Chapter 4.

Isolation and analysis of *prtB*, a putative *A. nidulans* aspartic protease gene.

4.1. Background.

The results of PAGE and inhibitor studies showed that the culture filtrate of *A. nidulans* exposed to nutrient-limiting conditions contains an intracellular serine protease, two extracellular serine proteases, and an extracellular metallo-protease (section 3.3., Cohen 1973). Prior to the commencement of this project, an extracellular serine protease gene, *prtA*, had been cloned from *A. nidulans* by members of our laboratory (Katz *et al.* 1994).

Biochemical studies have shown that many *Aspergillus* species are capable of producing a family of acid proteases known as aspergillopepsins (reviewed in Matsubara and Feder 1971). A gene, designated *pepA*, encoding an aspergillopepsin had been isolated from *A. niger* (Berka *et al.* 1990a and b). A gene, designated *pepB*, encoding a non-pepsin acid protease had also been isolated from *A. niger* (Inoue *et al.* 1991). It was not known if our inability to detect proteolytic activity corresponding to this class of enzymes was due to the unsuitability of the experimental methods, or was due to the fact that *A. nidulans* does not produce acid proteases (see section 3.4.).

4.2. Rationale and aims.

As a serine protease had already been cloned from *A. nidulans* (Katz *et al.* 1994), it was decided to try to clone some other extracellular protease structural genes. This would allow a comparison of regulatory regions to be conducted, with the intent of gaining insight into the mechanisms by which the extracellular proteases are regulated, and the identification of putative cis-acting regulatory regions. The serine protease gene, *prtA*, was cloned using a probe encoding the homologous gene of the *Aspergillus* species *A. oryzae* (Katz *et al.* 1994). As this had proved to be a successful approach, heterologous probing was chosen as the strategy for the cloning of additional extracellular protease structural genes.

The aspergillopepsins are a class of aspartic proteases which have been detected in many *Aspergillus* species (Matsubara and Feder 1971). As no proteolytic activity corresponding to that of an acid protease was observed in the culture filtrate of *A. nidulans* grown under carbon-, nitrogen-, or sulphur-nutrient-limiting conditions, we decided to use a molecular biology approach to determine if *A. nidulans* possessed an acid protease gene.

4.3. Results.

4.3.1. Heterologous probing of *A. nidulans* genomic DNA.

The probe used to clone an aspergillopepsin gene from *A. nidulans* was a 1064 bp PCR-generated fragment from the *pepA* gene of *A. niger* (Berka *et al.* 1990). The conditions used to screen the library were determined empirically by probing and washing Southern blots which contained genomic DNA from both *A. nidulans* and the *A. niger*. A variety of stringency levels for hybridisation and washing were tested. The hybridisation conditions which were tested included hybridisation solutions as described in the DIG detection protocol (Boehringer Mannheim), containing 50% formamide and 30% formamide. Hybridisations in 50% formamide were carried out at 37°C. Hybridisations using the 30% formamide solution were carried out at 37°C and 22°C. Washing solutions tested include: 5x SSC 0.1% SDS, and 2x SSC 0.1% SDS. Both solutions were tested at 65°C and 37°C for 2x 15 minute washes.

When Southern blots were probed with the *A. niger pepA* gene, a faintly hybridising band was detected on nylon membrane (Hybond-N) after fixing DNA by both baking and exposure to UV light. This binding was observed using hybridisation solution containing 50% formamide, and washing solution of 2x SSC 0.1% SDS at 37°C. These conditions were used to screen the λ library for an *A. nidulans* homologue of the *A. niger pepA* gene.

4.3.2. Screening of the *A. nidulans* genomic DNA λ library.

Approximately 10000 λ clones from a λGEM11 library (Katz *et al.* 1994) containing fragments of *A. nidulans* genomic DNA were screened with a heterologous probe, containing part of the coding region of the *A. niger*, aspergillopepsin gene, *pepA*. One clone, designated λPVK2, which hybridised to the *A. niger pepA* probe, was obtained.

4.3.3. Restriction mapping of λPVK2.

The positive clone identified in the library screen was characterised by restriction enzyme analysis, and a restriction map for this clone was constructed (fig. 4.1.). Southern blot analysis was used to localise the region hybridising to the *A. niger* probe to a single 6.5 kb *Xba*I fragment of λPVK2 (fig. 4.2.).

A Southern blot, of *Xba*I digested λPVK2 DNA and *A. nidulans* genomic DNA, was hybridised with the *A. niger pepA* probe. It was noted that the hybridising fragment in the λPVK2 lane was smaller than the hybridising fragment in the *A. nidulans* genomic lane (fig. 4.3.). As the restriction map did not place the λPVK2 hybridising *Xba*I fragment on the edge of the insert, this result suggested that a deletion or rearrangement of some sort had occured.

Other restriction enzymes were used in an effort to determine if this deletion or rearrangement had occured in the region of interest. The double restriction endonuclease digests, *Hind*III / *Bam*HI, *Bam*HI / *Pst*I, and *Pst*I / *Eco*RV, which cut within the hydridising region showed that all the hybridising fragments from λPVK2 were smaller than the corresponding *A. nidulans* genomic DNA fragment (for example see fig. 4.4.). This led me to believe that the most likely explaination for the observed difference in fragment size was a difference in the cleanliness of the two

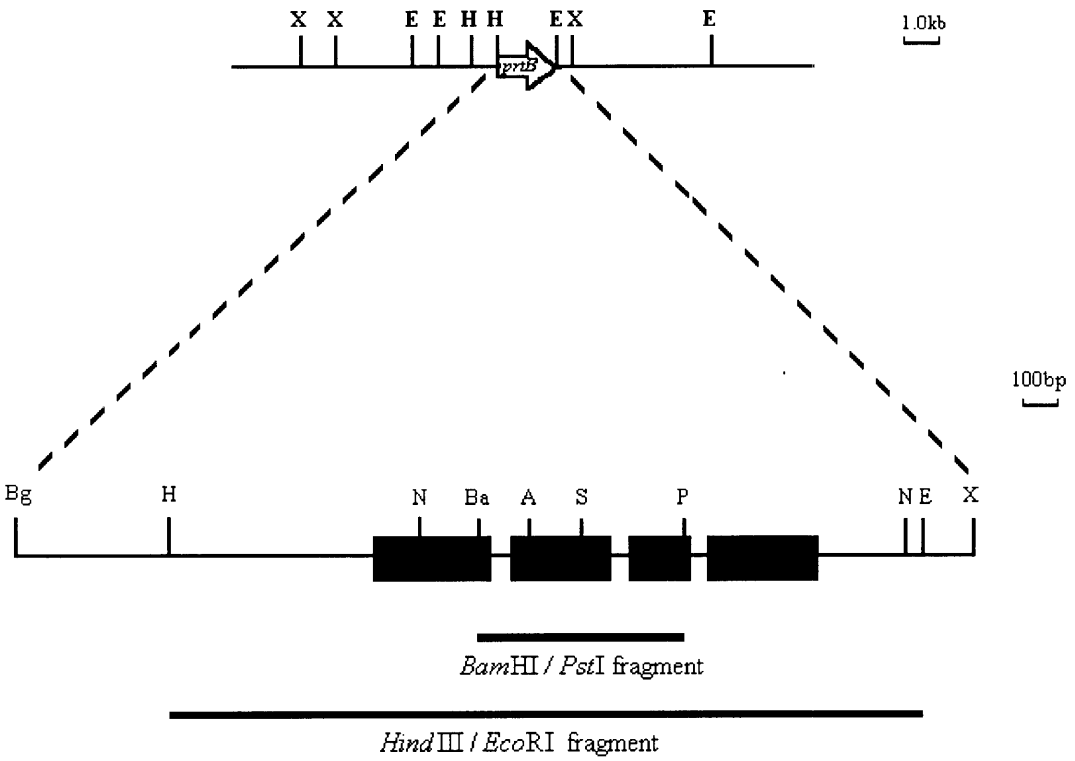


Figure 4.1. Partial restriction map of the 17kb *A. nidulans* genomic DNA insert of λPVK2 and the 2.75kb *Bgl*II/*Xba*I fragment containing the aspartic protease gene, *prtB*. The *Bgl*II/*Xba*I fragment encompasses the region of the clone which was sequenced. Solid blocks in the *Bgl*II/*Xba*I fragment represent the four exons which comprise the *prtB* coding region. The thick black lines labelled *Bam*HI/*Pst*I and *Hind*III/*Eco*RI show the fragments used to probe the blots shown in figures 4.4., 4.5., and 4.6. respectively. Restriction enzymes are represented by A = *Acc*I, Ba = *Bam*HI, Bg = *Bgl*II, E = *Eco*RI, H = *Hind*III, N = *Nsi*I, P = *Pst*I, S = *Sma*I, and X = *Xba*I.

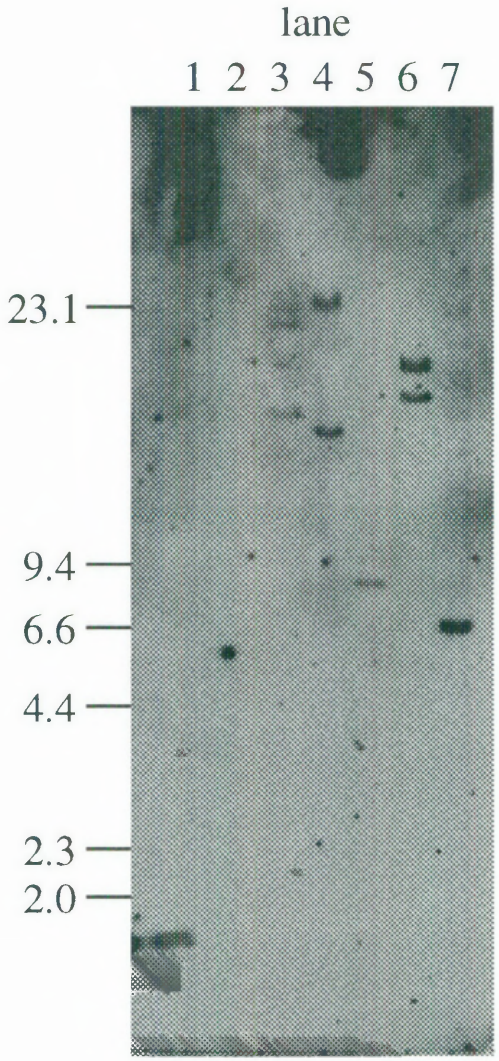


Figure 4.2. Southern blot showing that the *A. niger pepA* probe hybridises, under low stringency conditions, to single fragments of the λPVK2 clone. The lanes contain the following: λ *Hind*III standards (lane 1), and λPVK2 digested with *Asp*718 (lane 3), *Bam*HI (lane 4), *Eco*RI (lane 5), *Nru*I (lane 6), and *Xba*I (lane7). This blot was probed with a PCR-generated fragment containing most of the *A. niger pepA* coding region. Lane 2 contained no DNA. The probe was labelled with DIG (Boehringer Manheim). Hybridisation was carried out at 37°C in hybridisation solution containing 50% formamide (Boehringer Manheim). This blot was washed twice for 15 minutes in 2x SSC, 1% SDS at room temperature.

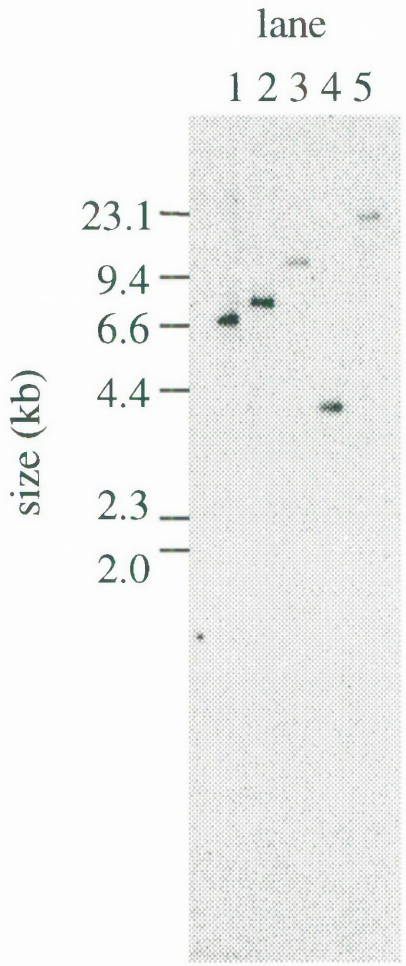


Figure 4.3. High stringency Southern blot of λPVK2 and *A. nidulans* genomic DNA. Note the hybridizing band from the λPVK2 *Xba*I digest (lane 1), differs in size from the hybridizing band in the genomic DNA digested with *Xba*I (lane 2). Lanes 3, 4, and 5 represent *A. nidulans* genomic DNA digested with *Bam*HI, *Eco*RI, and *Hind*III respectively. The *A. niger pepA* gene hybridised to a *Bam*I/*Pst*I fragment within the 6.6 kb XbaI fragment of λPVK2 (fig. 4.1.). This *Hind*III/*Eco*RI fragment of the λPVK2 clone was labelled with DIG (Boehringer Manheim), and used as a probe in this experiment.

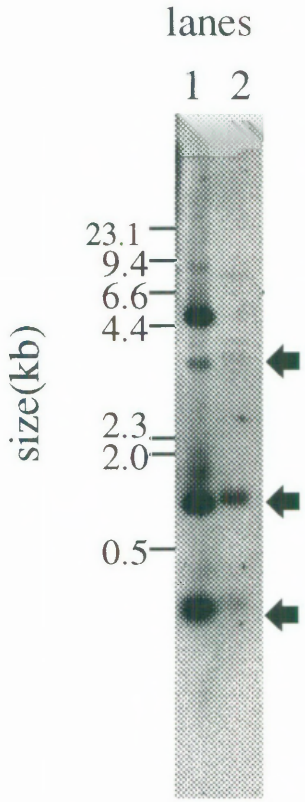


Figure 4.4. Southern blot demonstrating that the size difference between hybridising fragments from λPVK2 and *A. nidulans* genomic DNA is not specific to a single fragment, but is an artifact of the genomic DNA preparation. As there are two *Eco*RV sites within the hybridising region, *Hind*III/*Eco*RV double digests of both λPVK2 DNA (lane 1) and *A. nidulans* genomic DNA (lane 2) were expected to contain three hybridising bands. The extra faint bands indicate that the digests were not complete. Major (expected) bands are marked with arrows. It is of note that the λPVK2 bands appear proportionally smaller than their genomic DNA counterparts, suggesting that the size difference is a preparational artifact. This blot was probed with a DIG-labelled *Hind*III/*Eco*RI fragment containing most of the *prtB* coding region (figure 4.1.).

DNA preparations, *i.e.* the genomic DNA was contaminated by a factor which binds to the DNA retarding its mobility, resulting in all λPVK2 fragments appearing proportionally smaller than their genomic DNA counterpart. *A. nidulans* genomic DNA was prepared using a different method (Lee and Taylor, 1990), and the size of the hybridising genomic DNA fragment was shown to be identical in size to that of λ PVK2 DNA digested with the same restriction enzyme (fig. 4.5.).

4.3.4. Copy number analysis.

Southern blots of *A. nidulans* genomic DNA were probed with the *Bam*HI/*Pst*I fragment of λPVK2 (fig. 4.1), which hybridised to the *A. niger pepA* gene, and washed under low stringency conditions. No additional hybridising bands were revealed, indicating that *prtB* is a single copy gene in *A. nidulans* (fig. 4.6.).

4.3.5. DNA sequence analysis of *prtB*.

Southern blot analysis (section 4.3.3.) of λPVK2 was employed to further localise the region of interest within the 6.5 kb *Xba*I fragment to which the *A. niger pepA* probe hybridised. This region was sequenced and a search of the GenBank DNA sequence database using the FASTA program (ANGIS) indicated that this region shows similarity to *Aspergillus* genes which encode aspartic proteases known as aspergillopepsins. We have designated this aspartic protease gene *prtB*. The *A. nidulans* aspartic protease gene was found to be more similar to the *A. fumigatus pepF* gene than the *A. niger* homologue, based on nucleotide identity which was 67.7% and 54.8% respectively. The *prtB* coding region is interrupted by three introns which are found in locations which correspond to the positions in which introns are found in the other aspergillopepsin genes (fig. 4.7.).

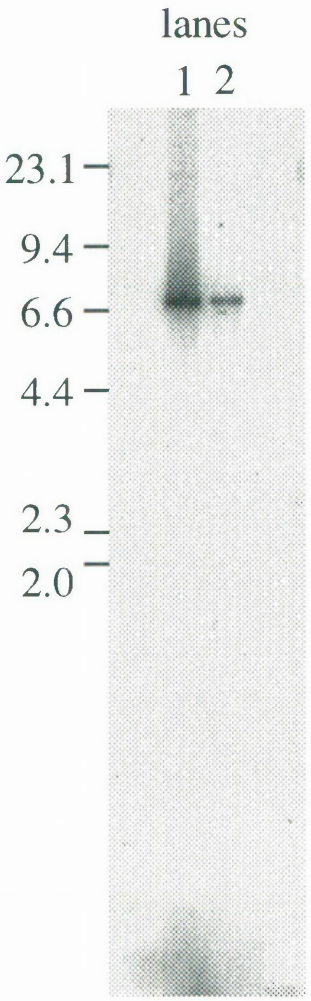


Figure 4.5. Southern blot showing that the λPVK2 fragment is identical in size to the *A. nidulans* genomic DNA fragment. The genomic DNA used in this experiment was obtained using the method of Lee and Taylor (1990). The DIG-labelled probe used in this experiment was a *Hind*III/*Eco*RI fragment containing the entire *prtB* coding region (fig. 4.1.). The samples are: λPVK2 (lane 1) and *A. nidulans* genomic DNA (lane 2). Both samples were digested with *Xba*I.

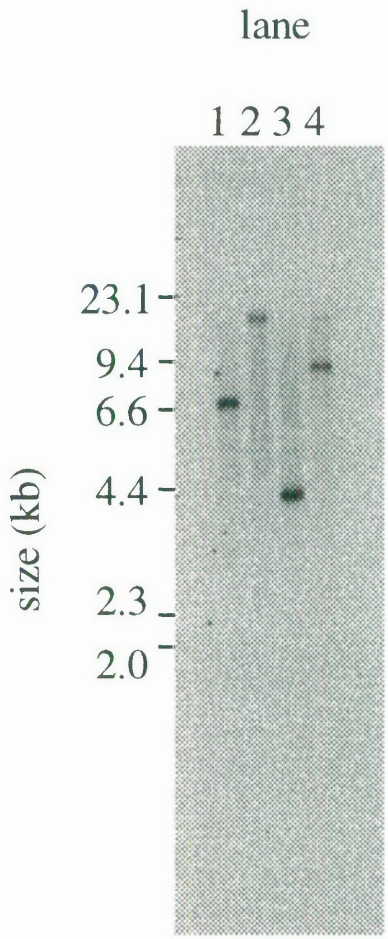
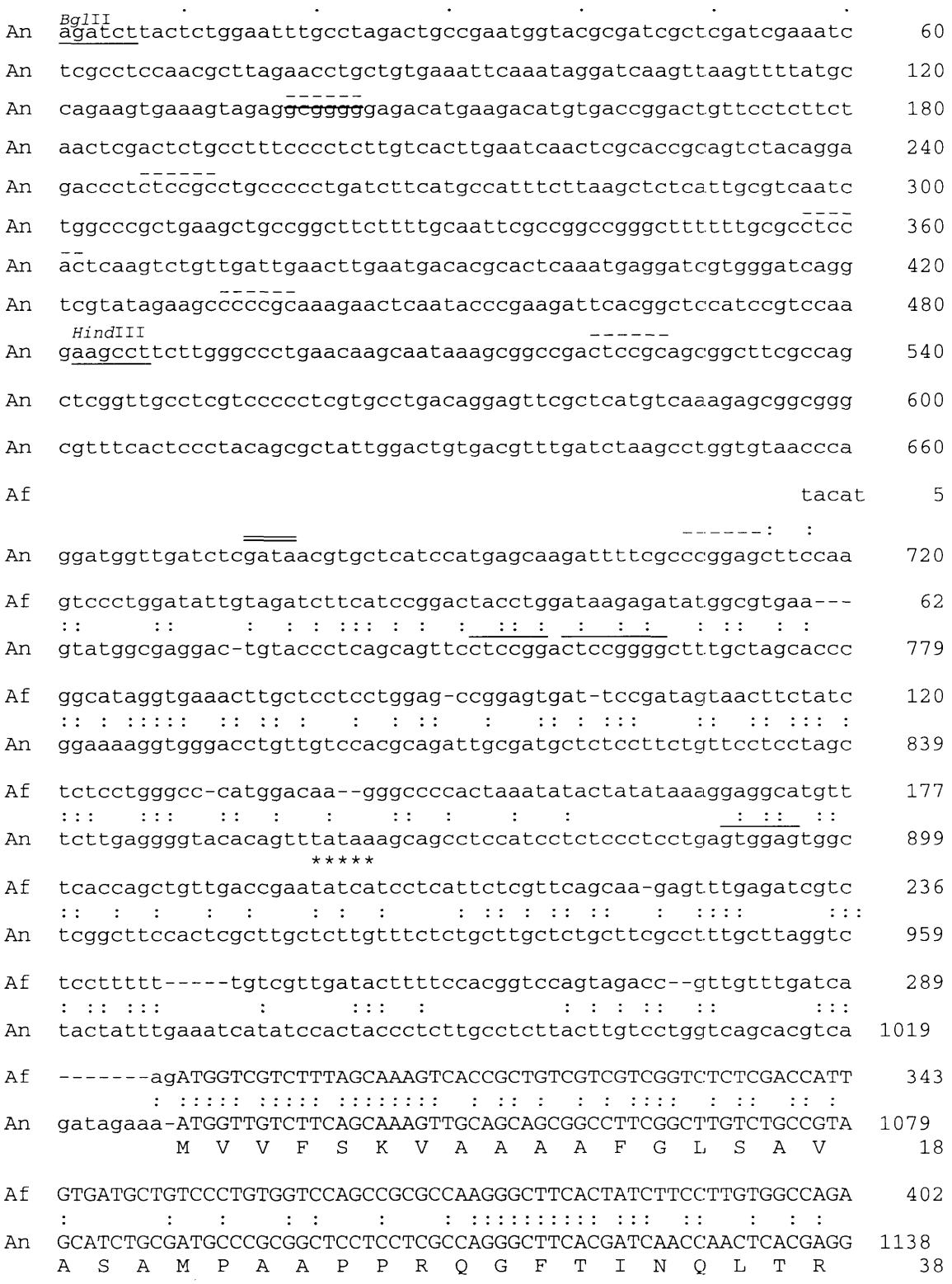
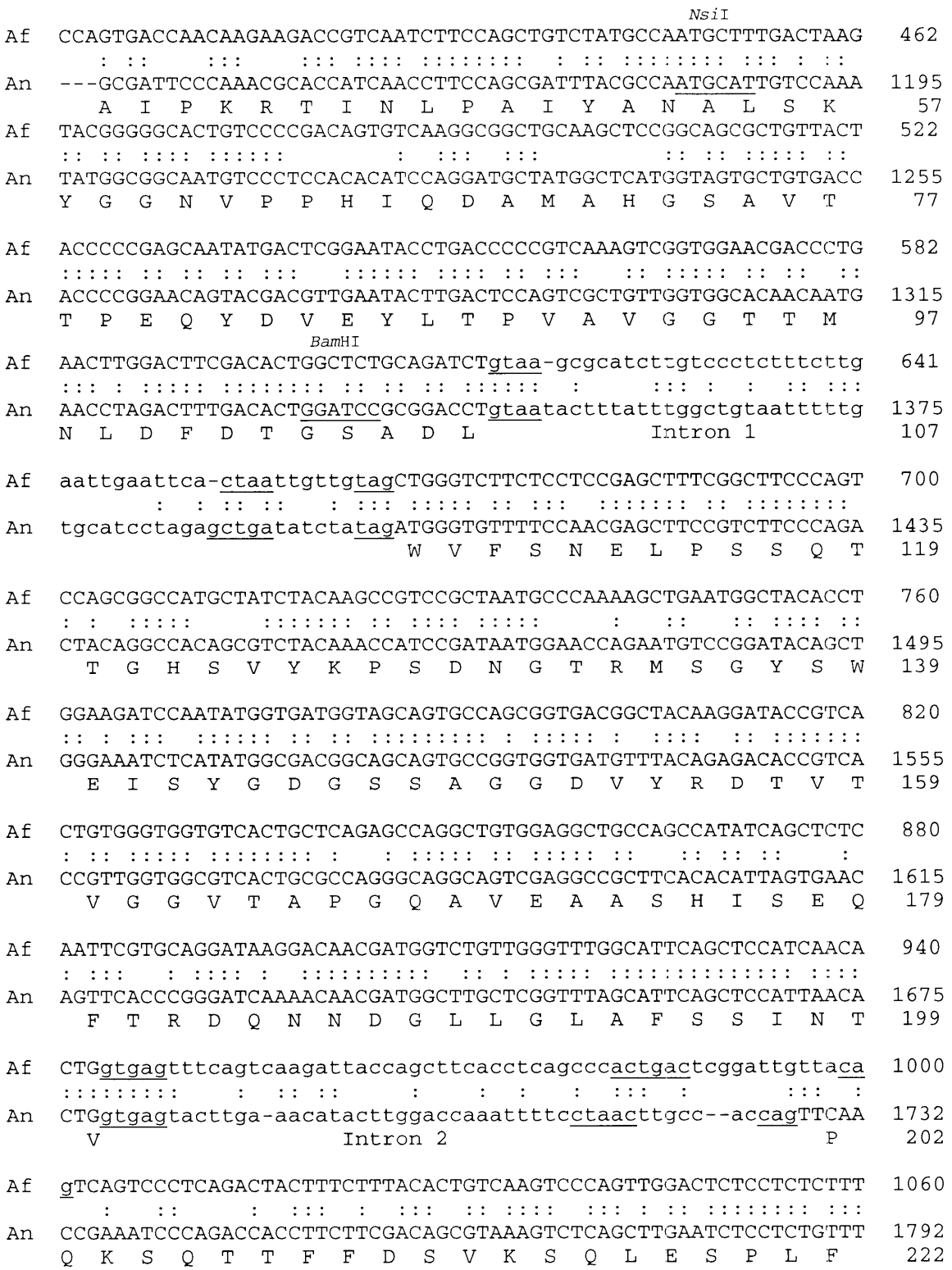


Figure 4.6. Low stringency Southern blot to determine if there are multiple copies of *prtB*. Genomic DNA was digested with *Xba*I (lane 1), *Hind*III (lane 2), *Eco*RI (lane 3), and *Bam*HI (lane 4) and probed with a DIG-labelled probe (Boehringer Manheim). This probe was an *BamH*I/*Pst*I fragment containing the entire *prtB* coding region (fig. 4.1.). Only a single hybridising band is visible for each digest, suggesting that *prtB* is a single copy gene. This blot was hybridised in a solution containing 0.25 M NaOP pH 7.2, 7% SDS, 1 mM EDTA, 30% formamide, at 37°C. This blot was washed once for 5 minutes in 2x SSC at room temperature, then twice, for 20 minutes, in 0.25 M NaOP pH7.2, 2% SDS 1 mM EDTA at 42°C.


BglII
HindIII


NsiI
Af CCAGTGACCAACAAGAAGACCGTCAATCTTCCAGCTGTCTATGCCAATGCTTTGACTAAG 462
A I P K R T I N L P A I Y A N A L S K 57
An ---GCGATTCCCAAACGCACCATCAACCTTCCAGCGATTTACGCCAATGCATTGTCCAAA 1195
Af TACGGGGGCACTGTCCCCGACAGTGTCAAGGCGGCTGCAAGCTCCGGCAGCGCTGTTACT 522
Y G G N V P P H I Q D A M A H G S A V T 77
An TATGGCGGCAATGTCCCTCCACACATCCAGGATGCTATGGCTCATGGTAGTGCTGTGACC 1255
Af ACCCCCGAGCAATATGACTCGGAATACCTGACCCCCGTCAAAGTCGGTGGAACGACCCTG 582
T P E Q Y D V E Y L T P V A V G G T T M 97
An ACCCCGGAACAGTACGACGTTGAATACTTGACTCCAGTCGCTGTTGGTGGCACAACAATG 1315
BamHI
Af AACTTGGACTTCGACACTGGCTCTGCAGATCTgtaa-gcgcatcttgtccctctttcttg 641
An AACCTAGACTTTGACACTGGATCCGCGGACCTgtaatactttatttggctgtaatttttg 1375
N L D F D T G S A D L Intron 1 107
Af aattgaattca-ctaattgttgtagCTGGGTCTTCTCCTCCGAGCTTTCGGCTTCCCAGT 700
An tgcatcctagagctgatatctatagATGGGTGTTTTCCAACGAGCTTCCGTCTTCCCAGA 1435
W V F S N E L P S S Q T 119
Af CCAGCGGCCATGCTATCTACAAGCCGTCCGCTAATGCCCAAAAGCTGAATGGCTACACCT 760
An CTACAGGCCACAGCGTCTACAAACCATCCGATAATGGAACCAGAATGTCCGGATACAGCT 1495
T G H S V Y K P S D N G T R M S G Y S W 139
Af GGAAGATCCAATATGGTGATGGTAGCAGTGCCAGCGGTGACGGCTACAAGGATACCGTCA 820
An GGGAAATCTCATATGGCGACGGCAGCAGTGCCGGTGGTGATGTTTACAGAGACACCGTCA 1555
E I S Y G D G S S A G G D V Y R D T V T 159
Af CTGTGGGTGGTGTCACTGCTCAGAGCCAGGCTGTGGAGGCTGCCAGCCATATCAGCTCTC 880
An CCGTTGGTGGCGTCACTGCGCCAGGGCAGGCAGTCGAGGCCGCTTCACACATTAGTGAAC 1615
V G G V T A P G Q A V E A A S H I S E Q 179
Af AATTCGTGCAGGATAAGGACAACGATGGTCTGTTGGGTTTGGCATTCAGCTCCATCAACA 940
An AGTTCACCCGGGATCAAAACAACGATGGCTTGCTCGGTTTAGCATTCAGCTCCATTAACA 1675
F T R D Q N N D G L L G L A F S S I N T 199
Af CTGgtgagtttcagtcaagattaccagcttcacctcagcccactgactcggattgttaca 1000
An CTGgtgagtacttga-aacatacttggaccaaattttcctaacttgcc--accagTTCAA 1732
V Intron 2 P 202
Af gTCAGTCCCTCAGACTACTTTCTTTACACTGTCAAGTCCCAGTTGGACTCTCCTCTCTTT 1060
An CCGAAATCCCAGACCACCTTCTTCGACAGCGTAAAGTCTCAGCTTGAATCTCCTCTGTTT 1792
Q K S Q T T F F D S V K S Q L E S P L F 222

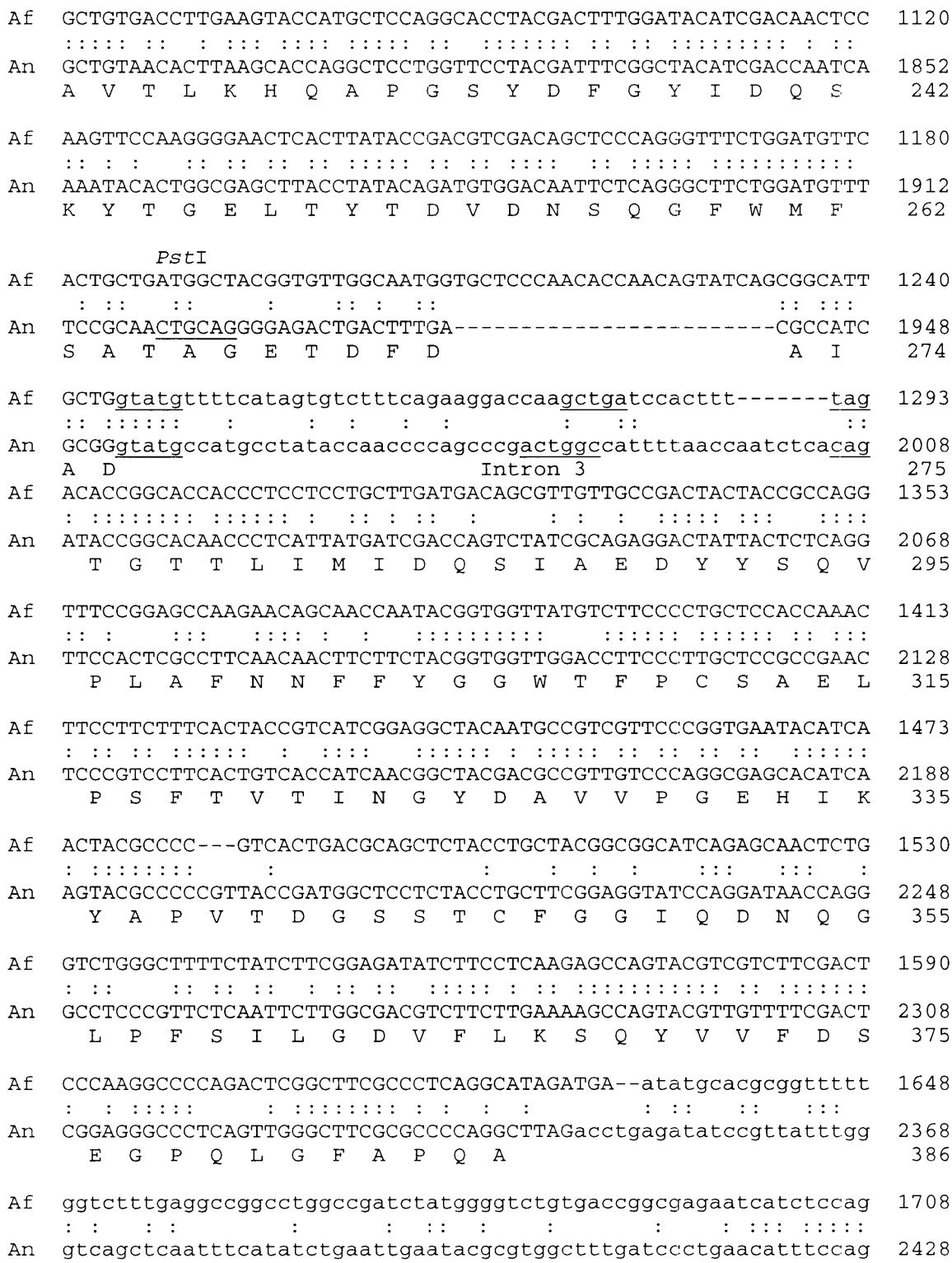

Af GCTGTGACCTTGAAGTACCATGCTCCAGGCACCTACGACTTTGGATACATCGACAACTCC 1120
An GCTGTAACACTTAAGCACCAGGCTCCTGGTTCCTACGATTTCGGCTACATCGACCAATCA 1852
A V T L K H Q A P G S Y D F G Y I D Q S 242
Af AAGTTCCAAGGGGAACTCACTTATACCGACGTCGACAGCTCCCAGGGTTTCTGGATGTTC 1180
An AAATACACTGGCGAGCTTACCTATACAGATGTGGACAATTCTCAGGGCTTCTGGATGTTT 1912
K Y T G E L T Y T D V D N S Q G F W M F 262
PstI
Af ACTGCTGATGGCTACGGTGTTGGCAATGGTGCTCCCAACACCAACAGTATCAGCGGCATT 1240
An TCCGCAACTGCAGGGGAGACTGACTTTGA----------CGCCATC 1948
S A T A G E T D F D A I 274
Af GCTGgtatgttttcatagtgtctttcagaaggaccaagctgatccacttt-------tag 1293
An GCGGgtatgccatgcctataccaaccccagcccgactggccattttaaccaatctcacag 2008
A D Intron 3 275
Af ACACCGGCACCACCCTCCTCCTGCTTGATGACAGCGTTGTTGCCGACTACTACCGCCAGG 1353
An ATACCGGCACAACCCTCATTATGATCGACCAGTCTATCGCAGAGGACTATTACTCTCAGG 2068
T G T T L I M I D Q S I A E D Y Y S Q V 295
Af TTTCCGGAGCCAAGAACAGCAACCAATACGGTGGTTATGTCTTCCCCTGCTCCACCAAAC 1413
An TTCCACTCGCCTTCAACAACTTCTTCTACGGTGGTTGGACCTTCCCTTGCTCCGCCGAAC 2128
P L A F N N F F Y G G W T F P C S A E L 315
Af TTCCTTCTTTCACTACCGTCATCGGAGGCTACAATGCCGTCGTTCCCGGTGAATACATCA 1473
An TCCCGTCCTTCACTGTCACCATCAACGGCTACGACGCCGTTGTCCCAGGCGAGCACATCA 2188
P S F T V T I N G Y D A V V P G E H I K 335
Af ACTACGCCCC---GTCACTGACGCAGCTCTACCTGCTACGGCGGCATCAGAGCAACTCTG 1530
An AGTACGCCCCCGTTACCGATGGCTCCTCTACCTGCTTCGGAGGTATCCAGGATAACCAGG 2248
Y A P V T D G S S T C F G G I Q D N Q G 355
Af GTCTGGGCTTTTCTATCTTCGGAGATATCTTCCTCAAGAGCCAGTACGTCGTCTTCGACT 1590
An GCCTCCCGTTCTCAATTCTTGGCGACGTCTTCTTGAAAAGCCAGTACGTTGTTTTCGACT 2308
L P F S I L G D V F L K S Q Y V V F D S 375
Af CCCAAGGCCCCAGACTCGGCTTCGCCCTCAGGCATAGATGA--atatgcacgcggttttt 1648
An CGGAGGGCCCTCAGTTGGGCTTCGCGCCCCAGGCTTAGacctgagatatccgttatttgg 2368
E G P Q L G F A P Q A 386
Af ggtctttgaggccggcctggccgatctatggggtctgtgaccggcgagaatcatctccag 1708
An gtcagctcaatttcatatctgaattgaatacgcgtggctttgatccctgaacatttccag 2428

Figure 4.7. Comparison of the DNA sequence of the *A. nidulans* (An) *prtB* gene and the *A. fumigatus* (Af) gene, *pepF* (Reichard *et al.* 1995). Identical nucleotides are connected with colons. A putative TATA box is marked by asterisks, and a potential polyadenylation signal is double underlined. Restriction sites and the fungal intron consensus sequence 5'-GTRNGT-RCTRAC-YAG-3' (Rambosek and Leach 1986), are underlined. Putative CreA consensus and AreA consensus binding sites are overlined and double overlined, respectively.

The region upstream of the putative *A. nidulans* aspartic protease gene sequence was examined for putative binding sites for the wide domain regulatory proteins, AreA and CreA. One consensus AreA binding sequence, and one putative CreA binding consensus sequence were observed (fig. 4.7.). No sequence with any homology to the "18 mer" observed upstream of both the *pepA* and *pepB* acid protease genes of *A. niger* was found in the corresponding region of the putative *prtB* promoter (Jarai and Buxton 1994).

4.3.6. Putative protein sequence of *prtB*.

4.3.6.1. Comparison to other fungal aspartic proteases.

The deduced *prtB* protein sequence possesses the active sites which are highly conserved between aspergillopepsins, and other fungal aspartic proteases including *Penicillium janthinellum* and *Rhizopus chinensis* (figs. 4.8. and 4.9.) (James and Sielecki 1983, Delaney *et al.* 1987). The predicted *A. nidulans* protein sequence has a stretch of eight amino acids missing from a region which is not highly conserved in aspergillopepsins. However, all but the *A. nidulans* aspergillopepsin have the same number of amino acids separating the two active sites (fig. 4.8.). The three-dimensional (3D) structure of two fungal aspartic proteases has been determined. Information regarding the 3D structure of the aspartic proteases of *Penicilium janthinellum* and *Rhizopus chinensis* was obtained from the Brookhaven Protein Database. The Rasmol program was used to highlight the location of the active sites and the amino acids which correspond to the region absent in the putative *prtB* protein (figs. 4.10. and 4.11.) Both models show that the deletion effects amino acids which would be expected to be found on the surface of the PrtB protein. It is unknown what effect this deletion has on the function of the *A. nidulans* aspartic protease.

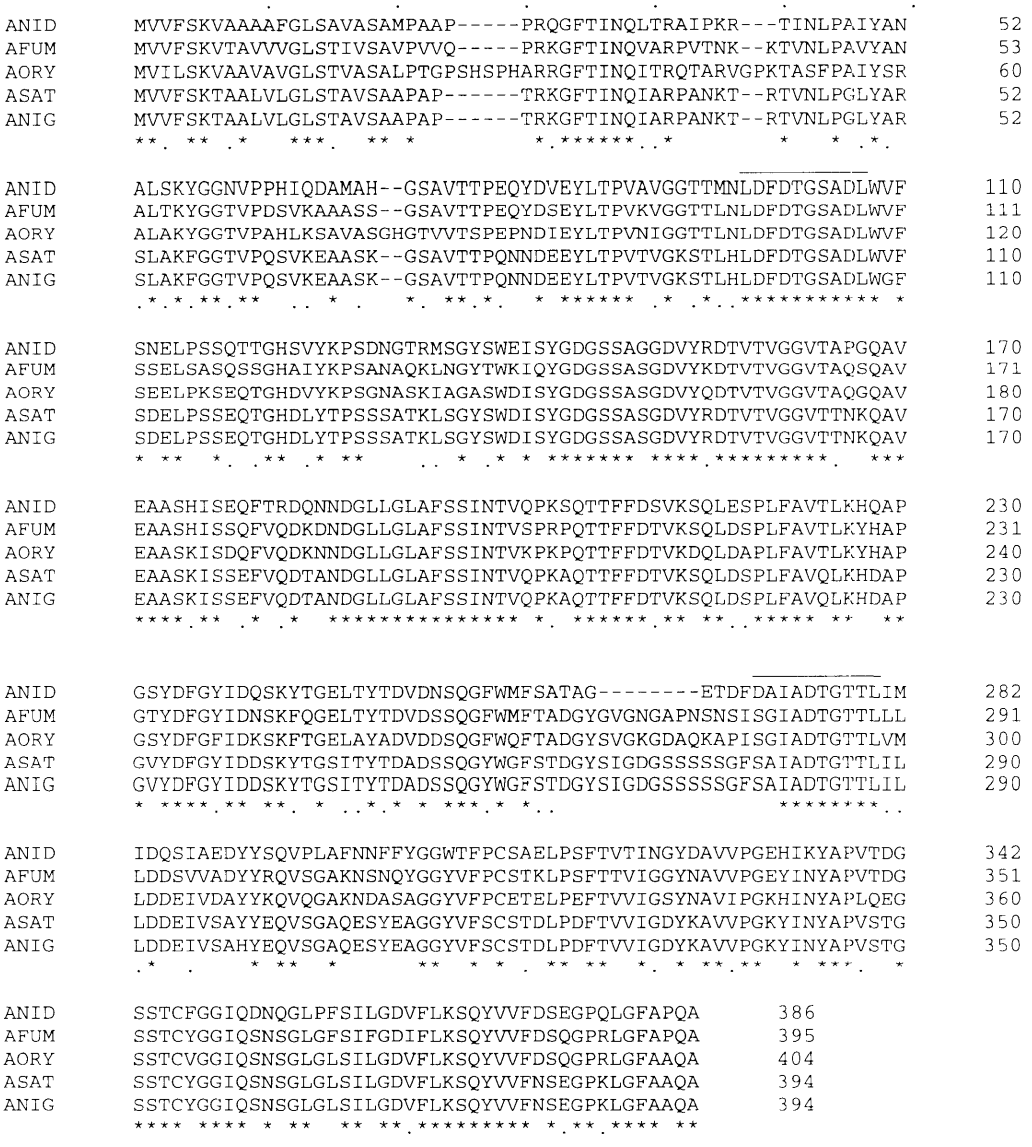


Figure 4.8. Comparison of the deduced amino acid sequences of *Aspergillus* aspartic proteases. The amino acid sequences of the *A. nidulans* aspartic protease (ANID), *A. fumigatus* aspartic protease PepF (AFUM) (Lee and Kolattukudy 1995, Reichard *et al.* 1995), *A. oryzae* aspartic protease PepA (AORY) (Gomi *et al.* 1993), *A. satoi* aspartic protease AspI (ASAT) (Shintani and Ichishima 1994), and *A.niger* aspartic protease PepA (ANIG) (Berka *et al.* 1990) were aligned using the Clustal W program. Asterisks mark amino acids which are identical in all five sequences, and conserved amino acids are marked with a period. Active sites are overlined. Of particular interest is the eight amino acid deletion in the *A nidulans* sequence, prior to the second active site.

```
AFUM      ------MVVFSKVTAVVVGLSTIVSAVPVVQ-----P-RKGFTINQVARPVTN--KKTVN      46
ANID      ------MVVFSKVAAAAFGLSAVASAMPAAP-----P-RQGFTINQLTRAIP---KRTIN      45
ANIG      ------MVVFSKTAALVLGLSTAVSAAPAPT-------RKGFTINQIARPAN--KTRTVN      45
AORY      ------MVILSKVAAVAVGLSTVASALPTGPSHSPHA-RRGFTINQITRQTARVGPKTAS      53
ASAT      ------MVVFSKTAALVLGLSTAVSAAPAPT-------RKGFTINQIARPAN--KTRTVN      45
CPAR      MSSPLKNALVTAMLAGGALSSPTKQHVGIPVNASPEVGPGKYSFKQR-NPNYKF----NG      55
GCIN      ------MTLITALTAGLALAS---SVIGAPTNNANEK---RFTVDQIKNPRYIR----NG      44
PJAN      ------MVVFSKITVVLAGLATVASAVPTGTS----R-KSTFTVNQKARPVAQ--AKAIN      47
PROQ      ------MVVFSQVTVALTCFSAIASAAAVRQE----P-PQGFTVNQVQKAVPG--TRTVN      47
RCHI      -------MKFTLISSCIAIAALAVAVDAAPG-------EKKISIPLAKNPNYK-----PS      41
RNIV      -------MKFTLISSCVALAAMTLAVEAAPN-------GKKINIPLAKNNSYK-----PS      41
SFIB      -------MLFSKSLLLSVLASLSFAAPVEKR-------EKTLTLDFDVKRIS------SK      40

AFUM      LPAVYANALTKYGGTVPDSVKAAAS-----------SGSAVTTPEQYDSEYLTPVKVG--      93
ANID      LPAIYANALSKYGGNVPPHIQDAMA-----------HGSAVTTPEQYDVEYLTPVAVG--      92
ANIG      LPGLYARSLAKFGGTVPQSVKEAAS-----------KGSAVTTPQNNDEEYLTPVTVG--      92
AORY      FPAIYSRALAKYGGTVPAHLKSAVASG---------HGTVVTSPEPNDIEYLTPVNIG--     102
ASAT      LPGLYARSLAKFGGTVPQSVKEAAS-----------KGSAVTTPQNNDEEYLTPVTVG--      92
CPAR      PLSVKK-TYLKYGVPIPAWLEDAVQNSTSGLAERSTGSATTTPIDSLDDAYITPVQIGTP     114
GCIN      PLALAK-AYRKYGKALPEDLSRVVANITSTGATKRATGSVAATPQDYDVEYLSPVQIGTP     103
PJAN      LPGMYASALSKYGAAVPASVKAAAES-----------GTAVTTPEANDVEYLTPVNVG--      94
PROQ      LPGLYANALVKYGATVPATVHAAAVS-----------GSAITTPEADDVEYLTPVTIG--      94
RCHI      AKNAIQKAIAKYNKHKINTSTGGIVP-------DAGVGTVPMTDYGNDVEYYGQVTIGTP      94
RNIV      AKNALNKALAKYNRRKVGS--GGIT--------TEASGSVPMVDYENDVEYYGEVTVGTP      92
SFIB      AKNVTVASSPGFRRNLRAASDAGVT----------------ISLENEYSFYLTTIEIGTP      84

AFUM      GTTLNLDFDTGSADLWVFSSELSASQSSGHAIYKPS--ANAQKLNGYTWKIQYGDGSSAS     115
ANID      GTTMNLDFDTGSADLWVFSNELPSSQTTGHSVYKPS--DNGTRMSGYSWEISYGDGSSAG     150
ANIG      KSTLHLDFDTGSADLWGFSDELPSSEQTGHDLYTPS--SSATKLSGYSWDISYGDGSSAS     150
AORY      GTTLNLDFDTGSADLWVFSEELPKSEQTGHDVYKPS--GNASKIAGASWDISYGDGSSAS     160
ASAT      KSTLHLDFDTGSADLWVFSDELPSSEQTGHDLYTPS--SSATKLSGYSWDISYGDGSSAS     150
CPAR      AQTLNLDFDTGSSDLWVFSSETTASEVDGQTIYTPSKSTTAKLLSGATWSISYGDGSSSS     174
GCIN      AQTLTLDFDTGSSDLWVFSTSTPSSQRNGQTVYDPSKSSTASRLTGATWSISYGDGSSSS     163
PJAN      GTTLNLDFDTGSADLWVFSSELSSSESTGHSLYKPS--SNATKLAGYSWSITYGDQSSAS     152
PROQ      SSTLNLDFDTGSADLWVFSSELTSSQQSGHDVYNVG--SLGTKLSGASWSISYGDGSSAS     152
RCHI      GKKFNLDFDTGSSDLWIASTLCTNCGSR-QTKYDPKQ-SSTYQADGRTWSISYGDGSSAS     152
RNIV      GIKLKLDFDTGSSDMWFASTLCSSCSNS-HTKYDPKK-SSTYAADGRTWSISYGDGSSAS     150
SFIB      GQKLQVDVDTGSSDLWVPGQGTSSLYGT----YDHTK-STSYKKDRSGFSISYGDGSSAR     139

AFUM      GDVYKDTVTVGGVTAQSQAVEAASHISSQFVQDKDNDGLLGLAFSSIN-TVSPRPQTTFF     210
ANID      GDVYRDTVTVGGVTAPGQAVEAASHISEQFTRDQNNDGLLGLAFSSIN-TVQPKSQTTFF     209
ANIG      GDVYRDTVTVGGVTTNKQAVEAASKISSEFVQDTANDGLLGLAFSSIN-TVQPKAQTTFF     209
AORY      GDVYQDTVTVGGVTAQGQAVEAASKISDQFVQDKNNDGLLGLAFSSIN-TVKPKPQTTFF     219
ASAT      GDVYRDTVTVGGVTTNKQAVEAASKISSEFVQDTANDGLLGLAFSSIN-TVQPKAQTTFF     209
CPAR      GDVYTDTVSVGGLTVTGQAVESAKKVSSSFTEDSTIDGLLGLAFSTLN-TVSPTQQKTFF     233
GCIN      GIVYKDTVSVGGSLSVTGQAVEAASKVSSSFSEESDLDGLLGLGFSSIN-TVSPTQQKTFF     222
PJAN      GDVYKDFVVVGGVKASPQAVEAASQISQQFVNDKNNDGLLGLAFSSIN-TVKPKSQTTFF     211
PROQ      GDVYKDTVTVGGVKATGQAVEAAKKISSQFLQDKNNDGLLGMAFSSIN-TVSPTPQKTFF     211
RCHI      GILAKDNVNLGGLLIKGQTIELAKREAASFANGP-NDGLLGLGFDTIT-TVRGVKTPMDN     210
RNIV      GILATDNVNLGGLLIKKQTIELAKRESSAFATDV-IDGLLGLGFNTIT-TVRGVKTPVDN     208
SFIB      GDWAQETVSIGGASITG--LEFGDATSQDVGQGLLGIGLKGNEASAQSSNSFTYDNLPLK     197

FUM       DTVKSQLDSPLFAVTLKYH---APGTYDFGYIDNSKFQGELTYTDVDSS--QG-------     258
ANID      DSVKSQLESPLFAVTLKHQ---APGSYDFGYIDQSKYTGELTYTDVDNS--QG-------     257
ANIG      DTVKSQLDSPLFAVQLKHD---APGVYDFGYIDDSKYTGSITYTDADSS--QG-------     257
AORY      DTVKDQLDAPLFAVTLKYH---APGSYDFGFIDKSKFTGELAYADVDDS--QG-------     267
ASAT      DTVKSQLDSPLFAVQLKHD---APGVYDFGYIDDSKYTGSITYTDADSS--QG-------     257
CPAR      DNAKASLDSPVFTADLGYH---AG-TYNFGFIDTTAYTGSITYTAVSTK--QG-------     280
GCIN      ETAKSKLDAYLFTADLKHN---TPGKYNFGYIDSSAYTGAITYVSIDNS--DG-------     270
PJAN      DTVKGQLDSPLFAVTLKHN---APGTYDFGFVDKNKYTGSLTYAQVDSS--QG-------     259
PROQ      DTVKSSLGEPLFAVTLQGTG--RPWHLRFGYIDSDKYTGTLAYADVDDS--DG-------     260
RCHI      LISQGLISRPIFGVYLGKASNGGGGEYIFGGYDSTKFKGSLTTVPIDNS--RG-------     261
RNIV      LISQGLISRPIFGVYLGKQSNGGGGEYIFGGYDSSKFKGSLTTVPIDNS--EG-------     259
SFIB      LKDQGLIDKAAYSLYLNSED-ATSGSILFGGSDSSKYSGSLATLDLVNIDDEGDSTSGAV     256
```

```
                                  ________
AFUM     FWMFTADGYGVGNGAPNSNSISGIADTGTTLLLLDDSVVADYYRQVS-GAKNSNQYGGYV     317
ANID     FWMFSAT--------AGETDFDAIADTGTTLIMIDQSIAEDYYSQVP-LAFNNFFYGGWT     308
ANIG     YWGFSTDGYSIGDGSSSSSGFSAIADTGTTLILLDDEIVSAHYEQVS-GAQESYEAGGYV     316
AORY     FWQFTADGYSVGKGDAQKAPISGIADTGTTLVMLDDEIVDAYYKQVQ-GAKNDASAGGYV     326
ASAT     YWGFSTDGYSIGDGSSSSSGFSAIADTGTTLILLDDEIVSAYYEQVS-GAQESYEAGGYV     316
CPAR     FWEWTSTGYAVGSGTFKSTSIDGIADTGTTLLYLPATVVSAYWAQVS-GAKSSSSVGGYV     339
GCIN     WWQFTSSGYSVGSASFTSTSLNGIADTGTTLLLLPQSVVTAYYAKIS-GAKYDSSQGGYT     329
PJAN     FWSFTADGYKIG--SKSGGSIQGIADTGTTLLLLPDNVVSDYYGQVS-GAQQDSSAGGYT     316
PROQ     FWSFTADSYKIGTG-AAGKSITGIADTGTTLLLLDSSIVTGLLQEGYPGSQNSSSAGGYI     319
RCHI     WWGITVDRATVG-TSTVASSFDGILDTGTTLLILPNNVAASVARAYG---ASDNGDGTYT     317
RNIV     FWGVTVKSTKIG-GTTVSASFDAILDTGTTLLLLPDDVAAKVARSYG---ASDNGDGTYS     315
SFIB     AFFVELEGIEAGSSSITKTTYPALLDSGTTLIYAPSSIASSIGREYG---TYSYSYGGYV     313
                               . *.****.     .                 *

AFUM     FPCSTK-LPSFTTVIGGYNAVVPGEYINYAPVTDGSSTCYGGIQSNSGLGFSIFGDIFLK     376
ANID     FPCSAE-LPSFTVTINGYDAVVPGEHIKYAPVTDGSSTCFGGIQDNQGLPFSILGDVFLK     367
ANIG     FSCSTD-LPDFTVVIGDYKAVVPGKYINYAPVSTGSSTCYGGIQSNSGLGLSILGDVFLK     375
AORY     FPCETE-LPEFTVVIGSYNAVIPGKHINYAPLQEGSSTCVGGIQSNSGLGLSILGDVFLK     385
ASAT     FSCSTD-LPDFTVVIGDYKAVVPGKYINYAPVSTGSSTCYGGIQSNSGLGLSILGDVFLK     375
CPAR     FPCSAT-LPSFTFGVGSARIVIPGDYIDFGPISTGSSSCFGGIQSSAGIGINIFGDVALK     398
GCIN     FPCSAT-VPSFTFGVGSARVTIPASYMNYAPVST--STCFGGLQSSSGIGINIFGDVALK     386
PJAN     VPCSAQ-LPDFTVTIGSYNAVVPGSLINYAPLQSGSSTCFGGIQSNSGLGFSIFGDIFLK     375
PROQ     FPCSAT-LPDFTVTINGYDAVVPGKYINFAPVSTGSSSCYGGIQSNSGIGFSIFGDIFLK     378
RCHI     ISCDTSRFKPLVFSINGASFQVSPDSLVFEEYQG---QCIAGFGYGN-FDFAIIGDTFLK     373
RNIV     ITCDTSKLQPLVFTLGSSTFEVPSDSLIFEKDGN---KCIAGFAAG--GDLAILGDVFLK     370
SFIB     TSCDAT-GPDFKFSFNGKTITVPFSNLLFQNSEG-DSECLVGVLSS-GSNYYILGDAFLR     370
           * .          .    . .         * *         * **  *.

AFUM     SQYVVFDSQGP-RLGFAPQA--       395
ANID     SQYVVFDSEGP-QLGFAPQA--       386
ANIG     SQYVVFNSEGP-KLGFAAQA--       394
AORY     SQYVVFDSQGP-RLGFAAQA--       404
ASAT     SQYVVFNSEGP-KLGFAAQA--       394
CPAR     AAFVVFNGATTPTLGFASK---       417
GCIN     AAFVVFDGASN-RLGWAAKTLS       406
PJAN     SQYVVFDANGP-RLGFAPQA--       394
PROQ     SQYVVFDSEGP-RLGFAAQA--       397
RCHI     NNYVVFNQGVP-EVQIAPVAQ-       393
RNIV     NNYVVFNQEVP-EVQIAPVAN-       390
SFIB     SAYVYYDIDNS-QVGIAQAKY-       390
           .* .       .  *
```

Figure 4.9. Comparison of the deduced amino acid sequences of fungal aspartic proteases. The amino acid sequences of the aspartic proteases of *A. fumigatus* (AFUM) (Lee and Kolattukudy 1995, Reichard *et al.* 1995), *A. nidulans* (ANID), *A.niger* (ANIG) (Berka *et al.* 1990), *A. oryzae* (AORY) (Gomi *et al.* 1993), *A. satoi* (ASAT) (Shintani and Ichishima 1994), *Cryphonectria parasitica* (CPAR) (Razanamparany *et al.* 1992), *Glomerella cingulata* (CGIN) (Clark *et al.* 1997), *Penicilium janthinellium* (PJAN) (James and Sielecki 1983), *Penicillium roqueforti* (PROQ) (Bruyant *et al.* 1995), *Rhizopus chinensis* (RCHI) (Delaney *et al.* 1987), *Rhizopus niveus* (RNIV) (Horiuch *et al.* 1988), and *Saccharomycopsis fibuligera* (SFIB) (Hirata *et al.* 1988) were aligned using the Clustal W program. Asterisks mark amino acids which are identical in all twelve sequences, and conserved amino acids are marked with a period. Active sites are overlined. Of particular interest is the eight amino acid deletion in the *A nidulans* sequence, prior to the second active site.

Figure 4.10. 3-dimensional structure of the aspartic protease of *Penicillium janthinellum*. The regions coloured red represent the active site regions, and the region coloured green indicates the amino acids which correspond to the 8 amino acids "releted" from the *A. nidulans* aspartic protease PrtB. This protein structure was obtained from the Brookhaven Protein Database, is based on the datum of James and Sielecki (1983), and was visualised using RasMol version 2.5 (Sayle, 1994).

Figure 4.11. 3-dimensional structure of the aspartic protease of *Rhizopus chinensis*. The regions coloured red represent the active site regions, and the region coloured green indicates the 8 amino acids "deleted" from the *A. nidulans* aspartic protease PrtB. This protein structure was obtained from the Brookhaven Protein Database, is based on the datum of Suguna *et al.* (1987), and was visualised using RasMol version 2.5 (Sayle, 1994).

In the media of Cove (1966) the carbon, nitrogen, and sulphur needs of *A. nidulans* are supplied by compounds which do not require metabolism by extracellular proteases. Therefore laboratory culture conditions would not be expected to exert selective pressure to maintain functional extracellular protease genes. It has been noted in laboratory strains of organisms such as *Drosophila*, that mutations may arise in genes which are not essential under laboratory conditions. An example of a spontaneous mutation arising in a laboratory stock is the original white eye mutation, *w1*, of *Drosophila melanogaster* which renders the flies blind (Beadle and Ephrussi 1937). As all laboratory strains of *A. nidulans* are derived from the one isolate, which has been in continuous culture for over 50 years, it was thought possible that the deletion in the *prtB* coding region may be specific to the laboratory strains. In order to determine if this eight amino acid deletion was specific to the laboratory strains of *A. nidulans*, or was also present in environmental strains, a laboratory strain and an *A. nidulans* environmental isolate which had recently been obtained by our laboratory were compared. The environmental *A. nidulans* isolate, was obtained from an emu chick exhibiting symptoms of aspergillosis. Genomic DNA from the two *A. nidulans* strains was used as the template for the amplification of part of the *prtB*, including the region encompassing the deletion. No difference in the size of the two products was observed (fig. 4.12.). This result indicated that the eight amino acid deletion is not specific to the laboratory strains of *A. nidulans*.

4.3.6.2. Phylogenetic analysis.

An alignment generated using Pileup (GCG) was used to generate phylogenic trees. Both parsimony and distance neighbour joining analysis generated identical trees with bootstrap values that indicated that all the branch points were valid (figs. 4.13. and 4.14.). Similarity between PrtB and the *Rhizopus* aspartic proteases was detected in a search of the SwissProt databank. The sequence from the yeast, *Saccharomycopsis fibuligera,* was not amoungst the top 20 matches identifed in this

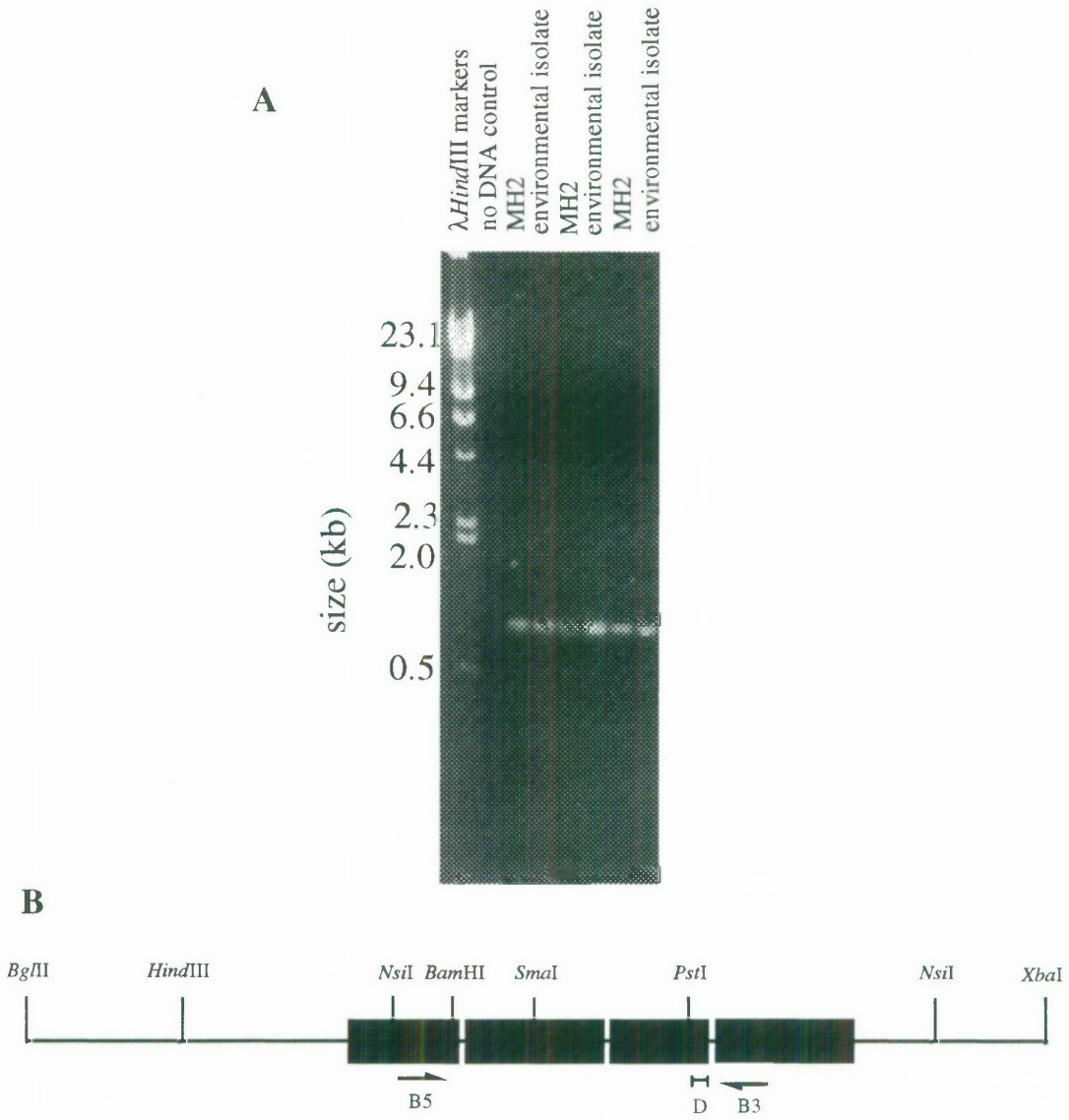


Figure 4.12. PCR products from the *prtB* gene amplified from DNA from an *A. nidulans* laboratory strain (MH2), and a strain of *A. nidulans* recently isolated from the environment. The schematic diagram (B) of the *prtB* gene shows the positions of the primers designated B5 and B3 (marked by arrows) which were used to amplify part of the *prtB* coding region. The products of this PCR experiment were separated on a 1.8% agarose gel (A). No size difference is apparent in the PCR products from the two *A. nidulans* strains. This result shows that the 24 bp deletion (location marked D on the schematic diagram) is not specific to the laboratory strain.

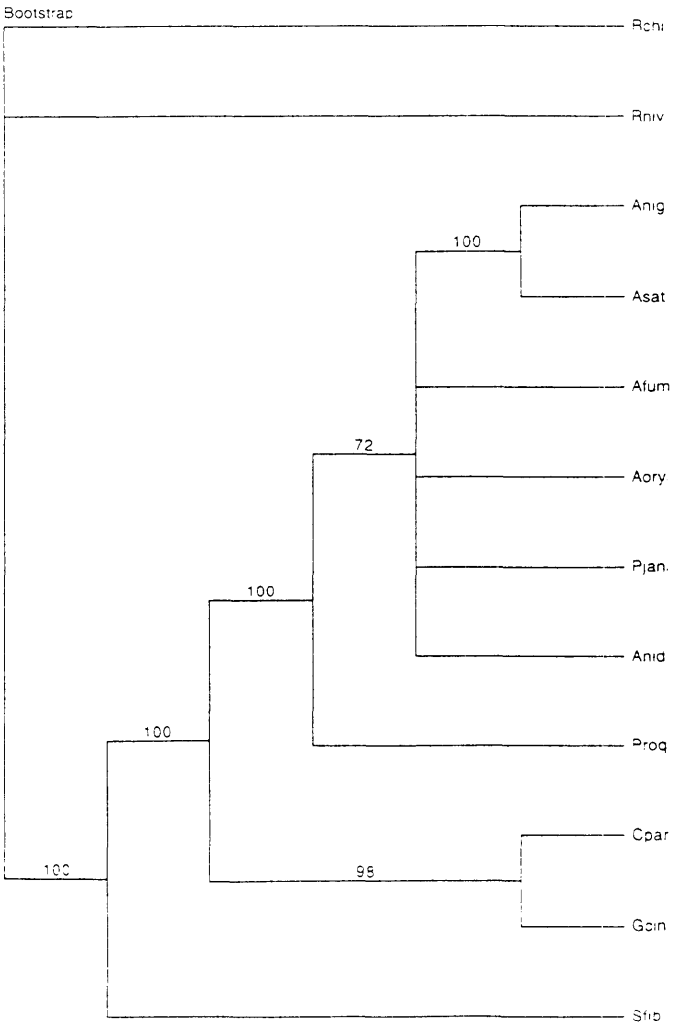


Figure 4.13. Phylogenic tree of the fungal aspartic proteases generated by unrooted parsimony analysis. The tree is based on the amino acid sequences of the aspartic proteases of *A. fumigatus* (AFUM) (Lee and Kolattukudy, 1995, Reichard *et al.*, 1995), *A. nidulans* (ANID), *A.niger* (ANIG) (Berka *et al.*, 1990), *A. oryzae* (AORY) (Gomi *et al.*, 1993), *A. satoi* (ASAT) (Shintani and Ichishima, 1994), *Cryphonectria parasitica* (CPAR) (Razanamparany *et al.*, 1992), *Glomerella cingulata* (CGIN) (Clark *et al.*, 1997), *Penicilium janthinellium* (PJAN) (James and Sielecki, 1983), *Penicillium roqueforti* (PROQ) (Bruyant *et al.*, 1995), *Rhizopus chinensis* (RCHI) (Delaney *et al.*, 1987), *Rhizopus niveus* (RNIV) (Horiuch *et al.*, 1988), and *Saccharomtcopsis fibuligera* (SFIB) (Hirata *et al.* 1988) was generated by Paup 3.1.1. (Swofford, 1992).

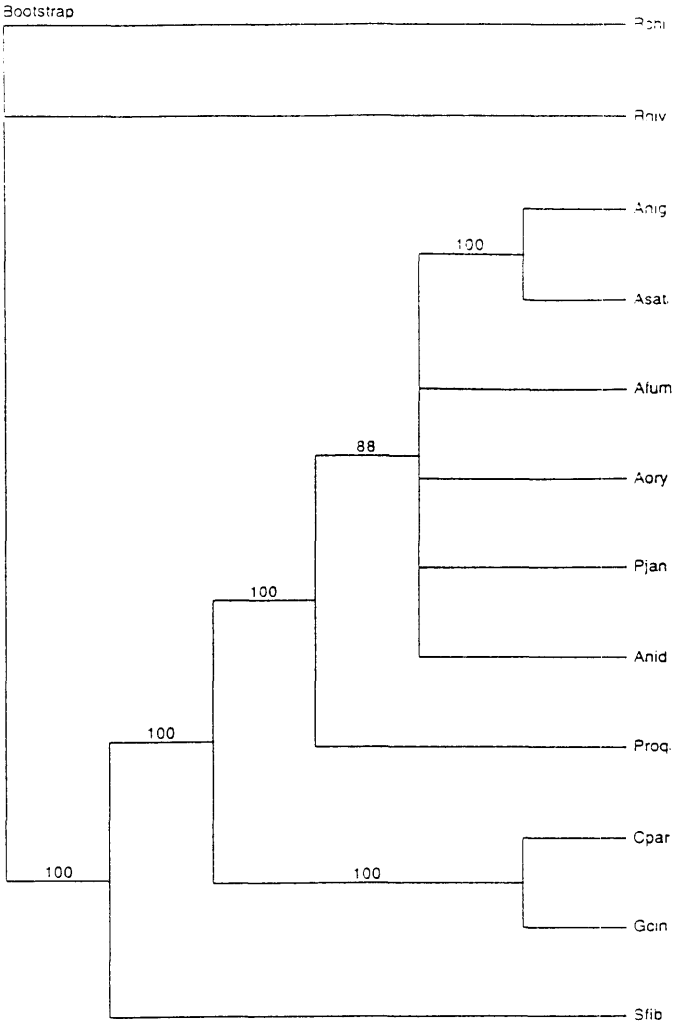


Figure 4.14. Phylogenic tree of the fungal aspartic proteases generated by unrooted distance neighbour joining analysis. The tree is based on the amino acid sequences of the aspartic proteases of *A. fumigatus* (AFUM) (Lee and Kolattukudy, 1995, Reichard *et al.*, 1995), *A. nidulans* (ANID), *A.niger* (ANIG) (Berka *et al.*, 1990), *A. oryzae* (AORY) (Gomi *et al.*, 1993), *A. satoi* (ASAT) (Shintani and Ichishima, 1994), *Cryphonectria parasitica* (CPAR) (Razanamparany *et al.*, 1992), *Glomerella cingulata* (CGIN) (Clark *et al.*, 1997), *Penicilium janthinellium* (PJAN) (James and Sielecki, 1983), *Penicillium roqueforti* (PROQ) (Bruyant *et al.*, 1995), *Rhizopus chinensis* (RCHI) (Delaney *et al.*, 1987), *Rhizopus niveus* (RNIV) (Horiuch *et al.*, 1988), and *Saccharomtcopsis fibuligera* (SFIB) (Hirata *et al.* 1988) was generated by Paup 3.1.1. (Swofford, 1992).

search, and was included in this analysis to function as an outgroup. It was initially surprising to find that in unrooted trees (figs 4.13. and 4.14) *Rhizopus* aspartic proteases were shown to be the outgroup. The dissimilarity of the *Rhizopus* protein sequences to the other fungal sequences can be understood in light of fungal taxonomy which classes *Rhizopus* species as lower fungi while all other aspartic proteases compared here are derived from species belonging to the higher fungi. It was unsuprising that the *Aspergillus* sequences clustered together. PrtB was positioned with the other aspergillopepsins, and this analysis has shown that the *A. nidulans* aspartic protease is most similar to the *A. fumigatus*, *A. oryzae*, and *P. janthinellum* aspartic proteases. It was surprising to see that the two *Penicillum* species were not positioned at the same node. This was the case for both trees, and the bootstrap values indicate that this branch-point is valid. Therefore the difference between these two sequences is significant. This may be indicative of these two *Penicillum* aspartic proteases undergoing divergent evolution.

4.3.7. Expression of *prtB*.

4.3.7.1. Northern analysis.

No *prtB* transcript was detected in Northern blots of total *A. nidulans* RNA, obtained from mycelia grown at pH 3 or pH 5 in the presence or absence of a nitrogen source for 4 hours. As a control, the same northern was probed with a PCR fragment generated from the constitutively expressed γ-actin gene (Fidel *et al.* 1988), and bands of the expected size were observed (fig. 4.15.).

4.3.7.2. RT-PCR analysis.

As no signal corresponding to the *prtB* transcript was detected on the Northern blot, a more sensitive technique, RT-PCR, was also used to detect *prtB* transcript in total RNA obtained from mycelia which were grown at either pH 3 or pH 8 in the presence

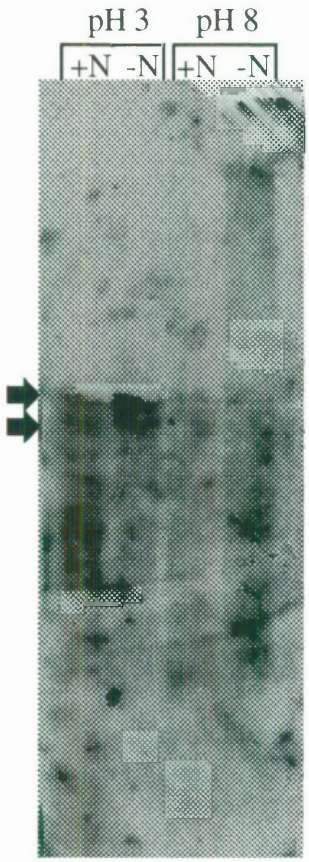


Figure 4.15. Northern blot probed with the γ-actin gene. The multiple γ-actin transcripts (Fidel *et al.*, 1988) are easily visible in total RNA obtained from wildtype (MH2) mycelia which was exposed to 4 hours repressing (not nutrient-limiting) conditions (+N), or nitrogen-limiting conditions(-N) at pH 3. Hybridising bands are not as easily visible in the lanes containing RNA from which was exposed to 4 hours repressing (not nutrient-limiting) conditions (+N), or nitrogen-limiting conditions(-N) at mycelia pH 8. This may be due to differences in the amount of RNA in the different samples. The γ-actin transcripts are marked with arrows. This blot was probed with ^{32}P labelled PCR fragment.

or absence of a nitrogen source for 4 hours or 24 hours. Though the γ-actin cDNA product was synthesised from RNA obtained from mycelia grown under all four conditions (fig. 4.16.), no product of the size expected for the *prtB* cDNA was observed in any of the reactions (fig. 4.16.). This experiment was repeated on the same RNA samples, and independently-generated RNA samples. For each sample, RT-PCR were carried out in duplicate using 1 and 10 μg of total RNA.

In a separate experiment mycelia from both *A. nidulans* and *A. niger* strains were grown for 4 hours in nitrogen starvation conditions in media buffered to pH 3 or pH 8. As *A. niger* is known to produce functional acid proteases, it was used to test if the conditions used in these experiments were suitable for the production of acid protease mRNA by *Aspergillus* species. In this experiment RNA from the environmental isolate of *A. nidulans* (section 4.3.7.2.) was also tested in case years of culture under laboratory conditions had affected the regulation of the *prtB* gene. Products corresponding to those expected for the cDNA product, and the genomic DNA product, of the *pepA* gene of *A. niger* were observed in RT-PCR reactions which used 1μg of total RNA (fig. 4.17.). RT-PCR products corresponding to the *pepA* cDNA product were not observed in the reactions which used 10μg of total RNA, presumably due to the inhibitory affect of high concentrations of template DNA on the PCR (McPherson *et al.* 1995). The template used was from mycelia transferred to nitrogen starvation conditions in media buffered to pH 3 or pH 8 for 4 hours. No cDNA product was observed when *A. nidulans* RNA was used as the template. As the *A. niger* homologue of *prtB* was expressed under the conditions examined it seems reasonable to expect that *prtB* would also be expressed under these conditions.

Examination of the promoter region of *prtB* showed only a single 5'-GATA-3' site, so it is possible that my inability to detect mRNA produced by this gene reflects a difference in the stimulus to which the aspergillopepsins of *A. nidulans* and *A. niger* respond. The promoter region of *prtB* was shown to possess a number of 5'-SYGGRG-3' sites, therefore *prtB* may be expressed at detectable levels under carbon derepressing conditions.

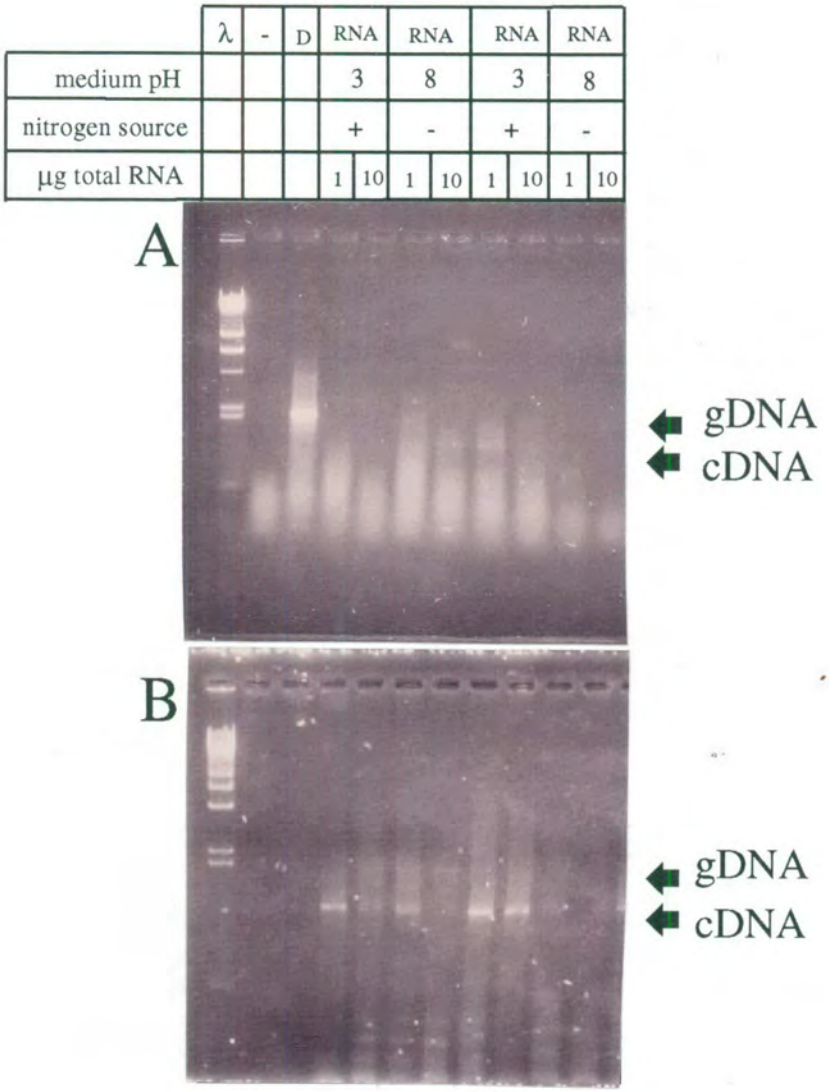


Figure 4.16. RT-PCR of the γ-actin gene of *A. nidulans*. Figure A and B show the results of two separate experiments in which the oligonucleotides A5 and A3 were used to amplify γ-actin transcripts. In figure A the DNA control produces a significantly larger product than the RT-PCR reactions. Though the DNA control did not work in the experiment labelled B, cDNA product was produced when RNA from all four conditions was used as the template. The genomic DNA sized product (gDNA) seen in some of the RT-PCR reactions in gel B, could either result from amplification from genomic DNA contaminating the RNA sample or amplification of unprocessed transcript. Reaction types are in the same position on both gels. Symbols: λ = λ *Hind*III standards, - = PCR control reaction containing no DNA or RNA, D = PCR control reaction containing *A. nidulans* genomic DNA. For culture conditions: nitrogen source; + = 10mM ammonium tartrate, - = no nitrogen source. Media was buffered to pH 3 with citric acid, and to pH 8 with Tris-HCl. RNA was obtained from mycelia which had been grown for 16 hours , and then transfered to the conditions specified for 4 hours. RT reactions used 1 or 10 µg total RNA.

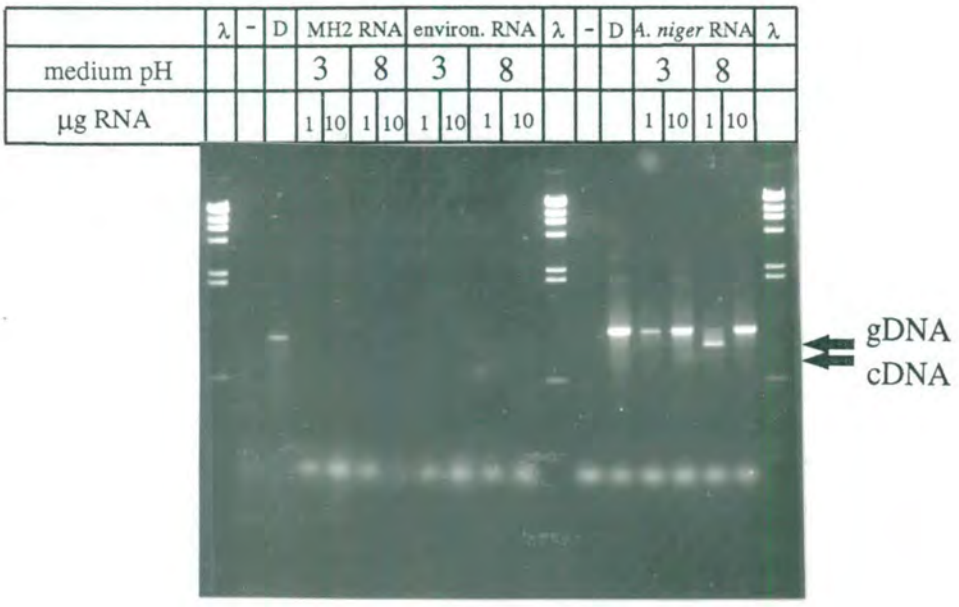


Figure 4.17. RT-PCR of the aspergillopepsin genes of *A. nidulans* and *A. niger*. RNA was obtained from mycelium subjected to 4 hours nitrogen-limitation in medium buffered to pH 3 and pH 8. RT-PCR reactions contained RNA from an *A. nidulans* laboratory strain (MH2), or an *A. nidulans* environmental isolate (environ.), or *A. niger*. The oligonucleotides designated B5 and B3, and MK21 and MK22, were used to amplify the *A. nidulans prtB* and *A. niger pepA* genes respectively. Genomic DNA (gDNA) and cDNA (cDNA) products are marked. λ *Hind*III markers (λ), -DNA (-), *A. nidulans* genomic DNA (D) are marked. The RT-PCR rections were performed using 1 μg or 10 μg total RNA.

4.4. Discussion.

The *A. nidulans* genome contains a single copy gene similar to aspartic protease genes from other *Aspergillus* species. This gene has been designated *prtB*. The *A. nidulans* aspartic protease gene sequence contains only one putative binding site for AreA and nine for CreA. It has been shown that the NIT-2 protein of *N. crassa*, which is homologous to AreA, requires at least two 5'-GATA-3' sequences within 30 bp for binding (Chiang and Marzluf 1994), suggesting that it is unlikely that the AreA binding site in the *prtB* promoter is functional. Binding of CreA also requires two consensus sequences in close proximity (Cubero and Scazzocchio 1994), therefore it is possible that the sites at nt 750 to 765 may comprise a functional CreA binding site. The "18 mer" motif (Jarai and Buxton 1994a), which is located between the putative TATA box and translation start point in the acidic protease genes of *A. satoi* (Shintani and Ichishima 1994), *A. oryzae* (Gomi *et al.* 1993), *pepA* (Berka *et al.* 1990a and 1990b), and *pepB* of *A. niger* (Inoue *et al.* 1991), is not present in *prtB* (fig. 4.18.). The effect of the presence or absence of the "18 mer" on expression of these genes is yet to be determined.

Overall, the PrtB amino acid sequence is highly conserved (70.7% identity to the *pepF* gene product of *A. fumigatus*). The deduced amino acid sequence of *prtB* contains the active sites conserved in aspergillopepsins. There appears to be a 24 bp deletion which results in a protein that is 8 amino acids shorter, but the coding regions remain in the correct reading frame. This deletion was also observed in a non-laboratory strain of *A. nidulans* which had recently been isolated. The effect of this sequence aberration is unknown, though it may alter the folding of the protein rendering it non-functional, effect its level of activity, or alter its specificity. A possible strategy to determine the effect of this deletion on protease activity, could involve the expression of this protein in *E. coli* and an *in vitro* assay of its activity. However, no *prtB* transcript was detected and, therefore, no *prtB* cDNA can be made to clone into an expression

```
ANIG(pepA)      TGGATCTTCTTGTTCATC      (-94 to -112)
ANIG(pepB)      TGGCTTTTCTTCTTCATC     (-135 to -117)
AORY(pepA)      TTGTCATTCTTTCACATC      (-68 to -50)
ASAT(pepI)      TGTCTCTTCTTGTTCATC      (-96 to -78)
                *     *****   ****
```

Figure 4.18. The "18 mer" -a DNA sequence motif common to a number of *Aspergillus* extracellular acid protease promoters. The "18 mer" was first identified in the 5' regions of the *A. niger* extracellular acid protease genes *pepA* and *pepB* [ANIG(*pepA*), and ANIG(*pepB*)](Jarai and Buxton 1994). This motif is also present in the 5' regions of the *A. oryzae* extracellular aspartic protease gene *pepA* [AORY(*pepA*)](Gomi *et al.* 1993), and the *A. satoi* extracellular aspartic protease gene *pepI* [ASAT(*pepI*)](Shintani and Ichishima 1994). Asterisks indicate nucleotides which are identical in all four sequences, and nucleotides in bold are common between at least three of the sequences.

RNA isolated from *A. nidulans* mycelia which had been subjected to nitrogen limiting conditions for 4 hours. We were able to detect *A. niger pepA* mRNA using RT-PCR from RNA isolated from mycelia subjected to nitrogen limiting conditions for 4 hours. It would therefore be expected that these conditons would also be suitable for the detection of *prtB* expression. These experiments indicate that, unlike the extracellular serine protease of *A. nidulans* and the *pepA* gene of *A. niger*, the extracellular aspartic protease gene of *A. nidulans* is not expressed in mycelia subjected to 4 hours of nitrogen-limiting conditions. Interestingly, a band corresponding to the cDNA of the *A. niger pepA* gene was observed after 4hr incubation in medium lacking a nitrogen source buffered to pH 8 (fig. 4.17.), suggesting nutrient starvation overrides pH regulation in *A. niger*, as it has been shown to do in *A. nidulans*. As the *prtB* promoter region contains two CreA consensus sequences in close proximity (nt 750 to 765), it may be that *prtB* is expressed under carbon-limiting conditions. Though it must be remembered that, at this time, there is no evidence to show that CreA regulates the extracellular proteases of *A. nidulans* directly.

The eight amino acid deletion found in PrtB could affect the structure of the mature protein and may effect its ability to function. If the *prtB* product is not functional, one might still expect to be able to detect *prtB* mRNA, as has been the case with some pseudogenes. The scarcity of putative binding sites for the global regulatory protein AreA, may be indicative of the inability of the promoter region of the *prtB* gene to function under nitrogen-limiting conditions. Fu and Marzluf (1990b) found that though NIT-2 is able to bind to single 5'-GATA-3' sequences, its affinity for sequences containing two copies of the 5'-GATA-3' sequence was much greater than it's afinity for sequences containing a single 5'-GATA-3' site. If this preference for sites containing two 5'-GATA-3' sequences is common to AreA, the putative binding site upstream of the *prtB* coding region would constitute a low affinity binding site at best. To our knowledge, no proteolytic activity that could be attributed to an acid protease has ever been detected in *A. nidulans*. It is unknown whether *prtB* is a non-functional

gene, is expressed at levels too low to detect, or is expressed under conditions different to those used in this study.