

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 (K_m , 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* DH5 α 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 μ g/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* MS941 was 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 KH₂PO₄, 3 ml of 10% (w/v) yeast extract, 10 ml of 20% (w/v) glucose, 0.1 ml of 2 mg/ml MnSO₄ and Prot medium to a final volume of 1 litre. Prot medium (pH 7.5) contained (g/l) NH₄Cl, 1; Tris-HCl, 12; KCl, 0.035; NaCl, 0.058; MgSO₄·7H₂O, 0.246; MgCl₂·5H₂O, 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 μ g/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:		
<i>Geobacillus</i> strain PA-9		Hawumba <i>et al.</i> (2002)
<i>Escherichia coli</i> DH5 α	<i>hsdR recA lacZYA</i> ϕ 80 <i>dlacZ</i> Δ M15	Stratagene
<i>Bacillus megaterium</i> MS941		F. Meinhardt and K-D. Wittchen ^a
Plasmids:		
pGEM [®] -T Easy	Cloning vector for PCR products, ColE1, Amp ^r , LacZ α peptide	Promega
pSVBI	<i>E. coli/B. megaterium</i> shuttle vector, Tet ^r , Amp ^r	F. Meinhardt and K-D. Wittchen ^a
Primers:		
FRD1	5'-TGTC AACAGTGGCATTGC-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'-GGCGATTAAAGTTGGGTAACG-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 imidazole-ZnSO₄ 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 ZnSO₄ 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 μ l of UHQ H₂O. 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* MS941 protoplasts

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 OD₅₄₆ 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 MgCl₂, 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 *Eco*RI 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 IE3_{II} 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

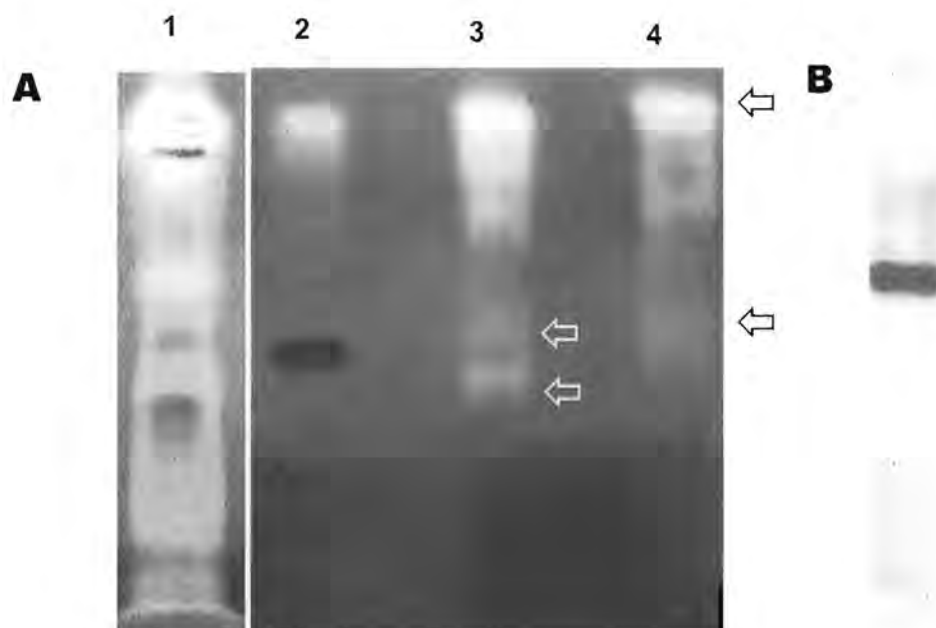


Fig. 4.1 Zymogram analysis of proteases purified from the culture supernatant of *Geobacillus* sp. PA-9 by ion-exchange chromatography. Lane 1, cell-free culture supernatant from a *Geobacillus* sp. PA-9 culture; lane 2, sample of fraction IE2II; lane 3, sample of fraction IE3II; lane 4, sample of fraction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

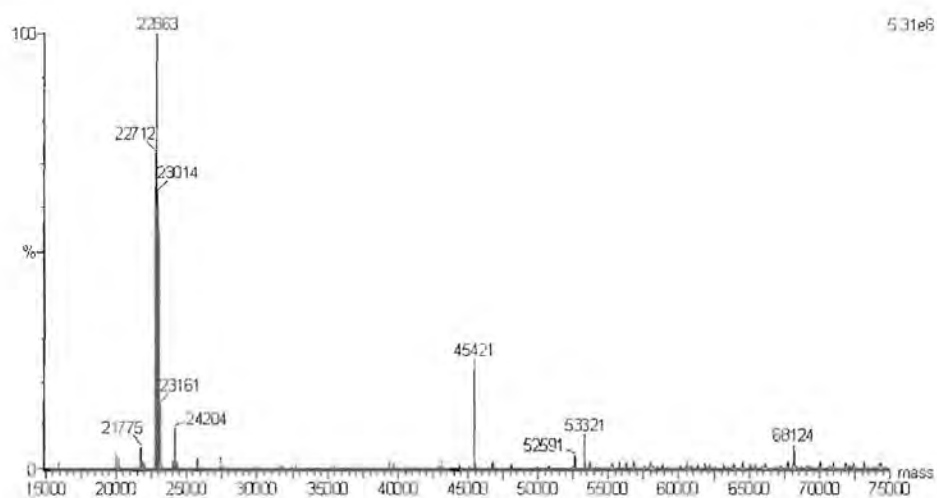


Fig. 4.2 Analytical spectrum of fraction IE2II following analysis by electrospray mass spectrophotometry. The protein sample was diluted 1:1 in 50% acetonitrile/water containing formic acid and 5 μ l of the diluted protein sample was injected into the mass spectrometer.

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* DH5 α

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 than 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* DH5 α 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.

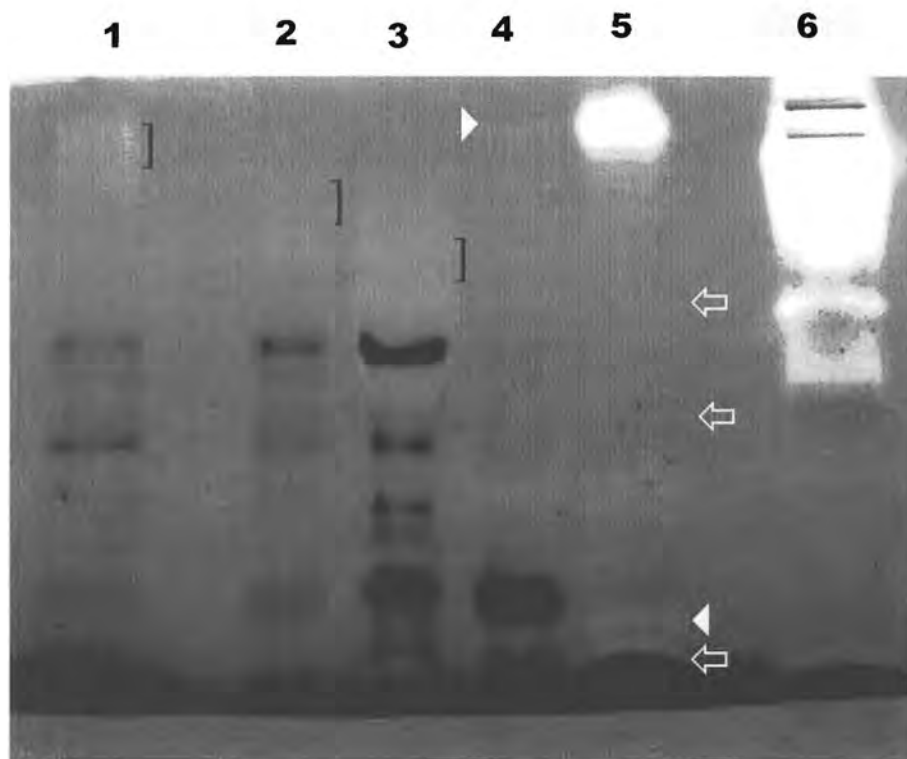


Fig. 4.3 Zymogram analysis of proteins purified from negative-stained SDS-PAGE gels. Five different protein bands (lanes 1 to 5), corresponding to caseinolytic proteases observed in a crude enzyme solution (lane 6) prepared from the supernatant of a *Geobacillus* sp. PA-9 culture, were eluted as single 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 (Dalbóge, 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 II₂ (obtained from primer set FRD1/FRD2) and II₃ (obtained from primer set RD1/RD2), respectively, were hybridised at 42°C to colony blots prepared from the transformation plates.

Under these conditions, probe II₂ exhibited non-specific hybridisation and was excluded from further use. Of the 1 244 clones screened, fifteen hybridised to probe II₃. 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 9I₁, 9I₂, and 9I₃, while those obtained by making use of primers RD1 and RD2 were designated 9II₁ through 9II₉ (Fig. 4.4). Six of the amplicons (9I₁, 9II₁, 9II₂, 9II₃, 9II₄ and 9II₅) 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 (9II₂) involved in the transcription of proteins, alanyl-tRNA synthetase (9II₃) involved in translation of proteins, hypothetical protein (9II₄) of unknown function, and ABC Fe³⁺ transporter/BP-2 periplasm binding protein (9II₅) involved in protein transport / export.

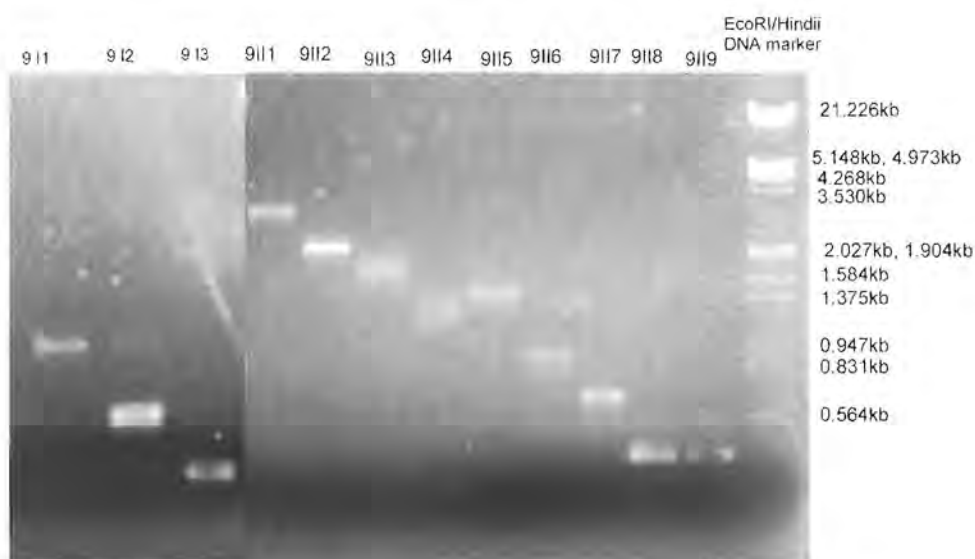


Fig. 4.4 Agarose gel electrophoretic analysis of purified amplicons. Genomic DNA of *Geobacillus* sp. PA-9 was used as template DNA in PCR reactions with either primers FRD1 and FRD2 to generate amplicons 9I₁ to 9I₃, or with primers RD1 and RD2 to generate amplicons 9II₁ through 9II₉. The sizes of the 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-9 are 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-Seuning, 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.
- Kuhn, 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., Altling, 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., Poltarau, 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 Maniatis, 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.

Uhlir, 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., Eijssink, 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-hydroxyphenylacetic 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 C₃ 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 C₃ 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).

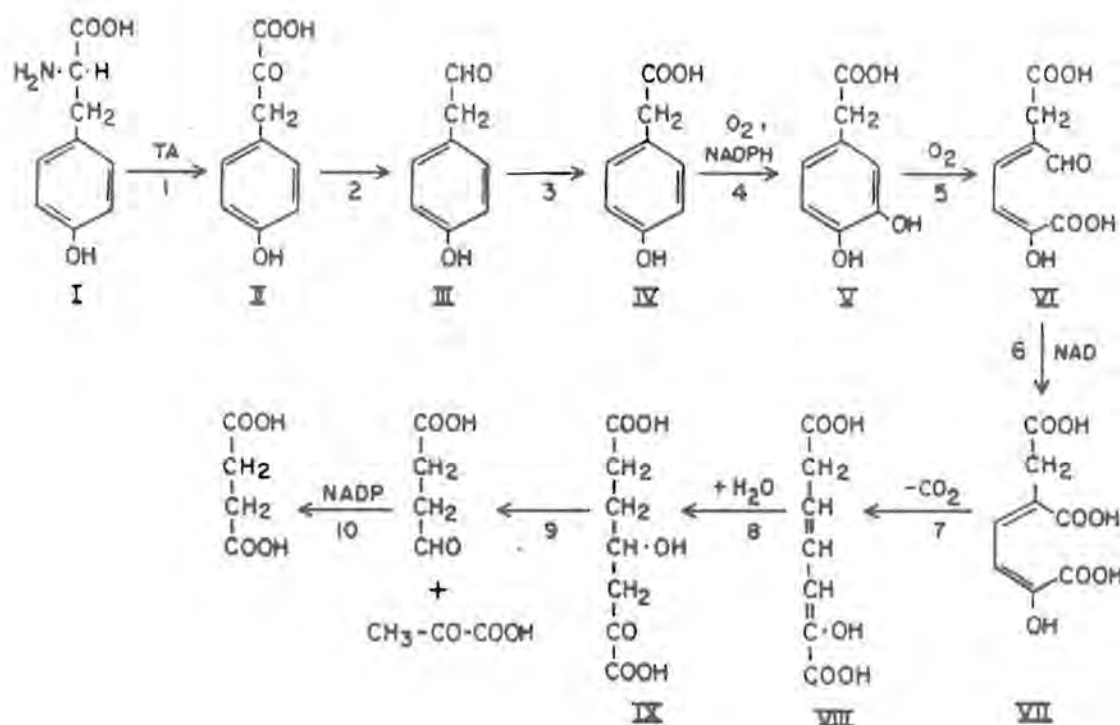


Fig. 5.1 Proposed steps in the degradation of L-tyrosine to pyruvic and succinic acids. Reactions 4 through 10 are those previously described for strains *Acinetobacter* and *Pseudomonas putida*. The key intermediates are as follows: 4-hydroxyphenylacetic acid (IV); 3,4-dihydroxyphenylacetic acid (V); 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 Biotechnology, 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 dendrogram 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:		
<i>Geobacillus</i> sp. PA-9		Hawumba <i>et al.</i> (2002)
<i>Escherichia coli</i> DH5 α	<i>hsdR recA lacZYA</i> ϕ 80 <i>dlacZ</i> Δ M15	Stratagene
<i>Escherichia coli</i> MC1061	<i>XylE</i> , Δ (<i>lac</i>)X74	W.H. van Zyl ^a
Plasmids:		
pMP220	<i>LacZ</i> ⁺ Tet ^r	V. Venturi ^b
pSVBI	<i>E. coli</i> / <i>B. megaterium</i> shuttle vector, Tet ^r , Amp ^r	F. Meinhardt and K-D. Wittchen ^c
pBluescript SK II (+)	Cloning vector, ColE1, Amp ^r , <i>LacZ</i> α peptide	Stratagene
pSVBI-R113	2.7-kb genomic DNA fragment of <i>Geobacillus</i> sp. PA-9 cloned into the <i>Hind</i> III site of pSVBI	This study, Chapter 4
pBluescript-R113	Insert from pSVBI-R113 cloned into the <i>Hind</i> 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 Biotechnology, 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
I13Fd	5'-GGTGACGGATGTAAAGAG-3'	This study
I13Rv	5'-GTATACAAGCGGACGAAG-3'	This study
PCR Amplification*:		
Fp-Promoter	5'- gaagatc TTCGGAATGTGACG-3'; <i>Bgl</i> II site incorporated	This study
Rp-Promoter	5'- gctctag GCCTTTCTGTTTGGC-3'; <i>Xba</i> 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 μ l) contained 1 $\times$ 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 thermocycler (Applied Biosystems, Hercules, CA, USA). Cycling was completed by a final elongation step at 72°C for 7 min. An aliquot (5 μ l) 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 *Eco*RI 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 μ g/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 μ g/ml) and 40 μ g/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- μ l volume: 387.5 μ l of 50 mM Tris-HCl buffer (pH 9), 50 μ l of 10 mM NADH, 2.5 μ l of 10 mM FAD, 250 μ l of 10 mM 4-HPA, and 10 μ l of 10 mM Fe^{2+} . The mixture was equilibrated at 50°C for 5 min after which 100 μ l of the cell extract was added, and the mixture incubated at 50°C for 30 min. The reaction was terminated by addition of 200 μ l 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* DH5 α 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-digoxigenase enzymes. The highest level of homology was observed with the putative 3,4-dihydroxyphenylacetate 2,3-digoxigenase of *Oceanobacillus iheyensis* HTE831 (GenBank Accession no. NP693790; 55% amino acid sequence identity over 54 amino acids) and of *Arthrobacter globiformis* (GenBank Accession 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-dihydroxyphenylacetate 2,3-digoxigenase.

To determine the presence of a promoter upstream of the start codon of *pheH*, a *lacZ* 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 *lacZ* 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 ATCCACGGGGAAAAAGTGGAGGATGTCACCGTTCATCCGCGTTCCGCAACGTCGTCCGC
I H G E K V E D V T V H P A F R N V V R 43

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

481 CCGACAACCGGGCAGCCGGTCCGGATGACGTTTCATTTCAGCCAACTACGATTGACGAGCTC
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 TCGCCTGATTATTTGAATGCCGAAGTCATGGCGATGGGCATCGCCAACGATTGTGTTGCC
S P D Y L N A E V M A M G I A N D L F A 123

661 GAAGACGATCCGATGTTTGCCGAAAATGCGAAAACTATTATGAATACGCGCGGGAACAC
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 GTCAGCGGCATCCGCCATTGCGGACGCAAGGCGGCATCACCGATGAAATTTTAGTGTTT
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 AATGTGTTTGTGCCGTGGGAGCGCGTGTGTTGTGTGCGGCAATTCATCGATTGCAACCGG
N V F V P W E R V F V C G N S S I C N R 283

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

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

1261 GAGTTCAGCATGTGAAAGACAAAGGGGCGGAAATCATGCTCGTCCTCGAGACGATGAAA
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 CAGCACGAGGAAATTGGCGACATCGTTTCGCCGGGCGATGCAAGGGGCGACGGTGGACGGC
    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 CGGCAGACGCATTACGAATATTACTTTTTTCGCCGACCCCGTTCGGATGGGGATGGCGTAT
    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 AAAAGCGTGTTTCATTCCGGCTGACAACAAGCATTTGACCGTTCATTCGCACGTTTCGCAAC
    K S V F I P A D N K H * 494

1801 ACGGATCGGGGAACAGCGGAGGGTGCGCGGCTGCGGCGTTTCCTAAGAGGCTGGTGATTT
1861 CGTGCAAAATATGACTGATTCCATCTGTTTCTCAATTAGAGAAACCTTGAGAATCAGCT
1921 TTCTTTTCTGAATGAAAATACCGCCGATTTCGTGTGTGACAGTCGTTTGTACCGCCCTTC
1981 TAAGGTTGCCGCCGCTCGAAAACAGGCGTTTCAGTGGCGCGCAAGCCGAGCGGATTTGGGC
                                pheH2
2041 GTGCTGCGAGATGTTTGAAGGAAAGATGTGGCCAGCAACAAGGAACGTTTTGTTCCAT
                                M W P A T R N V L F H 11

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

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

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

2281 TGACGGAAGCGGATGAAGATCATCTGTATTTGCGGGGACTTGAAGAATATCATCATCATA
    *

2341 GCCTTCTCTTGAAAAAGGCGCCCCGCCCTGCGGTGCAAGCGCTTAGCTACAAAGTCGGTT
                                pheC
2401 CCGAGCAGGAATTGGAAGCGCTTTATGAATGGTTTGCGGCGCAAACGCTGAGCCGAAATG
                                M 1

2461 GCTCGAAGCGGGAAGCCAGCGCGCCGTGCGCAAGGCGTTTCGCGTGAAGATCCATCCGGA
    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), *Deinococcus 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 *pheA1*, 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 dendrogram 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).

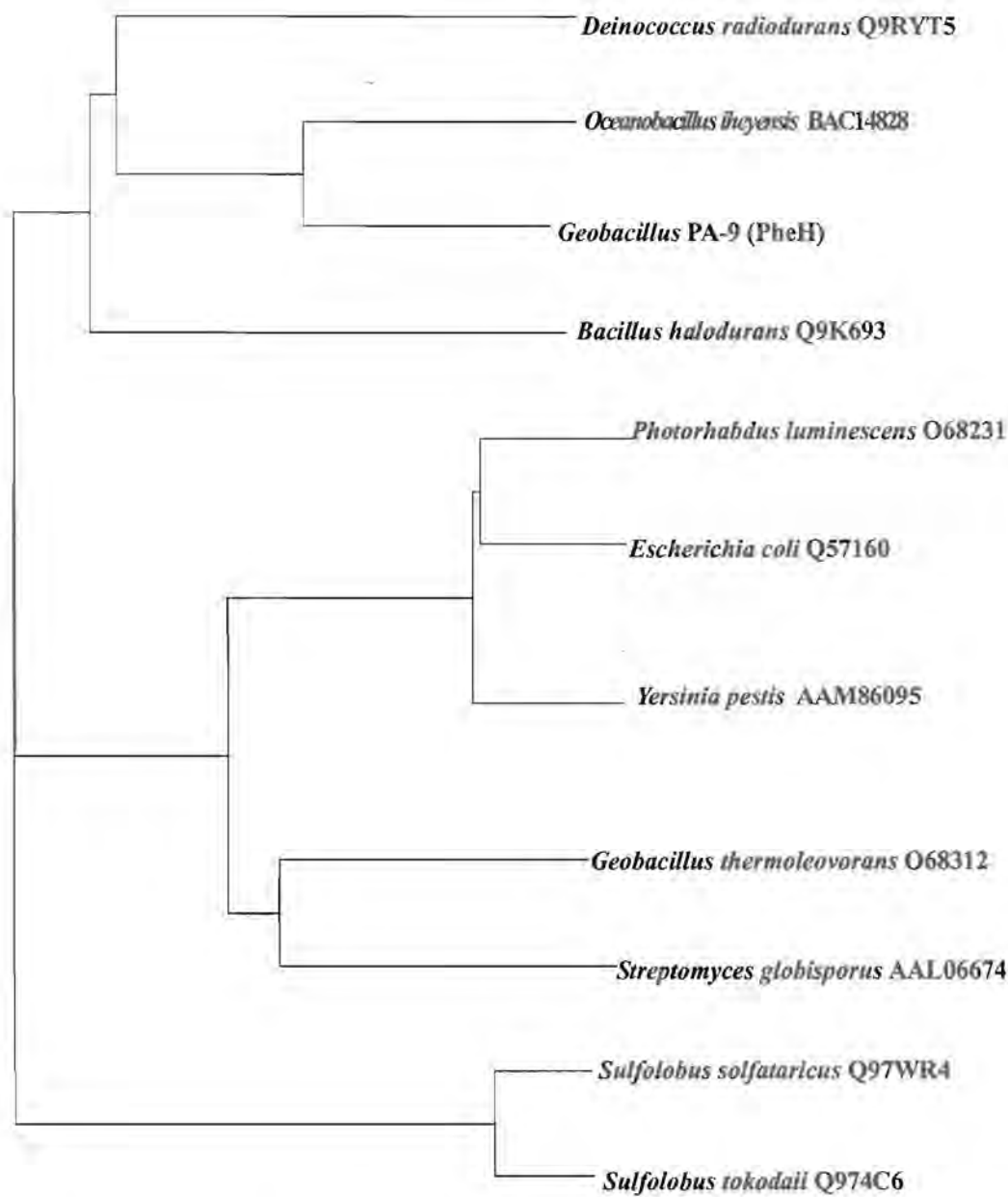


Fig. 5.3 Dendrogram resulting from the multiple amino acid sequence alignment of different 4-hydroxyphenylacetic acid 3-hydroxylases (4-HPA 3-hydroxylases) and phenol hydroxylases with PheH from *Geobacillus* PA-9. Q9RYT5, 4-HPA 3-hydroxylase of *Deinococcus radiodurans*; BAC14828, 4-HPA 3-hydroxylase 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-hydroxylase 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-dihydroxyphenylacetic acid was determined spectrophotometrically 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-hydroxyphenylacetic 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* DH5 α 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-dihydroxyphenylacetic acid was determined spectrophotometrically 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* DH5 α (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-hydroxylases 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.

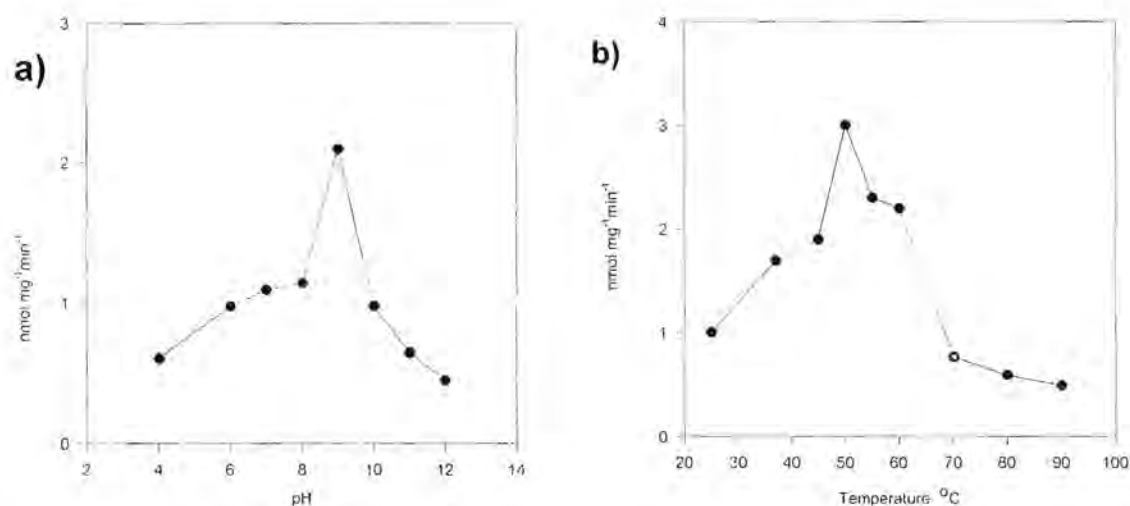


Fig. 5.4 Influence of pH (a) and temperature (b) on the specific activity of the 4-hydroxyphenylacetic acid 3-hydroxylase (4-HPA 3-hydroxylase) from *Geobacillus* isolate PA-9. For the pH profile, activity was measured at 50°C in buffers of different pH values. For the temperature profile, activity was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

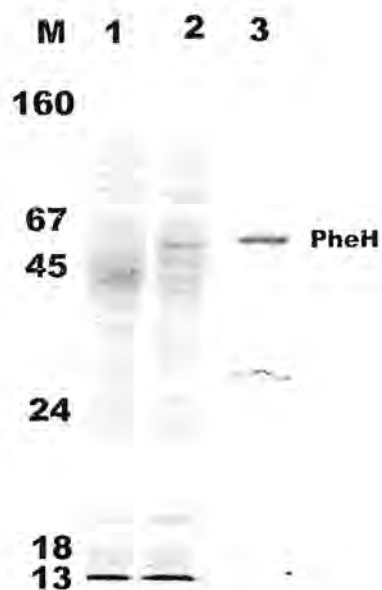


Fig. 5.5 SDS-PAGE analysis of the expression and purification of the 4-hydroxyphenylacetic acid 3-hydroxylase (4-HPA 3-hydroxylase) from *Geobacillus* isolate PA-9. Lane 1, cell extract sample from non-recombinant *E. coli* DH5α; lane 2, cell extract sample from *E. coli*/pSVBI-R113; lane 3, sample of 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 *pheA1* 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* (Gibello *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 thermophilic *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 *Pseudomonas 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 Maniatis, 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.
- Sparmins, V.L., Burbee, D.G. and Dagley, S. (1979). Catabolism of L-tyrosine in *Trichosporon cutaneum*. *Journal of Bacteriology* 138: 425-430.
- Sparmins, V.L., Chapman, P.J. and Dagley, S. (1974). Bacterial degradation of 4-hydroxyphenylacetic acid and homoprotocatechuic acid. *Journal of Bacteriology* 120: 159-167.
- Sparmins, 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 *Photobacterium*: 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 at 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 observed 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-hydroxylase (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.