et al.*, 2001), isoprenoid (respiratory) quinones (Yamamoto *et al.*, 1998; Hattori *et al.*, 2000), whole-cell protein (Karnauchow *et al.*, 1992; Agnew *et al.*, 1995), antigenic analysis (Suzuki *et al.*, 1983; Karnauchow *et al.*, 1992), isoenzyme analysis (Xu *et al.*, 1998), electron acceptors (Burggraf

et al., 1990; Wery *et al.*, 2001; Alain *et al.*, 2002; Seyfried *et al.*, 2002; Zavarzina *et al.*, 2002) and polyamines and histones (Vandamme *et al.*, 1996).

Many of these individual traits are insufficient as parameters for determining genetic relatedness, but when integrated, they may provide coherent descriptive information enabling the recognition of taxa (Martiensson *et al.*, 1996). Furthermore, features such as colour, colony morphology and form, whole-cell proteins and endospore formation (in some cases) are at times influenced by the culture conditions. Standardisation of methodologies is therefore a *sine qua none* in bacterial systematics.

2.2.1.2. Genotypic methods

Molecular methods such as DNA base composition, DNA-DNA hybridisation and ribosomal RNA-based nucleic acid sequence homology analyses have been, among others, useful in defining bacterial species, in phylogenetic placement of newly characterised bacterial strains (Vandamme *et al.*, 1996), and in the restructuring of existing taxons (Rőssler *et al.*, 1991; Ash *et al.*, 1991; Rainey *et a*l., 1994). Figure 2.1 summarises current methods used in bacterial systematics and their respective levels of resolution.

2.2.1.2.1. DNA base composition

Determination of DNA base composition, which is expressed as mole percent of guanine plus cytosine (mol % G+C), is considered part of the standard description of bacterial species (Vandamme *et al.*, 1996). Strains that are assigned to a single microbial genus tend to have closely similar or identical mol % G+C values (Stanier *et al.*, 1995). The mol % G+C varies greatly between genera, with values ranging from 24-25% for certain clostridia and mycoplasmas (Brian and Fergus, 1986; Lengeler *et al.*, 1999) to 76% for strains of certain *Actinomycetes* spp. (Lengeler *et al.*, 1999). While the mol % G+C-content of strains of a phylogenetically coherent species usually does not differ by more than *ca.* 3-5% (Brian and Fergus, 1986; Vandamme *et al.*, 1996), values between species of the genus may differ more. A range of more than 15% mol % G+C between species of the genus is usually taken as an indication of phylogenetic heterogeneity of the genus (Lengeler *et al.*, 1999). The mol % G+C of members of the genus *Bacillus*, for example, ranges between 32 to 69%, indicating that it is a very heterogeneous genus (Goodfellow and O'Donnel, 1994). However, the genus

Bacillus has since been subdivided, though the newly established genera have not yet been widely accepted. Therefore, *Bacillus* in this thesis should be taken to include all *Bacillaceae*.

2.2.1.2.2. DNA-DNA hybridisation

DNA-DNA hybridisation is widely used to delineate species (Vandamme *et al.*, 1996) and to resolve the finer taxonomic relationships between closely related organisms (Kim *et al.*, 1999). It is generally agreed that species should encompass strains, which show approximately 70% or more DNA-DNA relatedness under suitable experimental conditions (Kim *et al.*, 1999). A number of DNA hybridisation protocols have been developed and are being used in different laboratories. However, depending on the method used, different results may be obtained between laboratories (Vandamme *et al.*, 1996). Nevertheless, it remains a powerful method in the delineation of bacterial species.

2.2.1.2.3. Ribosomal RNA (rRNA)-based methods

Phylogenetic methods, which are based on the sequencing and cataloguing of rRNAs (5S, 16S, and 23S), have played a major role in modern systematics and phylogenetic studies. Comparative analysis of rRNA by means of T1 ribonuclease oligonucleotide catalogues led to the definition of a third primary kingdom, that of the archaebacteria, in addition to those of the eubacteria and eukaryotes (Jarsch and Bőck, 1985). Nevertheless, this analysis has since been replaced by methods involving comparison of full-length 16S rRNA sequences in order to resolve branching points between phylogenetically very distant groups. The analyses of full-length 16S rRNA sequences furthermore allow comparison of secondary structural features, in addition to primary structures (Jarsch and Bőck, 1985; Yang *et al.*, 1985). Both archaeal and bacterial domains have been phylogenetically described using rRNA-based methods. The strength of the rRNA sequencing approach has been illustrated in its application to the phylogenetic placement of unculturable microorganisms (Head *et al.*, 1998; Rondon *et al.*, 1999; Rondon *et al.*, 2000; Nogales *et al.*, 2001).

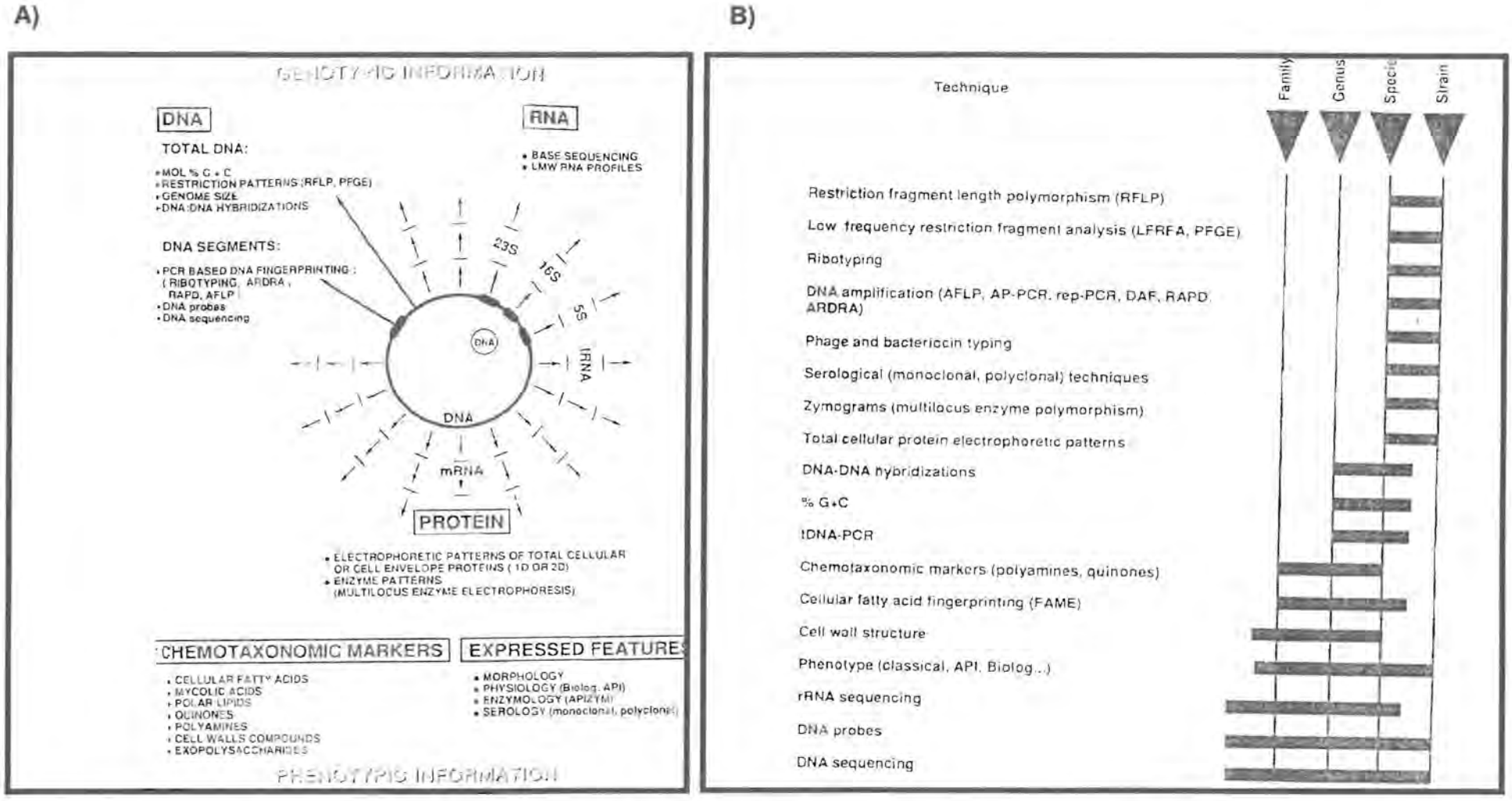


Fig. 2.1 Schematic overview of various cellular components and currently used techniques in systematics (A), as well as the taxonomic resolution of the respective techniques (B). (*Adapted from Vandamme et al., 1996*)

(Abbreviations: RFLP, restriction fragment length polymorphism; PFGE, pulse-field gel electrophoresis; AFLP, amplified fragment length polymorphism; AP-PCR, arbitrarily primed PCR; rep-PCR, repetitive extragenic palindromic PCR; RAPD, randomly amplified polymorphic DNA; ARDRA, amplified rDNA restriction analysis; 1D, one-dimensional SDS-polyacrylamide gel electrophoresis; 2D, two-dimensional SDS-polyacrylamide gel electrophoresis; FAME, fatty acid methyl esters analyses).

2.3. DISTRIBUTION OF THERMOPHILES IN THE ENVIRONMENT

2.3.1. Archaea

Archaebacteria have been isolated from very hot and acidic environments. It is thus not surprising that most of the thermophilic Archaea are both hyperthermophilic and anaerobic. Hydrothermal fluids resulting from sea floor fissures in axial valleys have frequently served as sources of viable anaerobic microorganisms with hyperthermophilic growth range. Moreover, most of the isolates from the sub-sea floors do not tolerate atmospheric concentrations of O_2 at temperatures that permit growth (Reysenbach and Cady, 2001).

Thermoacidophilic archaeabacteria, e.g. *Sulfolobus solfataricus* (Hemmi *et al.*, 2002) and *Thermococcus kodakareansis* KOD1 (Kannan *et al.*, 2001), have been isolated from acid marine solfataric environments. *Methanopyrus*, a rod-shaped methanogen, is endemic to marine vents (Reysenbach and Cady, 2001), while *Thermococcus siculi* sp. nov. has been isolated from a deep-sea thermal vent (1 394 m deep) at the mid-Okinawa trough (Grote *et al.*, 1999). These strains grow well at temperatures between 50 to 100°C with growth temperature optima ranging from 85 to 90°C (Kannan *et al.*, 2001). Thermophilic piezophiles, encompassing genera such as *Methanococcus*, *Palaeococcus*, *Pyrococcus* and *Thermococcus* (Abe and Horikoshi, 2001), have been isolated from high-pressure environments such as deep-sea trenches. Other archaeal genera, such as *Archaeoglobus*, *Pyrodictunum* and *Pyrococcus*, have been isolated from terrestrial and marine solfataric springs, as well as sub-marine and deep-sea hydrothermal vents around the world (Grote *et al.*, 1999).

Oil field reservoirs and surrounding environments are also sources of novel thermophilic archaeabacteria. For example, both *Methanobacterium thermoaggregans* and *M. thermoalcalphilum* have been isolated from oil field water and from sea water-flooded oil wells in California (Nilsen and Torsvik, 1996). Also, *Methanococcus thermolithotrophicus*, isolated from oil field reservoir water, is capable of growing at temperatures between 17 to 62°C, with an optimum growth temperature at 60°C (Nilsen and Torsvik, 1996).

2.3.2. Bacteria

Members of the domain Bacteria (eubacteria) have been isolated from diverse marine and terrestrial environments, which in the case of most spore-formers also include environments non-permissible for their growth and survival (Peinemann-Simon *et al.*, 1995; Summit and

Baross, 1998). Furthermore, many species of thermophilic eubacteria have been isolated from acidic (acidophiles/acidotolerant), alkaline (alkaliphiles/alkalitolerant), high salt (halophiles/halotolerant), and high-pressure (1000 bar) environments (Abe and Horikoshi, 2000).

2.3.2.1. Hyperthermophilic eubacteria

Hyperthermophilic eubacteria, such as *Thermotoga maritima*, *Aquifex* spp., *Staphylothermus* spp. and *Thermus* spp., have been isolated from terrestrial hot springs (Archenbach-Richter *et al.*, 1987; Nelson *et al.*, 1999), turf mats (Yamamoto *et al.*, 1998), deep-sea and shallow marine hydrothermal vents, as well as deep-sea hydrothermal chimneys (Gugliandolo and Maugeri, 1997; Reysenbach and Cady, 2000). Hyperthermophilic eubacteria, with the exception of the genus *Thermus*, are anaerobic and are closely related to hyperthermophilic archaeabacteria. The genus *Thermus* comprises Gram-negative, extremely thermophilic (hyperthermophilic) aerobes and is ubiquitous to many terrestrial hot springs (Gudni *et al.*, 1985; Peek *et al.*, 1992; Friedrich *et al.*, 2002). Representatives of this genus have a wide geographical distribution and have been isolated from hot springs in New-Zealand (Hudson *et al.*, 1987), Iceland (Gudni *et al.*, 1985) and from the Octopus hot spring in the Yellowstone National Park (USA) (Nold and Word, 1995).

2.3.2.2. Thermophilic eubacteria

Whereas most hyperthermophilic archaea and eubacteria are predominantly anaerobic, thermophilic eubacteria are represented almost equally among anaerobes and aerobes. Numerous genera, from a diverse array of both terrestrial and marine environments, have been described.

Anaerobic thermophiles, encompassing genera such as *Clostridium*, *Desulfotomaculum*, *Thermoanaerobacter*, *Thermoanaerobacterium*, *Caloramator*, *Thermoacetogenium* and *Anoxybacillus* have been isolated from both geothermal and non-geothermal environments, and comprise both sporulating and non-sporulating species (Patel *et al.*, 1986). For example, *Fervidobacterium nodosum* has only been isolated from geothermal environments (Patel *et al.*, 1985a), while *Clostridium thermocellum* has been isolated from non-geothermal environments. The sporulating *Clostridium thermosulfurogenes*, as well as non-sporulating *Thermoanaerobium brockii*, *Thermoanaerobacter ethanolicus* and *Thermobacteriodes*

acetoethylicus have all been isolated from geothermal environments that are rich in algal and bacterial biomass (Patel *et al.*, 1985a).

Additionally, thermophilic anaerobes, such as *Carboxydobranhium pacificum* (Sokolova *et al.*, 2001), *Caminicella sporogenes* (Alain *et al.*, 2002), *Caloranaerobacter azorensis* (Wery *et al.*, 2001), *Thermovenabulum ferrioganovorum* and *Caloramator veterbensis* (Zavarzina *et al.*, 2002; Seyfried *et al.*, 2002), *Petrotoga olearia* and *P. sibirica* (Haridon *et al.*, 2002) and *Thermacetogenium phaeum* (Hattori *et al.*, 2000), have been isolated from environments such as sub-marine and deep-sea hydrothermal vents, terrestrial hydrothermal springs, deep continental oil reservoirs and from a Kraft pulp-waste water methanogenic reactor.

Aerobic thermophilic eubacteria encompass both sporulating and non-sporulating genera. The genus *Rhodothermus* comprises obligately aerobic, non-sporulating, Gram-negative rod-shaped members (Alfredsson *et al.*, 1988). *Rhodothermus* strains have been isolated from shallow marine hot springs in Isafjardarjup Bay in Iceland and from Ribeira Quente, on the island of Miguel Azores (Alfredsson *et al.*, 1988; Nunes *et al.*, 1992a; 1992b). The sporulating genera include members of the *Bacillaceae*, *Thermoactinomyces* and *Streptomyces* (Stanier *et al.*, 1996).

Genera of the *Bacillaceae* have a very wide distribution, which ranges from air, hot springs, deep-sea hydrothermal vents, solfataras, geothermally heated soils, man-made thermal systems (e.g. hot water pipelines, heat exchangers), composting vegetation, and soil (Martiensson *et al.*, 1995; Maugeri *et al.*, 2001; Baker *et al.*, 2001). Members of these genera have been extensively studied, because of their relatively easier growth requirements. Their diverse distribution is attributed to their formation of heat-resistant endospores, which are spread to different environments where they remain viable for long periods (Baker *et al.*, 2001). Although most thermophilic *Bacilli* are thermophiles (Nazina *et al.*, 2001; Maugeri *et al.*, 2001), having optimal growth temperature ranging between ≥ 45 to $\leq 70^{\circ}C$, a few hyperthermophilic *Bacillus* spp. have been isolated from both natural hot springs and deep-sea hydrothermal vents (Marteinsson *et al.*, 1996; Baker *et al.*, 2001).

Streptomyces species, like *Bacillaceae* and *Thermoactinomyces*, are aerobic and form spores. The ability to form spores may account partly for their wide distribution in terrestrial habitats. Unlike *Bacillus*, however, *Streptomyces* spores are heat-labile. The lability of their spores

may, in part, account for them being distributed in less hostile environments. This is evidenced by their abundance in soil (Stanier *et al.*, 1995; Kim *et al.*, 1998; Kim *et al.*, 1999), lake sediments (Xu *et al.*, 1998) and, to a lesser extent, hot springs (Xu *et al.*, 1998). Members of this genus are moderately thermophilic, growing at temperatures between 25 to 55°C (Kim *et al.*, 1998; Kim *et al.*, 1999). However, Xu *et al.* (1998) described a new species, *Streptomyces thermogriseus,* which exhibited an upper growth limit ranging between 65 to 68°C. This is the highest growth limit so far recorded for a member of the *Streptomyces* genus.

Although a comprehensive analysis of thermophilic bacterial diversity is beyond the scope of this literature review, it is nevertheless very clear that thermophiles, like their mesophilic and psychrophilic counterparts, are widely distributed in the environment. As culturing methods become available, knowledge regarding their distribution and diversity will likewise increase.

2.4. PHYSIOLOGICAL ADAPTATIONS OF THERMOPHILES

The diverse environments of high temperatures ranging from ≥45 to ≤115°C (Huber *et al.*, 2000; Maugeri *et al.*, 2001), pH ranging from pH 2 to pH ≥10 (Darland and Brock, 1971; Horikoshi, 1999), and high salt environments (Santos and da Costa, 2002) have exposed thermophiles and hyperthermophiles to different physiological challenges, namely: (a) the requirement to meet the energy needs of the organism under both aerobic and anaerobic conditions, (b) the requirement to maintain an intracellular homeostatic state and (c) stabilisation of proteins. Consequently, these organisms have, in turn, acquired different physiological adaptations necessary to meet these challenges.

2.4.1. Respiratory / Energy needs

Chemolithotrophic thermophiles obtain energy from the oxidation of inorganic compounds and carbon dioxide (CO_2) serves as the sole carbon source (Reysenbach and Cady, 2001). A variety of inorganic compounds are used as electron donors and acceptors (Huber *et al.*, 2000). Under aerobic conditions, molecular hydrogen is converted to water, while elemental sulphur, sulphides or thiosulphates are oxidised to sulphuric acid in energy-yielding reactions. The above energy-yielding reactions have been identified in hyperthermophilic

genera such as *Aquifex*, *Sulfolobus*, *Metallosphaera* and *Acidianus* (Huber *et al.*, 1992; Huber *et al.*, 2000). Anaerobically, elemental sulphur, sulphides or thiosulphates, nitrates, nitrites and CO_2 may serve as suitable electron acceptors for growth, while molecular hydrogen may act as the electron donor in these energy-yielding reactions (Karnauchow *et al.*, 1992; Miroshnichenko *et al.*, 1998; Huber *et al.*, 2000). In some organisms, e.g. *Ferrioglobus placidus* (Huber *et al.*, 2000), ferrous ions serve as electron donors for nitrate reduction, while in *Thermovenabulum ferriorganovorum*, ferric ions act as electron acceptors (Zavarzina *et al.*, 2002).

Autotrophic thermophiles, on the other hand, obtain cellular carbon and energy through oxidative reduction of CO_2 (Reysenbach and Cady, 2001) via the tricarboxylic acid cycle or reductive Acetyl-CoA cycle, and the recently described 3-hydroxypropionate cycle (Huber *et al.*, 2000).

Heterotrophic thermophiles and hyperthermophiles satisfy their energy requirements through the oxidation of organic matter either aerobically or anaerobically. Many obligate aerobes described among genera such as *Bacillus*, *Thermus* and *Rhodothermus* oxidise organic carbon to CO_2 and H_2O to yield energy. Anaerobic species, such as *Clostridium* spp. (Rani and Nand, 2000), *Clostridium thermolacticum* and *Moorella thermoautotrophica* (Talabardon *et al.*, 2000), *T. phaeum* (Hattori *et al.*, 2000), *C. pacificum* (Sokolova *et al.*, 2001), *C. azorensis* (Wery *et al.*, 2001), *T. toyohensis* (Mori *et al.*, 2002) and *C. sporogenes* (Alain *et al.*, 2002), exhibit fermentative/chemoorganoheterotrophic respiration. While obligate fermentative strains do not require external electron acceptors (Alain *et al.*, 2002), facultative strains do. Electron acceptors, including thiosulphates, elemental sulphur, sulphates and ferrous or ferric ions, are reduced in a number of coupled oxidation reactions with a concomitant yield of energy (Wery *et al.*, 2001; Hattori *et al.*, 2000; Mori *et al.*, 2002). Since enzymes catalyse the above reactions and other metabolic processes, it requires that the cell's enzymes and other proteins be thermostable.

2.4.2. Stabilisation of proteins at high temperatures

Proteins from both thermophilic and hyperthermophilic bacteria are inherently thermostable. This has raised researcher's interest towards understanding of the factors that govern protein thermostability (Karshikoff and Landenstein, 1998; Kannan and Vishveshwara, 2000).

Comparative analytical studies of the primary structures of isofunctional enzymes from thermophilic, mesophilic and psychrophilic bacteria, for example, have revealed no substantial differences between them (Feller *et al.*, 1997; Lengeler *et al.*, 1999; Nèmeth *et al.*, 2002). Furthermore, analysis of 18 nonredundant protein families, containing pairs of high resolution protein structures from both thermophilic and mesophilic organisms, have indicated that both mesophile-thermophile homologous pairs were similar as regards their hydrophobicities, compactness, oligomeric states, main chain-main chain and main chain-side chain hydrogen bonds (Kumar *et al.*, 2000a; 2000b). Although, there appears to be a general increase in these factors among thermophilic members in a family of isofunctional proteins (Vieille and Zeikus, 1996), it has not been possible to develop general rules governing protein stabilisation (Jaenicke, 2000).

It has been established that a combination of cumulative small stabilising and destabilising factors operate either singularly or in synergy to stabilise thermophilic proteins. For example, few amino acid substitutions occur in regions occupied by thermolabile residues. Residue substitutions such as Lys to Arg, Ser to Ala and Gly to Ala, coupled with the general increase in the Arg to Arg+Lys ratio with increasing temperature, have been observed (Russel and Taylor, 1995; Feller *et al.*, 1997; Chung *et al.*, 2000). Such substitutions may contribute to the stabilisation of the tertiary structure of the enzyme through increased internal hydrophobicity, which is accompanied by a decrease in flexibility (Russel and Taylor, 1995). Furthermore, amino acids such as Arg and Tyr may be useful both in short-range local interactions and in long-range interactions. This, in some instances, results in an increase in the number of ion-pairs and ion-pair networks, while in other instances salt bridges are formed (Chang *et al.*, 1999; Kumar *et al.*, 2000a). Additionally, helix-destabilising residues such as Pro, His and Cys are avoided, while Arg and Ala, the helix-stabilising residues, are favoured (Kumar *et al.*, 2000b; Vieille and Zeikus, 1996).

While factors such as hydrophobicity, compactness, oligomeric states, main-chain and side-chain hydrogen bonds do not show consistent trends between mesophilic and thermophilic proteins, an increase in the number of salt bridges and side chain-side chain hydrogen bonds are commonly found in the majority of thermophilic proteins (Waldburger *et al.*, 1995; Kumar *et al.*, 2000a; 2000b). The stabilising nature of salt bridges lies in their networks and the co-operative nature of electrostatic interactions. The co-operativity of the network can resist unfolding, thereby opposing disorder and thus provide a mechanism to counteract

melting/unfolding (Kumar *et al.*, 2000b). This is through an increase in the electrostatic interactions that accompanies the structural plasticity in protein folds. The role of ion networks was also demonstrated in experiments with alcohol dehydrogenase from *Bacillus stearothermophilus* (Fiorentino *et al.*, 1998) whereby the elimination of ion-pair networks, through disruption of specific salt bridges in α-helices, decreased its thermostability. The main chain-main chain, main chain-side chain and side chain-side chain hydrogen bonds have been reported to increase among monomers of most thermophilic proteins and at their interface (Tanner *et al.*, 1996; Kumar *et al.*, 2000b). It has been suggested that hydrogen bonds may be favoured over salt bridges or neutral-neutral hydrogen bonds, because of the lower entropy and enthalpy costs for burying a charged-neutral hydrogen bond, which is lower than that associated with burying a salt bridge or a neutral-neutral hydrogen bond (Vieille and Zeikus, 1995; Kumar *et al.*, 2000a).

Thermophilic proteins also improve their thermostability through the reduction of possible conformations in the unfolded protein state by replacing amino acid residues with Gly and introducing Pro residues (Vieille and Zeikus, 1995), or by introducing covalent linkages such as disulfide bonds (Zhou *et al.*, 1993; Clarke and Fersht, 1993; DeDecker *et al.*, 1996; Li *et al.*, 1998; Németh *et al.*, 2002). As non-Gly residues present in proteins with a left-handed helical conformation often cause protein instability, replacement of such residues with Gly leads to a reduction in the conformation strain with an accompanying increase in stability. For example, Glu_{15} in the *Bacillus subtilis* DNA-binding protein HU and Lys_{95} in *E. coli* RNaseHI (both situated in turn regions) are substituted with Gly residues in their thermophilic counterparts (Vieille and Zeikus, 1995). When the contribution of the substitution was tested in *E. coli* DNA-binding protein HU by site-directed mutagenesis, the $Glu_{15} \rightarrow$ Gly mutation increased the stability of the protein, and it was found to account for 70% of the stability difference between the mesozymes and thermozymes. Disulfide bonds reduce the conformational strain through a decrease in the entropy of the unfolded state of the protein, while increasing their enthalpy state (Clarke and Fersht, 1993; Zhou *et al.*, 1993; Li *et al.*, 1998; Németh *et al.*, 2002). Most disulfide crosslinks occur in the flexible regions (terminal regions) rather than in the rigid regions (α-helices) and may therefore bring the terminal regions in close proximity to each other, and in effect improve on protein packing.

Aromatic clusters or enlarged aromatic networks (Kannan and Vishveshwara, 2000), deletion or shortening of loops (Vieille and Zeikus, 1996; Kumar *et al.*, 2000b), elimination/avoidance of potential sites for deamidation and isoaspartate formation (Chang *et al.*, 1999), and specific chaperones (Russel and Taylor, 1995) are all considered additional contributory factors to thermostability, and operate at a structural level to further strengthen the tertiary structure of the proteins (in the case of aromatic clusters and enlarged aromatic networks). In halophiles, protein stability and function are maintained by increased ion-binding and glutamic acid-content, both allowing the protein inventory to compete for water at high salt concentrations (Jaenicke, 2000).

2.4.3. Maintenance of a homeostatic intracellular environment

Survival at high temperatures requires maintenance of a homeostatic intracellular environment. This, in turn, requires that the membrane remains both stable and fluid to allow movement across, into and out of the cell, in addition to maintenance of a stable intracellular pH and ionic state. In order to achieve the former, saturated lipids are favoured in the membrane (Stanier *et al.*, 1996). An increase in saturated lipids with increasing temperature eliminates the leakage of ions, as an increase in the packaging of saturated lipids results in increased hydrophobic interactions. Thermophilic archaeabacteria, on the other hand, have solved the problem of membrane stability by producing ether lipids, which have a high melting temperature, and thus aid in maintaining the membrane integrity at high temperatures (Fiala *et al.*, 1998; Ko *et al.*, 1996; Pancost and Damsté, 2003).

Among acidophiles and alkalophiles, the internal cytoplasmic pH is maintained around neutral (Horikoshi, 1999; Jaenicke, 2000). This is achieved in part through an increase in the content of ionisable amino acids on proteins facing the outside extremes of pH and through transporter systems within the cell membrane. Furthermore, the ABC (ATP-binding cassette), the Na^+-dependent antiporter system and the solute-binding-protein-dependent transporter systems (Jaenicke, 2000; Driessen *et al.*, 2000), contribute to maintaining the internal homeostatic state through regulation of the rate of proton pumping at elevated temperatures to counter leakage (Van der Oost *et al.*, 1996).

2.5. THERMOSTABLE ENZYMES

Enzymes of industrial importance are broadly grouped into three major categories, namely: carbohydrate-degrading enzymes (amylases, cellulases, xylanases, chitinases, etc.), lipid-degrading enzymes (lipases and phopholipases) and protein-degrading enzymes (proteinases and proteases).

2.5.1. Carbohydrate-degrading enzymes

Carbohydrate-hydrolysing enzymes, commonly known as glycosidases or carbohydrases (Uhlig, 1998), hydrolyse glycosidic bonds between two or more carbohydrates (such as starch, cellulose, xylan, chitin, pectin) or between carbohydrate and non-carbohydrate moieties (Uhlig, 1998; Eichler, 2001). Thermostable glycosidases, e.g. chitinases, chitosanases and cellulase 103, have been isolated from thermophilic bacteria and are being applied in the medical and agricultural sectors, detergent formulations, and in the textile and pharmaceutical industries (Uhlig, 1998; Lengeler *et al.*, 1999; Yoon *et al.*, 2000; Schiraldi and De Rosa, 2002).

2.5.2. Lipid-degrading enzymes

Thermostable microbial lipases and esterases have also been isolated from thermophilic bacterial species, such as *Bacillus thermoleovorans* ID-1 and *Bacillus stearothermophilus* JY144, and are finding increasing industrial application in the synthesis of chiral compounds in non-aqueous solvents and in stereospecific enantioselectivity of enantiomers (Lee *et al.*, 2001; Kim *et al.*, 2002). Furthermore, thermostable lipases have been found to exhibit enhanced activity in the presence of organic solvents such as dimethylsulfoxide (DMSO), ethanol and β-mercaptoethanol (Eichler, 2001).

2.5.3. Proteases

Proteases are the most important industrial enzymes, accounting for about 60% of the total worldwide sales (Harwood, 1992; Mala *et al.*, 1998; Anwar and Saleemuddin, 1998; Mehrotra *et al.*, 1999). The wide range of applications of proteases, coupled with the need for enzymes that can operate under harsh conditions and in the presence of organic solvents and denaturing agents, has made thermophiles a promising source of such enzymes (Jaenicke *et al.*, 1996).

2.5.3.1. Classification of proteases

Proteases are classified based on (a) evolutionary history, (b) type of reaction catalysed, and (c) the nature of the active site (Mala *et al.*, 1998). The evolutionary history of proteases is hard to follow due to the lacunae in the data presented and appears, to some extent, contradictory. Depending on the type of reaction that is catalysed, proteases are classified as either exopeptidases (aminopeptidases and carboxypeptidases), which cleave peptide bonds proximal to the amino or carboxyl termini of the substrate, respectively, or endopeptidases, which cleave peptide bonds distant from both amino and carboxyl termini of the substrate. Based on the nature of the active site, proteases are classified as (a) serine proteases, i.e. a serine amino acid residue present in their active site, (b) aspartic proteases (acid proteases), which depend on the presence of an aspartic amino acid residue in the active site for their catalytic activity, (c) cysteine/thio-proteases, which depend on a catalytic dyad consisting of cysteine and histidine, and (d) metalloproteases, which require a divalent metal ion for their activity (Mala *et al.*, 1998).

2.5.3.2. Bacterial proteases of commercial importance and their synthesis in heterologous hosts

The development of bacterial enzymes for industrial and other biotechnological applications requires that the reaction kinetic properties, substrate specificities, catalytic activity, stability at different temperatures and pH, as well as stability in the presence of reducing agents and metal ions be characterised. The results from such studies have helped in the elucidation of catalytic mechanisms, which, in turn, has been used to guide bioengineering of the enzyme(s) for industrial/biotechnological applications. Furthermore, many bacterial species have been screened for enzyme production and the appropriate genes isolated, cloned and expressed in hosts such as *E. coli*, *B. subtilis* and *B. licheniformis* (in some cases) (Harwood, 1992).

The proteases of commercial importance are mainly neutral and alkaline proteases (Van Den Burg *et al.*, 1991; Kobayashi *et al.*, 1995). These proteases have been isolated from various mesophilic and thermophilic bacteria, with *Bacillus* spp. occupying the central position (Table 2.1). *Bacillus* spp. have an inherent ability to secrete a number of hydrolytic enzymes (e.g. carbohydrate-degrading and proteolytic enzymes) and other proteins (e.g. small peptides), into the extracellular milieu (Van der Laan *et al.*, 1991; Bolhuis *et al.*, 1999). This characteristic of *Bacillus* spp. has been exploited in the production of industrially important

enzymes such as alkaline proteases and neutral proteases (Harwood, 1992), and heterologous proteins such as human α-amylase (Bolhuis *et al.*, 1999), interferons (Henner, 1986) and streptolysin O (Yamada *et al.*, 1995).

Currently, *E. coli* and *B. subtilis* are the most widely used hosts for cloning and expression of recombinant proteins (Wu *et al.*, 1991; Van der Laan *et al.*, 1991; Pang *et al.*, 1991; Peek *et al.*, 1993; Bolhuis *et al.*, 1999; Chavagnat *et al.*, 1999). This is due to their well-characterised physiology and genetics, and a number cloning and/or expression vectors have been developed over the years for use in these expression hosts. Despite the success of the two organisms, there are some limitations to their general application. For example, *E. coli* is not an ideal system for expression of secreted proteins. Most of the proteins are not exported to the extracellular environment, but rather accumulate in the periplasmic space or remain attached to the outer or inner membranes (Mezes *et al.*, 1983; Ikemura *et al.*, 1987; Tan *et al.*, 2002). Recovery of such proteins involves cell disruption and aggregation of cell debris by flocculants, which leads to increased costs of downstream processing of the recombinant protein.

Bacillus subtilis, on the other hand, exports proteins efficiently to the extracellular milieu, but its high extracellular protease activity coupled with limited chaperone and signal peptidase (SipS and SipT) activities affect the efficiency of processing, export and stability of heterologous proteins, especially of eukaryotic origin (Bolhuis *et al.*, 1999). In some instances, the high extracellular protease activity renders many of the recombinant proteins functionless due to proteolysis. This disadvantage has been addressed by developing mutant *Bacillus* strains deficient in extracellular protease activity. *Bacillus subtilis* strain WB600, for example, is a mutant strain deficient for six of the seven proteases produced by *Bacillus subtilis* wild-type strains (Wu *et al.*, 1991). Another potentially useful host for the production of heterologous proteins is *B. brevis*. Like *B. subtilis*, it also secretes proteins directly into the culture medium where they accumulate to high levels in a relatively pure state (Yamada *et al.*, 1981; Yamagata *et al.*, 1985; Udaka and Yamagata 1993). *B. brevis*, however, has very low protease activity and a thin cell wall, which makes the export of recombinant proteins efficient, and the secreted proteins remain stable since they are not significantly degraded. Despite these advantages, its genetics is largely uncharacterised. Consequently, few vectors have been constructed for use in the cloning and expression of heterologous recombinant proteins in *B. brevis*.

2.5.3.3. Export of extracellular proteins

Many secreted proteins are exported as precursors (pre-proproteins). They consist of a signal peptide (pre-sequence) that is responsible for directing protein secretion across the membrane, and a peptide extension between the signal peptide and the mature protein that consists of varying number of amino acids residues, depending on the protein secreted (pro-sequence). The pro-sequence is essential for guiding appropriate folding of the secreted protein to yield an enzymatically active conformation (Wong and Doi, 1986; Ikemura *et al.*, 1987; Zhu *et al.*, 1989; Navarre and Schneewind, 1999; Tjalsma *et al.*, 2000). It is cleaved through an intra-molecular self-processing mechanism to produce a mature protein (Ikemura and Inouye, 1988; Zhu *et al.*, 1989). With the exception of the staphylococcal nuclease A gene (Takahara *et al.*, 1985; Peek *et al.*, 1993), successful expression, in their active conformation, of proteases originating from *Bacillus* spp. by hosts such as *E. coli* and *B. subtilis* has only been achieved by cloning the whole pro-proteinase gene (Takahara *et al.*, 1985; Peek *et al.*, 1993). Therefore, failure to identify foreign genes by screening for function in either *E. coli* or *B. subtilis* may often be a result of failure of processing the gene into a mature protein, because the prosequence was lost during random restriction digestion whilst preparing a genomic DNA library (Dalbøge, 1997).

2.5.3.4. Biotechnological applications of bacterial proteases

Enzymes, as opposed to chemical reagents, are increasingly being used as biocatalysts as a means of not only offsetting the impact of high production costs involved in using chemical catalysts but also to protect the environment from toxic recalcitrant industrial wastes. An obvious source of enzymes is microorganisms, i.e. viruses, bacteria and fungi, as they produce many enzymes that display a range of activities (Mala *et al.*, 1998).

Interest in the application of enzymes from hyperthermophiles for biotechnologically-related processes has rapidly accelerated (Eichler, 2001), since it is often advantageous to perform such processes at high temperatures. This is partly due to the fact that an increase in temperature may lead to a decrease in medium viscosity and an increase in the diffusion coefficient at elevated temperatures, thus resulting in higher reaction rates. Such considerations are relevant to a variety of processes including those involving hydrophobic compounds that normally display low solubilities. Initially, studies on hyperthermophiles were hampered by a lack of specialised growth vessels and protocols. However, as protocols

for the large-scale growth of such organisms are becoming available (Eichler, 2001), coupled with advances in recombinant DNA technology, the discovery of novel hyperthemophilic enzymes is greatly being enhanced.

2.5.3.4.1. Industrial uses

Currently, a number of enzymes, amongst them, alkaline and neutral proteases, are being used in different industrial processes. Table 2.1 shows various thermostable microbial proteases that have been studied with the aim of applying them in different industrial and biotechnological processes. For example, neutral proteases, e.g. thermolysin produced by *Bacillus thermoproteolyticus,* and alkaline proteases, e.g. subtilisins such as subtilisin Carlsberg produced by *B. licheniformis*, are constituents of most laundry detergents (Nagami and Tanaka, 1986; Tran *et al.*, 1991; Mala *et al.*, 1998). Takami *et al.* (1992a; 1992b) and Friedrich and Antranikian (1996) reported thermostable alkaline serine protease from *Bacillus* spp. No. AH-101 and from *Fervidobacterium pennavorans*, respectively, with keratinolytic activity. The properties of the two enzymes, i.e. their high thermostability, activity at alkaline pH (between pH 11-13) and resistance to denaturants, suggests that they may be suitable candidates for application in the leather tanning and poultry farming (waste management) industries. Another keratinolytic alkaline serine protease, produced by *Streptomyces albidoflavus*, which may also be applicable to the leather tanning industry, has been reported by Bressollier *et al.* (1999). The enzyme is active optimally at pH values ranging from 6 to 9.5 and at temperatures ranging from 40 to 70°C. Furthermore, proteolytic enzymes are being used in the dairy industry in the manufacture of cheese (Harwood, 1992), as well as in the bioconversion of proteinaceous waste into useful biomass (Anwar and Saleemuddin, 1998; Mala *et al.*, 1998).

2.5.3.4.2. Catabolism of aromatic compounds and xenobiotics

Degradation of aromatic compounds including phenolics, such as phenol, cresol, benzoate, catechol (Dong *et al.*, 1992; Duffner and Müller, 1998; Ali *et al.*, 1998; Duffner *et al.*, 2000), and substituted aromatics, such as 4-hydroxyphenylacetic acid (4-HPA) and dihydroxyphenylacetic acid (DHPA), has been widely studied among mesophiles and several degradation pathways have been elucidated. Mesophilic enzymes, with a few exceptions, have often been found very unstable, which has, in part, contributed to the failure of purifying a number of phenol-degrading enzymes. Thus far, successful purification of 4-

hydroxyphenylacetate 3-hydroxylase from *Pseudomonas putida* and *Pseudomonas putida* U, to homogeneity, has been reported (Raju *et al.*, 1988; Fernandez-Medarde and Luengo, 1997). The physiological role of these enzymes is to convert phenolic compounds into metabolic intermediates and the subsequent generation of energy through their catabolism. A number of phenol degradation pathways and their constituent enzymes have also been described among thermophilic *Bacillus* (Dong *et al.*, 1992; Duffner *et al.*, 2000). However, few genes have been cloned and sequenced (Dong *et al.*, 1992; Duffner *et al.*, 2000). Since it has been recognised that thermostable enzymes from thermophiles tend to be resistant to chemical denaturation (Gurujeyalakshmi and Oriel, 1989), thermophilic degradation conditions may be more advantageous in terms of higher metabolic rates, bio-availability of many organic pollutants, and increased enzyme stability (Duffner and Müller, 1998).

Table 2.1. Thermostable protease enzymes and their potential industrial relevance

Species	Protease type	Optimal pH	Optimal Temp. (°C)	Application	Reference
Bacillus sp. NKS-21	Alkaline serine protease	10-11	50-55	NS	Tsuchida *et al.* (1986)
B. stearothermophilus	Neutral protease	ND	ND	NS	Kubo and Imanaka (1988)
Bacillus sp. No. AH-101	Alkaline serine protease	12-13	80	Laundry detergent	Takami *et al.* (1989)
B. stearothermophilus	Neutral protease	ND	ND	NS	Nishiya and Imanaka *et al.* (1990)
B. caldolyticus	Neutral protease	ND	ND	NS	Van den Burg *et al.* (1991)
Bacillus sp. Ak.1	Alkaline serine protease	7.5	60-70	Laundry detergent	Peek *et al.* (1993)
Bacillus sp. KSM-K16	Alkaline serine protease	12.3	55	Laundry detergent	Hakamada *et al.* (1994)
Bacillus sp. EA1	Neutral protease	ND	ND	NS	Saul *et al.* (1996)
F. pennavorans	Alkaline serine protease (keratinolytic activity)	10.0	80	Keratin waste management	Friedrich and Antranikian (1996)
B. licheniformis	Alkaline serine protease	12.0	60	Laundry detergent	Ferrero *et al.* (1996)
Pimelobacter sp. Z-483	Alkaline serine protease	9.0	50	NS	Oyama *et al.* (1997)
S. albidoflavus	Alkaline serine protease	6-9.5	40-70	Leather industry	Bressollier *et al.* (1999)
Bacillus sp. BT1	Neutral protease	ND	ND	NS	Vecerek and Venema (2000)
T. kodakaerensis KOD1	Alkaline serine protease	9.5	80	NS	Kannan *et al.* (2001)
B. pumilus MK6-5	Alkaline serine protease	11.5	55-60	Ultrafitration membrane cleaning	Kumar (2002)

ND, not done; NS, no suggestion

2.6. REFERENCES

Abe, F. and Horikoshi, K. (2001). The biotechnological potential of piezophiles. *Trends in Biotechnology* 19: 102-108.

Adams, M. and Kelly, R. (1998). Finding and using hyperthermophilic enzymes. *Trends in Biotechnology* 16:329-332.

Agnew, M.D., Kovak, F.S. and Jarrel, F.K. (1995). Isolation and characterisation of novel alkaliphiles from Bauxite-processing waste and description of *Bacillus vedderi* sp. nov., a new obligate alkaliphile. *Systematic and Applied Microbiology* 18: 221-230.

Alain, K., Pignet, P., Zbinden, M., Quillevere, M., Duchiron, F., Donval, J-P., Lesongeur, F., Raguenes, G., Crassous, P., Querellou, J. and Cambon-Bonavita, M-A. (2002). *Caminicella sporogenes* gen. nov., sp. nov., a novel thermophilic spore-forming bacterium isolated from an East-Pacific Rise hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology* 52: 1621-1628.

Alfredsson, G.A., Kristjansson, J.K., Hjorleifsdottir, S. and Stetter, K.O. (1988). *Rhodothermus marinus*, gen. nov., sp. nov., a thermophilic bacterium from sub-marine hot springs in Iceland. *Journal of General Microbiology* 134: 299-306.

Ali, S., Fernandez-Lafuente, R. and Cowan, D.A. (1998). Meta-pathway degradation of phenolics by themophilic *Bacilli*. *Enzyme and Microbial Technology* 23: 462-468.

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

Anwar, A. and Saleemuddin, M. (1998). Alkaline proteases: A Review. *Bioresource Technology* 64: 175-183.

Archenbach-Richter, L., Gupta, R., Stetter, O.K. and Woese, R.C. (1987). Were the original eubacteria thermopiles? *Systematic and Applied Microbiology* 9: 34-39.

Ash, C., Farrow, J.A.E., Wallbank, S. and Collins, M.D. (1991). Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. *Letters in Applied Microbiology* 13: 202-206.

Baker, G.C., Gaffar, S., Cowan, D.A. and Suharto, A.R. (2001). Bacterial community analysis of Indonesian hot springs. *FEMS Microbiology Letters* 200: 103-109.

Bolhuis, A., Tjalsma, H., Smith, H.E., De Jong, A., Meima, R., Venema, G., Bron, S. and Van Dijl, J.M. (1999). Evaluation of bottlenecks in the late stages of protein secretion in *Bacillus subtilis*. *Applied and Environmental Microbiology* 66: 2934-2943.

Bressollier, P., Letourneau, F., Urdaci, M. and Verneuil, B. (1999). Purification and characterisation of a keratinolytic serine proteinase from *Streptomyces albidoflavus*. *Applied and Environmental Microbiology* 65: 2570-2576.

Brian, A. and Fergus, P. (1986). Modern Bacterial Taxonomy. Van Nostrand Reinhold Co Ltd., Berkshire, England, pp. 43-71.

Burggraf, S., Fricke, H., Neuner, A., Kristjansson, J., Rouvier, P., Mandelco, L., Woese, R.C. and Stetter, K.O. (1990). *Methanococcus igneus* sp. nov., a novel hyperthermophilic methanogen from a shallow sub-marine hydrothermal system. *Systematic and Applied Microbiology* 13: 263-269.

Chang, C., Park, B.C., Lee, D-S. and Suh, S.W. (1999). Crystal structures of thermostable xylose isomerases from *Thermus caldophilus* and *Thermus thermophilus*: Possible structural determinants of thermostability. *Journal of Molecular Biology* 288: 623-634.

Chavagnat, F., Casey, M.G. and Meyer, J. (1999). Purification, characterisation, gene cloning, sequencing, and overexpression of aminopeptidase N from *Streptococcus thermophilus* A. *Applied and Environmental Microbiology* 65: 3001-3007.

Chung, A.P., Rainey, A.F., Valente, M., Nobre, M.F. and da Costa, S.M. (2000). *Thermus igniterrae* sp. nov., and *Thermus antranikianii* sp. nov., two new species from Iceland. *International Journal of Systematic and Evolutionary Microbiology* 50: 209-217.

Clarke, J. and Fersht, A.R. (1993). Engineering disulfide bonds as probes of the folding pathway of Barnase: Increasing the stability of proteins against the rate of denaturation. *Biochemistry* 32: 4322-4329.

Dalbøge, H. (1997). Expression cloning of fungal enzyme genes; a novel approach for efficient isolation of enzyme genes of industrial relevance. *FEMS Microbiology Reviews* 21: 29-42.

Darland, G. and Brock, D.T. (1971). *Bacillus acidocaldarius* sp. nov., an acidophilic thermophilic spore-forming bacterium. *Journal of General Microbiology* 67: 9-15.

DeDecker, B.S., O'Brien, R., Fleming, P.J., Geiger, J.H., Jackson, S.P. and Sigler, P.B. (1996). The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *Journal of Molecular Biology* 264: 1072-1084.

Dong, L., Wang, L., Wang, C., Cheng, J., He, Z., Sheng, Z. and Shen, R. (1992). Molecular cloning and mapping of phenol degradation genes from *Bacillus stearothermophilus* FDTP-3 and their expression in *Escherichia coli*. *Applied and Environmental Microbiology* 58: 2531-2535.

Driessen, A.J.M., Rosen, P.B. and Koning, W.N. (2000). Diversity of transport mechanisms: Common structural principles. *Trends in Biochemical Science* 25: 397-401.

Duffner, F.M. and Müller, R. (1998). A novel phenol hydroxylase and catechol 2,3-dioxygenase from the thermophilic *Bacillus thermoleovorans* strain A2: Nucleotide sequence and analysis of the genes. *FEMS Microbiology Letters* 161: 37-45.

Duffner, F.M., Kirchner, U., Bauer, M.P. and Müller, R. (2000). Phenol/cresol degradation by the thermophilic *Bacillus thermoglucosidasius* A7: Cloning and sequence analysis of five genes involved in the pathway. *Gene* 256: 215-221.

Eichler, J. (2001). Biotechnological uses of archaeal extremozymes. *Biotechnology Advances* 19: 261-278.

Feller, G., Arpigny, L.J., Narinx, E. and Gerday, C. (1997). Molecular adaptations of enzymes from psychrophilic organisms. *Comparative Biochemistry and Physiology* 118A: 495-499.

Fernandez-Medarde, A. and Luengo, J.S. (1997). Purification and characterisation of the 4-hydroxyphenylacetic acid 3-hydroxylase from *Psuedomonas putida* U. *FEMS Microbiology Letters* 157: 47-53.

Ferrero, M.A., Castro, G.R., Abate, M.C., Baigori, D.M. and Siñeriz, F. (1996). Thermostable alkaline proteases of *Bacillus licheniformis* MIR 29: Isolation, production and characterisation. *Applied Microbiology and Biotechnology* 45: 327-332.

Fiala, G., Stetter, O.K., Jannasch, W.H., Langworthy, A.T. and Madon, J. (1986). *Staphylothermus marinus* sp. nov. represents a novel genus of extremely thermophilic sub-marine heterotrophic Archaebacteria growing at 98°C. *Systematic and Applied Microbiology* 8: 106-113.

Fiorentino, G., Cannio, R., Rossi, M. and Bartolucci, S. (1998). Decreasing the stability and changing the substrate specificity of the *Bacillus stearothermophilus* alcohol dehydrogenase by single amino acid replacement. *Protein Engineering* 11: 925-930.

Friedrich, A.B. and Antranikian, G. (1996). Keratin degradation by *Fervidobacterium pennavorans*, a novel thermophilic anaerobic species of the order *Thermotogales*. *Applied and Environmental Microbiology* 62: 2875-2882.

Friedrich, A., Prust, C., Hartsh, T., Henne, A. and Averhoff, B. (2002). Molecular analyses of the natural transformation machinery and identification of pilus structures in the extremely thermophilic bacterium *Thermus thermophilus* strain HB27. *Applied and Environmental Microbiology* 68: 745-755.

Goodfellow, M. and O'Donnell, A.G. (Eds.) (1994). Chemical Methods in Prokaryotic Systematics. John Willey and Sons Ltd., USA, pp. 463-517.

Grote, R., Li, L., Tamaoka, J., Kato, C., Horikoshi, K. and Antranikian, G. (1999). *Thermococcus siculi* sp. nov., a novel hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent at the Mid-Okinawa trough. *Extremophiles* 3: 55-62.

Gudni, A.A., Sigurdur, B. and Jacob, K.K. (1985). Nutritional diversity among *Thermus* spp. isolated from Icelandic hot springs. *Systematic and Applied Microbiology* 6: 308-311.

Gugliandolo, C. and Maugeri, T.L. (1997). Temporal variations in heterotrophic mesophilic bacteria from a marine shallow hydrothermal vent off the Island of Vulcano (Eolian Islands, Italy). *Microbial Ecology* 36: 13-22.

Gurujeyalakshmi, G. and Oriel, P. (1989). Isolation of phenol-degrading *Bacillus stearothermophilus* and partial characterisation of phenol hydroxylase. *Applied and Environmental Microbiology* 55: 500-502.

Hakamada, Y., Kobayashi, T., Hitomi, J., Kawai, S. and Ito, S. (1994). Molecular cloning and nucleotide sequence of the gene for an alkaline protease from the alkalophilic *Bacillus* sp. KSM-K16. *Journal of Fermentation and Bioengineering* 76: 105-109.

Haridon, L.S., Miroshnichenko, L.M., Fardeau, M-L., Bonch-Osmolovskaya, E.A., Stackebrandt, E. and Jeanthon, C. (2002). *Petrotoga olearia* sp. nov. and *Petrotoga sibirica* sp. nov., two thermophilic bacteria isolated from a continental petroleum reservoir in Western Siberia. *International Journal of Systematic and Evolutionary Microbiology* 52: 1715-1722.

Harwood, R.C. (1992). *Bacillus subtilis* and its relatives: Molecular biology and industrial workhorses. *Trends in Biotechnology* 10: 247-256.

Hattori, S., Kamagata, Y., Hanada, S. and Shoun, H. (2000). *Thermacetogenium phaeum* gen. nov., sp. nov., a strictly anaerobic, thermophilic, syntrophic acetate-oxidising bacterium. *International Journal of Systematic and Evolutionary Microbiology* 50: 1601-1609.

Head, I.M., Saunders, J.R. and Pickup, R.W. (1998). Microbial evolution, diversity, and ecology: A decade of ribosomal RNA analysis of unculturable microorganisms. *Microbial Ecology* 35: 1-21.

Hemmi, H., Ikejiri, S., Yamishita, S. and Nishino, T. (2002). Novel medium-chain prenyl diphosphate synthase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Journal of Bacteriology* 184: 615-620.

Henner D, (1986) Expression of heterologous genes in *Bacillus subtilis*. METHODS IN ENZYMOLOGY 185: 199-227

Horikoshi, K. (1999). Alikaliphiles: Some applications of their products for biotechnology. *Microbiology and Molecular Biology Reviews* 63: 735-750.

Huber, R., Wilharm, T., Huber, D., Trincone, A., Burggraf. S., König, H., Rachel, R., Rockinger, I., Fricke, H. and Stetter, O.L. (1992). *Aquifex pyrophilus* gen. nov.,. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. *Systematic and Applied Microbiology* 15: 340-351.

Huber, R., Huber, H. and Stetter, O.K. (2000). Towards the ecology of hyperthermophiles: Biotypes, new isolation strategies and novel metabolic properties. *FEMS Microbiology Reviews* 24: 615-623.

Hudson, J.A., Morgan, W.H. and Daniel, M.R. (1987). Numerical classification of some *Thermus* isolates from Icelandic hot springs. *Systematic and Applied Microbiology* 9: 218-223.

Ikemura, H., Takagi, H. and Inouye, M. (1987). Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli*. *The Journal of Biological Chemistry* 262: 7859-7864.

Jaenicke, R., Schurig, H., Beaucamp, N. and Ostendorp, R. (1996). Structure and stability of hyperstable proteins: Glycolytic enzymes from hyperthermophilic bacterium *Thermotoga maritima*. *Advances in Protein Chemistry* 48: 181-269.

Jaenicke, R. (2000). Stability and stabilisation of globular proteins in solution. *Journal of Biotechnology* 79: 193-203.

Jarsch, M. and Böck, A. (1985). Sequence of the 16S ribosomal RNA gene from *Methanococcus vannielli*: Evolutionary implications. *Systematic and Applied Microbiology* 6: 54-59.

Kannan, N. and Vishveshwara, S. (2000). Aromatic clusters: A determinant of thermal stability of thermophilic proteins. *Protein Engineering* 13: 753-761.

Kannan, Y., Koga, Y., Inoue, Y., Haruki, M., Takagi, M., Imanaka, T., Morikawa, M. and Kanaya, S. (2001). Active subtilisin-like protease from a hyperthermophilic Archaeon in a form with a putative prosequence. *Applied and Environmental Microbiology* 67: 2445-2452.

Karnauchow, M.T., Koval, F.S. and Jarrel, F.K. (1992). Isolation and characterisation of three thermophilic anaerobes from a St. Lucia hot spring. *Systematic and Applied Microbiology* 15: 296-310.

Karshikoff, A. and Landenstein, R. (1998). Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. *Protein Engineering* 11: 867-872.

Kim, I.C. and Oriel, P.J. (1995). Characterisation of a *Bacillus stearothermophilus* BR219 phenol hydroxylase gene. *Applied and Environmental Microbiology* 61: 1252-1256.

Kim, B.S., Falconer, C., Williams, E. and Goodfellow, M. (1998). *Streptomyces thermocarboxydovorans* sp. nov. and *Streptomyces thermocarboxydus* sp. nov., two moderately thermophilic carboxydotrophic species from soil. *International Journal of Systematic Bacteriology* 48: 59-68.

Kim, B., Sahin, N., Minnikin, E.D., Zakrzewska-Czerwinska, J., Mordarski, M. and Goodfellow, M. (1999). Classification of thermophilic streptomycetes, including the description of *Streptomyces thermoalcalitolerans* sp. nov. *International Journal of Systematic Bacteriology* 49: 1-17.

Kim, J-Y., Choi, G-S., Kim, Y-J., Ryu, Y-W. and Kim, G-J. (2002). A new isolate *Bacillus stearothermophilus* JY144 expressing a novel esterase with high enantioselectivity to (*R*)-ketoprofen ethyl ester: Strain selection and gene cloning. *Journal of Molecular Catalysis: Enzymatics* 18: 133-145.

Kobayashi, T., Hakamada, Y., Achida, S., Hitomi, J., Yoshimatsu, T., Koike, K., Kawai, S. and Ito, S. (1995). Purification and properties of an alkaline protease from *Bacillus* sp. KSM-K16. *Applied Microbiology and Biotechnology* 43: 473-481.

Ko, J.H., Jang, W.H., Kim, E.K., Lee, H.B., Park, K.D., Chung, J.H. and Yoo, O.J. (1996). Enhancement of thermostability and catalytic efficiency of AprA, an alkaline protease from *Psuedomonas* sp., by the introduction of a disulfide bond. *Biochemical and Biophysical Research Communications* 221: 631-635.

Kubo, M. and Imanaka, T. (1988). Cloning and nucleotide sequence of the highly thermostable neutral protease gene from *Bacillus stearothermophilus*. *Journal of General Microbiology* 134: 1883-1892.

Kumar, S., Ma, B., Tsai, C-J. and Nussinov, R. (2000a). Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. *Proteins: Structure, Function and Genetics* 38: 368-383.

Kumar, S., Tsai, C-J. and Nussinov, R. (2000b). Factors enhancing protein thermostability. *Protein Engineering* 13: 179-191.

Kumar, C.G. (2002). Purification and characterisation of a thermostable alkaline protease from alkalophilic *Bacillus pumilus*. *Letters in Applied Microbiology* 34: 13-17.

Lengeler, J.W., Drews, G. and Schlegel, H.G. (Eds.) (1999). Biology of the Prokaryotes: Blackwell Science Inc. Thieme Stuttgart, New York, pp. 652-915.

Lee, D-W., Kim, H-W., Lee, K-W., Kim, B-C., Choe, E-A., Lee, H-S., Kim, D-S. and Pyun, Y-R. (2001). Purification and characterization of two distinct thermostable lipases from the Gram-negative *Bacillus thermoleovorans* ID-1. *Enzymes and Microbial Technology* 29: 363-370.

Li, Y., Coutinho, P.M. and Ford, C. (1998). Effect on thermostability and catalytic activity of introducing disulfide bonds into *Aspergillus awamori* gluciamylase. *Protein Engineering* 11: 661-667.

Mala, B.R., Aparn, M.T., Mohini, S.G. and Vasanti, V.D. (1998). Molecular and biotechnological aspects of microbial poteases. *Microbiology and Molecular Biology Reviews* 62: 597-635.

Martiensson, T.V., Birrien, J., Kristjánsson, K.J. and Prieur, D. (1995). First isolation of thermophilic aerobic non-sporulating heterotophic bacteria from deep-sea hydrothermal vents. *FEMS Microbiology Ecology* 18: 163-174.

Martiensson, T.V., Birrien, J., Jeanthon, C. and Prieur, D. (1996). Numerical taxonomic study of thermophilic *Bacillus* isolated from three geographically separated deep-sea hydrothermal vents. *FEMS Microbiology Ecology* 21: 255-266.

Maugeri, L.T., Gugliandolo, C., Caccamo, D. and Stackebrandt, E. (2001). A polyphasic taxonomic study of thermophilic *Bacilli* from shallow, marine vents. *Systematic and Applied Microbiology* 24: 572-587.

Mehrotra, S., Pandey, P.K., Gaur, R. and Darmwal, N.S. (1999). The production of alkaline protease by *Bacillus* species isolate. *Bioresource Technology* 67: 201-203.

Mèzes, P.S.F., Yang, Y.Q., Hussain, M. and Lampen, J.O. (1983). *Bacillus cereus* 569/H ß-lactamase I: Cloning in *Escherichia coli* and signal sequence determination. *FEBS Letters* 161: 195-200.

Mori, K., Hanada, S., Maruyama, A. and Marumo, K. (2002). *Thermanaeromonas toyohensis* gen. nov., sp. nov., a novel thermophilic anaerobe isolated from a subterranean vein in the Toyoha Mines. *International Journal of Systematic and Evolutionary Microbiology* 52: 1675-1680.

Miroshnichenko, M.L., Rainey, A.F., Hippe, H., Chernyh, A.N., Kostrinika, A.N. and Bonch-Osmolovskaya, A.E. (1998). *Desulfurella kamchatkensis* sp. nov. and *Desulfurella propionica* sp. nov., new sulfur-respiring thermophilic bacteria from Kamchatka thermal environments. *International Journal of Systematic Bacteriology* 48: 475-479.

Nagami, Y. and Tanaka, T. (1986). Molecular cloning and nucleotide sequencing of a DNA fragment from *Bacillus natto* that enhances production of extracellular proteases and levansucrase in *Bacillus subtilis*. *Journal of Bacteriology* 166: 20-28.

Navarre, W.W. and Schneewind, O. (1999). Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and Molecular Biology Reviews* 63: 174-229.

Nazina, N.T., Tourova, P.T., Poltaraus, B.A., Novikova, V.E., Grigoryan, A.A., Ivanova, A.E., Lysenko, M.A., Petrunyaka, V.V., Ospov, A.G., Belyaev, S.S. and Ivanov, V.M. (2001). Taxonomic study of aerobic thermophilic bacilli: Descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustrophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustrophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. *International Journal of Systematic and Evolutionary Microbiology* 51: 433-446.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garret, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M. (1999). Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399: 323-329.

Németh, A., Komondi, S., Szilágyi, A., Magyar, C., Kovári, Z. and Závodsky, P. (2002). Increasing the thermal stability of cellulase C using rules learned from thermophilic proteins: A pilot study. *Biophysical Chemistry* 132: 229-241.

Nilsen, K.R. and Torsvik, T. (1996). *Methanococcus thermolithotrophicus* isolated from North Sea oil field reservoir water. *Applied and Environmental Microbiology* 62: 728-731.

Nishiya, Y. and Imanaka, T. (1990). Cloning and nucleotide sequence of the *Bacillus stearothermophilus* neutral protease gene and its transcription activator gene. *Journal of Bacteriology* 172: 4861-4869.

Nogales, B., Moore, E.R.B., Llobet-Brossa, E., Rossello-Mora, R., Amann, R. and Timmis, K.N. (2001). Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. *Applied and Environmental Microbiology* 67: 1874-1884.

Nold, S.C. and Ward, D.M. (1995). Diverse *Thermus* species inhabit a single hot spring microbial mat. *Systematic and Applied Microbiology* 18: 274-278.

Nunes, C.O., Donato, M.M., Manaia, M.C. and Da Costa, S.M. (1992a). The polar lipid and fatty acid composition of *Rhodothermus* strains. *Systematic and Applied Microbiology* 15: 59-62.

Nunes, C.O., Donato, M.M. and Da Costa, S.M. (1992b). Isolation and characterisation of *Rhodothermus* strains from S. Miguel, Azores. *Systematic and Applied Microbiology* 15: 92-97.

Pancost, D.R. and Damsté, S.S.J. (2003). Carbon isotopic compositions of prokaryotic lipids as tracers of carbon cycling in diverse settings. *Chemical Geology* 14134: 1-30.

Pang, A.S.H., Nathoo, S. and Wong, S-L. (1991). Cloning and characterisation of a pair of novel genes that regulate production of extracellular enzymes in *Bacillus subtilis*. *Journal of Bacteriology* 173: 46-54.

Patel, B.K.C., Morgan, W.H. and Daniel, M.R. (1985a). *Fervidobacterium nodosum* gen. nov. and sp. nov., a new chemo-organotrophic, caldoactive, anaerobic bacterium. *Archives Of Microbiology* 141: 63-69.

Patel, B.K.C., Morgan, W.H. and Daniel, M.R. (1986). Studies on some thermophilic glycolytic anaerobic bacteria from New Zealand hot springs. *Systematic and Applied Microbiology* 8: 128-136.

Peek, K., Daniel, R.M., Monk, C., Parker, L. and Coolbear, T. (1992). Purification and characterisation of a thermostable proteinase isolated from *Thermus* sp. strain Rt41A. *European Journal of Biochemistry* 207: 1035-1044.

Peek, K., Veitch, P.D., Prescott, M., Daniel, M.R., MacIver, B. and Bergquist, L.P. (1993). Some characteristics of a proteinase from a thermophilic *Bacillus* sp expressed in *Escherichia coli*:

Comparison with the native enzyme and its processing in *E. coli* and *in vitro*. *Applied and Environmental Microbiology* 59: 1168-1175.

Pienemann-Simon, S., Ludwig, W., Vogt, B. and Gottschalk, G. (1995). Taxonomic analysis of the thermophilic bacterium strain SP83 producing H_2 from starch at 75°C. *Systematic and Applied Microbiology* 18: 231-236.

Rainey, F.A., Fritze, D. and Stackebrandt, E. (1994). The phylogenetic diversity of thermophilic members of the genus *Bacillus* as revealed by 16S rDNA analysis. *FEMS Microbiology Letters* 115: 205-212.

Raju, G.S., Kamath, A.V. and Vaidyanathan, S.C. (1988). Purification and properties of 4-hydroxyphenylacetic acid 3-hydrxylase from *Pseudomonas putida*. *Biochemical and Biophysical Research Communications* 154: 537-543.

Rani, D.S. and Nand, K. (2000). Production of thermostable cellulase-free xylanase by *Clostridium absonum* CFR-702. *Process Biochemistry* 36: 355-362.

Reysenbach, A. and Cady, S. (2001). Microbiology of ancient and modern hydrothermal systems. *Trends in Microbiology* 9: 79-85.

Rondon, M.R., Goodman, M.R. and Handelsman, J. (1999). The Earth's bounty: Assessing and accessing soil microbial diversity. *Trends in Biotechnology* 17: 403-409.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, O.J., Handelsman, J. and Goodman, R.M. (2000). Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology* 66: 2541-2547.

Rőssler, D., Ludwig, W., Schleifer, H.K., Lin, C., McGill, T.J., Wistzkey, D.J., Jurtshuk, P. and Fox, G. (1991). Phylogenetic diversity in the genus *Bacillus* as seen by 16S rRNA sequencing studies. *Systematic and Applied Microbiology* 14: 266-269.

Russel, R.J.M. and Taylor, G.L. (1995). Engineering thermostability: Lessons from thermophilic proteins. *Current Opinion in Biotechnology* 6: 370-374.

Santos, H. and da Costa, M.S. (2002). Compatible solutes of organisms that live in hot saline environments. *Environmental Microbiology* 4: 501-509.

Saul, J.D., Williams, C.L., Toogood, S.H., Daniel, M.R. and Bergquist, L.P. (1996). Sequence of the gene encoding a highly thermostable neutral proteinase from *Bacillus* sp. strain EA1: Expression in *Escherichia coli* and characterisation. *Biochimica et Biophysica Acta* 1038: 74-80.

Schiraldi, C. and De Rosa, M. (2002). The production of biocatalysts and biomolecules from extremophiles. *Trends in Biotechnology* 20: 515-521.

Seyfried, M., Lyon, D., Rainey, A.F. and Wiegel, J. (2002). *Caloramator viterbensis* sp. nov., a novel thermophilic, glycerol-fermenting bacterium isolated from a hot spring in Italy. *International Journal of Systematic and Evolutionary Microbiology* 52: 1177-1184.

Sokolova, G.T., González, M.J., Kostrinika, A.N., Chernyh, A.N., Tourova, P.T., Kato, C., Bonch-Osmolovskaya, E.A. and Robb, T.F. (2001) *Carboxybrachium pacificum* gen. nov., sp. nov., a new anaerobic, thermophilic, CO-utilising marine bacterium from Okinawa Trough. *International Journal of Systematic and Evolutionary Microbiology* 51: 141-149.

Stanier, R.Y., Ingraham, J.L., Wheelis, M.L. and Painter, P.R. (1995). General Microbiology (5th Ed.). MacMillan Press Ltd., Prentice-Hall Inc., New Jersey, USA, pp. 204-210.

Steffan, J.R., Goksøyr, J., Bej, A.K. and Atlas, R.M. (1998). Recovery of DNA from soil and sediments. *Applied and Environmental Microbiology* 54: 2908-2915.

Summit, M. and Baross, A.J.(1998). Thermophilic sub-sea floor microorganisms from the 1996 North Gord Ridge eruption. *Deep-sea Research II* 45: 2751-2766.

Suzuki, Y., Kishigami, T., Inoue, K., Mizoguchi, Y., Eto, N., Takagi, M. and Abe, S. (1983). *Bacillus thermoglucosidasius* sp. nov., a new species of obligate thermophilic bacilli. *Systematic and Applied Microbiology* 4:487-495.

Talabardon, M., Schwitzguébel, J-P. and Përinger, P. (2000). Anaerobic thermophilic fermentation for acetic acid production from milk permeate. *Journal of Biotechnology* 76: 83-92.

Takami, H., Akiba, T. and Horikoshi, K. (1989). Production of extremely thermostable alkaline protease from *Bacillus* sp. no. AH-101. *Applied Microbiology and Biotechnology* 30: 120-124.

Takami, H., Akiba, T. and Horikoshi, K. (1992a). Cloning, expression, and characterisation of a minor alkaline protease from *Bacillus* sp. no. AH-101. *Bioscience Biotechnology and Biochemistry* 56: 510-511.

Takami, H., Kobayashi, T., Aono, R. and Horikoshi, K. (1992b). Molecular cloning, nucleotide sequence and expression of the structural gene for a thermostable alkaline protease from *Bacillus* sp. no. AH-101. *Applied Microbiology and Biotechnology* 38: 101-108.

Takahara, M., Hibler, D.W., Barr, P.J., Gerit, J.A. and Inouye, M. (1985). The *ompA* signal peptide-directed secretion of staphylococcal nuclease A by *Escherichia coli*. *Journal of Biological Chemistry* 260: 2670-2674.

Tan, S.N., Ho, B. and Ding, J.L. (2002). Engineering a novel secretion signal for cross-host recombinant protein expression. *Protein Engineering* 15: 337-345.

Tanner, J.J., Hecht, R.M. and Krause, K.L. (1996). Determinants of enzyme thermostability observed in the molecular structure of *Thermus aquaticus* D-glyceraldehye-3-phosphate dehydrogenase at 2.5Å resolution. *Biochemistry* 35: 2597-2609.

Tjalsma, H., Bolhuis, A., Jongbloed, H.D.J., Bron, S. and Van Dijl, J.M. (2000). Signal peptide-dependent protein transport in *Bacillus subtilis*: A genome-based survey of the secretome. *Microbiology and Molecular Biology Reviews* 64: 515-547.

Tran, L., Wu, X. and Wong, S. (1991). Cloning and expression of a novel protease gene encoding an extracellular neutral protease from *Bacillus subtilis*. *Journal of Bacteriology* 173: 6364-6372.

Tsai, Y-L. and Olson, B.H. (1991). Rapid method for direct extraction of DNA from soil and sediments. *Applied and Environmental Microbiology* 57: 1070-1074.

Udaka, S. and Yamagata, H. (1993). High-level secretion of heterologous proteins by *Bacillus brevis*. *Methods in Enzymology* 217: 23-33.

Uhlig, H. (1998). Industrial Enzymes and their Applications. John Wiley and Sons, New York, USA pp. 37-72.

Vandamme, P., Pot, B., Gillis, M., De Vos, P., Kersters, K. and Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological Reviews* 60: 407-438.

Van den Burg, B., Enequist, G.H., Van Der Haar, M.E., Eijsink, G.H.V., Stulp, B.K. and Venema, G. (1991). A highly thermostable neutral protease from *Bacillus caldolyticus*: Cloning and expression of the gene in *Bacillus subtilis* and characterisation of the gene product. *Journal of Bacteriology* 173: 4107-4115.

Van der Laan, J.C., Gerritse, G., Mulleners, L.J.S.M., Van der Hoek, R.A.C. and Quax, W.J. (1991). Cloning, characterisation, and multiple chromosomal integration of a *Bacillus* alkaline protease gene. *Applied and Environmental Microbiology* 57: 901-909.

Van der Oost, J., de Vos, W.M. and Antranikian, G. (1996). Extremophiles. *Trends in Biotechnology* 14: 415-417.

Vecerek, B. and Venema, G. (2000). Expression of the neutral protease gene from a thermophilic *Bacillus sp.* BT1 strain in *Bacillus subtilis* and its natural host: Identification of a functional promoter. *Journal of Bacteriology* 182: 4101-4107.

Vieille, C. and Zeikus, J.G. (1996). Thermozymes: Identifying molecular determinants of protein structural and functional stability. *Trends in Biotechnology* 14: 183-190.

Waldburger, C.D., Schildbach, J.F. and Sauer, R.T. (1995). Are buried salt bridges important for protein stability and conformation specificity? *Structural Biology* 2: 122-128.

Wang, W., Mézes, P.S.F., Yang, Y.Q., Blacher, R.W. and Lampen, J.O. (1985). Cloning and sequencing of the ß-lactamase I gene of *Bacillus cereus* 5/B and its expression in *Bacillus subtilis*. *Journal of Bacteriology* 163: 487-492.

Wang, L-F., Brückner, R. and Doi, R.H. (1989). Construction of a *Bacillus subtilis* mutant deficient in three extracellular proteases. *Journal of General and Applied Microbiology* 35: 487-492.

Wery, N., Moricet, J-M., Cueff, V., Jean, J., Pignet, P., Lesonguer, F., Cambon-Bonavita, M-A. and Barbier, G. (2001). *Caloranaerobacter azorensis* gen. nov., sp. nov., an anaerobic thermophilic bacterium isolated from a deep-sea hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology* 51: 1789-1796.

Woese, R.C., Kandler, O. and Wheelis, L.M. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eukarya. *Proceedings of the National Academy of Sciences, USA* 87: 4576-4579.

Wong, S. and Doi, R.H. (1986). Determination of the signal peptidase cleavage site in the preprosubtilisin of *Bacillus subtilis*. *Journal of Biological Chemistry* 261: 10176-10181

Wu, X-C., Lee, W., Tran, L. and Wong, S-L. (1991). Engineering a *Bacillus subtilis* expression-secretion system with a strain deficient in six extracellular proteases. *Journal of Bacteriology* 173: 4952-4958.

Xu, L-H., Tiang, Y-Q., Zhang, Y-F., Zhao, L-X. and Jiang, C-X. (1998). *Streptomyces thermogriseus*, a new species of the genus *Streptomyces* from soil, lake and hot spring. *International Journal of Systematic Bacteriology* 48: 1089-1093.

Yamada, H., Tsukagoshi, N. and Udaka, S. (1981). Morphological alterations of cell wall concomitant with protein release in a protein-producing bacterium, *Bacillus brevis* 47. *Journal of Bacteriology* 148: 322-332.

Yamada, S., Seki, M., Kato, H. and Yamashita, S. (1995). Expression of streptolysin O gene in *Bacillus subtilis*. *Bioscience, Biotechnology and Biochemistry* 59: 363-366.

Yamagata, H., Nakagawa, K., Tsukagoshi, Z. and Udaka, S. (1985). A stable plasmid vector and control of its copy number in *Bacillus brevis* 47, a protein-producing bacterium. *Applied and Environmental Microbiology* 49: 1076-1079.

Yamamoto, H., Hiraishi, A., Kato, K., Chiura, H.X., Maki, Y. and Shimizu, A. (1998). Phylogenetic evidence for the existence of novel thermophilic bacteria in hot spring sulfur-turf microbial mats in Japan. *Applied and Environmental Microbiology* 64: 1680-1687.

Yamamoto, K., Marukami, R. and Takamura, Y. (1998). Isoprenoid quinone, cellular fatty acid composition and diaminopimelic acid isomers of newly classified thermophilic anaerobic Gram-positive bacteria. *FEMS Microbiology Letters* 161: 351-358.

Yang, D., Kaine, P.B. and Woese, R.C. (1985). The phylogeny of Archaebacteria. *Systematic and Applied Microbiology* 6: 251-256.

Yoon, H-G., Kim, H-Y., Lim, Y-H., Kim, H-K., Shin, D-H., Hong, B-S. and Cho, H-Y. (2000). Thermostable chitosanase from *Bacillus* sp. strain CK4: Cloning and expression of the gene and characterisation of the enzyme. *Applied and Environmental Microbiology* 66: 3727-3734.

Zavarzina, G.D., Tourova, P.T., Kuznetsov, B.B., Bonch-Osmolovskaya, A.E. and Slobodkin, I.A. (2002). *Thermovenabulum ferrioganovorum* gen. nov., sp. nov., a novel thermophilic, anaerobic, endospore-forming bacterium. *International Journal of Systematic and Evolutionary Microbiology* 52: 1737-1743.

Zhou, N.E., Kay, C.M. and Hodges, R.S. (1993). Disulfide bond contribution to protein stability: Positional effects of substitution in the hydrophobic core of the two-stranded α-helical coiled-coil. *Biochemistry* 32: 3178-3187.

Zhu, X., Ohta, Y., Jordan, F. and Inouye, M. (1989). Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* 339: 483-484.

CHAPTER THREE*

THERMOPHILIC PROTEASE-PRODUCING *Geobacillus* FROM BURANGA HOT SPRINGS IN WESTERN UGANDA

ABSTRACT

Two thermophilic aerobic bacterial strains, PA-9 and PA-5, were isolated from hot spring environments in western Uganda. The cells were rods, approximately 9.7 to 12 μm in length. Whereas isolate PA-9 grew at temperatures between 38 and 68°C (optimum, 62°C), isolate PA-5 grew at temperatures between 37 and 72°C (optimum, 60°C). Both isolates grew optimally at pH 7.5 - 8.5. The 16S rRNA gene sequencing of isolates PA-9 and PA-5 indicated that they belong to the genus *Bacillus*. During growth on casein, the isolates produced enzymes with caseinolytic activity. Zymogram analysis of the crude enzyme extracts revealed the presence of two extracellular enzymes for isolate PA-5, and at least eight for isolate PA-9. The optimum temperature and pH for casein-degrading activity were 70°C, pH 6.5 for isolate PA-9, but caseinolytic activity could also be observed at 2°C. In the case of isolate PA-5, optimal activity was observed over a temperature and pH range of 50 - 70°C and pH 5 - 10, respectively.

This chapter has been published in:

Hawumba, J.F., Theron, J. and Brözel, V.S. (2002). Thermophilic protease-producing *Geobacillus* from Buranga hot springs in western Uganda. Current Microbiology 45, 144-150.

3.1. INTRODUCTION

The use of enzymes as diagnostic tools, as analytical reagents and as industrial catalysts is an important and rapidly expanding technology (Cowan, 1992). This expansion is driven by the requirements of some industries for high levels of specificity in the biosynthesis of high value products (as in the pharmaceutical industries) and by the success of some enzyme products as speciality reagents, e.g. the DNA polymerases in the polymerase chain reaction.

Research on thermo- and extremophiles, as promising sources for highly stable enzymes, has remained an active research subject (Davis, 1998). Thermophiles represent an obvious source of thermostable enzymes, being reasonable to assume that such character will confer their proteins a high thermal stability (Adams and Kelly, 1998). Enzymes isolated from these microorganisms are not only thermostable and active at high temperature, but are also often resistant to and active in the presence of organic solvents and detergents (Jaenicke *et al.*, 1996). Despite being sources of novel biochemical pathways, enzymes and bioproducts, the technological use of thermophiles still faces several challenges since knowledge on physiology, genetics and diversity of such organisms is generally poor (Prieur, 1992). Thermophiles or the genes derived from them are, however, still the preferred source for thermostable enzymes (Adams and Kelly, 1998).

By now, a high number of thermostable enzymes from thermophiles have been reported; most of them belonging to eubacterial and archaebacterial kingdoms. Hyperthermophilic archaebacteria and eubacteria, such as *Pyrococcus furiosus* and *Thermotoga* spp., have yielded several biotechnologically relevant enzymes (Adams and Kelly, 1998). However, *Bacillus* spp. have remained major sources of enzymes of industrial and commercial value and a number of thermophilic *Bacillus* spp. have been reported to produce useful enzymes (Mala *et al.*, 1998).

In this part of the study, two thermophilic protease-secreting, Gram-positive, spore forming aerobes were isolated from the Buranga hot springs in western Uganda and were characterised morphologically and physiologically. The preliminary characterisation of extracellular proteases with caseinolytic activity produced by these isolates is furthermore reported.

3.2. MATERIALS AND METHODS

3.2.1. Isolation of protease-secreting thermophiles

Soil and water samples collected from the Buranga hot springs in western Uganda, were serially diluted, plated onto milk agar plates (pH 8.5), and incubated at 55°C for 48 h. Individual colonies that produced clear zones were selected and purified via subculturing, and two isolates producing the largest hydrolysis zones were selected and designated as PA-5 and PA-9, respectively.

3.2.2. Culture conditions

The isolates were grown at 55°C for 48 h in modified Castenholz medium D (Castenholz, 1969) containing 5 g peptone, 2 g yeast extract, 2 g glucose, 100 ml Solution A, 1 ml Solution B and 10 ml Solution C per litre. Solution A contained (g/l) $CaCl_2$, 0.6; $MgSO_4.7H_2O$, 1; NaCl, 0.8; $NaNO_3$, 3; KNO_3, 1; Na_2HPO_4, 0.11; $FeCl_3$, 0.03 and $NH_4CH_3COO^-$, 0.36. Solution B contained (g/l) $MnCl_2$, 0.513; $ZnSO_4.7H_2O$, 0.25; H_3BO_3, 0.25; $(NH_4)_2MoO_4$, 0.0125; $CoCl_2$, 0.023 and $CuSO_4$, 0.0125. Solution C contained 10 mg/ml each of nicotinamide, riboflavin and *p*-aminobenzoic acid.

3.2.3. Morphological properties and growth limits

Differences in cellular morphology were examined by microscopic analysis of Gram-stained cells (Zeiss Axioskop, Germany), and images were captured using a COHU monochrome CCD camera (RS-170, Cohn Inc., San Diego, CA, USA) at a magnification of ×1000. The optimal pH for growth of each isolate was determined in 50 ml of the medium buffered with 10 mM benzoic acid at pH 4, 10 mM pyridine at pH 5, 10 mM phosphate buffer at pH 5, 10 mM Tris-HCl at pH 7 to 9, 10 mM $Na_2HPO_4.2H_2O$ at pH 10, and 10 mM methylamine-HCl at pH 11. The optimum growth temperature of each isolate was determined after inoculation of 15 ml of the medium, pre-equilibrated to temperatures ranging from 37 to 72°C. The tubes were incubated in a temperature gradient incubator (Scientific Industries, Mineola, USA) and the bacterial growth was followed by measuring optical density at 600 nm every 30 min for 10 h. At each of the investigated temperatures, the growth rate (μ) was calculated for the period of their exponential increase.

3.2.4. Phylogenetic analysis by 16S rDNA sequence analysis

Genomic DNA of both isolates PA-5 and PA-9 was isolated using the method described by Marmur and Doty (1962) with the following modifications. Cells from overnight cultures of the two isolates were collected by centrifugation and the pellet suspended in 2 ml of 50 mM TE buffer (pH 8.0), containing lysozyme (5 mg/ml). Following incubation at 37°C for 1 h, an equal volume 10% (w/v) SDS was added and the cell suspensions were incubated at 60°C for 20 min. The DNA was subsequently extracted with phenol/chloroform/isoamylalcohol (25:24:1) after which the chromosomal DNA was precipitated from the aqueous phase through the addition of 2 volumes 95% ethanol and sodium acetate (pH 5.2) to a final concentration of 0.3 M. The genomic DNA was recovered by spooling, washed in 70% ethanol and suspended in ddH_2O.

The 16S rDNA gene of each isolate was subsequently amplified by using the universal primers fD1 and rP2 (Weisburg *et al.*, 1991). The purified PCR product was directly sequenced with an ABI Prism model 377 DNA sequencer and the resulting sequence was edited to a total length of 1 347 nucleotides. The phylogenetic relationship of isolates PA-5 and PA-9 was determined by comparing the sequencing data with sequences of related *Bacillus* and *Geobacillus* spp. (GenBank database of the National Center for Biotechnology Information, www.ncbi.nlm.nih.gov/GenBank/) by using the neighbour-joining method of Saitou and Nei (1987) in combination with the bootstrap method (Felsenstein, 1985).

3.2.5. Preparation of crude extracellular protease extracts

The isolates were cultured at 55°C in medium containing 0.5% (w/v) meat extract and soybean-casein digest, and 1% (w/v) of either glucose or glycerol. Following incubation, cells were harvested by centrifugation (13 000 rpm, 5 min) and the supernatants were dialysed against 5 mM Tris buffer (pH 8) containing 2 mM $CaCl_2$ at 4°C overnight. The dialysates were concentrated by lyophilization and resuspended in the same buffer.

3.2.6. Proteolytic activity assays

For cup plate assays, 100 µl of crude enzyme extract was added to wells punched into casein or milk agar plates. After incubation at 55°C for 12 h, zones of hydrolysis around the wells were measured as an indicator of proteolytic activity. Proteolytic activity was also assayed by using azocasein (Sigma, St. Louis, MO, USA) as a substrate according to the method of

Smibert and Krieg (1994). The protein concentrations were determined according to the method of Bradford (1976) by using the Pierce assay reagent (Pierce, Rockford, Ill., USA) and bovine serum albumin as the standard.

3.2.7. Influence of pH and temperature on protease activity

The influence of the pH on caseinolytic activity was determined by using the azocasein protocol described above, except for the use of different reaction buffers (benzoic acid, pyridine, phosphate buffer, Tris-HCl, $Na_2HPO_4.2H_2O$, methylamine-HCl) to obtain values from pH 4 to pH 11. All the buffers were pH-adjusted for the incubation temperature and all of the assays were performed at 55°C. To determine the influence of temperature on the caseinolytic activity, samples were incubated for 10 min at temperatures of 2, 20, 37, 40, 50, 60, 70, 80, 90, and 94°C, respectively, in 10 mM Tris-HCl buffer (pH 7.5).

3.2.8. Electrophoresis and zymograms

The method used for SDS-PAGE analysis was essentially the method described by Laemmli, (1970). For zymogram analysis, 2% (w/v) caseinate co-polymerised with the gels was used. Samples were loaded into the gel without prior heating, and electrophoresis was performed at 4°C at 25 V/gel. Following electrophoresis, gels were washed successively, first with 2.5% (v/v) TritonX-100 in water and then with 50 mM Tris-HCl (pH 8) buffer containing 2.5% (v/v) TritonX-100, each for 10 min at room temperature. The gels were equilibrated for 10 min in 50 mM Tris-HCl buffer (pH 7.5) and then incubated for 5 min at 55°C. Finally, gels were stained with 2% Coomassie brilliant blue and destained in deionized water, to reveal zones of substrate hydrolysis. For electrophoretic analysis of low-molecular-weight proteins, a Tricine gel was used according to the method described by Judd (1996).

3.2.9. Nucleotide sequence accession numbers

The 16S rRNA sequences obtained in this study have been deposited in GenBank under accession no. AF267508 and AF267509 for isolates PA-5 and PA-9, respectively.

3.3. RESULTS AND DISCUSSION

Microbial studies of hot spring communities may provide an unique and wide-ranging source of novel microorganisms, many of which may show significant variance from known terrestrial phylotypes. These organisms, each containing a catalogue of enzymes and other bioproducts, could provide a highly valuable resource for biotechnological developments and applications. In this study, protease-secreting thermophilic bacteria were isolated from Buranga hot springs in western Uganda. The temperature and pH of the hot springs ranged between 45°C and 101°C and from pH 6.8 to 8.5, respectively. Two isolates that demonstrated high protease activity on milk and casein agar plates, designated as PA-5 and PA-9, respectively, were selected and characterised.

3.3.1. Morphology

Both isolates PA-9 and PA-5 are aerobic, sporulating, Gram-positive, rod-shaped bacteria. The cell morphology of isolate PA-9 differed from that of PA-5 grown under similar conditions (Fig. 3.1). The cells of isolate PA-9 were long straight rods (12 µm in length and 3 µm in width), whereas cells of isolate PA-5 were slightly shorter (10 µm in length by 3 µm in width). There was a general tendency to increase in size with time. After 24 h of stationary broth culture, isolate PA-9 grew as single cells and as filament-like chains at least 53 µm long. In contrast, isolate PA-5 grew as single cells, and short chains (19 µm long) were observed only after 48 h of incubation.

3.3.2. Effect of temperature and pH on growth

The growth rates of isolates PA-9 and PA-5 were determined within the temperature range of 36 to 72°C. For isolate PA-5, the optimum growth rate was at 60°C (Fig. 3.2a), while isolate PA-9 had the highest maximum growth rate at 62°C (Fig. 3.2b). At the temperatures for maximum growth rates, the generation time was approximately 35 and 21 min for isolate PA-5 and PA-9, respectively. Isolate PA-5 was unable to grow at 72°C or below 38°C, and isolate PA-9 was unable to grow above 68°C or below 36°C. Regarding pH growth limitation, isolates PA-9 and PA-5 were grown in modified Castenholz medium that had a range of pH values between 4 and 12. A pH range around neutrality (7.5 - 8.5) favoured optimal growth of the two isolates. The isolates did not grow at a pH lower than pH 4. However, reduced growth at pH 11 and pH 12 was noted for isolates PA-5 and PA-9, respectively.

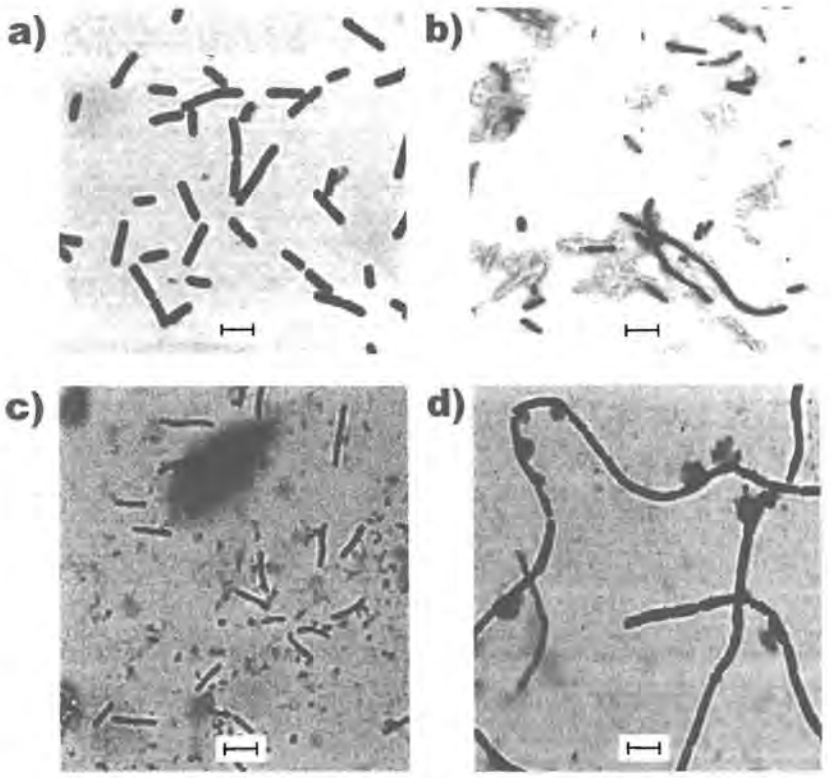


Fig. 3.1 Micrographs showing typical cell morphologies of isolates PA-5 after 10 h (a) and 48 h (b) of incubation, and of isolate PA-9 after 8 h (c) and 75 h (d) of incubation. The bar corresponds to 2 µm and applies to all micrographs.

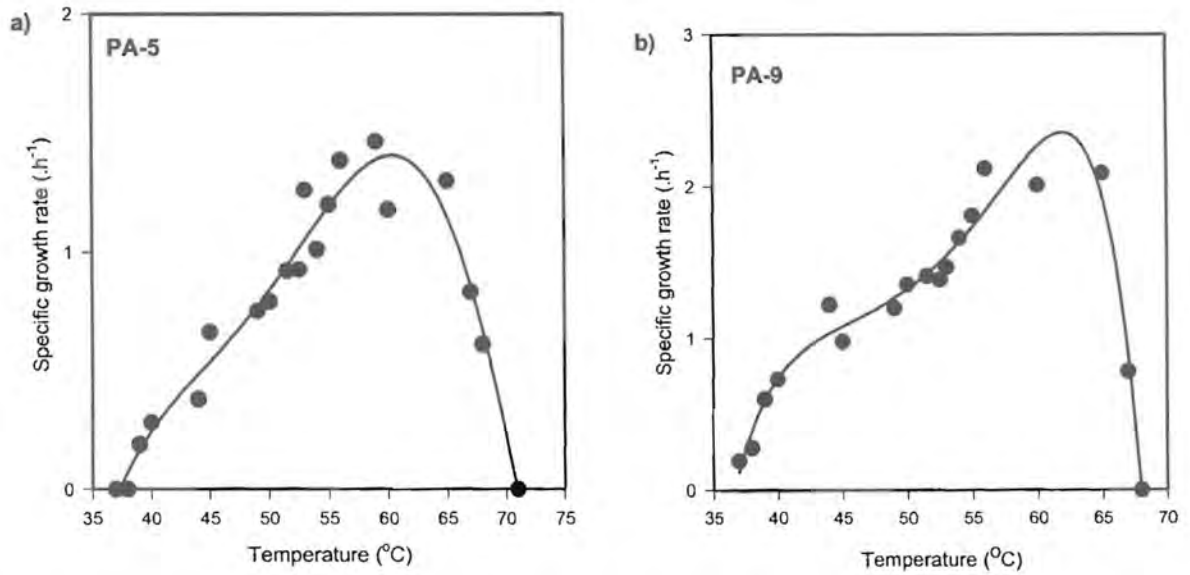


Fig. 3.2 Influence of temperature on maximal specific growth rates of isolates PA-5 (a) and PA-9 (b) cultivated in modified Castenholz medium D. All values are the average of two separate experiments.

3.3.3. Phylogenetic analysis

Based on their morphologies, PA-5 and PA-9 were tentatively identified as *Bacillus* species. To confirm their identity, PCR amplification and sequencing of the 16S rRNA gene of these isolates allowed us to determine approximately 96% of the complete sequence. A phylogenetic tree (Fig. 3.3) was constructed by the neighbor-joining method and indicated that both isolates PA-5 and PA-9 were part of the cluster within the thermophilic group of the genus *Bacillus*. All of these species, like the presently obtained isolates, grow at high temperatures and neutral pH (Rainey *et al.*, 1994). Among the described species, the closest relatives of isolate PA-5 were *B. thermoleovorans* B23 and DSM5366, while those for isolate PA-9 were *Geobacillus uzenensis* strain X and *B. kaustophilus* NCIMB8547. It has recently been proposed that the thermophilic *Bacillus* growing at neutral pH be transferred to a newly proposed genus, *Geobacillus*, with *Geobacillus stearothermophilus* (formally *B. stearothermophilus*) as the type species (Nazina *et al.*, 2001). The 16S rDNA sequence analysis indicated that isolates PA-5 and PA-9 were phylogenetically related to a low degree (96%) when compared with the relatedness between other known thermophilic *Bacilli*. A more definitive classification in the future, however, should take additional characteristics, especially DNA-DNA hybridisation data, into account (Stackebrandt and Goebel, 1994).

3.3.4. Caseinolytic characteristics of the different *Bacillus* isolates

Isolates PA-5 and PA-9 were cultured at 55°C in broth containing either glycerol or glucose as the sole carbon source. Analysis of extracellular protease extracts by plate cup assays at various time intervals indicated that maximal levels of extracellular enzymatic activity could be obtained after 3 days for isolate PA-9, and after 6 days for isolate PA-5. In the case of isolate PA-9, similar levels of protease production was observed when either glucose or glycerol was added to the medium. In contrast, for isolate PA-5, production of proteases was significantly favoured by glycerol in the medium (data not shown). Thus, glucose and glycerol was used as the sole carbon source for isolate PA-9 and PA-5, respectively, in all subsequent experiments.

To investigate the activity of the extracellular proteases, their ability to hydrolyze azocasein at pH 4 to 11 was investigated spectrophotometrically. For isolate PA-9, high levels of protease activity were detected from pH 5 to 9, and with two optima, one at pH 6.5 and a

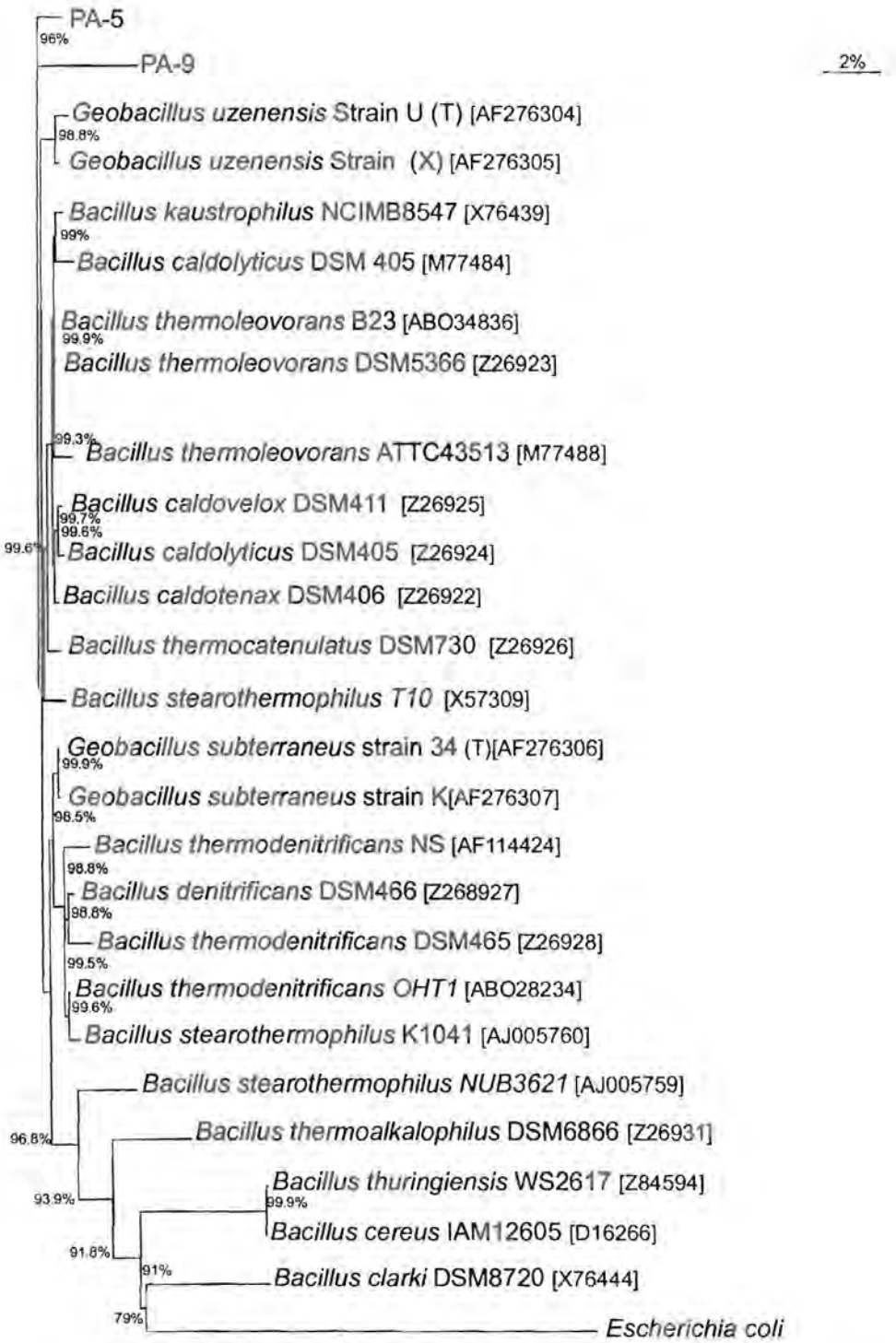


Fig. 3.3 Neighbor-joining tree showing the phylogenetic affiliation of isolates PA-9 and PA-5 to a selected number of members of *Bacillus* spp., of which the GenBank Accession no. are indicated in brackets. Bootstrap values greater than 50% are indicated. The scale bar represents 2% of the estimated substitutions per nucleotide position.

second at pH 8.0 (Fig. 3.4a). The protease extract of isolate PA-5 displayed optimal activity between pH 5 and 10, albeit low when compared to that of isolate PA-9.

The optimum temperature for caseinolytic activity of the crude extracellular enzyme extracts prepared from isolates PA-5 and PA-9 was investigated at various temperatures from 2 to 94°C at pH 7.5. For isolate PA-9, the optimum reaction temperature of the enzyme solution was 70°C (Fig. 3.4b). The enzymes remained active over a wide range of temperatures varying from 37 to 80°C. In the case of isolate PA-5 (Fig. 3.4b), the optimal temperature of the enzyme solution was between 50 and 70°C. It is noteworthy that the extracellular enzyme preparation of PA-9 displayed significant activity at 2°C. Furthermore, the enzyme preparations prepared from both isolates were active at 94°C.

3.3.5. Zymogram analysis of secreted proteases

The extracellular enzyme solutions prepared from isolates PA-5 and PA-9 had different caseinolytic activities as measured by the azocasein method. Zymogram analysis was therefore carried out to compare the different proteases secreted by the respective isolates. Like many reported *Bacillus* species (Priest, 1977), isolates PA-5 and PA-9 also produced multiple extracellular proteases.

The zymogram obtained for isolate PA-5 (Fig. 3.5), indicated the presence of two bands with protease activity. A larger zone of hydrolysis was observed for the 25-kDa band compared to the 24-kDa band. For isolate PA-9, eight proteins with protease activity could be identified in the extracellular enzyme extract and were found to have molecular masses of approximately 97, 72, 50, 27, 22, 17 and 12 kDa (resolved as two bands by Tricine gel electrophoresis) (Fig. 3.6b). Whether each of the observed proteases in fact represent distinct protease species or whether some of these may represent unprocessed forms of other proteases must await primary amino acid sequence comparisons. Following incubation at 55°C for 10 min, the two protein bands of approximately 97 and 72 kDa, respectively, displayed larger zones of hydrolysis and thus appear to be heat-activated proteases (Fig. 3.6a, lane 2). The 12-kDa protease appeared to be responsible for the observed protease activity at low temperature (2°C) (Fig. 6a, lane 1). The use of a Tricine gel revealed that the small molecular weight protease band of 12 kDa could be resolved as two separate bands (Fig. 3.6b), each displaying caseinolytic activity.

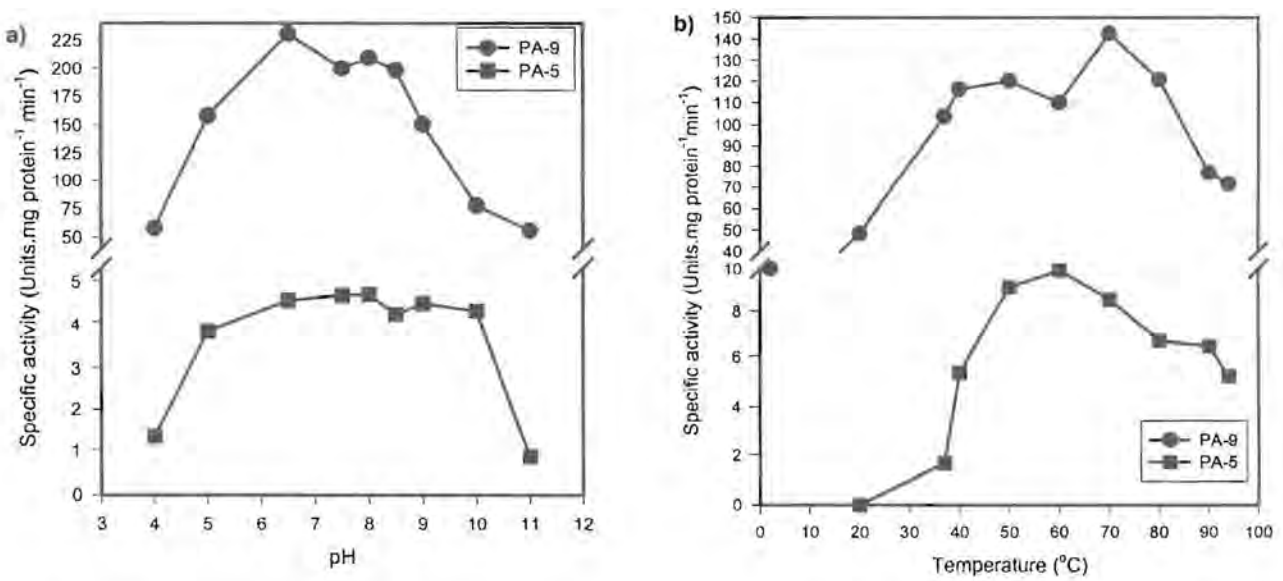


Fig. 3.4 Influence of (a) pH and (b) temperature on the activity of crude extracellular enzyme preparations prepared from the culture supernatants of isolates PA-5 and PA-9. For the pH profile, activity was measured at 55°C in 10 mM Tris buffer adjusted to the correct pH. For the temperature profile, activity was measured in 10 mM Tris buffer (pH 7.5) at different temperatures. Values are the means of results of duplicate experiments.

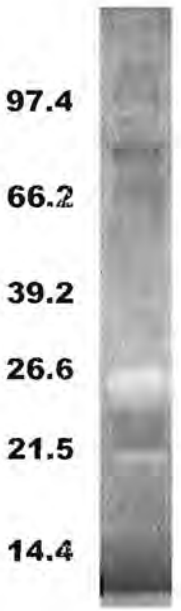


Fig. 3.5 Analysis of the presence of caseinolytic proteases in cell-free supernatants of isolate PA-5. Aliquots of culture supernatants from cells grown for 6 days were used to analyze the presence of caseinolytic proteases by zymograms (lane 1). The sizes of the molecular mass standards (in kilodaltons) are indicated to the left.

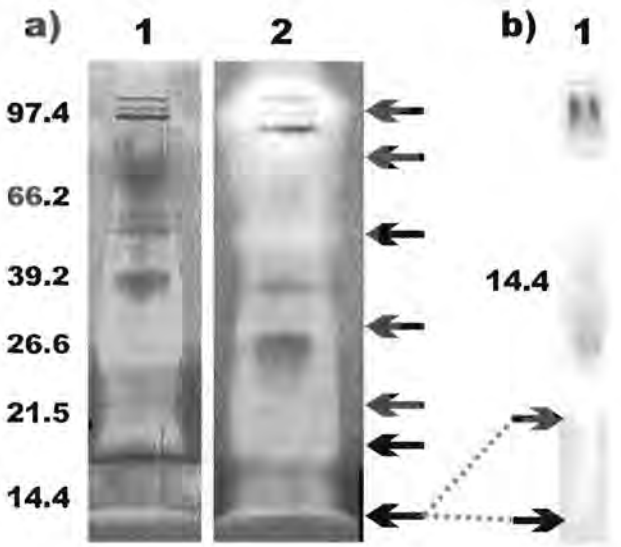


Fig. 3.6 Analysis of the presence of caseinolytic proteases in cell-free supernatants of isolate PA-9. (a) Aliquots of culture supernatants from cells grown for 3 days were used to analyze the presence of caseinolytic proteases by zymograms without (lane 1) and with incubation at 55°C for 5 min (lane 2). (b) Zymogram of a Tricine gel indicating two low-molecular-weight proteases. The positions of the proteases are indicated by arrows, and the sizes of the molecular mass standards (in kilodaltons) are indicated to the left.

The observation that proteases produced by isolate PA-9 displaying activity at 2°C may be of significance. The genes of cold-active proteases have been cloned from two psychrotrophic *Bacillus* strains, strains TA39 (Narinx *et al.*, 1997) and TA41 (Davail *et al.*, 1992). Both of these proteases are members of the subtilisin subfamily. Such enzymes have also been isolated from Gram-negative psychrotrophs such as *Shewanella* strain Ac10, as well as a few mesophiles (Kulakova *et al.*, 1999). The two small proteases produced by the thermophilic *Bacillus* PA-9 therefore constitute a novel finding with potential for the development of recombinant cold-active proteases. There is a growing biotechnological interest in cold-adapted enzymes as they are particularly suited to low temperature biotransformations where substrates and/or co-factors are thermolabile, where low temperature operation is considered as a means of reducing microbial growth, or simply to save on heating expenses (Margesin and Schinner, 1994).

3.4. REFERENCES

Adams, M. and Kelly, R. (1998). Finding and using hyperthermophilic enzymes. *Trends in Biotechnology* 16: 329-332.

Bradford, M. (1976). A rapid and sensitive method for quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72: 248-254.

Castenholz, D.W. (1969). Thermophilic blue-green algae and the thermal environment. *Bacteriological Reviews* 33: 476-504.

Cowan, D.A. (1992). Enzymes from thermophilic archaebacteria: Current and future applications in biotechnology. *Biochemistry Society Symposium* 58: 149-169.

Davail, S., Feller, G., Narinx, E. and Gerday, C. (1992). Sequence of the subtilisin-encoding gene from an antarctic psychrotroph *Bacillus* TA41. *Gene* 119: 1430-1440.

Davis, M. (1998). Making a living at the extremes. *Trends in Biotechnology* 16: 102-104.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783-791.

Jaenicke, R., Schurig, H., Beaucamp, N. and Ostendorp, R. (1996). Structure and stability of hyperstable proteins: Glycolytic enzymes from hyperthermophilic bacterium *Thermotoga maritima*. *Advances in Protein Chemistry* 48: 181-269.

Judd, R.C. (1996). SDS-polyacrylamide gel electrophoresis of peptides. In: *The Protein Protocols Handbook*. Walker, J.M. (Ed.). Humana Press, Totowa, New Jersey, USA pp. 101-107.

Kulakova, L., Galkin, A., Kurihara, T., Yoshimura, T. and Esaki, N. (1999). Cold-active serine alkaline protease from the psychrotrophic bacterium *Shewanella* strain Ac10: Gene cloning and enzyme purification and characterization. *Applied and Environmental Microbiology* 65: 611-617.

Laemmli, U.K. (1970). Cleavage associated with the maturation of the head of bacteriophage T4. *Nature* 227: 680-685.

Mala, B.R., Aparn, M.T., Mohini, S.G. and Vasanti, V.D. (1998). Molecular and biotechnological aspects of microbial proteases. *Microbiology and Molecular Biology Reviews* 62: 597-635.

Margesin, R. and Schinner, F. (1994). Properties of cold-adapted microorganisms and their potential role in biotechnology. *Journal of Biotechnology* 33: 1-4.

Marmur, J. and Doty, P. (1962). Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *Journal of Molecular Biology* 5: 109-118.

Narinx, E., Baise, E. and Gerday, C. (1997). Subtilisin from psychrophilic antarctic bacteria: Characterization and site-directed mutagenesis of residues possibly involved in the adaptation to cold. *Protein Engineering* 10: 1271-1279.

Nazina, N.T., Tourova, P.T., Poltaraus, B.A., Novikova, V.E., Grigoryan, A.A., Ivanova, A.E., Lysenko, M.A., Petrunyaka, V.V., Ospov, A.G., Belyaev, S.S. and Ivanov, V.M. (2001). Taxonomic study of aerobic thermophilic bacilli: Descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustrophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustrophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. *International Journal of Systematic and Evolutionary Microbiology* 51: 433-446.

Priest, F.G. (1977). Extracellular enzyme synthesis in the genus *Bacillus*. *Bacteriological Reviews* 41: 711-753.

Prieur, D. (1992). Physiology and biotechnological potential of deep-sea bacteria. In: *Molecular Biology and Biotechnology of Extremophiles*. Herbert, R.A. and Sharp, R. (Eds.). Blackie, London, England, pp. 163-202.

Rainey, F.A., Fritze, D. and Stackebrandt, E. (1994). The phylogenetic diversity of thermophilic members of the genus *Bacillus* as revealed by 16S rDNA analysis. *FEMS Microbiology Letters* 115: 205-212.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425.

Smibert, R.M. and Krieg, N.R. (1994). Phenotypic characterization. In: *Methods for General and Molecular Bacteriology*. Gerhardt, P., Murray, R.G.E., Wood, W.A. and Krieg, N.R. (Eds.). American Society for Microbiology Press, Washington DC, USA, pp. 607-654.

Stackebrandt, E. and Goebel, B.M. (1994). Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology* 44: 846-849.

Weisburg, W.G., Barns, S.M., Pelletier, D.A. and Lane, D.J. (1991). 16S Ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology* 173: 697-703.

CHAPTER FOUR

ISOLATION OF PROTEASE-ENCODING GENE(S) OF *Geobacillus* sp. PA-9

ABSTRACT

In an attempt to isolate and characterise protease-encoding genes from *Geobacillus* sp. PA-9, proteolytic enzymes were purified from the cell-free culture supernatant by gel-filtration and ion-exchange chromatography or, alternatively, directly from negative-stained SDS-polyacrylamide gels. Five protease-active proteins, as verified by zymogram analysis, were selected for N-terminal amino acid sequence determination. However, an unambiguous sequence of the respective proteins could not be obtained due to the presence of contaminating proteins. The results obtained from the protein purification procedures nevertheless suggested that protease activity of the enzyme solution might be due to more than one enzyme, one of which was a large multimeric protein. Two other strategies were subsequently used as means to isolate the protease-encoding genes. The first approach involved the construction of a genomic DNA library followed by functional screening of the clones in *Bacillus megaterium* and in *Escherichia coli* on casein agar plates and in broth. The second approach relied on the use of oligonucleotide primers, which had been designed following multiple sequence alignment of various alkaline protease genes. DNA fragments resulting from the use of the newly designed primer sets in polymerase chain reactions (PCR) with genomic DNA from *Geobacillus* sp. PA-9 were either labeled and used as probes to screen the genomic DNA library or the amplicons were cloned. Nucleotide sequencing and sequence analysis of the inserts in hybridisation-positive clones and of the cloned amplicons indicated that these approaches were unsuccessful in isolating protease-encoding genes. These results suggest that the protease(s) from PA-9 do not share a significant level of homology to other proteases identified thus far.

4.1. INTRODUCTION

Although some enzymes from mesophilic bacteria are known to withstand temperatures approaching 100°C, e.g. α-amylase from *B. licheniformis*, such enzymes are rare (Uhlig, 1998; Adams and Kelly, 1998). The instability and lability of the majority of mesophilic enzymes at elevated temperatures limit their application in a number of industrial processes. In contrast, thermophilic bacteria have been regarded as a promising source of thermostable enzymes and the enzymes are also often resistant to and active in the presence of organic solvents and detergents (Jaenicke *et al.*, 1996). Consequently, such enzymes have been used in the detergent, leather tanning and food processing industries, and in the management of waste through bioconversion of proteinaceous waste into biomass (Kobayashi *et al.*, 1995; 1996; Hameed *et al.*, 1996; Anwar and Saleemuddin, 1998).

In the past two decades, many thermostable enzymes from different thermophiles, amongst them, the extracellular proteases from *Geobacillus* spp., *Alicyclobacillus* spp. and *Thermobacillus* spp., have been characterised and their structural genes identified, cloned and overexpressed in appropriate host organisms (Nishiya and Imanaka, 1990; Van der Laan *et al.*, 1991; Nazina *et al.*, 2001). These studies commonly rely on the isolation and purification of the enzyme to enable analytical studies regarding its molecular properties (molecular weight, number of subunits, isoelectric point), catalytic properties (Km, optimum temperature and pH, effect of inhibitors) and catalytic mechanisms. Consequently, a number of separation and purification methods have been developed whereby enzymes may be purified. In addition to different electrophoresis techniques (Poulsen *et al.*, 1989; Davril *et al.*, 1993; Králová, 1999), the most widely used methods include gel-filtration, size-exclusion, ion-exchange and affinity chromatography and/or a combination of these (Martín-Hernández *et al.*, 1994; Kobayashi *et al.*, 1995; Ferrero *et al.*, 1996; Oyama *et al.*, 1997; Bressollier *et al.*, 1999). Problems are, however, often encountered in purifying an enzyme of interest to homogeneity in instances where the microorganisms produce multiple enzymes (Dalbøge, 1997) or minute quantities of the desired protein(s) (Lila *et al.*, 1996), as well as in instances where the protein is strongly complexed with other proteins and cannot be readily dissociated (Kuln and Wagner, 1989). To overcome these problems, various approaches have been reported whereby genes encoding for the proteins of interest can be isolated. The approaches rely on the functional screening of constructed genomic DNA libraries (Nishiya and Imanaka, 1990), PCR amplification of the genes of interest using primers, which have been designed

based on the nucleotide sequences of previously characterized gene sequences (Sung *et al.*, 1991; Vecerek and Venema, 2000), and the construction of genomic DNA libraries in expression-cloning vectors (Dalbøge, 1997).

Two thermophilic protease-producing *Geobacillus* strains, PA-5 and PA-9, isolated from Buranga hot springs in western Uganda, were characterised (Chapter 3). While strain PA-5 produces at least two proteases, strain PA-9 was found to produce at least eight extracellular proteases. Notably, caseinolytic activity displayed by proteases from strain PA-9 was observed over a wide spectrum of temperatures (2-90°C). The possibility of discovering a novel enzyme(s) with a broad temperature range therefore warranted further investigation of strain PA-9. In this part of the study, different approaches that were used to isolate the protease-encoding gene(s) from *Geobacillus* sp. PA-9 are described.

4.2. MATERIALS AND METHODS

4.2.1. Bacterial strains, plasmids and culture conditions

The plasmids and bacterial strains used in this part of the study are listed in Table 4.1. *Escherichia coli* DH5α was used as the host in the cloning procedures and was grown at 37°C in Luria-Bertani (LB) liquid medium (0.5% [w/v] yeast extract, 1% [w/v] tryptone, 1% [w/v] NaCl; pH 7.4) or on LB agar plates. LB agar plates supplemented with ampicillin (100 μg/ml) were used to screen for recombinants. Plasmids pGEM®-T Easy (Promega, Madison, WI, USA) and pSVBI (obtained from F. Meinhardt and K-D. Wittchen, Westfälische Wilhelms-Universität, Germany) were used as cloning vectors. *Geobacillus* sp. PA-9 was grown in modified Castenholtz medium as described previously (Section 3.3.2). *Bacillus megaterium* MS941was routinely cultured in LB broth at 37°C. For preparation of protoplasts, *B. megaterium* was grown in Hyp liquid medium, which contained 2.05 ml of 0.5 M KH_2PO_4, 3 ml of 10% (w/v) yeast extract, 10 ml of 20% (w/v) glucose, 0.1 ml of 2 mg/ml $MnSO_4$ and Prot medium to a final volume of 1 litre. Prot medium (pH 7.5) contained (g/l) NH_4Cl, 1; Tris-HCl, 12; KCl, 0.035; NaCl, 0.058; $MgSO_4.7H_2O$, 0.246; $MgCl_2.5H_2O$, 4.26; and 136 ml of 50% (w/v) sucrose. For protoplast regeneration, Hyp-Tc agar plates were used consisting of Hyp agar medium supplemented with 12.5 μg/ml tetracycline, and Hyp-top agar, which consisted of Hyp medium containing 0.8% (w/v) agar (Meinhardt *et al.*, 1989).

Table 4.1: Bacterial strains, plasmids and primers used

Strains, plasmids and primers	Relevant properties	Source
Bacterial strains:		
Geobacillus strain PA-9		Hawumba *et al.* (2002)
Escherichia coli DH5α	*hsdR recA lac*ZYA ϕ 80d*lac*ZΔ*M15*	Stratagene
Bacillus megaterium MS941		F. Meinhardt and K-D. Wittchen[a]
Plasmids:		
pGEM®-T Easy	Cloning vector for PCR products, ColE1, Amp^r, LacZα peptide	Promega
pSVBI	*E. coli*/*B. megaterium* shuttle vector, Tet^r, Amp^r	F. Meinhardt and K-D. Wittchen[a]
Primers:		
FRD1	5'-TGTCAACAGTGGCATTGC-3'	This study
FRD2	5'-TCCTTGTGGAAGCAGCCT-3'	This study
RD1	5'-ACGCTGGCGAATTTGGTT-3'	This study
RD2	5'-GCAGCAGAAAGAAGGAAGCA-3'	This study
F-pSVBI	5'-GGGATCAACTTTGGGAGAG-3'	This study
R-pSVBI	5'-GGCGATTAAGTTGGGTAACG-3'	This study
M13 Forward	5'-GTAAAACGACGGCCAGTG-3'	This study
M13 Reverse	5'-CAGGAAACAGCTATGAC-3'	This study

[a] F. Meinhardt and K-D. Wittchen, Westfälische Wilhelms-Universität Münster, Institut für Mikrobiologie, Münster, Germany

4.2.2. Preparation of crude extracellular enzyme extracts

Isolate PA-9 was grown in liquid medium consisting of 0.5% (w/v) meat extract and soybean-casein digest, and 1% (w/v) glucose in the mineral base of modified Castenholtz medium (Section 3.2.5). For preparation of extracellular enzyme extracts to be used in enzyme purification, 200-ml cultures were grown at 55°C for 72 h in 1-L Erlenmeyer flasks, without shaking. Following incubation, bacterial cells were removed from the culture supernatant by centrifugation at 7 000 rpm for 10 min, and the cell-free culture supernatants were pooled (2 L) and concentrated by lyophilisation. The pellet was finally dissolved in 100 ml of 10 mM Tris-HCl (pH 8).

4.2.3. Purification of protease enzymes

The crude enzyme solution (20 ml) was loaded onto a Sephadex G-100 gel-filtration column (Pharmacia, Sweden; 1.2 × 33 cm), equilibrated with 10 mM Tris-HCl buffer (pH 8). The

flow rate was maintained at 1 ml/min and 3-ml fractions were collected every 3 min using a Model 2112 Redirac fraction collector (LKB, Sweden). The absorbance of each fraction was determined at 280 nm using a U-2000 spectrophotometer (Hitachi, Japan). Fractions constituting an elution peak were pooled (Appendix 1) and the proteolytic activity of the combined fraction was determined by cup-plate assays as previously described (Section 3.2.6). The active enzyme fractions were applied to a DEAE-Sephacel ion-exchange column (Pharmacia, Sweden; 1.2 × 15 cm). The adsorbed enzyme was eluted with a step-wise gradient of 0 - 0.5 M NaCl in 50 mM Tris-HCl (pH 8) at a flow rate of 0.15 ml/min and fractions of 1.2 ml were collected. The protease-containing fractions, analysed as described above, were applied to a DEAE-Sepharose CL-6B ion-exchange column (Sigma, St. Louis, MO, USA; 1.2 × 15 cm), equilibrated in 10 mM Tris-HCl (pH 8). Elution was done with the same buffer as above using a flow rate of 2 ml/min and fractions of 4 ml each were collected. Protease-active fractions, as confirmed by cup-plate assays, were analysed by denaturing and non-denaturing PAGE gels, as well as by zymogram analyses (Section 4.2.5). For determination of molecular mass by electrospray mass spectrometry, a fraction containing an apparently pure enzyme was kindly analysed by Dr M.J. van der Merwe, Department of Chemistry, University of Stellenbosch, South Africa (see legend of Fig. 4.2 for details).

4.2.4. Purification of proteases from SDS-PAGE gels

The crude enzyme solution was separated on a 10% SDS-PAGE gel and the protein bands were visualised by negative staining using the immidazole-$ZnSO_4$ method (Fernandez-Patron *et al.*, 1995). Briefly, after electrophoresis, gels were rinsed in distilled water for 30 s and then soaked in 100 ml of a 0.2 M imidazole, 0.1% (w/v) SDS solution for 10 min. The solution was discarded and the gels immersed in 100 ml of 0.2 M $ZnSO_4$ until the gel background became deep white, resulting in transparent protein bands. The reaction was stopped by rinsing the gels in distilled water after which the protein bands of interest were carefully excised from the gel using a sharp scalpel blade. The proteins were eluted from the gel slices according to the method of Lila *et al.* (1996). Briefly, the excised protein bands were placed in microfuge tubes containing 1 ml 2% (w/v) citric acid. After incubation for 10 min, the protein bands were crushed by forcing the gel pieces through the opening of a 1-ml syringe into a second microfuge tube. Following addition of 1-1.5 volumes of a 25 mM Tris (pH 8.3), 500 mM glycine solution, the slurry was vortexed for 10 min. The gel pieces were pelleted by centrifugation at 15 000 rpm for 5 min and the supernatant transferred to a clean

microfuge tube. The pelleted gel pieces were subjected to a second round of extraction using identical procedures to those described above. The supernatants were pooled, dialysed against distilled water, and the dialysate was concentrated by lyophilization and finally resuspended in 200 µl of UHQ H_2O. The protein recovery and activity were subsequently determined by SDS-PAGE and zymogram analysis, respectively.

4.2.5. SDS-PAGE electrophoresis and zymogram analysis

The method used for SDS-PAGE analysis was essentially the method described by Laemmli (1970). For zymogram analysis, 2% (w/v) caseinate co-polymerized with the gels was used. Samples were loaded into the gel without prior heating, and electrophoresis was performed at room temperature at 35 V/gel for the first 30 min and then at 40 V/gel for 4 h. Electrophoresis of the crude enzyme solution, which displayed proteolytic activity at 2°C, was performed at 25 V/gel at 4°C. Following electrophoresis, gels were washed successively, first with 2.5% (v/v) TritonX-100 in water and then with 50 mM Tris-HCl (pH 7.5) buffer containing 2.5% (v/v) TritonX-100, each for 10 min at room temperature. The gels were equilibrated for 10 min in 50 mM Tris-HCl buffer (pH 7.5) and then incubated for 5-10 min at 55°C. The gels were finally stained with 2% Coomassie brilliant blue for 20 min and destained in deionized water, to reveal zones of substrate hydrolysis.

4.2.6. Protein blotting and N-terminal amino acid sequence determination

Following electrophoresis, gels were fixed for 1 h in a solution containing 40% (v/v) methanol and 10% (v/v) acetic acid. The gel and an Immobilon-P polyvinylidene difluoride (PVDF) membrane (Millipore, Bedford, MA, USA), cut to the size of the gel, were soaked in the transfer buffer (25 mM Tris, 192 mM glycine or 20 mM sodium phosphate buffer) for 30 min after which the proteins were transferred to the membrane at 120 mA for 2 h at 4°C, using a Trans-blot cell (BioRad, Hercules, CA, USA). For blotting of low-molecular-weight proteins, electrophoresis was performed using a Tricine gel (Judd, 1996) and methanol (10% [v/v]) was added to the transfer buffer. The protein(s) was visualised on the membrane by staining in 0.03% (w/v) Coomassie brilliant blue prepared in 40% methanol, and then dried between filter papers. The protein bands that corresponded to the protein-active bands on the zymogram were cut from the membrane. The N-terminal sequences were determined by Pascal Cosette, University of Rouen, France, on a Procise 492 automatic sequencer (Applied Biosystems, Courtabeuf, France).

4.2.7. Construction of a *Geobacillus* sp. PA-9 genomic DNA library in *Escherichia coli*

Genomic DNA of *Geobacillus* strain PA-9 was isolated as previously described (Section 3.2.4). The purified genomic DNA was digested to completion with *Hind*III (Roche Diagnostics, Germany) at 37°C overnight. Following electrophoresis on a 1% (w/v) agarose gel, restriction fragments ranging between 1 to 8 kb were purified from the gel by the freeze-squeeze method (Benson, 1984). Plasmid pSVBI was linearized with *Hind*III, dephosphorylated, purified and ligated to the restriction fragments for 16 h at 16°C in the presence of 1 unit T4 DNA ligase and 1 × ligase buffer (Roche, Germany). Competent *E. coli* DH5α cells were prepared and transformed by the method of Chung and Miller (1993), and transformants were screened on LB agar plates supplemented with ampicillin (100 µg/ml), X-Gal (40 µg/ml) and IPTG (20 µg/ml). The recombinant transformants from each plate were inoculated into 100 ml LB broth, and cultured at 37°C overnight after which glycerol stocks were prepared. The stocks were stored at -70°C and when needed were diluted 10-fold in LB-broth and spread onto LB agar plates supplemented with 100 µg/ml of ampicillin.

4.2.8. Preparation of recombinant plasmids for transformation of *B. megaterium*

Transformants on transformation plates obtained above were divided into four quadrants. All colonies in a quadrant were inoculated into 100 ml LB broth supplemented with ampicillin (100 µg/ml) and cultured at 37°C overnight. Plasmid DNA was isolated by the alkaline-lysis method (Sambrook *et al.*, 1989). To remove any trace of SDS, plasmid DNA preparations were dialysed extensively against deionised water using a Slide-a-Lyzer mini-dialysis unit with a 10-kDa cut-off (Pierce, Rockford, IL USA).

4.2.9. Preparation of *B. megaterium* MS941protoplasts

Protoplasts were prepared according to the method of Meinhardt *et al.* (1989). A starter culture was prepared by inoculating a single colony of *B. megaterium* MS941 into 10 ml LB broth and incubating at 37°C overnight with shaking (220 rpm). An aliquot (1.5 ml) of starter culture was then inoculated into 50 ml of Hyp medium and cultured at 37°C until an OD_{546} of 0.9, whereafter the culture was incubated on ice for 30 min to stop further growth. After incubation, the cells were harvested at 5 000 rpm for 7 min at 4°C and the cell pellet washed once in 5 ml Hyp medium. The pellet was suspended in 4 ml Hyp medium and 1 ml of

lysozyme (0.5 mg/ml) was added. The suspension was incubated at 37°C until the protoplast to cell ratio reached at least 60:40, which typically occurred after 2-3 h of incubation. The protoplast suspension was then incubated on ice for 10 min. The protoplasts were collected by centrifugation at 4 000 rpm for 4 min at 4°C, washed three times in 5 ml Hyp medium each, and then gently resuspended in 2 ml of Hyp medium.

4.2.10. Transformation of *B. megaterium* MS941 protoplasts

The prepared protoplasts were transformed using a PEG method as described by Meinhardt *et al.* (1989). Briefly, 0.5 ml of the protoplast suspension was mixed with 4-8 µl of recombinant plasmid DNA (0.5-1 µg). After addition of 1.5 ml of 40% (w/v) PEG-6000, the mixture was incubated at room temperature for 4 min. Following incubation, 5 ml of Hyp medium was added and the protoplasts collected by centrifugation at 4 000 rpm for 4 min at 4°C. The collected protoplasts were suspended in 1 ml Hyp medium and incubated at 37°C for 2 h in a shaking incubator (100 rpm), to allow regeneration of the protoplasts. Aliquots (100-200 µl) were then added to 5 ml Hyp-top agar, mixed and spread evenly over the surface Hyp agar plates supplemented with tetracycline (12.5 µg/ml). The agar plates were incubated at 37°C for 24 to 48 h.

4.2.11. Screening for protease-positive recombinants in *B. megaterium* MS941

Transformed *B. megaterium* cells were screened for protease activity after streaking the colonies onto LB-casein (0.4% [w/v]) agar plates followed by incubation at 37°C for at least 48 h. Protease-positive transformants were identified by halo(s) around the colonies.

4.2.12. Screening for protease-positive recombinants in *E. coli* DH5α

Screening of the library constructed in *E. coli* DH5α was performed in liquid medium and on agar plates. In broth, 250 µl of LB broth supplemented with 0.25% (w/v) casein was pipetted into the wells of microtitre plates. Each well was inoculated with a recombinant colony, and the microtitre plates were cultured at 37°C in a shaking incubator for 48 h. Aliquots (100 µl) of each culture were transferred into new wells and stored under sterile conditions. The remaining cultures (150 µl) were lysed by 2 pulses of 15 s each using an Ultrasonic Homogeniser (Cole-Parmer Instruments Co., Chicago, Ill, USA), at an output of 60%. The lysate was heat activated at 70°C for 0.5-1 h and subsequently spotted onto 1% (w/v) casein agar plates. The plates were incubated at 55°C overnight and observed for proteolytic

activity. Crude enzyme solution was spotted onto the casein agar plates and served as a positive control. Screening for protease activity on 1% (w/v) casein agar plates was performed by streaking recombinant colonies and incubating the agar plates at 37°C. The agar plates were checked daily for transformants displaying proteolytic activity, as indicated by the presence of a hydrolysis halo surrounding the colony.

4.2.13. Isolation of protease genes by PCR amplification

Sequences of alkaline serine proteases were retrieved from the protein and nucleotide GenBank databases by means of the Entrez server available at the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) by using keywords "protease, serine protease, alkaline protease and serine alkaline protease". Multiple sequence alignments were performed using ClustalW (Thompson *et al.*, 1994). The alkaline serine proteases could be divided into two groups, with members in each group displaying high levels of homology (see Appendix II). Based on the obtained sequence alignments, two sets of oligonucleotide primers were subsequently designed, namely FRD1/FRD2 and RD1/RD2 (Table 4.1).

The newly designed primer sets were used in PCR reactions containing 50 ng of *Geobacillus* sp. PA-9 genomic DNA as template, 50 pmol of each the sense and antisense primer, 1 × PCR buffer, 1.5 mM $MgCl_2$, 0.25 mM of each deoxynucleoside triphosphate (dNTP) and 1 U of *Taq* DNA polymerase (Southern Cross Biotechnology, South Africa). The reaction tubes were placed in a Perkin-Elmer GeneAmp 2400 thermal cycler. Following initial denaturation of 3 min at 96°C, the reactions were subjected to 30 cycles of denaturation for 30 s at 94°C, annealing for 30 s at 52°C or 53°C and elongation for 45 s or 1 min at 72°C. After the last cycle, the reactions were kept at 72°C for 7 min to complete synthesis of all strands. For control purposes, reaction mixtures containing UHQ water and all other reagents but no template DNA were included. Aliquots of the PCR reaction mixtures were subsequently analyzed by agarose gel electrophoresis on 1% (w/v) agarose gels in the presence of an appropriate DNA molecular weight marker (phage λ DNA digested with both *EcoR*I and *Hind*III). The amplicons were purified from the agarose gels by the freeze-squeeze method (Benson, 1984), and then labeled with DIG-dUTP using the DIG-High Prime kit (Roche, Germany), according to the manufacturer's instructions, and used as probes to screen the genomic library. Alternatively, amplicons were purified from the gel using a Wizard SV Gel

and PCR Clean-Up system (Promega, Madison, WI, USA) and then cloned into the pGEM®-T Easy vector system (Promega).

4.2.14. Sequencing of cloned genes

The nucleotide sequence of the cloned amplicons was determined in a Hitachi 3100 capillary array automated DNA sequencer using an ABI PRISM™ BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer Applied Biosystems, Foster City, CA, USA). The amplified DNA fragments were sequenced in both directions using 12.5 pmol of either the universal pUC/M13 forward and reverse sequencing primers, or primers F-pSVBI and R-pSVBI (Table 4.1). The sequence data were edited using the Sequence Analysis 3.1 and Sequencing Navigator 1.0.1 programmes supplied in the ABI PRISM™ software package (Perkin-Elmer Applied Biosystems, Foster City, CA, USA). Homology searches were conducted using the BLAST programme (Altschul *et al.*, 1997) available at the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/BLAST/).

4.3. RESULTS AND DISCUSSION

4.3.1. Enzyme purification

Zymogram analysis of the crude enzyme solution prepared from the culture supernatant of *Geobacillus* sp. PA-9 had revealed that some of the low-molecular-mass proteins (< 14 kDa) had activity at low temperature (2°C), and may possibly reach maximum activity at 50°C (Chapter 3). Proteases that display such a broad range of temperature activity may be of potential biotechnological interest, as they have potential for the development of recombinant cold-active and heat-active proteases. Thus, to isolate the protease-encoding genes, an approach was followed whereby degenerate oligonucleotide probes, for use in screening of a genomic DNA library of isolate PA-9, could be designed based on the N-terminal amino acid sequence of the purified proteases (Kaneko *et al.*, 1989; Van Den Burg *et al.*, 1991; Takami *et al.*, 1992a; 1992b; Saul *et al.*, 1996).

To purify the proteolytic enzymes, a crude enzyme solution was prepared from the cell-free culture supernatant of *Geobacillus* sp. PA-9 and then subjected to gel-filtration followed by ion-exchange chromatography, as described in Materials and Methods (Section 4.2.4). Following ion-exchange chromatography, three protein fractions were obtained that displayed

caseinolytic activity. Whereas one protease-active fraction eluted at 0.2 M NaCl, designated as $IE2_{II}$, two protease-active fractions were eluted at 0.3 M NaCl, designated as IE3II and IEIII-3, respectively. Zymogram analysis of the respective protein fractions revealed that a caseinolytic protein of *ca.* 66 kDa was present in all three of the fractions. Furthermore, both fractions $IE3_{II}$ and IEIII-3 contained a protease-active band of identical size, but an additional, unique protease-active band could be observed in fraction $IE3_{II}$ (Fig. 4.1a). Thus, except for fraction $IE2_{II}$, which appeared to contain a single purified polypeptide, as verified by electrophoresis of a sample on a nondenaturing polyacrylamide gel (Fig. 4.1b), both of the protein fractions eluted at 0.3 M NaCl contained a mixture of caseinolytic enzymes. To more accurately determine the molecular mass of the protein contained in fraction $IE2_{II}$, the sample was subsequently analysed by electrospray mass spectroscopy. The result (Fig. 4.2) indicated that the monomeric protein has a molecular mass 22.86 kDa, and can form a dimer (45.42 kDa) and trimer (68.12 kDa), which could also be seen in SDS-polyacrylamide gels electrophoresed under denaturing conditions. However, the presence of an additional contaminant protein of 53.32 kDa was noted, which was not detected in the denaturing and nondenaturing polyacrylamide gels. The low-molecular-mass protease-active proteins that had previously been observed in zymograms of crude enzyme extracts (Fig. 4.1, lane 1) could, however, not be detected in any of the fractions displaying caseinolytic activity. The failure to purify these low-molecular-mass proteases suggested that they could either have been produced in minute quantities or have been diluted from the samples during the initial gel-filtration chromatography step.

4.3.2. Purification of proteins from negative-stained SDS-PAGE gels

Since it was not possible to purify the proteases by chromatography, an alternative strategy was adopted whereby individual proteins displaying proteolytic activity were eluted from the gel pieces excised from negative-stained SDS-PAGE gels. Although zymogram analyses had previously revealed that at least eight caseinolytic enzymes were produced by *Geobacillus* sp. PA-9, it was not possible to observe all of the corresponding protein bands on a duplicate SDS-PAGE gel. This suggested either that the concentration of some proteins were low or that the proteins stained poorly with Coomassie brilliant blue and could thus not be detected. Negative staining has been reported to be a very sensitive staining method, and is capable of

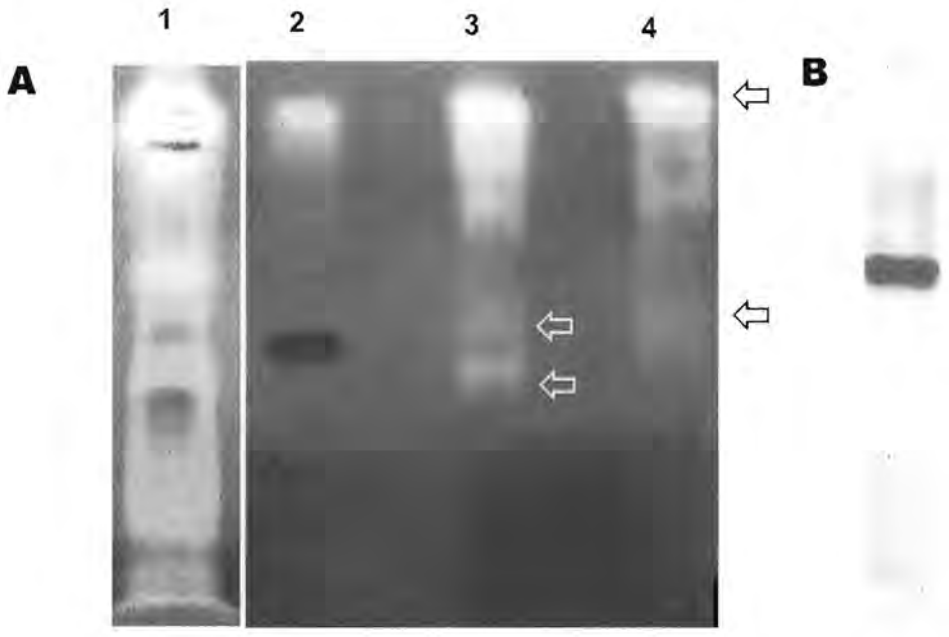


Fig. 4.1 Zymogram analysis of proteases purified from the culture supernatant of *Geobacillus* sp. PA-9 by ion-exchange chromatography. Lane 1, cell-free culture supernatant from a *Geobacillus* sp. PA-9 culture; lane 2, sample of fraction IE2II; lane 3, sample of fraction IE3II; lane 4, sample of fraction IEIII-3. Caseinolytic proteases present in the respective fractions are indicated by arrows. (B) Analysis of fraction IE2II by nondenaturing polyacrylamide gel electrophoresis.

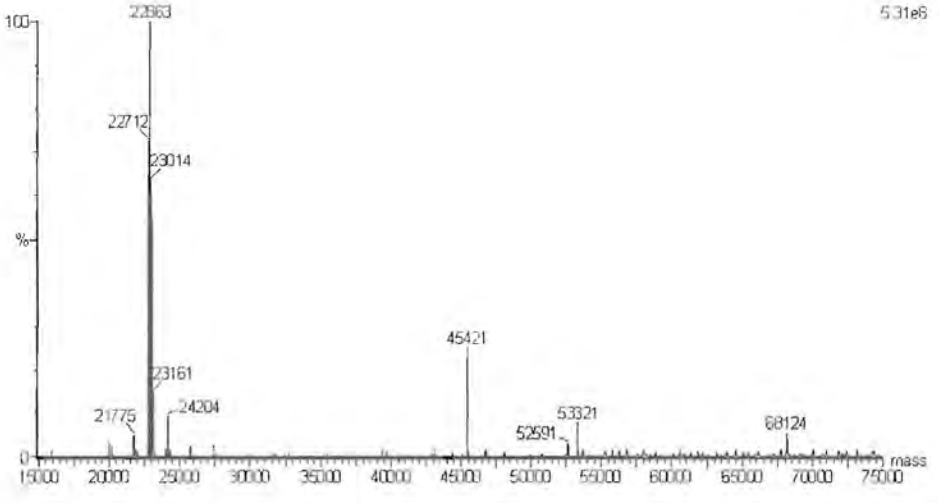


Fig. 4.2 Analytical spectrum of fraction $IE2_{II}$ following analysis by electrospray mass spectrophotometry. The protein sample was diluted 1:1 in 50% acetonitrile/water containing formic acid and 5 µl of the diluted protein sample was injected into the mass spectrometer.

detecting proteins in quantities of 1 to 10 ng per band, which is normally undetectable by conventional staining procedures (Fernandez-Patron *et al.*, 1995; Lila *et al.*, 1996).

Using this approach and the procedures outlined in Materials and Methods (Section 4.2.6), five proteins were purified from several negative-stained SDS-PAGE gels and concentrated to obtain sufficient amounts of the respective proteins for further characterisation. Zymography of the respective proteins was performed to confirm their casein-degrading activity (Fig. 4.3). Despite having been resolved and excised as single bands from the SDS-PAGE gels, all five the protein samples showed heterogeneity, consisting of more than one protein. Not surprisingly, an ambiguous amino acid sequence could not be determined from these protein samples. The caseinolytic activity of the purified enzymes was observed as smears at the top of the gel (Fig. 4.3, lanes 1-4) and as distinct bands (Fig. 4.3, lane 5). A low-molecular-mass protein of *ca.* 14-18 kDa, which was eluted from near the bromophenol front, yielded five protein bands with caseinolytic activity (Fig. 4.3, lane 5). The sizes of these bands were in agreement with those of proteins that could be observed in the crude enzyme solution. This suggested that the large caseinolytic protein of *ca.* 66 kDa was probably a multimer comprising separate monomers that could associate in more than one caseinolytically active combination. Alternatively, the results may indicate self-autolysis of the major enzyme thereby yielding differently sized truncated forms of the protein of which some may exhibit activity.

4.3.3. Construction of a *Geobacillus* sp. PA-9 genomic DNA library and functional screening of the library in *B. megaterium* MS941 and *E. coli* DH5α

As an alternative to the strategies based on enzyme purification towards isolating the protease-encoding genes of *Geobacillus* sp. PA-9, a genomic DNA library of isolate PA-9 was constructed in *E. coli* using the *E. coli*/*B. megaterium* shuttle vector pSVBI as cloning vector. The genomic DNA library was subsequently screened on casein-containing media for clones displaying proteolytic activity in both *B. megaterium* and *E. coli* as hosts.

Although transformation of *B. megaterium* protoplasts was not very efficient (an estimated transformation efficiency of 20% was obtained), screening in a *Bacillus* host has the advantage of the recombinant protein being potentially secreted to the extracellular milieu, thereby greatly facilitating identification of protease-active clones and simplifying

subsequent purification of the extracellular enzyme. However, the *Bacillus megaterium* strain MS941 used in this study possesses a residual 14% protease activity, which is due to the presence of a minor protease (Wittichen and Meinhardt, 1995), which manifested as a background activity that was difficult to differentiate from recombinant protease-active clones. Thus, all of the 408 transformants, representing less that 10% of the total *Geobacillus* sp. PA-9 genome, screened in *B. megaterium* as host were regarded as negative for heterologous protease production, as they did not display significantly higher levels of proteolytic activity than the nonrecombinant host strain on casein agar plates. In addition, subcultures of transformants grew poorly or not at all and the use of *B. megaterium* as a screening host was therefore suspended.

Due to the problems encountered in using *B. megaterium* MS941 as host for screening of the genomic DNA library, further screening for protease-encoding genes of *Geobacillus* PA-9 was performed in *E. coli* DH5α using LB broth containing casein and on casein agar plates. A total of 10 000 transformants were screened, comprising 4 383 that were screened in broth and 5 617 that were screened on agar medium. None of the transformants exhibited protease activity. Screening on agar medium, however, yielded eight recombinant colonies that were initially thought to be protease-positive, but upon closer investigation they showed activity against 4-hydroxyphenylacetate (Chapter 5).

Taking into account that the *Bacillus* spp. genome is *ca.* 4.2 Mb in size and that the library was comprised of fragments with an average length of 4 kb, it was calculated, using the formula of Clarke and Carbon (Old and Primrose, 1994), that between 3 000 and 5 000 independent recombinants would be required to achieve a probability of 95 and 99%, respectively, of finding a recombinant clone displaying protease activity. This implied that even if the genomic DNA library was only 50% representative, then 6 000 to 10 000 recombinant clones would have to be screened to isolate a protease-positive recombinant clone. For an organism producing at least eight different protease enzymes and presumably encoded by different genes, it was expected that at least one protease gene would be obtained upon screening at most 1 000 recombinant clones. No protease gene(s) was, however, obtained, suggesting an absence of a restriction fragment with a complete protease gene, i.e. with functional promoter and signal sequences. The limit digest used during construction the library may have resulted in a slightly less than random library and could also have resulted in the internal cleavage of the full-length protease gene(s), thereby affecting the outcome.

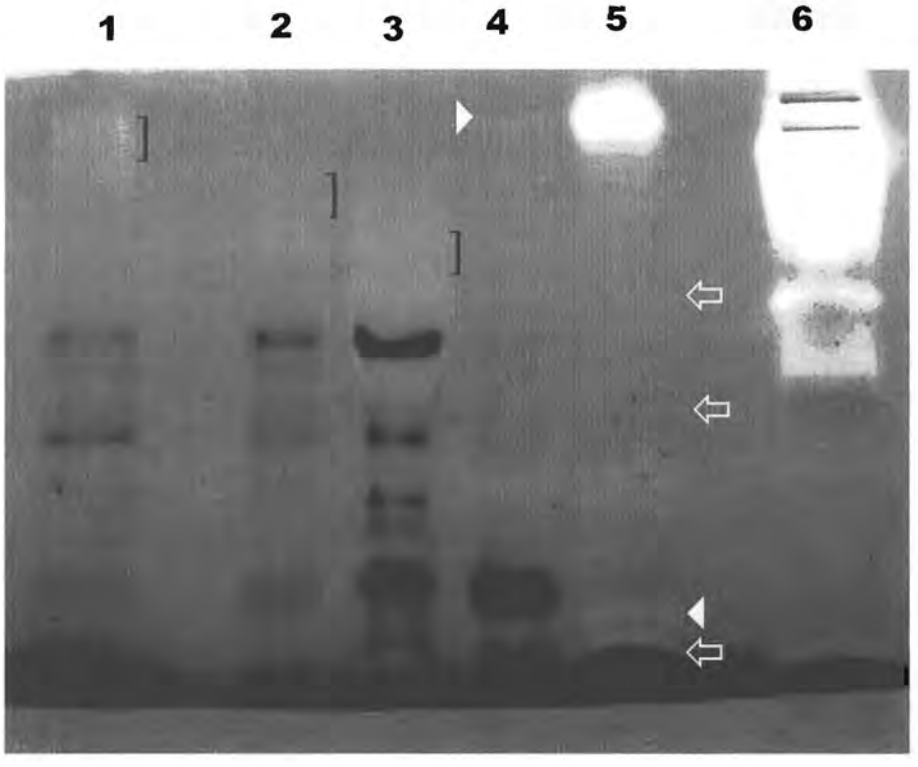


Fig. 4.3 Zymogram analysis of proteins purified from negative-stained SDS-PAGE gels. Five different protein bands (lanes 1 to 5), corresponding to caseinolytic proteases observed in a crude enzyme solution (lane 6) prepared from the supernatant of a *Geobacillus* sp. PA-9 culture, were eluted as single bands from a SDS-PAGE gel after negative staining. Arrows or brackets indicate positions on the gels at which caseinolytic activity was observed.

The importance of promoter and signal sequences in the efficient processing and export of extracellular and cell wall-bound proteins in prokaryotes is well documented (Finkelstein *et al.*, 1983; Ikemura *et al.*, 1987; Zhu *et al.*, 1989; Peek *et al.*, 1993), and it has been reported that a signal peptide is required in the efficient processing of heterologous extracellular proteins and enzymes when using *E. coli* as host (Peek *et al.*, 1993). Other approaches, which may in future studies be used to overcome the obstacles, include the construction and screening of a library harbouring large gene fragments (8-23-kb fragments) in cosmids (Van Den Burg *et al.*, 1991), use of oligonucleotide primers that had been designed based on previously characterised protease genes to screen library by hybridisation or to PCR-amplify the desired cognate genes (O'Donohue *et al.*, 1994; Vecerek and Venema, 2000), and the use of an expression-cloning system consisting of an appropriate promoter, pre-prosequence, prosequence, signal sequence, transcriptional start site and a termination signal (Dalbøge, 1997).

4.3.4. Screening of the *Geobacillus* sp. PA-9 genomic DNA library using PCR-generated probes

To design oligonucleotide primers, which could be used to PCR-amplify the alkaline serine protease-encoding genes from the genomic DNA of *Geobacillus* sp. PA-9, multiple sequence alignment of the amino acid sequences of several alkaline serine proteases was performed. The results indicated that the proteases could be grouped into two distinct groups; one group represented by *B. stearothermophilus* (GenBank Accession no. AY028615) and the second group by *Bacillus* sp. (GenBank Accession no. D-13158) (see Appendix II). Two sets of primers, FRD1/FRD2 and RD1/RD2 (Table 4.1), were subsequently designed based on the nucleotide sequences of the pre-prosequence regions of the respective proteases. These primer sets were then used in PCRs together with genomic DNA of *Geobacillus* sp. PA-9 at an annealing temperature of 53°C for 1 min. Amplification using the FRD1 and FRD2 primers yielded a single amplicon of 1.2 kb, while primers RD1 and RD2 generated a single amplicon of 1.5 kb. The respective sizes of the two amplicons were in agreement with the expected size range, i.e. 0.8 to 2.0 kb. The respective amplicons were presumed to represent alkaline protease genes and were subsequently labeled with DIG-dUTP and used as probes to screen the genomic DNA library. The labeled probes, designated II_2 (obtained from primer set FRD1/FRD2) and II_3 (obtained from primer set RD1/RD2), respectively, were hybridised at 42°C to colony blots prepared from the transformation plates.

Under these conditions, probe II_2 exhibited non-specific hybridisation and was excluded from further use. Of the 1 244 clones screened, fifteen hybridised to probe II_3. The 15 hybridisation-positive clones were subcultured, plasmid DNA isolated and characterised by digestion with *Hind*III to confirm the presence of a cloned DNA insert and to estimate their relative sizes (result not shown). Of the 15 clones, 7 did not contain a cloned DNA insert or the insert DNA might have been too small to be resolved on the 1% (w/v) agarose gel (clones 5, 8 through 12 and 14). Clones 1, 2, 6, 7 and 13 each contained a DNA fragment of *ca.* 5.0 kb, while clones 3, 4, and 15 contained fragments of 1.9, 6.0 and 5.7 kb, respectively. The nucleotide sequence of the DNA inserts in clones 1, 3, 7, 13 and 15 was determined in pSVBI using the F-pSVBI and R-pSVBI sequencing primers (see Appendix III). All the clones contained DNA fragments, which displayed homology to enzymes of the glycolytic pathway. These included phosphoglycerate mutase (clones 1, 3, 7 and 15) and enolase (clone 13). The hybridisation to glycolytic genes suggested that the probe had a high specificity to genes in this pathway. Alternatively, it suggested that the proteases of PA-9 might not share high homology with the known protease genes. Screening using the above probes was therefore suspended.

4.3.5. Identification of protease-encoding genes by PCR

Since annealing of the two primer sets to template DNA at 53°C for 1 min resulted in the amplification of a single amplicon each, the PCR reaction conditions were relaxed to allow for a greater degree of non-specific annealing of the primers to template DNA. Consequently, the annealing conditions were adjusted to 52°C for 45 s. Using these conditions, several amplicons of different sizes were generated. Amplicons obtained using primers FRD1 and FRD2 were designated $9I_1$, $9I_2$, and $9I_3$, while those obtained by making use of primers RD1 and RD2 were designated $9II_1$ through $9II_9$ (Fig. 4.4). Six of the amplicons ($9I_1$, $9II_1$, $9II_2$, $9II_3$, $9IL_4$ and $9II_5$) were purified and cloned into the pGEM®-T Easy vector and the nucleotide sequence of each cloned insert DNA was determined (Appendix IV). Their identity was determined by searching the GenBank database for homologues using BLAST-N (Table 4.2). The PA-9 amplicons shared homologies with different proteins, which included: sigma factor/hypothetical regulatory protein ($9II_2$) involved in the transcription of proteins, alanyl-tRNA synthetase ($9II_3$) involved in translation of proteins, hypothetical protein ($9IL_4$) of unknown function, and ABC Fe^{3+} transporter/BP-2 periplasm binding protein ($9II_5$) involved in protein transport / export.

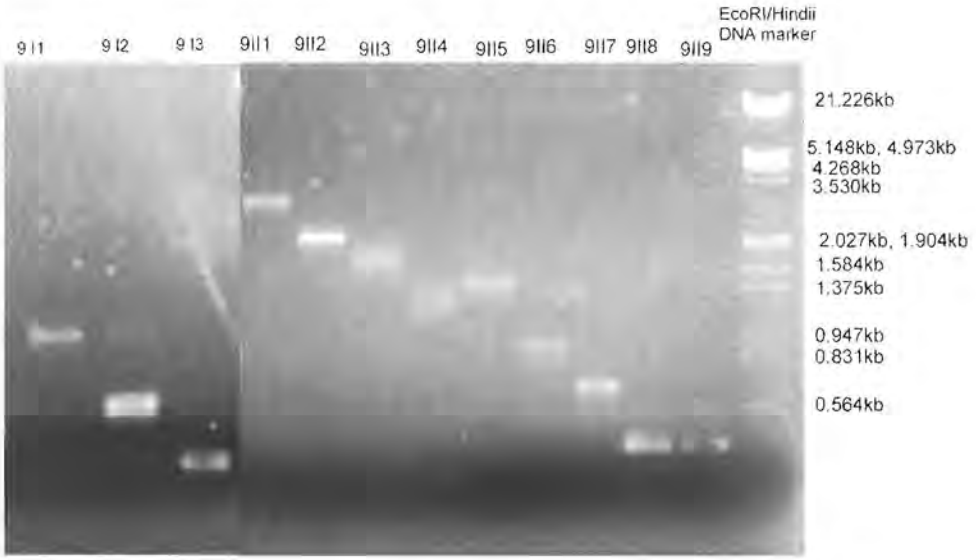


Fig. 4.4 Agarose gel electrophoretic analysis of purified amplicons. Genomic DNA of *Geobacillus* sp. PA-9 was used as template DNA in PCR reactions with either primers FRD1 and FRD2 to generate amplicons $9I_1$ to $9I_3$, or with primers RD1 and RD2 to generate amplicons $9II_1$ through $9II_9$. The sizes of the DNA molecular weight marker, phage λ DNA digested with both *Eco*RI and *Hind*III, are indicated to the right of the figure.

Table 4.2: Summary of the amplicons sequenced, their homologous proteins and putative function of the homologous proteins

Amplicon	Length of sequence (nt)	BLAST-N homologue	Predicted function
$9I_1$	679	Hypothetical protein	Unknown
$9II_1$	601	5-methyltetrahydrofolate 5-homocystein synthetase	C_1 carrier in metabolism
$9II_2$	677	Sigma factor / hypothetical regulatory protein	RNA polymerase binding
$9II_3$	808	Alanyl-tRNA synthetase	Translation
$9II_4$	1 223	Hypothetical protein	Unknown
$9II_5$	1 355	ABC Fe^{3+} transporter / BP-2 periplasm binding protein	Membrane transport

From the foregoing, it could be concluded that the oligonucleotide primers were not specific for the protease-encoding genes of *Geobacillus* sp. PA-9, as none of the PCR-amplified genomic DNA fragments displayed homology to known protease gene(s). These results suggested either that the protease(s) from PA-9are unique and share no or very low levels of homology to any protease identified so far, or that the regions from which the primers were designed are conserved in a number of non-protease genes. The most likely explanation for the results obtained, however, is that the primers were not specific for alkaline serine protease genes only. Using BLAST-N, the primers displayed homology to non-protease genes, though to a lower level compared to alkaline protease-encoding genes.

Additionally, the result of the partially purified enzyme and the accompanying zymogram and SDS-PAGE analyses (Figs. 4.1, 4.2 and 4.3) appeared to point to the production of, among others, a multimeric protease, capable of dissociating into either smaller subunits or truncated forms of a large protein, some of which were active on their own (Fig. 4.3). Multimeric enzymes have been described among the metalloenzymes (Dos Santos *et al.*, 2000), thermostable β-mannanase and α-galactosidase from *Bacillus stearothermophilus* (Talbot and Sygusch, 1990), and L-proline dehydrogenase from *Thermococcus profundus* (Sakuraba *et al.*, 2001). However, whether separate subunits possess activity has not been highlighted in literature, but oligomerisation has been observed among thermophilic enzymes as a means to enhance thermostability (Kumar *et al.*, 2000). Only once the enzyme(s) has been purified to homogeneity and gene(s) encoding for the protease have been isolated and sequenced, will the true nature of this protease be deciphered.

4.4. REFERENCES

Adams, M. and Kelly, R. (1998). Finding and using hyperthermophilic enzymes. *Trends in Biotechnology* 16: 329-332.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25: 3389-402.

Anwar, A. and Saleemuddin, M. (1998). Alkaline proteases: A Review. *Bioresource Technology* 64: 175-183.

Benson, S.A. (1984). Rapid procedure of isolation of DNA fragments from agarose gel. *BioTechniques* 2: 66-67.

Bressollier, P., Letourneau, F., Urdaci, M. and Verneuil, B. (1999). Purification and characterisation of a keratinolytic serine proteinase from *Streptomyces albidoflavus*. *Applied and Environmental Microbiology* 65: 2570-2576.

Chung, T.C. and Miller, H.R. (1993). Preparation and storage of competent *Escherichia coli* cells. *Methods in Enzymology* 218: 621-627.

Dalbøge, H. (1997). Expression cloning of fungal enzyme genes: A novel approach for efficient isolation of enzyme genes of industrial relevance. *FEMS Microbiology Reviews* 21: 29-42.

Davril, M., Ducourouble, M. and Van-Seuningen, I. (1993). Western blotting of basic proteins after nondenaturing electrophoresis in acid conditions using PhastSystem. *Electrophoresis* 14: 876-880.

Dos Santos, G.W., Pacheco, I., Liu, M., Teixeira, M., Xavier, V.A. and LeGall, J. (2000). Purification and characterisation of an iron superoxide dismutase and a catalase from the sulphate-reducing bacterium *Desulfovibrio gigas*. *Journal of Bacteriology* 182: 796-804.

Fernandez-Patron, C., Hardy, E., Sosa, A., Seoane, J. and Lila, C. (1995). Double staining of Coomassie blue-stained polyacrylamide gels by imidazole-sodium dodecyl sulfate-zinc reverse staining: Sensitive detection of Coomassie blue-undetected proteins. *Analytical Biochemistry* 224: 263-269.

Ferrero, M.A., Castro, G.R., Abate, M.C., Baigori, D.M. and Siñeriz, F. (1996). Thermostable alkaline proteases of *Bacillus licheniformis* MIR 29: Isolation, production and characterisation. *Applied Microbiology and Biotechnology* 45: 327-332.

Finkelstein, V.A., Bendzko, P. and Rapoport, A.T. (1983). Recognition of signal sequences. *FEBS* 161: 176-179.

Hameed, A., Natt, M.A. and Evans, C.S. (1996). Production of alkaline protease by a new *Bacillus subtilis* isolate for use as a bating enzyme in leather treatment. *World Journal of Microbiology and Biotechnology* 12: 289-291.

Hawumba, F.J., Theron, J. and Brözel, S.V. (2002). Thermophilic protease-producing *Geobacillus* from Buranga hot springs in western Uganda. *Current Microbiology* 45: 144-150.

Ikemura, H., Takagi, H. and Inouye, M. (1987). Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli*. *The Journal of Biological Chemistry* 262: 7859-7864.

Jaenicke, R., Schurig, H., Beaucamp, N. and Ostendorp, R. (1996). Structure and stability of hyperstable proteins: Glycolytic enzymes from the hyperthermophilic bacterium *Thermotoga maritima*. *Advances in Protein Chemistry* 48: 181-269.

Judd, R.C. (1996). SDS-polyacrylamide gel electrophoresis of peptides. In: *The Protein Protocols Handbook*. Walker, J.M. (Ed.). Humana Press, Totowa, New Jersey, USA, pp. 101-107.

Kaneko, R., Koyama, N., Tsai, Y., Juang, R., Yoda, K. and Yamasaki, M. (1989). Molecular cloning of the structural gene for alkaline elastase YaB, a new subtilisin produced by an alkalophilic *Bacillus* strain. *Journal of Bacteriology* 171: 5232-5236.

Kobayashi, T., Hakamada, Y., Achida, S., Hitomi, J., Yoshimatsu, T., Koike, K., Kawai, S. and Ito, S. (1995). Purification and properties of an alkaline protease from *Bacillus* sp. KSM-K16. *Applied Microbiology and Biotechnology* 43: 473-481.

Kobayashi, T., Hakamada, Y., Hitomi, J., Koike, K. and Ito, S. (1996). Purification of alkaline proteases from a *Bacillus* strain and their possible interrelationship. *Applied Microbiology and Biotechnology* 45: 63-71.

Králová, B. (1999). Electrophoretic methods for isolation and characterisation of enzymes. *Analytica Chimica Acta* 383: 109-117.

Kuln, R. and Wagner, H. (1989). Free-flow electrophoresis as a method for the purification of enzymes from *E. coli* cell extract. *Electrophoresis* 10: 165-172.

Kumar, S., Tsai, C-J. and Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein Engineering* 13: 179-191.

Laemmli, U.K. (1970). Cleavage associated with the maturation of the head of bacteriophage T4. *Nature* 227: 680-685.

Lila, R., Castellanos-Serra, G., Fernandez-Patron, C., Hardy, E. and Huerta, V. (1996). A procedure for protein elution from reverse-stained gels applicable at the low picomole level: An alternative route to the preparation of low abundance proteins for microanalysis. *Electrophoresis* 17: 1561-1572.

Martín-Hernández, C.M., Alting, C.A. and Exterkate, A.F. (1994). Purification and characterisation of the mature, membrane-associated cell-envelope proteinase of *Lactobacillus helveticus* L89. *Applied Microbiology and Biotechnology* 40: 828-834.

Meinhardt, F., Stahl, U. and Ebeling, W. (1989). Highly efficient expression of homologous and heterologous genes in *Bacillus megaterium*. *Applied Microbiology and Biotechnology* 30: 343-350.

Nishiya, Y. and Imanaka, T. (1990). Cloning and nucleotide sequences of the *Bacillus stearothermophilus* neutral protease gene and its transcriptional activator gene. *Journal of Bacteriology* 172: 4861-4869.

Nazina, N.T., Tourova, P.T., Poltaraus, B.A., Novikova, V.E., Grigoryan, A.A., Ivanova, A.E., Lysenko, M.A., Petrunyaka, V.V., Ospov, A.G., Belyaev, S.S. and Ivanov, V.M. (2001). Taxonomic study of aerobic thermophilic bacilli: Descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustrophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustrophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. *International Journal of Systematic and Evolutionary Microbiology* 51: 433-446

O'Donohue, J.M., Roques, P.B. and Beaumont, A. (1994). Cloning and expression in *Bacillus subtilis* of the *npr* gene from *Bacillus thermoproteolyticus* Rokko coding for the thermostable metalloprotease thermolysin. *Biochemistry Journal* 300: 599-603.

Old, R.W. and Primrose, S.B. (1994). Principles of Gene Manipulation: An Introduction to Genetic Engineering (5th Edn.), Blackwell Scientific Publications, Osney Mead, Oxford, U.K, pp. 100-120.

Oyama, H., Kinjoh, M., Watari, M. and Murao, S. (1997). Purification and characterisation of an alkaline proteinase produced by *Pimelobacter* sp. Z-483. *Journal of Fermentation and Bioengineering* 84: 351-353.

Peek, K., Veitch, P.D., Prescott, M., Daniel, M.R., MacIver, B. and Bergquist, L.P. (1993). Some characteristics of a proteinase from a thermophilic *Bacillus* sp. expressed in *Escherichia coli*: Comparison with the native enzyme and its processing in *E. coli* and *in vitro*. *Applied and Environmental Microbiology* 59: 1168-1175.

Poulsen, M.O., Jacobsen, T., Hau, J., Jensen, B., Vodder, L. and Andersen, F. (1989). Enzyme activity electrophoresis: Development and applications. *Electrophoresis* 10: 857-864.

Sambrook, J., Fritsch, E.F. and Maniates, T. (1989). Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sakuraba, H., Takamatsu, Y., Satomura, T., Kawakami, R. and Ohshima, T. (2001). Purification, characterisation, and application of a novel dye-linked L-proline dehydrogenase from a hyperthermophilic Archaeon, *Thermococcus profundus*. *Applied and Environmental Microbiology* 67: 1470-1475.

Saul, J.D., Williams, C.L., Toogood, S.H., Daniel, M.R. and Bergquist, L.P. (1996). Sequence of the gene encoding a highly thermostable neutral proteinase from *Bacillus* sp. strain EA1: Expression in *Escherichia coli* and characterisation. *Biochimica et Biophysica Acta* 1038: 74-80.

Sung, M., Tanizawa, K., Tanaka, H., Kuramitsu, S., Kagamiyama, H., Hirotsu, K., Okamoto, A., Higuchi, T. and Soda, K. (1991). Thermostable aspartate aminotransferase from a thermophilic *Bacillus* species. *Journal of Biological Chemistry* 266: 2567-2572.

Takami, H., Akiba, T. and Horikoshi, K. (1992a). Cloning, expression, and characterisation of a minor alkaline protease from *Bacillus* sp. no. AH-101. *Bioscience Biotechnology and Biochemistry* 56: 510-511.

Takami, H., Kobayashi, T., Aono, R. and Horikoshi, K. (1992b). Molecular cloning, nucleotide sequence and expression of the structural gene for a thermostable alkaline protease from *Bacillus* sp. no. AH-101. *Applied Microbiology and Biotechnology* 38: 101-108.

Talbot, C. and Sygusch, J. (1990). Purification and characterisation of thermostable β-mannanase and α-galactosidase from *Bacillus stearothermophilus*. *Applied and Environmental Microbiology* 56: 3505-3510.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673-4680.

Uhlig, H. (1998). Industrial Enzymes and their Applications. John Wiley and Sons, New York, USA, pp. 37-72.

Van Den Burg, B., Enequist, G.H., Van Der Haar, M.E., Eijsink, G.H.V., Stulp, B.K. and Venema, G. (1991). A highly thermostable neutral protease from *Bacillus caldolyticus*: Cloning and expression of the gene in *Bacillus subtilis* and characterisation of the gene product. *Journal of Bacteriology* 173: 4107-4115.

Van der Laan, J.C., Gerritse, G., Mulleners, L.J.S.M., Van Der Hoek, R.A.C. and Quax, W.J. (1991). Cloning, characterisation, and multiple chromosomal integration of a *Bacillus* alkaline protease gene. *Applied and Environmental Microbiology* 57: 909-915.

Vecerek, B. and Venema, G. (2000). Expression of the neutral protease gene from a thermophilic *Bacillus* sp. BT1 strain in *Bacillus subtilis* and its natural host: Identification of a functional promoter. *Journal of Bacteriology* 182: 4101-4107.

Wittichen, K.D. and Meinhardt, F. (1995). Inactivation of the major extracellular protease from *Bacillus megaterium* DSM319 by gene replacement. *Applied Microbiology and Biotechnology* 42: 871-877.

Zhu, X., Ohta, Y., Jordan, F. and Inouye, M. (1989). Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* 339: 483-484.

CHAPTER FIVE*

NUCLEOTIDE SEQUENCE AND BIOCHEMICAL CHARACTERISATION OF A 4-HYDROXYPHENYLACETIC ACID 3-HYDROXYLASE FROM THERMOPHILIC *Geobacillus* sp. PA-9

ABSTRACT

A recombinant clone encoding the hydroxylase involved in the *meta*-cleavage pathway of 4-hydroxylphenylacetic acid (4-HPA) was isolated from a genomic DNA library of *Geobacillus* sp. PA-9. The 2.7-kb DNA fragment was completely sequenced and three open reading frames (ORFs), designated as *pheH*, *pheH2* and *pheC*, could be identified. Whereas PheH (56 269 Da) exhibited homology to several members of the 4-hydroxyphenylacetate 3-hydroxylase (HpaB) family of proteins, PheH2 (8 035 Da) did not show significant sequence homology to previously characterised proteins. The deduced amino acid sequence of the third C-terminal truncated ORF (*pheC*) exhibited homology to 3,4-dihydroxyphenylacetate 2,3-dioxygenase enzymes. Biochemical characterisation of the crude 4-hydroxyphenylacetic acid 3-hydroxylase (4-HPA 3-hydroxylase) indicated that it is a moderately thermophilic intracellular enzyme. The enzyme was alkalophilic, with a maximum activity at pH 9.0. The enzyme was active in the temperature range of 45-65°C, with a maximum at 50°C. The PheH protein was purified by affinity chromatography and SDS-PAGE analysis revealed a molecular mass of 56 kDa, which is in agreement with the theoretical size of the monomeric protein. However, the purified PheH did not display hydroxylase activity, suggesting that the *Geobacillus* 4-HPA 3-hydroxylase is composed of two proteins with PheH being the hydroxylase and PheH2 serving as a "helper" protein required for productive hydroxylation of the substrate.

***This chapter has been submitted for publication:**

Hawumba, J.F., Brözel, V.S. and Theron, J. (2003). Nucleotide sequence and biochemical characterization of a 4-hydroxyphenylacetic acid 3-hydroxylase from thermophilic *Geobacillus* sp. PA-9. Submitted for publication: *Applied and Environmental Microbiology*.

5.1. INTRODUCTION

Degradation of aromatic compounds including phenolics, e.g. phenol, cresol, benzoate and catechol, as well as substituted aromatics, e.g. 4-hydroxyphenylacetic acid (4-HPA) and 3,4-dihydroxyphenylacetic acid (3,4-DHPA), has been widely studied among mesophiles, and several degradation pathways have been elucidated (Cooper and Skinner, 1980; Kim and Oriel, 1995; Ali *et al.*, 1998; Duffner *et al.*, 2000). In contrast, there is relatively little information on the degradation of these compounds by thermophilic microorganisms. Thermophilic degradation conditions may be more advantageous in terms of higher metabolic rates, higher solubility and bio-availability of many organic pollutants, lower sludge production and increased enzyme stability (Duffner and Müller, 1998).

Catabolism of phenolics proceeds through cleavage of the aromatic ring via the *ortho*-(Intradiol) or *meta*-(Extradiol) pathways (Cooper and Skinner, 1980; Que *et al.*, 1981). Gram-positive bacteria (Sparnins and Chapman, 1976), including thermophilic *Bacilli* (Ali *et al.*, 1998), utilise the *meta*-pathway in the catabolism of phenolic compounds. Degradation by the *meta*-cleavage pathway involves (a) hydroxylation of phenol to catechol by phenol hydroxylase; (b) ring fission via catechol 2,3-dioxygenase to 2-hydroxymuconic semialdehyde; and (c) either oxidation to 4-oxalocrotonate by NAD^+-dependent dehydrogenase, or hydrolysis to 2-oxopent-4-enoate (Buswell, 1974; Ali *et al.*, 1998). Hydroxylation at C_3 of the benzene nucleus of L-tyrosine (Fig. 5.1), 4-HPA and phenylacetate yields 3,4-dihydroxyphenylacetic acid (3,4-DHPA), which is also referred to as homoprotocatechuic acid (Sparnins *et al.*, 1974; Sparnins and Chapman, 1976; Cooper and Skinner, 1980). 3,4-DHPA is subsequently metabolised to carbon dioxide, pyruvate and succinate as the final products via 5-carboxymethyl 2-hydroxymuconic semialdehyde. If 3,4-DHPA is not metabolised further, its intracellular accumulation may lead to the production of a brown oxidation product that leaks from the cells, and results in the colonies displaying a brown colour (Cooper and Skinner, 1980). Alternatively, oxidation of the hydroxyl group at C_3 may be accompanied by a shift of the acetic acid side chain, resulting in the formation of homogentisic acid (2,5-dihydroxyphenylacetic acid) as the major metabolic intermediate (Blakley, 1972; Hareland, 1975).

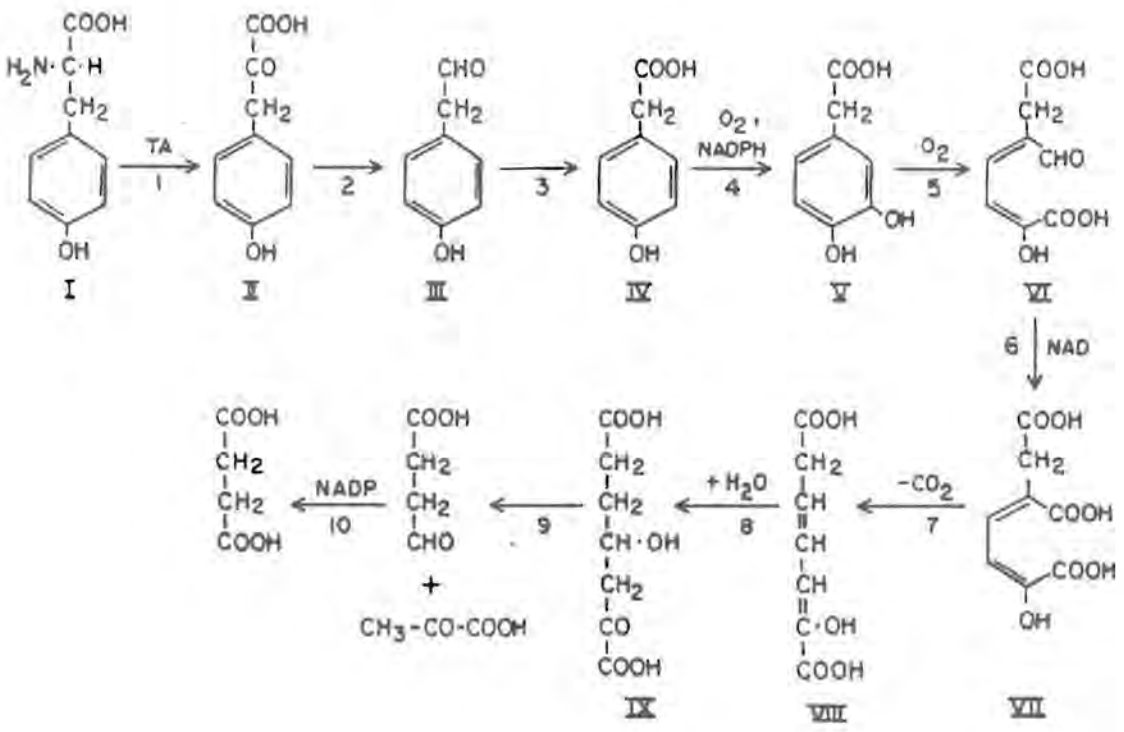


Fig. 5.1 Proposed steps in the degradation of L-tyrosine to pyruvic and succinic acids. Reacions 4 through 10 are those previously described for strains *Acinetobacter* and *Pseudomonas putida*. The key intermediates are as follows: 4-hydroxyphenylacetic acid (IV); 3,4-dihydroxyphenylacetic acid (V); 5-carboxymethyl-2-hydroxymuconic acid semialdehyde (VI); 2-hydroxyhepta-2,4-diene-1,7-dioic acid (VII); 4-hydroxy-2-ketopimelic acid (IX). *(Adapted from Cooper and Skinner, 1980)*

Owing to their exceptional stability, the 4-hydroxyphenylacetate 3-hydroxylase enzymes from *Pseudomonas putida* and *Pseudomonas putida* U, respectively, have been purified to homogeneity (Raju *et al.*, 1988; Fernández-Medarde and Luengo, 1997). A number of phenol degradation pathways and their constituent enzymes have also been described among thermophilic *Bacillus* spp. (Dong *et al.*, 1992; Kim and Oriel, 1995; Duffner *et al.*, 2000), but few genes have been cloned and sequenced to date (Dong *et al.*, 1992; Duffner *et al.*, 2000). In this part of the study, the identification, cloning and characterisation of a 4-hydroxyphenylacetic acid 3-hydroxylase (4-HPA 3-hydroxylase) from the thermophile *Geobacillus* sp. PA-9 is reported and its relationship with previously published sequences is discussed.

5.2. MATERIALS AND METHODS

5.2.1. Bacterial strains, plasmids and growth conditions

The bacterial strains and plasmids used in this part of the study are listed in Table 5.1. *Geobacillus* sp. PA-9 was cultured at 55°C in modified Castenholtz medium as described previously (Section 3.3.2). *Escherichia coli* strains DH5α and MC1061 were used as the hosts for the cloning procedures, and were grown in Luria-Bertani (LB) liquid medium (0.5% [w/v] yeast extract, 1% [w/v] tryptone, 1% [w/v] NaCl; pH 7.4) or on LB agar plates at 37°C. The media were supplemented when necessary with 100 µg/ml ampicillin or 10 µg/ml tetracycline to select recombinants. The vectors pBluescript SK II (+) (Stratagene) and pSVBI (obtained from F. Meinhardt and K-D. Wittchen, Westfälische Wilhelms-Universität Münster, Institut für Mikrobiologie, Münster, Germany) were used for cloning and DNA sequencing, and the vector pMP220, containing a promoterless *lacZ* gene (obtained from V. Venturi, Bacteriology Group, International Centre for Genetic Engineering and Biotechnolgy, Trieste, Italy), was used to screen for promoter activity.

5.2.2. Preparation of plasmid DNA

Plasmid DNA was prepared by the alkaline lysis method as described by Sambrook *et al.* (1989).

5.2.3. Origin of a clone displaying 4-HPA hydroxylase activity

The construction of a genomic library from *Geobacillus* sp. PA-9 had been described previously (Section 4.2.7). Briefly, the *Geobacillus* sp. PA-9 genomic DNA was digested with *Hind*III and 1- to 8-kb fragments were ligated with *Hind*III-digested pSVBI DNA. During screening of the transformants (*E. coli* as host) for proteolytic activity, several clones producing a brown pigment were identified (Section 4.3.5). Plasmid pSVBI-R113 represents one on the clones isolated from the genomic library and harbours a 2.7-kb insert, as confirmed by restriction enzyme analysis with *Hind*III. The insert was recloned into the *Hind*III site of plasmid pBluescript SK II (+), to yield plasmid pBluescript-R113.

5.2.4. DNA sequencing

Plasmids pSVBI-R113 and pBluescript-R113 were used as the templates for DNA sequencing. The nucleotide sequence of both strands of the 2 715-bp insert was determined by automated sequencing with an ABI PRISM™ BigDye™ Terminator Cycle Sequencing Ready Reaction mixture (Perkin-Elmer Applied Biosystems, Foster City, CA, USA), in a Hitachi 3100 capillary array automated DNA sequencer. From the sequence obtained several insert-specific primers were designed (Table 5.2), and used to determine the sequence of the full-length insert and to obtain good overlaps in both strands.

5.3.5. Nucleotide and deduced amino acid sequence analysis

The nucleotide sequence was edited using the Sequence Analysis 3.1 and Sequencing Navigator 1.0.1 programmes, included in the ABI PRISM™ software package (Perkin-Elmer Applied Biosystems). The deduced amino acid sequence was analysed using the complete nonredundant protein sequence database (NCBI) to search for protein homology and domain composition. Sequence alignments were done using ClustalX (Thompson *et al.*, 1997), and signal peptide identification was performed through SignalP software (Henrik *et al.*, 1997). The physico-chemical properties of the deduced amino acid sequence were determined using the sequence analysis tools available at the ExPASY server (http://www.expasy.org/), while the presence of defined protein patterns were determined using the SMART database (at http://www.smart.embl_heidelberg.de). A dendogram was constructed using the neighbour-joining algorithm (Saitou and Nei, 1987) and the tree was displayed with NJPLOT.

Table 5.1: Bacterial strains and plasmids used in the study

Strains and plasmids	Relevant properties	Source
Bacterial strains:		
Geobacillus sp. PA-9		Hawumba *et al.* (2002)
Escherichia coli DH5α	*hsdR recA lacZYA* ϕ 80d*lacZΔM15*	Stratagene
Escherichia coli MC1061	*XylE*⁻, Δ(*lac*)X74	W.H. van Zyl[a]
Plasmids:		
pMP220	*LacZ*⁺ Tet^r	V. Venturi[b]
pSVBI	*E. coli*/*B. megaterium* shuttle vector, Tet^r, Amp^r	F. Meinhardt and K-D. Wittchen[c]
pBluescript SK II (+)	Cloning vector, ColE1, Amp^r, LacZα peptide	Stratagene
pSVBI-R113	2.7-kb genomic DNA fragment of *Geobacillus* sp. PA-9 cloned into the *Hind*III site of pSVBI	This study, Chapter 4
pBluescript-R113	Insert from pSVBI-R113 cloned into the *Hind*III site of pBluescript SK II (+)	This study

[a] W.H. van Zyl, Department of Microbiology, University of Stellenboch, South Africa
[b] V. Venturi, Bacteriology Group, International Centre for Genetic Engineering and Biotechnolgy, Area Science Park, Trieste, Italy
[c] F. Meinhardt and K-D. Wittchen, Westfälische Wilhelms-Universität Münster, Institut für Mikrobiologie, Münster, Germany

Table 5.2: Primers used in this study

Oligonucleotide primer	Nucleic acid sequence	Source
Nucleic acid sequencing:		
F-pSVBI	5'-GGGATCAACTTTGGGAGAG-3'	This study
R-pSVBI	5'-GGCGATTAAGTTGGGTAACG-3'	This study
T3	5'-AATTAACCCTCACTAAAGGG-3'	Stratagene
T7	5'-GTAATACGACTCACTATAGGG-3'	Stratagene
IT3	5'-GGAGTATATGGAGCGGC-3'	This study
IT7	5'-GTAGCTAAGCGCTTCGAG-3'	This study
EXT3	5'-GCCATTCCGAACAATACGC-3'	This study
EXT7	5'-GACTGTCACACACGAATCG-3'	This study
113Fd	5'-GGTGACGGATGTAAAGAG-3'	This study
113Rv	5'-GTATACAAGCGGACGAAG-3'	This study
PCR Amplification*:		
Fp-Promoter	5'-**gaagatc**TTCGGAATGTGACG-3'; *Bgl*II site incorporated	This study
Rp-Promoter	5'-**gctctag**GCCTTTCCTGTTTGGC-3'; *Xba*I site incorporated	This study

* In primer sequences, the restriction endonuclease sites are indicated in bold lower case letters

5.2.6. Isolation of the putative promoter region

No typical -10 and -35 promoter regions of the putative gene encoding the 4-hydroxyphenylacetic acid 3-hydroxylase could be identified upstream of the start codon. Consequently, to verify the presence of a promoter in this region, oligonucleotide primers Fp-promoter (containing a *Bgl*II site) and Rp-promoter (containing a *Xba*I site) (Table 5.2) were designed to amplify a 280-bp region of the upstream sequence. The reaction mixture (50 µl) contained 1 × PCR buffer, 1.5 mM $MgCl_2$, 50 ng of pSVBI-R113 plasmid DNA as template, 50 pmol of each primer, 50 mM of each dNTP, and 1 U of *Taq* polymerase (Southern Cross Biotechnology, South Africa). After initial denaturation at 96°C for 2 min, the tubes were subjected to 30 cycles of denaturation (94°C, 30 s), annealing (58°C, 30 s), and extension (72°C, 45 s) in a Perkin-Elmer 2400 themocycler (Applied Biosystems, Hercules, CA, USA). Cycling was completed by a final elongation step at 72°C for 7 min. An aliquot (5 µl) of the reaction mixture was electrophoresed on a 1% (w/v) agarose gel in the presence of an appropriate DNA molecular weight marker (phage λ DNA digested with both *EcoR*I and *Hind*III).

5.2.7. Cloning of the promoter fragment

Promoter probe vector pMP220 was used to assay promoter activity of the 280-bp PCR-amplified fragment. The amplicon was purified using the Wizard SV Gel and PCR Clean-Up system (Promega Corporation, Madison, WI, USA), digested with restriction enzymes *Bgl*II (Roche) and *Xba*I (Promega), and ligated into the *Bgl*II/*Xba*I sites, upstream of the promoterless *lacZ* gene, of pMP220. The ligation reaction mixture was used to transform competent *E. coli* MC1061 cells, prepared as described by Chung and Miller (1993), and the transformation mixtures were plated onto LB agar plates supplemented with 10 µg/ml of tetracycline. Putative recombinant transformants were randomly selected, cultured overnight at 37°C, and the extracted plasmid DNA was characterised by restriction digestion with both *Xba*I and *Bgl*II followed by agarose gel electrophoresis.

5.2.8. Assay for promoter activity

To assay for expression of β-galactosidase, *E coli* MC1061 strains harbouring recombinant pMP220 plasmid DNA were streaked onto LB agar plates supplemented with tetracycline (10 µg/ml) and 40 µg/ml X-Gal. Production of β-galactosidase was indicated by the appearance

of blue colonies. As a negative control, *E. coli* MC1061 transformed with non-recombinant pMP220 plasmid DNA was included in the assay.

5.2.9. Preparation of cell extracts and fractions

To investigate the cellular localisation of the enzyme, recombinant *E. coli*/pSVBI-R113 was cultured in LB broth until late exponential phase (16-20 h). At this stage, the culture was beginning to turn brown. For preparation of cell fractions, cells were harvested by centrifugation at 7000 rpm for 10 min at 4°C, and both the supernatant and the cell pellet kept. The pellet was washed twice in 50 mM phosphate buffer (pH 7.2), suspended in the same buffer and the cells were lysed by 15-s pulses for 5 min using an Ultrasonic Homogenizer at an output of 80% (Cole-Palmer Instruments Co., Chicago, Ill, USA). The cell lysate was cleared by centrifugation (7 000 rpm, 10 min) and the supernatant, considered as the cytoplasmic fraction, recovered for further analysis. The remaining pellet, consisting mainly of membrane proteins, was subsequently suspended in 50 mM phosphate buffer (pH 7.2). The three fractions (cell-free culture medium, cytoplasmic extract and cell debris) were assayed for enzyme activity at optimal conditions of temperature and pH as described below.

5.2.10. Assay for 4-hydroxyphenylacetic acid 3-hydroxylase activity

Enzyme activity was determined by measuring the liberation of 3,4-dihydroxyphenyl acetic acid (3,4-DHPA) from 4-hydroxyphenylacetic acid (4-HPA; Sigma, St. Louis, MO, USA) according to the method of Anrow (1937). The reaction mixture contained the following in a 800-µl volume: 387.5 µl of 50 mM Tris-HCl buffer (pH 9), 50 µl of 10 mM NADH, 2.5 µl of 10 mM FAD, 250 µl of 10 mM 4-HPA, and 10 µl of 10 mM Fe^{2+}. The mixture was equilibrated at 50°C for 5 min after which 100 µl of the cell extract was added, and the mixture incubated at 50°C for 30 min. The reaction was terminated by addition of 200 µl of 8% (w/v) trichloroacetic acid. After centrifugation (15 000 rpm, 5 min), the supernatant was transferred to a clean test tube and the liberated 3,4-DHPA determined by the method of Anrow (1937). Briefly, 1 ml of each 0.5 N HCl, 2% (w/v) nitrite-molybdate reagent and 1 M NaOH were added, in this order, to the supernatant and vortexed after each addition. The volume was adjusted to 5 ml with distilled water and the absorbance at 510 nm was read against a reagent blank using a Genesys spectrophotometer (Spectronic Instruments, USA). The absorbance of 1 ml of 3,4-DHPA (5 mg %; Sigma) was also determined and the moles of 3,4-DHPA liberated were calculated in relation to that of the standard.

5.2.11. Effect of pH and temperature on enzyme activity

The influence of pH on enzyme activity was determined in the following buffers: pH 4, 50 mM acetate buffer; pH 6, 50 mM pyridine buffer; pH 7, 50 mM phosphate buffer; pH 8-9, 50 mM Tris-HCl buffer; pH 10-11, 50 mM methylamine-HCl; and pH 12, 50 mM phosphate buffer. The reaction mixtures at the above pH values were incubated at 50°C for 30 min. To determine the influence of temperature on enzyme activity, samples were incubated in 50 mM Tris-HCl buffer (pH 9) at various temperatures ranging from 25 to 90°C. In all cases, the enzyme activity was determined as described above.

5.2.12. Preparation of 4-HPA-coupled amino-agarose

Amino-agarose (ICN Biochemicals Inc., Ohio, USA) and 4-HPA were coupled according to the method of Cuatrecasas (1970). Briefly, a Büchner funnel containing a moistened filter paper was prepared. Amino-agarose (suspended in 8 ml of water) was poured into the filter funnel and washed with 50 mM phosphate buffer (pH 7.2). The funnel with the washed amino-agarose was removed from the filtering flask and its outlet covered tightly with Parafilm. The ligand (4-HPA; 5 mg % in 50 mM cold phosphate buffer, pH 7.2) and equal to the volume of the amino-agarose was added, and the suspension was mixed immediately with a stirring glass rod, followed by washing with phosphate buffer. The entire procedure of washing, adding ligand solution and mixing was limited to 2 min. The suspension was then transferred to a glass beaker containing a stirrer magnetic bar and gently stirred for 16 h at 4°C. The efficiency of coupling was tested as described by Raju *et al.* (1988) by formation of a violet colour after treating an aliquot of the matrix with diazotized *p*-nitroaniline, followed by 10 % (w/v) NaOH.

5.2.13. Enzyme purification

The enzyme was purified by affinity chromatography from the cytoplasmic extract prepared from a 100-ml recombinant *E. coli* culture, grown in LB broth overnight at 37°C in a shaking incubator. The enzyme extract (10 ml) was loaded onto a 4-HPA-coupled amino-agarose column (8 mm × 152 mm) and unbound proteins were removed by washing the column with five volumes of the 50 mM phosphate buffer (pH 7.2). The bound enzyme was eluted from the column using 50 mM sodium acetate buffer (pH 5.2). The purified enzyme was tested for activity using the standard enzyme activity assay in 50 mM Tris-HCl (pH 9) at 50°C as

described above (Section 5.2.10), and its molecular mass determined using SDS-PAGE gel electrophoresis.

5.2.14. Electrophoresis

After affinity chromatography, the purified protein was analysed by 12% SDS-PAGE gels, essentially as described by Laemmli (1970). The following molecular mass markers (ICN Biochemicals Inc., Ohio, USA) were used: cytochrome C, 13 kDa; myoglobin, 18kDa; chymotrypsinogen A, 24 kDa; ovalbumin, 45 kDa; bovine serum albumin, 67 kDa; gamma (γ) globulin, 160 kDa.

5.3. RESULTS AND DISCUSSION

5.3.1. Characterisation of a clone encoding a putative 4-HPA hydroxylase

Recombinant *E. coli* colonies producing a diffusible brown pigment were serendipitously identified while screening a genomic library of the newly characterised *Geobacillus* isolate PA-9 for proteolytic enzymes (Section 4.3.5). This brown pigment has been identified as a polymer similar to melanin and may result from the formation of quinones after subsequent polymerisation (Gibello *et al.*, 1995). Among the aromatic compounds able to produce quinones by spontaneous oxidation are dihydroxylated aromatic compounds (catechol, resorcinol, 3,4-dihydroxybenzoate and 3,4-dihydroxyphenylacetic acid) and DL-dihydroxyphenylalanine (Gibello *et al.*, 1995). The brown pigment production by the clones thus suggested that a component in LB medium acts as a substrate for the enzyme(s) synthesised by the genes contained on the cloned genomic DNA fragment. To confirm that production of the pigment was indeed plasmid encoded, the extracted recombinant plasmid DNA, designated as pSVBI-R113, was digested with *Hind*III and an excised 2.7-kb fragment was subsequently cloned into pBluescript SK II (+) vector. The resultant recombinant plasmid (pBluescript-R113) was used to transform competent *E. coli* DH5α cells. Transformed *E. coli* colonies also produced the brown pigment, and its production appeared to be independent of the orientation of the cloned DNA fragment relative to the *lac* promoter of the pBluescript SK II (+) vector. These results therefore suggested that the gene(s) on the cloned DNA fragment was transcribed from its own promoter and possibly contained the full-length gene(s).

5.3.2. Nucleotide sequence analysis of the 4-HPA hydroxylase-encoding region of *Geobacillus* isolate PA-9

In order to identify the gene(s) on pSVBI-R113, the sequences of both strands of the 2.7-kb *Hind*III fragment were determined by the primer walking method using newly constructed oligonucleotides (Table 5.2). Analysis of the nucleotide sequence of the 2 715-bp *Hind*III fragment by using the ORF Finder program revealed the existence of three co-linear open reading frames (ORFs) (Fig. 5.2).

The amino acid sequence deduced from the largest ORF exhibited homology to known 4-hydroxyphenylacetic acid 3-hydroxylase enzymes. The *Geobacillus* isolate PA-9 enzyme was thus designated PheH, and the corresponding gene designated *pheH*. The *pheH* gene starts at an ATG start codon at nucleotide 292 and terminates at a TGA stop codon at nucleotide 1 776, yielding an ORF of 1 484 nucleotides that encode a predicted protein of 494 amino acids, with a molecular mass of 56 269 Da and a pI of 6.18. A putative ribosomal binding sequence (Shine and Dalgarno, 1974), AACGGA, is present upstream of the predicted start codon. Although no typical -10 and -35 promoter regions were identified upstream of the ATG codon of the *pheH* gene, the upstream region of the ORF was shown to contain a promoter (see below). No inverted repeat sequence for a possible transcription terminator was found downstream of the terminator codon. The G+C content of *pheH* is 51%, and the codon usage is typical for a *Geobacillus* gene with an average G+C content of 69% in the third codon position (Duffner and Müller, 1998; Nakamura *et al.*, 2000).

The second ORF is located downstream of the *pheH* gene and is in the same orientation as that of *pheH*. The ORF was assigned as *pheH2* and encodes for a protein of 70 amino acid residues initiating with an ATG at nucleotide position 2 068 and terminating at position 2 283 by a following TGA codon. The predicted protein has a molecular mass and pI of 8 035 Da and 11.1, respectively. The ORF was preceded by a potential Shine-Dalgarno sequence (TGAAGG) and no inverted repeat sequence that resembled a putative transcription termination site could be identified downstream of the stop codon. Like the upstream *pheH* gene, *pheH2* exhibits an average G+C content of 57% and a high frequency of G or C in the third position (55%). The intergenic region between *pheH* and *pheH2* did not conspicuously show conserved promoter structures, suggesting that the *pheH* and *pheH2* ORFs may be part of the same transcription unit. Database searches with the deduced amino acid sequence of

pheH2 did not allow for a function to be assigned to the protein. The protein showed homology (29-32% amino acid sequence identity over 67 amino acids) to the hypothetical proteins of *Magnetospirillum magnetotacticum* (GenBank Accession no. ZP00056133) and *Microbulbifer degradans* (GenBank Accession no. ZP00067553), respectively. It is, however, important to note that several 4-hydroxyphenylacetic acid 3-hydroxylase enzymes require a second smaller protein to function as part of a two-component system for optimum activity in the degradation of aromatic compounds (Duffner *et al.*, 2000). Thus, it may be possible that PheH2 may play such a role in the 4-HPA 3-hydroxylase activity of *Geobacillus* isolate PA-9.

The third ORF, designated as *pheC*, started at nucleotide position 2 458, but is still incomplete as no TGA stop codon could be identified. Preceding the ATG start codon, a putative ribosome binding sequence (AGCAGG), as well as potential -10 (GAATAT) and -35 (ATGAAG) sequences were identified. Despite the deduced amino acid sequence derived from this ORF being truncated at its C-terminus, it nevertheless exhibited significant levels of homology to the amino acid sequence of 3,4-dihydroxyphenylacetate 2,3-digoxygenase enzymes. The highest level of homology was observed with the putative 3,4-dihydroxyphenylacetate 2,3-digoxygenase of *Oceanobacillus iheyensis* HTE831 (GenBank Accesion no. NP693790; 55% amino acid sequence identity over 54 amino acids) and of *Arthrobacter globiformis* (GenBank Accesion no. I39588; 45% amino acid sequence identity over 60 amino acids). Based on the results obtained by these database searches, it can be proposed that PheC of *Geobacillus* isolate PA-9 may be a 3,4-dihyroxyphenylacetate 2,3-digoxygenase.

To determine the presence of a promoter upstream of the start codon of *pheH*, a *lac*Z fusion plasmid was constructed in a promoterless promoter probe vector pMP220, as described in Materials and Methods (Section 5.2.6). The expression of the *lac*Z fusion in *E. coli* MC1061, of which the *lac* genes had been completely deleted, indicated that a 280-bp region upstream of the *pheH* gene contains a promoter, as colonies turned deep blue in colour after 12 hours of incubation on X-gal-containing agar plates (results not shown). *E. coli* MC1061 containing only the pMP220 vector did not turn blue. However, a more precise characterisation of the putative promoter and regulatory elements awaits further analysis by primer extension assays and mutagenesis analysis.

pheH

241 GAAAACGGAGGGGATGAAAATG
M K M 3

301 CCGGCCAAAACAGGAAAGGAGTATATGGAGCGGCTCAAGCAGGCGAAAAGCAGCGTGTAC
P A K T G K E Y M E R L K Q A K S S V Y 23

361 ATCCACGGGGAAAAAGTGGAGGATGTCACCGTTCATCCCGCGTTCCGCAACGTCGTCCGC
I H G E K V E D V T V H P A F R N V V R 43

421 TCGATGGCGGCGCTTTACGACCGGCAATACGAAAAGCCGGAGAAGATGTTGTACCGGTCG
S M A A L Y D R Q Y E K P E K M L Y R S 63

481 CCGACAACCGGGCAGCCGGTCGGGATGACGTTCATTCAGCCAACTACGATTGACGAGCTC
P T T G Q P V G M T F I Q P T T I D E L 83

541 ATCGCCCGCCGCGAGGCGACGCAAGAGTGGGCGCGGATGTCGGCCGGAATGATGGGGCGC
I A R R E A T Q E W A R M S A G M M G R 103

601 TCGCCTGATTATTTGAATGCCGAAGTCATGGCGATGGGCATCGCCAACGATTTGTTTGCC
S P D Y L N A E V M A M G I A N D L F A 123

661 GAAGACGATCCGATGTTTGCCGAAAATGCGAAAAACTATTATGAATACGCGCGGGAACAC
E D D P M F A E N A K N Y Y E Y A R E H 143

721 GACATCAGCTTGACGCATACGCTCATCCATCCGCAAATGAACCGCGCCAAGGCGCTGCAC
D I S L T H T L I H P Q M N R A K A L H 163

781 GAACAAAACGATGCCGATGTGCCGCTTCATTTGGTGGAACGGCGCAAAGACGGGATCATC
E Q N D A D V P L H L V E R R K D G I I 183

841 GTCAGCGGCATCCGCCTATTGGCGACGCAAGGCGGCATCACCGATGAAATTTTAGTGTTT
V S G I R L L A T Q G G I T D E I L V F 203

901 CCCTCGACCGTGAAAAAAGCGACATCCGGCGAAGACCCATATGCGCTTGCCTTTGCCATT
P S T V K K A T S G E D P Y A L A F A I 223

961 CCGAACAATACGCCAGGCGTGAAGTTCATTTGCCGCGAGGCGTTTGACTACGGGCGGAGC
P N N T P G V K F I C R E A F D Y G R S 243

1021 GCGTGGGATCATCCGCTGGCATCGCGTTTTGAGGAAGGCGATGCGATCGTTTCGTTTGAA
A W D H P L A S R F E E G D A I V S F E 263

1081 AATGTGTTTGTGCCGTGGGAGCGCGTGTTTGTGTGCGGCAATTCATCGATTTGCAACCGG
N V F V P W E R V F V C G N S S I C N R 283

1141 ACGTTCCGGGAAACGAACGCGGTTGTTCATATGTCCCATCAAGTCGTGGCGAAAAACATC
T F R E T N A V V H M S H Q V V A K N I 303

1201 GTCAAAACGGAGTTTTTGCTTGGCGTCACCCTTTGCCTCATCGAAGCGATCGGCATCGGC
V K T E F L L G V T L C L I E A I G I G 323

1261 GAGTTCCAGCATGTGAAAGACAAAGGGGCGGAAATCATGCTCGTCCTCGAGACGATGAAA
E F Q H V K D K G A E I M L V L E T M K 343

1321 AGCCATTTGTACCGGGCCGAGCACAATGCGAAGCGAGACCGTTGGGGAACGATGACGCCC
S H L Y R A E H N A K R D R W G T M T P 363

1381 GATTTTGCCGCGCTCGATGCCGCGCGCAACTGGTATCCGCGCATTTATCCGCGCCTGGCG
D F A A L D A A R N W Y P R I Y P R L A 383

1441 GAGATCATCCGCATTTTAGGGGCATCGGGGTTGATGGCGATTCCGACGGAGGCGGATTTT
E I I R I L G A S G L M A I P T E A D F 403

```
1501  CAGCACGAGGAAATTGGCGACATCGTTCGCCGGGCGATGCAAGGGGCGACGGTGGACGGC
       Q  H  E  E  I  G  D  I  V  R  R  A  M  Q  G  A  T  V  D  G  423

1561  TATGAGCGCGTCCAGCTGTTCCGGCTCGCCTGGGATCTCACGATGAGCGCGTTCGGCGCC
       Y  E  R  V  Q  L  F  R  L  A  W  D  L  T  M  S  A  F  G  A  443

1621  CGGCAGACGCATTACGAATATTACTTTTTCGGCGACCCCGTTCGGATGGGGATGGCGTAT
       R  Q  T  H  Y  E  Y  Y  F  F  G  D  P  V  R  M  G  M  A  Y  463

1681  TTCGATGGCTACGAAAAAGAGCCGTATAAACAGTTCGTGCGCGAATTTTTGCGCGGCGCC
       F  D  G  Y  E  K  E  P  Y  K  Q  F  V  R  E  F  L  R  G  A  483

1741  AAAAGCGTGTTCATTCCGGCTGACAACAAGCATTGACCGTTCATTCGCACGTTCGCAAAC
       K  S  V  F  I  P  A  D  N  K  H  *                          494

1801  ACGGATCGGGGAACAGCGGAGGGTGCGCGGCTGCGGCGTTTCCTAAGAGGCTGGTGATTT
1861  CGTGCAAAATATGACTGATTCCATCTGTTTCTCAATTAGAGAAACCTTGCAGAATCAGCT
1921  TTCTTTTCTGAATGAAAATACCGCCGATTCGTGTGTGACAGTCGTTTGTCACCGCCCTTC
1981  TAAGGTTGCCGCCGCTCGAAAACAGGCGTTCAGTGGCGCGCAAGCCGAGCGGATTTGGGC
                                  pheH2
2041  GTGCTGCGAGATGTTTGAAGGAAAGGATGTGGCCAGCAACAAGGAACGTTTTGTTCCAT
                                  M  W  P  A  T  R  N  V  L  F  H   11

2101  GTTCGCCAGCGGAAACGACGGCTTCTTTCTGTCGGCCGCCCAACGCTAGATGAAGGAAGG
       V  R  Q  R  K  R  R  L  L  S  V  G  R  P  T  L  D  E  G  R  30

2161  TTGCGAAAGGGGGAAGAGAACGATGGCATTTTCCATTATTCGCTGCGCCCGCGTGGTGCT
       L  R  K  G  E  E  N  D  G  I  F  H  Y  S  L  R  P  R  G  A  50

2221  GCATGTGACCGATTTGGCCCGCCGCCCGCGATTTTTACGAGCGGGTGCTTGGGTTTGTCG
       A  C  D  R  F  G  P  P  P  A  I  F  T  S  G  C  L  G  L  S  70

2281  TGACGGAAGCGGATGAAGATCATCTGTATTTGCGGGGACTTGAAGAATATCATCATCATA
       *

2341  GCCTTCTCTTGAAAAAGGCGCCCCGCCCTGCGGTCGAAGCGCTTAGCTACAAAGTCGGTT
                                                          pheC
2401  CCGAGCAGGAATTGGAAGCGCTTTATGAATGGTTTGCGGCGCAAACGCTGAGCCGAAATG
                                                              M   1

2461  GCTCGAAGCGGGAAGCCAGCGCGCCGTCGGCAAGGCGTTTCGCGTGAAGATCCATCCGGA
       A  R  S  G  K  P  A  R  R  R  Q  G  V  S  R  E  D  P  S  G  21

2521  CTGCCGCTTGAGTTTTTTGCCCATATGGAGAAAACCGAGCGGTTGTTGCAGCGGTATGAC
       L  P  L  E  F  F  A  H  M  E  K  T  E  R  L  L  Q  R  Y  D  41

2581  TTGTACCGCGGCGCCCGCGTGCAGCGCATCGATCATTTCAACTGCGCGGTGACGGATGAA
       L  Y  R  G  A  R  V  Q  R  I  D  H  F  N  C  A  V  T  D  E  61

2641  AGAGCGCATATGACTTTATGTCCGTGGTCTGGGCTTTGCCTGCTCAGAATATACGGAAAC
       R  A  H  M  T  L  C  P  W  S  G  L  C  L  L  R  I  Y  G  N  81

2701  AGAGGATGGGAAGCT
       R  G  W  E  A
```

Fig. 5.2 The nucleotide sequence of the 2715-bp *Hind*III fragment containing the genes of the 4-HPA 3-hydroxylase-encoding region of *Geobacillus* isolate PA-9 and the deduced amino acid sequence in one-letter code is shown. The ATG start and TGA stop codons of the predicted ORFs are shown in bold, and putative Shine-Dalgarno sequences are underlined. Putative -35 and -10 nucleotides upstream of the *pceC* ORF are indicated in italics.

5.3.3. Analysis of the deduced amino acid sequence of the putative 4-HPA 3-hydroxylase of *Geobacillus* isolate PA-9

To gain a better understanding regarding the possible function of the predicted PheH enzyme of *Geobacillus* sp. PA-9, the protein was analysed using various internet-based bioinformatic programmes, as indicated in Materials and Methods (Section 5.3.5). The calculated instability index (Guruprasad *et al.*, 1990) and aliphatic index (Ikai, 1980) were 37.91 and 77.19, respectively, suggesting that the protein is stable. The protein lacks an N-terminal signal peptide sequence (Henrik *et al.*, 1997) and is predicted to be localised to the cytoplasmic membrane (Klein *et al.*, 1985). Analysis of the sequence with SMART revealed that PheH possesses an HpaB domain, which spans amino acid residues 108 to 482, and is conserved among the HpaB family of 4-hydroxyphenylacetic acid 3-hydroxylase enzymes. These enzymes catalyse the hydroxylation of 4-hydroxyphenylacetic acid (4-HPA), leading to the production of 3,4-DHPA (Sparnins *et al.*, 1974; Sparnins and Chapman, 1976; Cooper and Skinner, 1980).

Comparison of the deduced amino acid sequence of *pheH* by a BLAST-P search to the sequences in the GenBank database revealed similarity to several 4-HPA hydroxylase enzymes, and the percentage identity between the closest matching sequences was calculated in pairwise alignments for the full-length proteins using LALIGN (Pearson *et al.*, 1997). PheH displayed homology to putative 4-hydroxyphenylacetate 3-hydroxylases of several bacteria of which the completed genome sequences have recently been published. PheH showed the highest amino acid sequence identity to the enzyme of *Oceanobacillus iheyensis* HTE831 (67%; GenBank Accession No. NP693794), while lower identities were found to the enzymes of *Bacillus halodurans* (42%; GenBank Accession no. BAB07555), *Deinoccocus radiodurans* R1 (41%; GenBank Accession no. AAF12410), *Yersinia pestis* KIM (30%; GenBank Accession no. AAM86095) and *Sulfolobus solfataricus* (28%; GenBank Accession no. AAK42238). In addition to these hydroxylases, PheH also displayed 30% amino acid sequence identity with the hypothetical phenol hydroxylase of *Sulfolobus tokodaii* (GenBank Accession no. BAB65734).

A comparison of PheH to previously characterised aromatic ring hydroxylases revealed that PheH shares 33% and 30% amino acid sequence identity, respectively, with the phenol 2-hydroxylase of *Geobacillus thermoglucosidasius* A7 (encoded by *pheA*1, GenBank Accession

no. AAF66546) (Duffner *et al.*, 2000) and *Geobacillus thermoleovorans* A2 (encoded by *pheA*, GenBank Accession no. AAC38324) (Duffner and Müller, 1998). Furthermore, PheH displayed 30% amino acid sequence identity with the 4-hydroxyphenylacetate hydroxylases of *Photorhabdus luminescens* (encoded by *hpaB*, GenBank Accession no. AAO17197) (Waterfield *et al.*, 2001), *E. coli* W ATCC 11105 (encoded by *hpaB*, GenBank Accession no. CAA86048) (Prieto *et al.*, 1996), *Klebsiella pneumoniae* (encoded by *hpaA*, GenBank Accession no. Q48440) (Gibello *et al.*, 1997) and 29% amino acid sequence identity with ORF4, which encodes the chlorophenol 4-monooxygenase of *Streptomyces globisporus* (GenBank Accession no. AAL06674) (Liu *et al.*, 2002). Notably, all these enzymes, with the exception of PheA of *G. thermoleovorans* A2, are two-component aromatic hydroxylases requiring a smaller component for optimal hydroxylase activity. A dendogram resulting from multiple sequence alignment of the *pheH* deduced amino acid sequence with the above enzymes, is indicated in Fig. 5.3.

As with the 4-HPA 3-hydroxylases of *E. coli* W and *K. pneumoniae*, and the phenol hydroxylases of *G. thermoglucosidasius* A7 and *G. thermoleovorans* A2, the FAD and NAD binding signature sequences, GXGXXG (Wierenga *et al.*, 1986) and [TM]XXXX[IVAL][YWF][IVAL][IVA]GD (Eggink *et al.*, 1990), respectively, were not detected in the PheH sequence of *Geobacillus* sp. PA-9. However, the activity of the 4-HPA hydroxylases from *K. pneumoniae* (Martin *et al.*, 1991) and *E. coli* (Prieto and Garcia, 1994) has been reported to be stimulated by FAD and consequently, the possibility that the above enzymes may contain atypical FAD and NAD binding sites cannot be excluded.

BLAST-P homology searches revealed that the deduced amino acid sequence of *pheH2* does not show similarity to characterised proteins. However, pairwise alignments between the small proteins (HpaC) of the 4-HPA 3-hydroxylases from *E. coli* (GenBank Accession no. Z29081), *K. pneumoniae* (GenBank Accession no. Q48411) and *P. luminescens* (GenBank Accession no. AAO17198) demonstrated homology (10-11% amino acid sequence identity) with the gene product of *pheH2* from *Geobacillus* sp. PA-9. In addition, the gene product from *pheA2* of *G. thermoglucosidasius* A7 (GenBank Accession no. AAF66547) shared 15% amino acid sequence identity with the *pheH2* gene product of *Geobacillus* sp. PA-9. These low levels of homology between the different proteins may not be surprising, as they appear to be diverse and show little sequence conservedness between them (Duffner *et al.*, 2000).

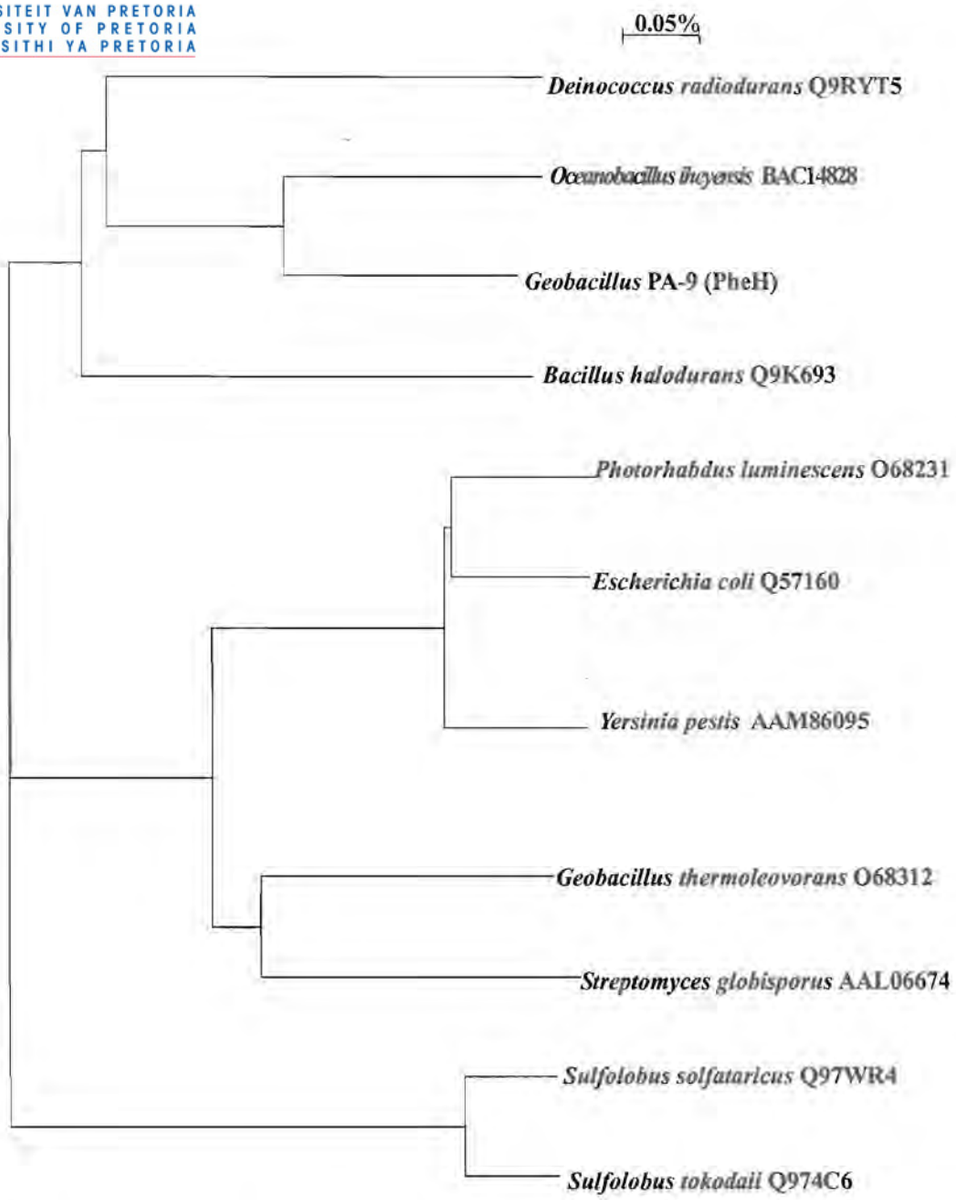


Fig. 5.3 Dendogram resulting from the multiple amino acid sequence alignment of different 4-hydroxyphenylacetic acid 3-hydroxylases (4-HPA 3-hydroxylases) and phenol hydroxylases with PheH from *Geobacillus* PA-9. Q9RYT5, 4-HPA 3-hydroxylase of *Deinococcus radiodurans*; BAC14828, 4-HPA 3-hydroxylase of *Oceanobacillus iheyensis*; Q9K693, 4-HPA 3-hydroxylase of *Bacillus halodurans*; O68231, 4-HPA 3-hydroxylase of *Photorhabdus luminescens*; Q57160, 4-HPA 3-hydroxylase of *E. coli*; AAM86095, putative 4-HPA 3-hydroxlase of *Yersinia pestis* KIM; O68312, phenol hydroxylase of *Geobacillus thermoleovorans*; AAL06674, chlorophenol 4-monooxygenase of *Streptomyces globisporus*; Q97WR4, 4-HPA 3-hydroxylase of *Sulfolobus solfataricus*; and Q974C6, putative phenol hydroxylase of *Sulfolobus tokodaii*. The GenBank Accession numbers for these enzymes are supplied in the text. The scale bar represents 0.05% of the estimated substitutions per amino acid position.

assays were performed on freshly prepared crude cytoplasmic enzyme extracts.

5.3.4.2. Effect of temperature and pH on enzyme activity

The effect of pH and temperature on the 4-HPA hydroxylase from *Geobacillus* sp. PA-9 was determined using a crude cytoplasmic enzyme extract and 4-HPA as substrate. The rate of liberation of 3,4-dihdroxyphenylacetic acid was determined spetrophotometrically at 510 nm, and one unit of activity was defined as the amount of enzyme that liberates one nanomole of product under the assay conditions. The activity of the 4-HPA 3-hydroxylase from

5.3.4. Biochemical features of the 4-HPA hydroxylase from *Geobacillus* strain PA-9

5.3.4.1. Cellular location

In order to study the cellular location of the 4-HPA hydroxylase, recombinant *E. coli*/pSVBI-R113 was used. Consequently, the cell-free culture medium, cellular (cytoplasmic) extract and cell debris of the recombinant culture were assayed for 4-hydroxyphenlyacetic acid 3-hydroxylase activity in 50 mM Tris-HCl buffer (pH 9) at 50°C. The highest activity was observed in the cytoplasmic fraction. No activity was detectable in the culture medium and low activity was recovered in the cell debris. By contrast, a control non-recombinant *E. coli* DH5α culture did not show hydroxylase activity in identically prepared fractions (results not shown). These results were in agreement with those obtained by *in silico* analysis of the respective proteins (Section 5.3.2), and thus strongly suggested that the 4-HPA 3-hydroxylase was probably intracellular showing some kind of association with the cytoplasmic membrane. Therefore, the cytoplasmic fraction of recombinant *E. coli*/pSVBI-R113 was selected for further enzyme characterisation and for enzyme purification. However, the enzyme lost activity when stored in solution at 4 or -21°C overnight. Due to these observations, all of the assays were performed on freshly prepared crude cytoplasmic enzyme extracts.

5.3.4.2. Effect of temperature and pH on enzyme activity

The effect of pH and temperature on the 4-HPA hydroxylase from *Geobacillus* sp. PA-9 was determined using a crude cytoplasmic enzyme extract and 4-HPA as substrate. The rate of liberation of 3,4-dihdroxyphenylacetic acid was determined spetrophotometrically at 510 nm, and one unit of activity was defined as the amount of enzyme that liberates one nanomole of product under the assay conditions. The activity of the 4-HPA 3-hydroxylase from *Geobacillus* strain PA-9 was measured following adjustment of the reaction pHs from 4 to 12 with various buffers (Fig. 5.4a). The enzyme showed maximal activity at pH 9. However, the enzyme exhibited 69% of the maximal activity at pH 8 and 47% at pH 10. The enzyme did not display appreciable activity at pH 4 and 12 (lower than 25% of the maximal activity). The optimum temperature of the enzyme was determined by varying the reaction temperature at pH 9 (Fig. 5.5b). The enzyme had an optimum temperature of 50°C. The enzyme, however, remained active over a range of temperatures varying from 45 to 65°C, with approximately 63% and 57% relative activity at 45 and 65°C, respectively. From 70°C onwards, the activity decreased sharply (26% residual activity), suggesting that the enzyme was not very thermostable.

5.3.4.3. Purification and SDS-PAGE analysis of the purified enzyme

SDS-PAGE analysis of the crude cell extracts from *E. coli*/pSVBI-R113 showed the presence of an additional band to those detected in control extracts from *E. coli* DH5α (Fig. 5.5, lanes 1 and 2, respectively). The enzyme was subsequently purified by affinity chromatography on a 4-HPA-coupled amino-agarose column. SDS-PAGE analysis of the purified enzyme (Fig. 5.5, lane 3) indicated that the enzyme was almost purified to homogeneity, and the molecular mass was estimated to be 56 kDa. This value is in agreement with the calculated molecular mass (56.269 kDa) of PheH. No protein corresponding in size to PheH2 could be detected in the stained gel. The purified protein, by contrast to the crude cytoplasmic enzyme extracts (Section 5.3.3.1), did not display activity toward the 4-HPA substrate under optimal assay conditions. This result suggests that for optimal activity of 4-HPA 3-hydroxylase from *Geobacillus* isolate PA-9, both PheH and PheH2 are needed and that the *Geobacillus* PA-9 enzyme is most probably a two-protein component system similar to the cognate enzymes from *E. coli* (Prieto and Garcia, 1994), *K. pneumoniae* (Gibello *et al.*, 1997), *G. thermoglucosidasius* A7 (Duffner *et al.*, 2000) and *P. luminescens* (Waterfield *et al.*, 2001).

In summary, the molecular and biochemical characterisation of the 4-HPA 3-hydroxylase from *Geobacillus* sp. PA-9 revealed that the enzyme belongs to a family of aromatic hydroxylases that require two protein components to catalyse the hydroxylation of its substrate. Only a few hydroxylases of this type have been characterised so far. In *Geobacillus* sp. PA-9, the proteins are encoded by different genes, *pheH* and *pheH2*, which may be translated from a single transcription unit. Although the molecular mass of the PheH component (56 kDa) is in agreement with that of the cognate subunit of the 4-HPA 3-hydroxylases from other bacteria, the deduced molecular mass of *pheH2* (8 kDa) is different, being lower. The corresponding small proteins of the 4-HPA 3-hydroxylasaes from *E. coli* (Prieto and Garcia, 1994), *K. pneumoniae* (Gibello *et al.*, 1997), *G. thermoglucosidasius* A7 (Duffner *et al.*, 2000) and *P. luminescens* (Waterfield *et al.*, 2001) display only 10-15% amino acid sequence identity to PheH2. This is in agreement with the low amino acid sequence identity observed between the large components of these enzymes (30%). From these results it is clear that the 4-HPA 3-hydroxylase from *Geobacillus* sp. PA-9 is novel, and significantly different from currently known 4-HPA 3-hydroxylases.

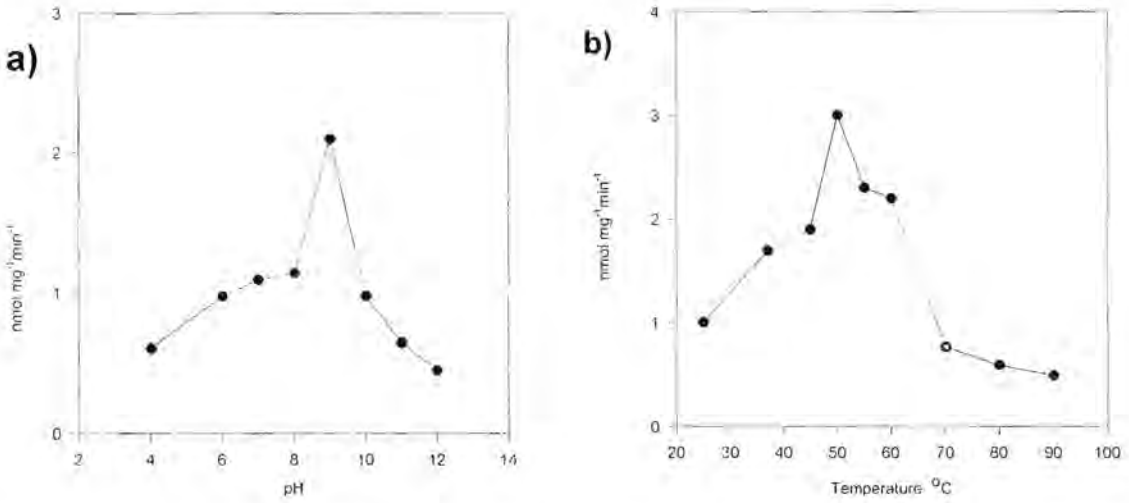


Fig. 5.4 Influence of pH (a) and temperature (b) on the specific activity of the 4-hydroxyphenylacetic acid 3-hydroxylase (4-HPA 3-hydroxylase) from *Geobacillus* isolate PA-9. For the pH profile, activity was measured at 50°C in buffers of different pH values. For the temperature profile, activity was measured in 50 mM Tris-HCl buffer (pH 9) at different temperatures. Values are the means of results of three separate experiments.

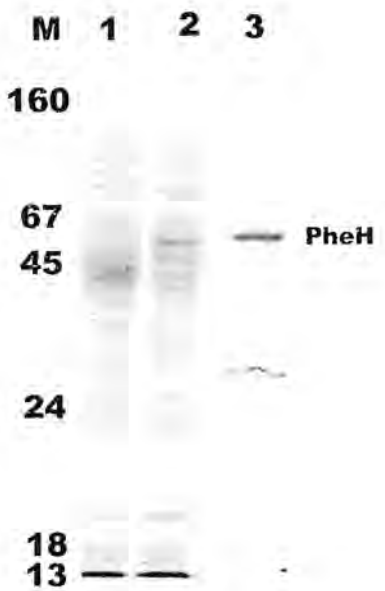


Fig. 5.5 SDS-PAGE analysis of the expression and purification of the 4-hydroxyphenylacetic acid 3-hydroxylase (4-HPA 3-hydroxylase) from *Geobacillus* isolate PA-9. Lane 1, cell extract sample from non-recombinant *E. coli* DH5α; lane 2, cell extract sample from *E. coli*/pSVBI-R113; lane 3, sample of the affinity chromatography-purified PheH protein. The sizes of the molecular mass markers (in kDa) are shown to the left of the figure.

Interestingly, the PheH protein from *Geobacillus* sp. PA-9 displayed homology (30-33% amino acid sequence identity) to the phenol 2-hydroxylases encoded by *pheA* of *Geobacillus thermoleovorans* A2 (Duffner and Müller, 1998) and *pheA*1 of *G. thermoglucosidasius* A7 (Duffner *et al.*, 2000). Furthermore, the brown pigment production observed for *E. coli*/pSVBI-R113 cells indicated that a component in LB medium, in the absence of 4-HPA, acts as a substrate for the 4-HPA 3-hydroxylase of *Geobacillus* sp. PA-9. This suggests that the 4-HPA 3-hydroxylase of *Geobacillus* sp. PA-9 might recognize other substrates distinct from 4-HPA and could thus have a broad substrate range for activity, similar to what has been reported for the 4-HPA 3-hydroxylases of *E. coli* (Prieto *et al.*, 1993) and *K. pneumoniae* (Gibello *et al.*, 1997).

The precise role of the smaller coupling protein still remains unclear, but it has been proposed to have a flavin reductase function (Galan *et al.*, 2000) and to prevent the wasteful oxidation of NADH in the absence of a substrate to be hydroxylated (Arunachalan *et al.*, 1992). Other relaxed substrate specificity hydroxylases have developed similar mechanisms of control, e.g. the methane monooxygenase, which is composed of three proteins: a reductase, a hydroxylase and a "regulatory" protein that is required for the coupling between NADH oxidation and substrate hydroxylation (Rosenzweig *et al.*, 1993). The 4-HPA 3-hydroxylases from *E. coli* (Prieto and Garcia, 1994), *K. pneumoniae* (Gibrello *et al.*, 1997) and *Pseudomonas putida* (Arunachalan *et al.*, 1992) are all FAD-dependent enzymes and the absence of the smaller subunit results in either a loss or drastically reduced activity of the 4-HPA 3-hydroxylases (Arunachalan *et al.*, 1992; Takizawa *et al.*, 1995; Gibello *et al.*, 1997; Duffner *et al.*, 2000). In addition, the two protein components of 4-HPA 3-hydroxylases are not very stable complexes, since they can be easily separated by ammonium sulfate fractionation (Arunachalan *et al.*, 1992) or during purification of the protein complex (Prieto *et al.*, 1993; Prieto and Garcia, 1994). Whether the lack of activity displayed by the purified PheH component of the 4-HPA 3-hydroxylase of *Geobacillus* sp. PA-9, observed in this study, is a consequence of the absence of the smaller protein has to be confirmed by future biochemical analysis. However, taking into account the brown phenotype shown by the *E. coli*/pSVBI-R113 cells and the hydroxylating activity of their cytoplasmic extracts, it is tempting to speculate that *pheH* encodes the hydroxylase and *pheH2* codes for the coupling protein, which serves as an effector protein of the hydroxylase reaction. Similar to the above 4-HPA 3-hydroxylases, the enzyme from *Geobacillus* sp. PA-9 also lacked the conserved

FAD and NADH signature sequences and might thus contain uncommon FAD and NADH binding sites, which need to be further characterized.

It has been recognised that enzymes derived from thermophiles tend to be resistant to chemical denaturation, suggesting that pathways exist whereby thermophilic biotransformation of xenobiotics in the environment can occur at concentrations that may be toxic to mesophiles (Gurujeyalakshmi and Oriel, 1989). Consequently, a number of possible applications of aromatic-degrading organisms and their enzymes have been sought. The degradation of 4-hydroxyphenylacetic acid (4-HPA) is environmentally important, because it is a product of aromatic amino acid catabolism (Sparnins *et al.*, 1979) and lignin decomposition (Crawford *et al.*, 1982), and is found as an industrial pollutant in waste water from olive oil production (Balice and Cera, 1984). The 4-HPA 3-hydroxylase of *Geobacillus* sp. PA-9 has optimum activity at 50°C and pH 9, making it an attractive enzyme to engineer for future biotechnological application in degrading 4-HPA.

5.4. REFERENCES

Ali, S., Fernandez-Lafuente, R. and Cowan, D.A. (1998). *Meta*-pathway degradation of phenolics by themophilic *Bacilli*. *Enzyme and Microbial Technology* 23: 462-468.

Anrow, E.L. (1937). Colorimetric determination of the components of 3,4-dihydroxyphenylalanine-tyrosine mixture. *Journal of Biological Chemistry* 118: 531-537.

Arunachalam, U., Massey, V. and Vaidyanathan, C.S. (1992). *p*-Hydroxyphenylacetate 3-hydroxylase. A two-protein component enzyme. *Journal of Biological Chemistry* 267: 25848-25855.

Balice, V. and Cera, O. (1984). Acidic phenolic fraction of the olive vegetation water determined by a gas chromatographic method. *Grasas y Aceites* 35: 178-180.

Blakley, E.R. (1972). Microbial conversion of *p*-hydroxyphenylacetic acid to homogentisic acid. *Canadian Journal of Microbiology* 18: 1247-1255.

Buswell, J.A. (1974). The *meta*-cleavage pathway of catechol by a thermophilic *Bacillus* species. *Biochemical and Biophysical Research Communication* 60: 934-941.

Chung, T.C. and Miller, H.R. (1993). Preparation and storage of competent *Escherichia coli* cells. *Methods in Enzymology* 218: 621-627.

Cooper, R.A. and Skinner, M.A. (1980). Catabolism of 3- and 4-hydroxyphenylacetate by the 3,4-dihydroxyphenylacetate pathway in *Escherichia coli*. *Journal of Bacteriology* 143: 302-306.

Crawford, D.L., Barder, M.J., Pometto, A.L. and Crawford, R.L. (1982). Chemistry of softwood lignin degradation by *Streptomyces viridosporus*. *Archives of Microbiology* 131: 140-145.

Cuatrecasas, P. (1970). Protein purification by affinity chromatography: Derivatizations of agarose and polyacrylamide beads. *Journal of Biological Chemistry* 245: 3059-3065.

Dong, L., Wang, L., Wang, C., Cheng, J., He, Z., Sheng, Z. and Shen, R. (1992). Molecular cloning and mapping of phenol degradation genes from *Bacillus stearothermophilus* FDTP-3 and their expression in *Escherichia coli*. *Applied and Environmental Microbiology* 58: 2531-2535.

Duffner, F.M. and Müller, R. (1998). A novel phenol hydroxylase and catechol 2,3-dioxygenase from the thermophilic *Bacillus thermoleovorans* strain A2: Nucleotide sequence and analysis of the genes. *FEMS Microbiology Letters* 161: 37-45.

Duffner, F.M., Kirchner, U., Bauer, M.P. and Müller, R. (2000). Phenol/cresol degradation by the thermophilic *Bacillus thermoglucosidasius* A7: Cloning and sequence analysis of five genes involved in the pathway. *Gene* 256: 215-221.

Eggink, G., Engel, H., Vriend, G., Terpstra, P. and Witholt, B. (1990). Rubredoxin reductase of *Pseudomonas oleovorans*. Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. *Journal of Molecular Biology* 212: 135-142.

Fernández-Medarde, A. and Luengo, J.S. (1997). Purification and characterisation of the 4-hydroxyphenyllacetic acid-3-hydroxylase from *Psuedomonas putida* U. *FEMS Microbiology Letters* 157: 47-53.

Galan, B., Diaz, E., Prieto, M.A. and Garcia, J.L. (2000). Functional analysis of the small component of the 4-hydroxyphenylacetate 3-monooxygenase of *Escherichia coli* W: A prototype of a new flavin:NAD(P)H reductase subfamily. *Journal of Bacteriology* 182: 627-636.

Gibello, A., Ferrer, E., Sanz, J. and Martin, M. (1995). Polymer production by a *K. pneumoniae* 4-HPA hydroxylase gene cloned in *E. coli*. *Applied and Environmental Microbiology* 61: 4167-4171.

Gibello, A., Suarez, M., Allende, J.L. and Martin, M. (1997). Molecular cloning and analysis of the genes encoding the 4-hydroxyphenylacetate hydroxylase from *Klebsiella pneumoniae*. *Archives of Microbiology* 167: 160-166.

Gurujeyalakshmi, G. and Oriel, P. (1989). Isolation of phenol-degrading *Bacillus stearothermophilus* and partial characterisation of phenol hydroxylase. *Applied and Environmental Microbiology* 55: 500-502.

Guruprasad, K., Reddy, B.V.B. and Pandit, M.W. (1990). Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Engineering* 4: 155-161.

Hareland, W.A., Crawford, R.L., Chapman, P.J. and Dagley, S. (1975). Metabolic function and properties of 4-hydroxyphenylacetic acid 1-hydroxylase from *Pseudomonas acidovorans*. *Journal of Bacteriology* 121: 272-285.

Hawumba, F.J., Theron, J. and Brözel, S.V. (2002). Thermophilic protease-producing *Geobacillus* from Buranga hot springs in western Uganda. *Current Microbiology* 45: 144-150.

Henrik, N., Engelbrecht, J., Brunak, S. and Von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* 10: 1-6.

Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *Journal of Biochemistry* 88: 1895-1898.

Kim, I.C. and Oriel, P.J. (1995). Characterisation of a *Bacillus stearothermophilus* BR219 phenol hydroxylase gene. *Applied and Environmental Microbiology* 61: 1252-1256.

Klein, P., Kanehisa, M. and DeLisi, C. (1985). Finding transmembrane regions. *Biochimica et Biophysica Acta* 815: 468.

Laemmli, U.K. (1970). Cleavage associated with the maturation of the head of bacteriophage T4. *Nature* 227: 680-685.

Liu, W., Christenson, S.D., Standage, S. and Shen, B. (2002). Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* 297: 1170-1173.

Martin., M, Gibello, A., Fernandez, J., Ferrer, E. and Garrido-Pertierra, A. (1991). Catabolism of 3- and 4-hydroxyphenylacetic acid by *Klebsiella pneumoniae*. *Journal of General Microbiology* 132: 621-628.

Nakamura, Y., Gojobori, T. and Ikemura, T. (2000). Codon usage tabulated from the international DNA sequence databases: Status for the year 2000. *Nucleic Acids Research* 28: 292.

Pearson, W.R., Wood, T., Zhang, Z. and Miller, W. (1997). Comparison of DNA sequences with protein sequences. *Genomics* 46: 24-36.

Prieto, M.A., Diaz, E. and Garcia, J.L. (1996). Molecular characterization of the 4-hydroxyphenylacetate catabolic pathway of *Escherichia coli* W: Engineering a mobile aromatic degradative cluster. *Journal of Bacteriology* 178: 111-120.

Prieto, M.A. and Garcia, J.L. (1994). Molecular characterisation of phenol 4-hydroxyphenylacetate 3-hydroxylase of *Escherichia coli*. A two-protein component system. *Journal of Biological Chemistry* 269, 22823-22829.

Prieto, M.A., Perez-Aranda, A. and Garcia, J.L. (1993). Characterization of an *Escherichia coli* aromatic hydroxylase with a broad substrate range. *Journal of Bacteriology* 175: 2162-2167.

Que, L., Widom, J. and Crawford, R.L. (1981). 3,4-Dihydroxyphenylacetate 2,3-dioxygenase: A manganese (II) dioxygenase from *Bacillus brevis*. *The Journal of Biological Chemistry* 256: 10941-10944.

Raju, G.S., Kamath, A.V. and Vaidyanathan, S.C. (1988). Purification and properties of 4-hydroxyphenylacetic acid 3-hydroxylase from *Pseudomonas putida*. *Biochemical and Biophysical Research Communications* 154: 537-543.

Rosenzweig, A.C., Frederick, C.A., Lippard, S.J. and Nordlund, P. (1993). Crystal structure of a bacterial non-haem iron hydroxylase that catalyses the biological oxidation of methane. *Nature* 366: 537-543.

Saitou, N. and Nei, M. (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425.

Sambrook, J., Fritsch, E.F. and Maniates, T. (1989). Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Shine, J. and Dalgarno, L. (1974). The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences USA* 71: 1342-1346.

Sparnins, V.L., Burbee, D.G. and Dagley, S. (1979). Catabolism of L-tyrosine in *Trichosporon cutaneum*. *Journal of Bacteriology* 138: 425-430.

Sparnins, V.L., Chapman, P.J. and Dagley, S. (1974). Bacterial degradation of 4-hydroxyphenylacetic acid and homoprotocatechuic acid. *Journal of Bacteriology* 120: 159-167.

Sparnins, V.L. and Chapman, P.J. (1976). Catabolism of L-tyrosine by the homoprotocatechuate pathway in Gram-positive bacteria. *Journal of Bacteriology* 127: 362-366.

Takizawa, N., Yokoyama, H., Yanagihara, K., Hatta, T. and Kiyohara, H. (1995). A locus of *Pseudomonas pickettii* DTP0602, *had*, that encodes 2,4,6-trichlorophenol-4-dechlorinase with hydroxylase activity, and hydroxylation of various chlorophenols by the enzyme. *Journal of Fermentative Bioengineering* 80: 318-327.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997). The Clustal-X windows interface-flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acid Research* 25: 4876-4882.

Waterfield, N.R., Bowen, D.J., Fetherston, J.D., Perry, R.D. and French-Constant, R.H. (2001). The *tc* genes of *Photorhabdus*: A growing family. *Trends in Microbiology* 9: 185-191.

Wierenga, R.K., Terpstra, P. and Hol, W.G.J. (1986). Prediction of the occurrence of the ADP-binding $\beta\alpha\beta$-fold in proteins using an amino acid sequence fingerprint. *Journal of Molecular Biology* 187: 101-107.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

Two protease-producing bacterial isolates, PA-5 and PA-9, from Buranga hot springs in western Uganda were characterised morphologically, physiologically and their phylogenetic position established by 16S rRNA sequencing analyses (Chapter 3). The two isolates could be grouped into the newly described genus *Geobacillus*. Isolate PA-5 was related to *Geobacillus thermoleovorans* B23 and DSM5366 (formerly known as *Bacillus thermoleovorans* B23 and DSM5366, respectively), while isolate PA-9 was more closely related to *Geobacillus uzenensis* strain X and *Geobacillus kaustophilus* NCIM8547 (formerly known as *Bacillus kaustophilus*). Between themselves, the two isolates are phylogenetically related to a low degree (96%). A more definitive classification in future, however, should include DNA-DNA hybridisation analysis in order to delineate the species of the isolates. Furthermore, it can be suggested that Buranga Hot Springs constitute a habitat for a diverse array of microorganisms whose diversity is unknown. Bio-prospecting of microorganisms in this region for production of novel products that may be of industrial use is recommended.

Like many members of *Bacillus* and related genera, including, among others, *Geobacillus*, isolates PA-5 and PA-9 were found to produce extracellular proteases. Preliminary zymogram analysis had indicated that isolate PA-5 produced at least two proteases, and PA-9 produced and least eight proteases (Chapter 3). Isolate PA-9 was further studied for its proteases (Chapter 4). Though purification of the proteases to homogeneity could not be achieved by gel-filtration and ion-exchange chromatography methods, partial purification from negative-stained SDS-PAGE gels suggested the existence of a multimeric protein whose monomers could possibly associate in more than one caseinolytic-active combination. However, whether the oberved active small molecular weight proteins with proteolytic activity were truncated forms of a large protease, could not be ruled out. Furthermore, the presence of contaminants in the eluted proteins suggested the existence of strong protein-protein interaction that could not be overcome by varying pH or by using different column matrices. It is recommended that the use of detergents, such as Triton-X100, and denaturing agents, such as urea and β-mercaptoethanol, be included in future purification strategies.

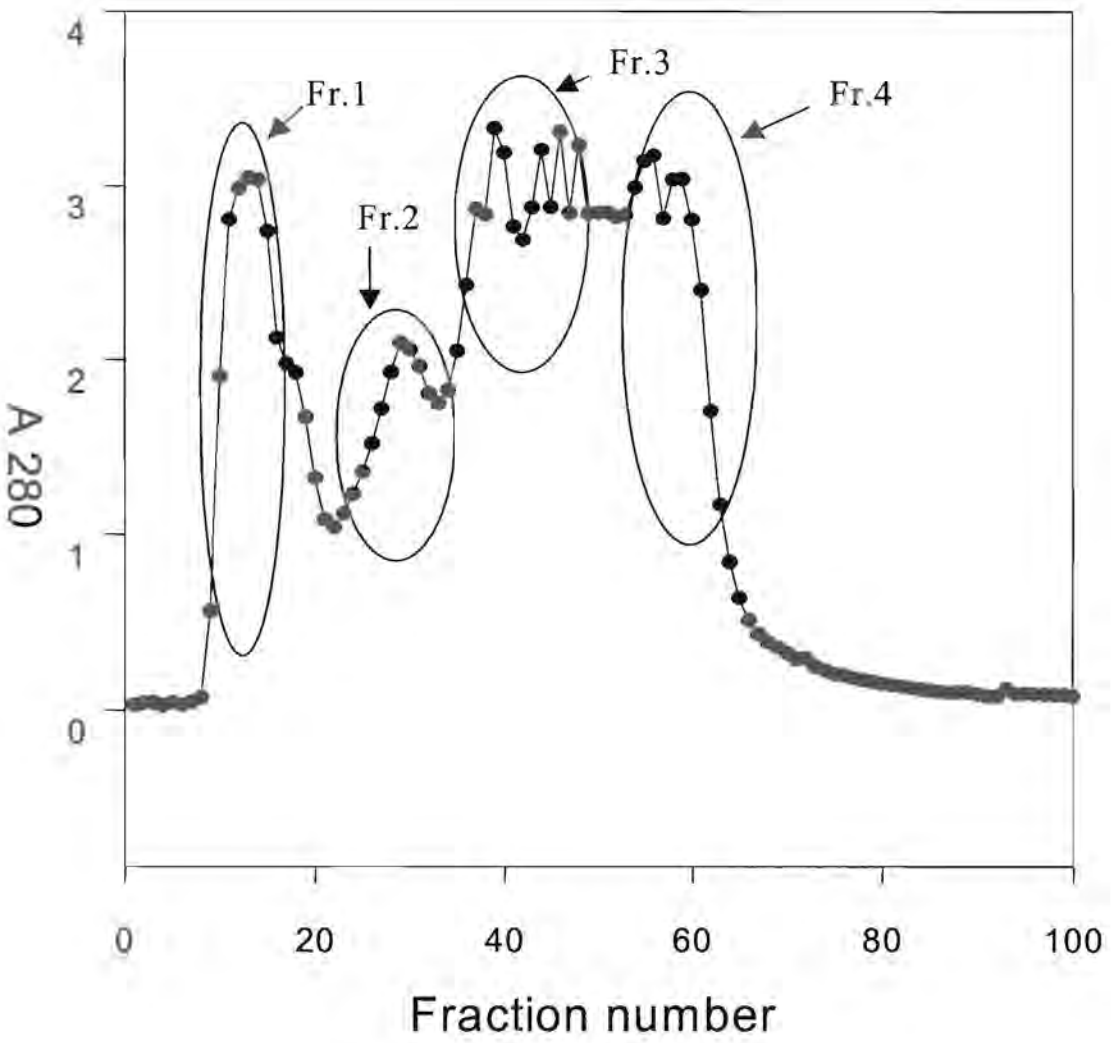
Failure to purify to homogeneity any proteases led to a switch from the strategy of identification of the gene through its N-terminal amino acid sequence, to a strategy whereby a genomic DNA library could be screened at a gene-level. This approach did not lead to the identification of protease-active recombinant clones, but brown pigment-producing recombinants were identified. Furthermore, direct amplification of the alkaline protease

gene(s) using primers, which had been designed based on the sequences of known proteases, followed by sequencing, did not yield amplicons resembling protease-encoding genes. This led to the following conclusions: that the protease(s) from PA-9 is unique and shared no or very low levels of homology to proteases identified thus far or that the regions from which the primers were designed are conserved in a number of non-protease genes. It had been thought that the shotgun approach failed, because the restriction fragments might have lacked the necessary control elements to allow for their expression in *Escherichia coli*. To isolate protease-encoding genes, other approaches can be investigated, especially the use of an expression-cloning system harbouring the appropriate control elements such as a functional promoter, pre-prosequence, prosequence and signal sequence. The nature of the protease(s) of PA-9 remains unresolved. It is highly recommended that efforts be directed at purifying the enzyme or individual monomers and to fully characterise the enzyme.

Analysis of a clone obtained from the genomic DNA library of *Geobacillus* isolate PA-9 that was capable of producing a brown pigment led to the discovery of a novel gene belonging to the tyrosine/phenylalanine catabolic pathway, i.e. 4-hydroxyphenylacetic acid 3-hyrdoxylase (PheH) (Chapter 5). This enzyme may be potentially useful in the environmental degradation of aromatic compounds and xenobiotics. However, for meaningful management of environmental wastes, it is desirable that other enzymes in the pathway be characterised. Since the 2.7-kb fragment revealed the existence of three open reading frames, *pheH*, *pheH2* and *pheC*, of which *pheC* displayed high homology with 3,4-dihydroxyphenylacetic acid 2,3-dioxygenases, it is recommended that the 2.7-kb fragment be labeled and used as a probe to screen the genomic library in order to identify other genes forming part of this pathway.

APPENDICES

Appendix I: Graph showing the elution profile of different fractions by gel-filtration chromatography.



Graph of absorbance against fraction number is shown. The fractions constituting an elution peak are circled. Four fractions represented as Fr.1 to Fr.4 were collected and assayed for caseinolytic activity by the cup-plate assay. Fractions: Fr.2, Fr.3 and Fr.4, which showed activity, were further purified by ion-exchange chromatography.

Appendix II: Multiple amino acid sequence alignment of different alkaline serine proteases.

```
Alignment: A:\sq.136762.pir

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
M22407          NSSEKEVIVV YKN-KAGKET IL---DS--- ---------- ----DADVEQ
P16396          NSSEKEVIVV YKN-KAGKET IL---DS--- ---------- ----DADVEQ
D13158          SEEKKEYLIV VEPDEVSAQS VE---ES--- ---------- ----YDVDVIH
D26542          SEEKKEYLIV VEPDEVSAQS VE---ES--- ---------- ----YDVDVIH
BAA25184        EEAKEKYLIG FKEQEVMSQF VD---QIDGD EYSISSSQVE D--VEIDLLH
M65086          EEAKEKYLIG FNEQEAVSEF VE---QVEAN DEVAILSEEE E--VEIELLH
M64743          SSTEKKYIVG FKQTMSAMSS AK---KKD-- -------VIS E--KGGKVQK
AAB47045        ---------- ---------- ---------- ---------- ----------
BAA93474        SESEKSYIVG FK--ASATTN SS---KKQ-- -------AVT Q--NGGKLEK
M13760          ---------- ---------- ---------- ---------- ----------
D10730          ---------- ---------- ---------- ---------- ----------
JW0075          DYVPNQLIVK FKQNASLSNV QS---FHKSV G--------A N--VLSKDDK
JC4802          DYVPGELIVK FKDGISAQST QS---IHAQY G--------A K--SIEKSKY
AY028615        AYVQGEVIVQ FKEQVNAEEK AK---ALKEV G--------A T--AVPDNDR
AAK29176        AYVQGEVIVQ FKEQVNAEEK AK---ALKEV G--------A T--AVPDNDR
Q45670          AYVQGEVIVQ FKEQVNAEEK AK---ALKEV G--------A T--AVPDNDR

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        60         70         80         90        100
M22407          QYKHLPAVAV TADQETV--- ---------- ---------- ----KELKQD
P16396          QYKHLPAVAV TADQETV--- ---------- ---------- ----KELKQD
D13158          EFEEIPVIHA ELTPEEL--- ---------- ---------- ----KKLQND
D26542          EFEEIPVIHA ELTKKEL--- ---------- ---------- ----KKLQKD
BAA25184        EFDFIPVLSV ELDPQDV--- ---------- ---------- ----EALELD
M65086          EFETIPVLSV ELSPEDV--- ---------- ---------- ----DALELD
M64743          QFKYVNAAAA TLDEKAV--- ---------- ---------- ----KELKKD
AAB47045        ---------- ---------- ---------- ---------- ----------
BAA93474        QYRLINAAQV KMSEQAA--- ---------- ---------- ----KKLEHD
M13760          ---------- ---------- ---------- ---------- ------MNGE
D10730          ---------- ---------- ---------- ---------- ------MR-K
```

```
JW0075      LGFEVVQFSK GTVKEKI--- ---------- ---------- ----KSYKNN
JC4802      LGFEVVKFD- GSVEKMI--- ---------- ---------- ----EKYKNN
AY028615    VKSKFNVLKV GNVEAVV--- ---------- ---------- ----KALNHN
AAK29176    VKSKFNVLKV GNVEAVV--- ---------- ---------- ----KALNHN
Q45670      VKSKFNVLKV GNVEAVV--- ---------- ---------- ----KALNNN

                    110        120        130        140        150
M22407      PDILYVENNV S------FTA ADSTDFKVLS DGTDTSDNFE QWNLEP-IQV
P16396      PDILYVENNV S------FTA ADSTDFKVLS DGTDTSDNFE QWNLEP-IQV
D13158      PNVKAIEENA E------VTI SQT------- ---------V PWGISR-IRT
D26542      PNVKAIEENA E------VTI SQT------- ---------V PWGISR-INT
BAA25184    PAISYIEEDA E------VTT MQT------- ---------V PWGINR-VQA
M65086      PAISYIEEDA E------VTT MAQS------ ---------V PWGISR-VQA
M64743      PSVAYVEEDH I------AHE YAQS------ ---------V PYGISQ-IKA
AAB47045    ---------- ---------- -AQS------ ---------V PYGISQ-IKA
BAA93474    PSIAYVEEDH K------AEA YAQT------ ---------V PYGIPQ-IKA
M13760      IRLIPYVTNE Q------IMD VNE------- ---------L PEGIKV-IKA
D10730      FRLIPYKQVD K------VSA LSE------- ---------V PMGVEI-VEA
JW0075      PDVEYAEPNY Y------VHA FWTPNDPYFN ----------N QYGLQK-IQA
JC4802      PNVEYVEPNH Y------VHI MWTPND-LTS ---------R QWGPQK-VQA
AY028615    PLVEYAEPNY L------FNA AWTPNDTYYQ G--------Y QYGPQN-TYT
AAK29176    PLVEYAEPNY L------FNA AWTPNDTYYQ G--------Y QYGPQN-TYT
Q45670      PLVEYAEPNY L------FNA AWTPNDTYYQ G--------Y QYGPQN-TYT

                    160        170        180        190        200
M22407      KQAWKAGLTG KNIKIAVIDS GISP------ --------HD ----------
P16396      KQAWKAGLTG KNIKIAVIDS GISP------ --------HD ----------
D13158      QQAHNRGIFG NGARVAVLDT GIAS------ --------HP ----------
D26542      QQAHNRGIFG NGAKVAVLDT GIAS------ --------HP ----------
BAA25184    PIAQSRGFTG TGVRVAVLDT GISN------ --------HA ----------
M65086      PAAHNRGLTG SGVKVAVLDT GIST------ --------HP ----------
M64743      PALHSQGYTG SNVKVAVIDS GIDSS----- --------HP ----------
AAB47045    PALHSQGYTG SNVKVAVIDS GIDSS----- --------HP ----------
BAA93474    PAVHAQGYKG ANVKVAVLDT GIHAA----- --------HP ----------
M13760      PEMWAKGVKG KNIKVAVLDT GCDTS----- --------HP ----------
D10730      PAVWKASAKG AGQIIGVIDT GCQVD----- --------HP ----------
```

```
JW0075        PQAWDSQRSD PGVKVAIIDT GVQGS----- --------HP ----------
JC4802        PQAWDVTRSS SSTVIAIVDT GVQTN----- --------HP ----------
AY028615      AYAWDVTKGS SGQEIAVIDT GVDYT----- --------HP ----------
AAK29176      AYAWDVTKGS SGQEIAVIDT GVDYT----- --------HP ----------
Q45670        DYAWDVTKGS SGQEIAVIDT GVDYT----- --------HP ----------

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            210        220        230        240        250
M22407        --DLS--IAG GYSAVSYTSS ------YKDD NGHG------ -----THVAG
P16396        --DLS--IAG GYSAVSYTSS ------YKDD NGHG------ -----THVAG
D13158        --DLR--IAG GASFISSEPS ------YHDN NGHG------ -----THVAG
D26542        --DLR--IAG GASFISSEPS ------YHDN NGHG------ -----THVAG
BAA25184      --DLR--IRG GASFVPGEPN ------ISDG NGHG------ -----THVAG
M65086        --DLN--IRG GASFVPGEPS ------TQDG NGHG------ -----THVAG
M64743        --DLN--VRG GASFVPSETN P-----YQDG SSHG------ -----THVAG
AAB47045      --DLN--VRG GASFVPSETN P-----YQDG SSHG------ -----THVAG
BAA93474      --DLN--VAG GASFVPSEPN A-----TQDF QSHG------ -----THVAG
M13760        --DLKNQIIG GKNFSDDDGG KEDA--ISDY NGHG------ -----THVAG
D10730        --DLAERIIG GVNLTTDYGG VETN--FSDN NGHG------ -----THVAG
JW0075        --DLASKVIY GHDYVDNDNT ------SDDG NGHG------ -----THCAG
JC4802        --DLQGKIVQ GYDFVDNDSN ------PQDG NGHG------ -----THCAG
AY028615      --DLDGKVIK GYDFVDNDYD ------PMDL NNHG------ -----THVAG
AAK29176      --DLDGKVIK GYDFVDNDYD ------PMDL NNHG------ -----THVAG
Q45670        --DLDGKVIK GYDFVDNDYD ------PMDL NNHG------ -----THVAG

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            260        270        280        290        300
M22407        IIGAK-HN-- GYGIDGIAPE AQIYAVKALD QN----GSGD LQSLLQGIDW
P16396        IIGAK-HN-- GYGIDGIAPE AQIYAVKALD QN----GSGD LQSLLQGIDW
D13158        TIAAL-NN-- SIGVLGVAPS ASLYAVKVLG AS----GSGS ISGIAQGLQW
D26542        TIAAL-NN-- SIGVLGVAPS ASLYAVKVLG AS----GSGS ISGIAQGLQW
BAA25184      TIAAL-NN-- SIGVLGVAPN VDLYGVKVLG AS----GSGS ISGIAQGLQW
M65086        TIAAL-NN-- SIGVLGVAPN AELYAVKVLG AS----GSGS VSSIAQGLEW
M64743        TIAAL-NN-- SIGVLGVSPS ASLYAVKVLD ST----GSGQ YSWIINGIEW
AAB47045      TIAAL-NN-- SIGVLGVSPS ASLYAVKVLD ST----GSGQ YSWIINGIEW
BAA93474      TIAAL-DN-- TIGVLGVAPS ASLYAVKVLD RY----GDGQ YSWIISGIEW
M13760        TIAAN-DS-- NGGIAGVAPE ASLLIVKVLG GEN---GSGQ YEWIINGINY
D10730        TVAAA-ET-- GSGVVGVAPK ADLFIIKALS GD----GSGE MGWIAKAIRY
```

```
JW0075      ITGALTNN-- SVGIAGVAPQ TSIYAVRVLD NQ----GSG- TDAVAQGIRE
JC4802      IAAAVTNN-- GTGIAGMAPN ASIMPVRVLN NS----GSGT MAAVANGIAY
AY028615    IAAAETNN-- ATGIAGMAPN TRILAVRALD RN----GSGT LSDIADAIIY
AAK29176    IAAAETNN-- ATGIAGMAPN TRILAVRALD RN----GSGT LSDIADAIIY
Q45670      IAAAETNN-- ATGIAGMAPN TRILAVRALD RN----GSGT LSDIADAIIY

                ....|....| ....|....| ....|....| ....|....| ....|....|
                      310        320        330        340        350
M22407      SIANR----- -MDIVNMSLG TTSDSKILHD AVNKAYEQGV LLVAASGNDG
P16396      SIANR----- -MDIVNMSLG TTSDSKILHD AVNKAYEQGV LLVAASGNDG
D13158      AINNN----- -MHIINMSLG STSGSSTLEL AVNRANNAGI LLVGAAGNTG
D26542      AINNN----- -MHIINMSLG STSGSSTLEL AVNRANNAGI LLVGAAGNTG
BAA25184    AANNG----- -MHIANMSLG SSAGSATMEQ AVNQATASGV LVVAASGNSG
M65086      AGNNG----- -MHVANLSLG SPSPSATLEQ AVNSATSRGV LVVAASGNSG
M64743      AISNN----- -MDVINMSLG GPSGSTALKT VVDKAVSSGI VVAAAAGNEG
AAB47045    AISNN----- -MDVINMSLG GPTGSTALKT VVDKAVSSGI VVAAAAGNEG
BAA93474    AVANN----- -MDVINMSLG GPNGSTALKK AVDTANNRGV VVVAAAGNSG
M13760      AVEQK----- -VDIISMSLG GPSDVPELEE AVKNAVKNGV LVVCAAGNEG
D10730      AVDWRGPKGE QMRIITMSLG GPTDSEELHD AVKYAVSNNV SVVCAAGNEG
JW0075      AADSG----- -AKVISLSLG APNGGTALQQ AVQYAWNKGS VIVAAAGNAG
JC4802      AAQNG----- -ADVISLSLG GTSGSSALQS AVQQAWNSGA VVVAAAGNSS
AY028615    AADSG----- -AEVINLSLG CDCHTTTLEN AVNYAWNKGS VVVAAAGNNG
AAK29176    AADSG----- -AEVINLSLG CDCHTTTLEN AVNYAWNKGS VVVAAAGNNG
Q45670      AADSG----- -AEVINLSLG CDCHTTTLEN AVNYAWNKGS VVVAAAGNNG

                ....|....| ....|....| ....|....| ....|....| ....|....|
                      360        370        380        390        400
M22407      NGK----PVN YPAAYSSVVA VSATNEKNQL ASFSTTGD-- --EVEFSAPG
P16396      NGK----PVN YPAAYSSVVA VSATNEKNQL ASFSTTGD-- --EVEFSAPG
D13158      RQG----VN YPARYSGVMA VAAVDQNGQR ASFSTYGP-- --EIEISAPG
D26542      RQG----VN YPARYSGVMA VAAVDQNGQR ASFSTYGP-- --EIEISAPG
BAA25184    AGN-----VG FPARYANAMA VGATDQNNNR ASFSQYGA-- --GLDIVAPG
M65086      AGS-----IS YPARYANAMA VGATDQNNNR ASFSQYGA-- --GLDIVAPG
M64743      SSGS-SSTVG YPAKYPSTIA VGAVNSSNQR ASFSSAGS-- --ELDVMAPG
AAB47045    SSGS-TSTVG YPAKYPSTIA VGAVNSSNQR ASFSSAGS-- --ELDVMAPG
BAA93474    STGS-TSTVG YPAKYDSTIA VANVNSNNVR NSSSSAGP-- --ELDVSAPG
M13760      DGDERTEELS YPAAYNEVIA VGSVSVAREL SEFSNANK-- --EIDLVAPG
D10730      DGREDTNEFA YPAAYNEVIA VGAVDFDLRL SDFPNTNE-- --EIDIVAPG

AAB47045    VSIQSTLPGG TYGAYNGTSM ATPHVAGAAA LILSKHP--- --TWTNAQVR
BAA93474    TSILSTVPSS GYTSYTGTSM ASPHVAGAAA LILSKYP--- --NLSTSQVR
M13760      ENILSTLPNK KYGKLTGTSM AAPHVSGALA LIKSYEEESF QRKLSESEVF
D10730      VGIKSTYLDS GYAELSGTSM AAPHVAGALA LIINLAKDAF KRTLSETEIC
JW0075      SNIYSTYKGS TYQSLSGTSM ATPHVAGVAV LLANQGY--- ----SNTQIR
JC4802      SNIYSTYLNS SYASLSGTSM ATPHVAGLAA LLASQGR--- ----SNSQIR
AY028615    VDIVSTITGN RYAYMSGTSM ASPHVAGLAA LLASQGR--- ----NNIEIR
AAK29176    VDIVSTITGN RYAYMSGTSM ASPHVAGLAA LLS-QGR--- ----NNIEIR
Q45670      VDIVSTITGN RYAYMSGTSM ASPHVAGLAA LLASQGR--- ----NNIEIR

                ....|....| ....|....| ....|....| ....|....| ....|....|
```

```
JW0075      QIIESTTDKI SGTG---TYW KNGRVNAYKA VQYAKQLQEN KAS-------
JC4802      AAIENTADKI SGTG---TYF QHGRINAYKA VNY------- ----------
AY028615    QAIEQTADKI SGTG---TYF KYGRINSYNA VTY------- ----------
AAK29176    QAIEQTADKI SGTG---TYF KYGRINSYNA VTY------- ----------
Q45670      QAIEQTADKI SGTG---TYF KYGRINSYNA VTY------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       510        520        530        540        550
M22407      AQIDINKARE LISQLPNSDA KTALHKRLDK VQSYRNVKDA KDKVAKAEKY
P16396      AQIDINKARE LISQLPNSDA KTALHKRLDK VQSYRNVKDA KDKVAKAEKY
```

APPENDIX III: Nucleotide sequence of hybridisation-positive clones from genomic DNA library of *Geobacillus* sp. PA-9

1. Blot 15 sequence, 634 nucleotides (forward)

AACCGGGCCCGTCGACTCTAGAGCTCTGCAGGCATGCGATATCCCACTGGAAGCTTCCGTTCTTTAAAT
CCGTCTACAAGAATAACAGGAAGCTGATGATATATCGCAACCTCTTCCACCGTTTCGATCGCCCGCGC
ATACGGACTGGAAATGACGACATCGATTTGTTCTTTGTTCAAGATTTGCGCGACGCGCTTAGCATCTTG
TTTTCCCCGCTCCGATAACGGCCGGTGCCGTTCATCAGGCGTAAAAATCGAGTGGGCGTGCCTCACCA
AATATACGTTTGTGTTCATCTTGCCTCATTCCTTTTTAACCAGTCTAACGTAAAATCGACCAACTCCCGT
TGTTTCATCAGGCGGCATTCGGCATTGCCAACATAGCCGATGTGTCTTTTCAAAGTGCTTGCCAATTTG
CACAAAACGCTCTTCATCGGTTTCGACATCGATAAATGTCTTCCACACCCGCCGGCCATTTTCCATTAT
GGCAGCTCCTTGCAAAAATGTTTTCTTGCCTAACAGCGGATTTCCCGACTCCATCGAAAATCCGCCGCC
CCCAAACGCGATGATTTGCCGCATTCTCCCCCCTCATGCCAGCCGTTTCACGAGTTTCGTAATAAaATC
CCTTCTGGCATCCC

2. Blot 1 sequence, 618 nucleotides (forward)

AATTCCGGGGATCCACGCGTCTTAAGGcGGCCGCGGTACCGGGCCCGTCGACTCTAGAGCTCTGCAGG
CATGCGATATCCCATGGAAGCTTcCGTTCTTTAAATCCGTCTACAAGAATAACAGGAAGCTGATGATAT
ATCGCAcACCTCTTCACCGTTTCGATCGCCCGCGCATACGGACTGGAAATGACGACATCGATTTGTTCT
TTGTTCAAGATTTGCGCGACGCGCTTAGCATCTTGTTTTCCCCGCTCGATAAcCGGCCGGTGCCGTTAAT
CAGGCGTAAAAATCGAGTGGGCGTGCCTCACCAAAcTATACGcTTTGTGTTcATCTTGCCTCAcTTcCTTT
TTAACCAGTCTAACGTAAAAcTcCGACCAAcCTCCCGTTGTtTTCAcTCAaGGCGGCATTCGGCAaTTGCC
AACATAGCCGATGTGTCTTTTCAAAGTGCTTGcCAATTtTGCACcAAAAcCGCTCTTCcTCGGTTTCcGAC
ATCGATAAAcTGTCcTTCCACcACCCGCCGGcCcATTTCAaTTATGGCAGCTCTTGcAAAAATGTTTTcTTG
cCCTAACnAGCGGGAtTTTcCCgACTCCAttCgAAAATCCGCCCGCCCCCaAACCGA

3. Blot 13 sequence, 635 nucleotides (forward)

TAACAGGAATTCCCGGGGATCCACGCGTCTTAAGGCGGCCGCGGTACCGGGCCCGTCGACTCTAGAGC
TCTGCAGGCATGCGATATCCATGGAAGCTTCCGTGAAGCCGCTGCCGCATGGGCTGCAGAAATTTTCC
ATACGCTTAAAAGCTGTGTTAAAAGCGAAAGGCTACAACACGGCTGTCGGTGACGAAGGCGGATTTG
CCTCCGAACTTAAAATCGAACGAAGAAGCGCTGCACACACGATCCACTTGAAGCGATCGAAAAAGCC
GGCCTACAAACCAGGCCGAACAACGTGATGCCTCGCTACTGGACGTTGCTTCGTCGGAGCTGTACACA
CAAAGAAGATGGCAAATATCATTTGGAAGGCGAAGGCGTCGTCAAAACATCACGAAGAAATGGTTGC
TTCGGTATGAAGAGCTTGTGTCGAAATATCCCGATCATCTCGATCGAAGACGGACTTGACGAAAATGA
CTGGGAAGGCCATAAACTGCTTACTGACGCCGCCTTGGCACAAAGTGCAGCTCCGTCCGGTGACCGAC
TTTGTTTTGTAAACGAACCACCCAAAAAACTGGCCCGAAGGCATTGAAAAAAGGCGTCCGGCAACCTC
GAATTTTTAATTAAAGTGAACCCAAA

4. **Blot 3 sequence, 640 nucleotides (forward)**

TTTTTATAACACGAATTCCCGGGGATCCACGCGTCTTAAGGCGGCGCGGTACCGGGCCCGCTCGACTC
TAGAGCTCTGCACGGCATGCGATATCCCATGGAAGCTTCGTTCTTTAAATCCGTCTACAAGAATAACA
GGAAGCTGATGATATATCGCAACCTCTTCCACCGTTTCGATCGCCCGCGCATACGGACTGGAAATGAC
GACACTCGATTTGTTCTTTGTTCAAGATTTGCGCGACGCGCTTAGCATCTTGTTTTCCCGCTCCGATAAC
GGCCGGTGCCGTTCATCAGGCGTAAAAATCGAGTGGGCGTGCCTCACCAAATATCGTTTGTGTTCATCT
TGCCTCATTCCTTTTTAACCAGTCTAACGTAAAATCGACCAACCTCCCGTTGTTTCATCAGGCGGCATT
CGGCATTGCCCAACATAGCCGATGTGTCTTTTCCAAAGTGCTTGCCAATTTGCACAAAACCGCTCTTCA
TCCGGTTTCCGACAATCGATAAATGTCCTTCCCACACCCGCCGGCATTTTCCATTATGGCAGCCTCCTT
GCCAAAAATGTTTTCTTTGCCTAAACACGCGGATTTCCCCGACTCCATCGAAAATCCGCCGCCCCAAA
ACGCGATGATTTGCCGCAATTC

APPENDIX IV: Nucleotide sequence of PCR clones

1. 9I1-amplicon, 679 nucleotides (forward/reverse consensus)

TCCTTGTGGAAGCAGCCTCCGTTCCAAGCGAATCCCCTCGAGCGGGGGCTCGTTCGGCGCGAGCGGAT
GGGTCCGCCGCGTTTCTTTATAGCGCGCGACCAGCGTCGAATCTTGCGCATCCAAAAACAAAATTTGC
GGAACAATCGCCGCTTGTTCAGACAGCTCATCGAGCGCCGCAAACAAATGGTCGAAAAAATCGCGGC
TGCGCAAGTCCATCACAAGAGCGACTTtGTtCATTttGTtCCCcGATtCcTTCaCAaGCTCcAAAAaCTTtGgCA
AtAGCGTCGgCGgCAAATtGTCAaCGCAAAAAAaCCGAGgTCTtCAAaGCTTtGAaTGgCCACCGTTTtCCCcG
CCCcGGACATGCcGGTGATAaTGACGAGCTGAATCGGCGGCGCTCCCCcGTTTtGCCCCATCTCCGCACC
TCcTCTCcGTCAGCTCGGGTCAaGCCGATAGGAAAGCAACTCAAAATCTTCCGTATACACAAACGTTCC
ATACAATATCCCATCGCCTTTGACCGCATACTCGACAATATGATAATCGCCCGGAGCCATCGGAAGCG
ACGGGAGCGCTTCGAGTTCATGCCAAGCGAGCGTTCCTTCTTCGGATGAGGCGACGTTTTCCCCGATG
AACTCATCGGCAAAAAACGTAAACATCATCCACTCTGACACCGTCTCGCTTC

2. 9II1-amplicon, 601 nucleotides (reverse)

CACTAGTGATTGCAGCAGAAAGAAGGAAGCACGAAAGGAAAAGTGATTTTGGCGACGGTGAAGGGCG
ATGTGCATGACATGGCAAAAACTTGGTCGACATTATTTTAAGCAACAACGGCTATGAAGTGATCGACC
TTGGCATCAAAGTCGCCCCCCAGCAGCTCATCGAAGCGGTGCGCGAGCATAAACCGGACATCATCGGC
CTGTCGGGCCTGCTCGTCAAATCGGCGCAACAAATGGTCGTCACCGCCCAAGACTTGCGGCAAGCCGG
TGTTTCGACCCCGATTTTAGTCGGCGGCGCCGCCTTGACGCGTAAGTTTACGGAAAACAAAATCGCGC
CCGAATACGACGGCATCGTCCTGTACGCGAAGGACGCCATGGATGGGCTCGCCCTTGCCAACCAAATT
CGCCAGCGTAATCGAATTCCCGCGGCCGCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGC
CCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCG
TTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCG

3. 9II2-Amplicon, 677 nucleotides (reverse)

GGAACGTTCGCCGACCATCGATGAAATCGCCGACCATCTAGGCTTATCAGCGGAGGAGACGATTGAA
ATTTTGGCCGGCCGCGACCATTACCAGCCTGTTTCGATCGACGCCCCCGTGCAAGACAGCGAAAAGGA
CGGAACGACCATCGGAGAATTTATCGCTGATCAGACAAATGAGGTTGAAGCGCTGATCGAACGTTTGG
ATTTGCAGCAGGCGATCGACAAGCTGAGCGAGCAGGAACGGCTCGTCATTGACGCCGTTTTCCGCCGC
GGAGAAACGCAGCGCTCGCTCGCCGAACGGCTCGGCGTCTCACAAATGACGATCAGCCGCATCCAAA
AGCGAGCGATCGACAAACTGAAACGGCAGCTTGCCGCCCATCCGTCCTGATCTTCACACCGTCAAGCC
AAACGAAGGAAGAAAGAGGGATCATGCTGTTTGGCGAAGAAAAGACGGCGGCTGCGCGTTGCTCCGC
ATTCGAACCATCGCTCGCCGAACCAAAAACGGCGCCGCCGAGTGCGGACAGGCACTAACGGCGGCGC
AGCAGCAGCAATGTCTGATCATCCTTTTGTTCGTATCGGTGATATGCCCGAATTTGCTCATATAGCTCG
TGAACGATTTGCGCAGCGGGCTGATCAAGATTGACGCGGGCGAGCATCGAAGCAAGAACGGAAAACC
G

4. 9II3-Amplicon, 808 nucleotides (forward)

TCGACTNCTATAGGGCGAATTGGGCCNCGACGTCGCATGCTCCCGGCCGCCATGGCTTGGCCcGCGGG
AATTCGATTGCAGCAGAAAGAAGGAAGCACCGTCCTTCCGGGGGAAGAGGCGTTTCGCTTGTATGATA
CGTACGGCTTCCCGCTCGAGCTGACGGAAGAGTACGCCGCCGAGGCAGGAATGACGGTCGATCACGC
CGGTTTTGAGCGCGAGATGGAGCGCCAGCGCGAACGGGCCCGCGCCGCTCGCCAAGACGTCGATTCG
ATGCAAGTGCAAGGCGGCGTGCTCGGCGATTTGAAAGACGAAAGCCGGTTTGTCGGCTACGATGAGCT
TGTTGCCTCGTCGACGGTCCTCGCCATCGTTAAAGACGGACGGCTTGTCGAGGAAGTGAAAGCCGGTG
AAGAGGCGCAAATCATCGTCGACGTCACGCCGTTTTACGCCGAAAGCGGCGGTCAAGTCGCCGACCA

AGGCGTATTTGAAAGCGAGACCGGGAGAGCGGTCGTCAAAGATGTGCAAAAAGCGCCGAACGGCCAG
CACCTTCACTCGATTGTCGTCGAGCGCGGCGCGGCGAAAAAAGGAGCCCGCTATACGGCGCGCGTTGA
TGAGGCGAAGCGGGCGCGCATCGTGAAAAACCATACGGCGACCCATTTGCTTCATCAAGCGTTAAAA
GACGTGCTCGGCCGCCATGTCACcAAGCGGGGTCGCTCGTCGCTCCGGATCGGCTGCGCTTTGACTTCA
CCCATTTGGCCAAGTGAAGCCGGACGAGCTCGAGCGCATCGAGGCGATCGTCATGACAATTTGG

5. **Consensus sequence of PRC amplicon 9II4 (ca. 1.2 kb in size)**

AAGCAGCACGATTGATGATTGGGAAGCGGTGCATGAGGAATATTATTTATTCCCATCCGATGAAGCGG
AAGAAAACCACCGCGACGATTGGCTAAAATTCAACAACTTTAAATCGGACGTGGAAGATTTCTTTCCG
AACTACACTGGCATTGTGGGTCGAGGATTGTACGTCAATGACCAACTTCAGAAGCTGACGATCAACAT
AACGATGCCATTTTACGGCAAGGCGGAAGTGATTGGCTTCACCCAATATGTGACCGGTTTGGTGATGG
AAAAATTTCCGGATTACATCACCGTCAACGTCTACATCTCTTCAGCCGGCCAACCCGAAAGCCTCATC
GTGCGCCAAGCGCAAGCAGATGAGCCGTTTGTGCATATTTATCAATAAGTTAGCAGGCCACAACCGTG
GGGCAGCATGAGGGTGTCCCAAAAGGACCGGGATGCCCTCTCTCAACGCTATATATAGACACTAACCA
TCGCCTCCATACTGCTGTGTCTTTCTTGAGGGGAGATCGTCTTTTGGGTCATCCCCATGCTTTGTTGGTG
AGGAACCATTGCCAAGCGAAGCACcGTCCTGCTCTGACTTGCATCAGCGATGGgATGGTTTGTtGATAG
AACGCGTTGGAAAAGGTGGTGAAAAACCGCTGTTTGCCTTGgATCGGTGATGGAACGCGTGAACATAG
AGGCTGATGCAGCAAGCTTTTtCTACTTGAGAAaCGATCCTAaGAaGTGGTGGCGCGAGTAAATCACACC
CCTCCTAGAGGGTTGGTGAAAAACGGTTGTTTCTTTTGAGCAATGGGGGAGCGTTTGAAAGACGATCC
CGATGCGACAAGGCTTTTCTATTTGAGAAACAGGCTTGATAAGCGACTTGCTCCCTTAAATCACCGGC
CTCCTAGAGGGTTGGTGACAAACGGCTGCCAAGAGCAAATCTGCGGCATTTCGCCAAAAAAGAAGGC
TGATTTCACCCGGCTTTTCTAACGGAGAAGCAGATGGAATCGCAGGATGTTGTACGAAATCACCGCTC
CTCTAGTGAAAGAATCGGCACAGGGGCTCGGAGCGTTTTTCCCCGCTGCGGCATTGTGCGCATTTACG
GCGGCGCGCTCGCTGTTTCAGCACTGACAAAAGAGGTGGGGCTGCCGCGCATATCAGGCTGGCCTTTC
ATTCACACGCCGCTGGCCAAATGGACGGTGCCATAGCGATTCGTCAAGCCGGTGCTTCCTTCTTTCTGC
TGC

6. **Consensus sequence of PCR amplicon 9II5 (ca. 1.35 kb in size)**

GCAGCAGAAAGAAGGAAGCAAGCTACACCGTCGAACACGCCATGGACACAACCGAAATCAAAGGCA
CGCCCAAACGGGTCGTTGTATTGACAAACGAAGGAACAGAAGCCCTGCTTGCGCTAGGTGTGAAACC
GGTCGGCGCCGTCAATCGTGGACGGGCGATCCGTGGTATGACCACATTAAAGACAAAATGGACGGCG
TCAAAGAGCTCGGGTTGGAATCGGAGCCGAACGTTGAAGCGATCGCTGCTTTAAAACCGGACTTGATC
ATCGGCAACAAAATGCGCCATGAAAAAATTTATGAACAGTTGAAACAAATCGCTCCAACCGTTTTCGC
TGAAACGCTGCGTGGCAACTGGAAAGACAACTTTACGCTCTATGCGAAAGCGGTGAATCAAGAAGAG
AAAGGAAAACAAGTCATTGCCGAATACGATCAGCACATTGAAGACTTAAAAGCGAAACTCGGCGACA
AGCTGAAAATGAAAGTGTCGGTCGTCCGCTTCATGGCTGGTGACGTCCGCATCTACCATAAAGACTCG
TTCTCCGGCGTCATCTTGGACCAGCTCGGCTTCGCCCGTCCGGAATCGCAAACGTCAACGACTTCGCGG
AACCGGCGTGACGAAGAACGCATCCGGCCATGGACGGCGACATCCTGTTCTCTTTACGTATGAACAGG
CGACGGCAAgcAAgCttgAAAcgTCGcgAAAcAaGgcAAtGGaTgTACgAtAAGCGAGCGACACGATTTGGAAC
ACAGCTGGCGGTGTGCTCGCTGCTCATCTGATGCTGGACGATATCGAGAAATATTTCTTGCAGGAACA
ATAATCTTACGTAAAGGCTGTGCGAAAGATGGCATACCGAATCTTCTGCACAGCCTTTGTCTTTGTGAT
CTTTTTCCGTTTCCACTCCCCTACCCAAAAAAGCAAGTCCATTGCCGTTTAACGCAGATTGTGATACAT
TCAAACTATGAAATGAACGAAAGAAAGGGAGAAGACGATTGAACCATCCAGCACGGATCGCAGCATG
CCTGTTGTCGTTCAGCATTTTTTTGTCAGGCTGCACTGGTGGAAACAACCAAGAACAACACCACTCGCA
TATGACGAAAACCGCCACTGGCGATATTCGCGAGACAACGAGATCGATCGAGCATTTGCCAAGCTTTT
TAGGCGCATTCGAAGAGGAAATGGCTGTGTTGTACCAACAGGCCGCGGAACACCGCGAACTGCTTGA
AAACATCCCATGCTATTGCGGCTGCGGACAATCGGCCGGCCATAAAAACAACTATGACTGCTTTGTCT
ATGAAAATAAAAAAGACGGATCGATTGTCTGGGATGATCATGCAACCAAATTCGCCAGCGT