V. RESULTS

5.1. Total RNA isolation

Total RNA was prepared from 5 g of 3-day-old yeast cells pelleted from a 500-ml culture of *P. marneffei*. The recovery of total RNA was 7.5 mg (1.5 mg/g of yeast cells). A total RNA preparation had a A260/280 ratio of 1.9. Formaldehyde agarose gel electrophoresis of the total RNA sample is shown in Figure 11, lane 2. From the gel, two predominant bands of 3.8 and 2.0 kb, representing 28S and 18S ribosomal RNAs were presented. The bands were intact, with a visualized band intensity ratio (28S:18S) higher than 1.5, indicating no degradation of RNA. The band below 0.2 kb was mainly tRNA. The mRNA population can be seen as faintly diffuse smears along the gel lane.

5.2. Poly $(A)^+$ RNA enrichment

Penicillium marneffei mRNA was purified from total RNA by using the oligo(dT) cellulose column (Oligotex mRNA purification kit; QIAGEN). 20 μg of poly $(A)^+$ RNA was obtained from 1 mg of total RNA starting material (accounting for 2% of total RNA). The concentration was 0.6 μg/μl, and the A260/280 ratio was 2.0. Gel analysis of the isolated poly$(A)^+$ RNA is shown in the Figure 11, lane 3. High molecular weight mRNA species was enriched from the total RNA while bands of rRNA and tRNA species were clearly diminished.

5.3. cDNA library construction, titering, cloning efficiency, and insert size determination

The λZipLox-based cDNA library was constructed from about 10 μg of poly $(A)^+$ RNA isolated from yeast cells of *P. marneffei*. Quality control during the synthesis of the cDNA library, an alkaline analysis of a first strand yield, was shown in Figure 12. The picture shows a smear of cDNA molecules ranging from 0.5 to 9 kb in size, indicating the conversion of the mRNA to cDNA was properly occurred.

After the library construction, a primary cDNA library titer was $2.5x10^5$ pfu/ml, with up to 98 % of recombinant plaques determined by calculation from the blue/white plaque assay. Subsequent library amplification increased the titer to 5 x 10^9 pfu/ml with the same percentage of recombinant clones.

Ten random plaques were picked individually from the primary library and amplified library to check the size of inserts by PCR. The amplified products were analyzed on 1 % agarose gel as exemplified on Figure 13. The insert size was the size determined from the gel minus to a 130-bp of vector portion. Among 10 clones of the primary library, insert lengths varied from 0.7 to 2.5 kb (Figure 13A), and from 0.5 to 1.5 kb in amplified library (Figure 13B). An average value of insert length was 1.1 ± 0.4 kb in both primary and amplified library.

5.4. Isolation of actin-encoding clones from the cDNA library

To generate an actin probe for library screening, the PCR using ACT1 and ACT2 primers was performed on a genomic DNA of *P. marneffei.* Several bands were amplified (data not shown). However a 530-bp product of the expected size was found. The 530-bp DNA was purified from a gel and sequenced. The obtained sequence of 494 bases was subjected to a BLAST search which found significant similarity at both nucleotide (Figure 14A) and protein levels (Figure 14B) to actin-encoding genes. These data confirmed the isolation of an actin gene fragment. The PCR product was further labeled and used as a probe in DNA hybridization to select the actin-encoding clones from the cDNA library.

Two positive clones with insert size of 1,890 and 600 bp were isolated from plaque hybridization assay of a 10,000-pfu derived from the cDNA library. The clone with a longer insert (1,890-kb) was PCR and sequenced from the 5'-end. The sequencing result (Appendix A) was shown that this clone encoded an actin. This clone contained a full-length actin gene as shown by an alignment result of the obtained sequence to the first ATG of *Histoplasma capsulatum* actin mRNA (Figure 14C).

5.5. Isolation of Hsp70-encoding clones from the cDNA library

Three positive clones were isolated from antibody screening of the cDNA library with a monoclonal antibody to the hsp70 of *Histoplasma capsulatum*. PCR amplification to determine the insert sizes gave 1600, 1700, and 1,970-bp products. The 1,970-bp PCR product was subjected to 5' end DNA sequencing. The sequencing result of 323-bp (Appendix A) was analyzed for the similarity against all nucleotide and protein sequences in the databases. BLASTN and BLASTX similarity searches revealed that this clone contained a gene encoding Hsp70. Alignment of the P. marneffei hsp70 is shown in Figure 15. The obtained sequence was not included the first ATG.

5.6. Antibody responses of *P. marneffei*-infected patients' sera

Fifteen sera obtained from the *P. marneffei*-infected AIDS patients were used in western immunoblotting assay with the 8-day-old *P. marneffei* yeast antigens. The result is shown in Figure 16. The immune sera from different patients have the different blot patterns (Figure 16, lane 1-15). This result was consistent to the previous report by Vanittanakom *et al.* (1997). The sera with the different patterns of immunoreactivity (no. 9, 11, 13, 14, 15) were pooled and used in the cDNA library screening assay.

5.7. Antibody screening for the antigenic protein-encoding genes from the cDNA library

Up to 100,000 pfu of the yeast phase cDNA library of *P. marneffei* were screened with the pooled sera. The signals of positive reaction on the primary screening were clearly distinguishable from the background produced by negative pfu (Figure 17A). However, the surrounding negative plaques were included during the selection of the positive pfu. The repeat of the screening process with lower plating density was required until homogeneous positive signals were obtained (Figure 17B and 17C). Typically, second and third screenings were performed to purif the positive clones.

Twenty-eight positive plaques were selected and purified to homogeneity. They were subjected to PCR using T7 and SP6 primers to determine the size of the cDNA inserts. The PCR products were separated on a 1% agarose gel, and their sizes were compared with the DNA molecular weight standards. The size of the cDNA inserts, which were the apparent size minus the 130-bp of the flanking vector sequences, are shown in Table 3.

Table 3. Size of cDNA inserts of the antigenic protein-encoding clones.

Clone	Size (kb)	Clone	Size (kb)
P1	2.5	**P15**	1.0
P2	1.25	**P16**	1.6
P3	3.2	**P17**	0.9
P4	1.3	**P18**	1.0
P5	0.9	**P19**	3.0
P6	1.2	**P20**	1.7
P7	1.6	**P21**	0.8
P8	1.0	**P22**	0.01
P9	1.8	**P23**	1.1
P10	1.0	**P24**	1.5
P11	2.0	**P25**	1.2
P12	1.4	**P26**	1.0
P13	0.9	**P27**	1.3
P14	1.6	**P28**	1.6

5.8. Grouping of the clones encoding antigenic proteins.

To identify the cDNA clones sharing the same gene, the PCR products containing the cDNA inserts from all 28 clones were immobilized on Hybond N^+ membrane and hybridized with an alkaline phosphatase-labeled insert of each clone. High stringent washing and signal generation was then performed. The positive signals (spots) generated on the film by an individual probe indicated that the cDNA inserts of these clones shared common DNA sequence. In the other words, these clones were derived from the same gene. Some examples of the dot blot hybridization assay are shown in Figure 18. Grouping of the cDNA clones is summarized in Table 4.

Table 4. Grouping of the antigenic protein-encoding clones

Group number	Clone number
1	1, 2, 4, 8, 16, 19, 20
2	3
3	5, 23
4	6
5	7
6	9
7	10
8	11
9	12, 25, 27
10	13
11	14
12	15, 18
13	17
14	21
15	22
16	24
17	26
18	28

5.9. DNA sequence analysis of the clones encoding antigenic proteins.

The clones of groups 1-5 were chosen for whole-clone sequencing, while the remaining clone grouping were only partially sequenced. Whole-clone sequencing was also performed on clones P1, P2, P4, and P8 of group 1 (catalase-peroxidase group). The sequencing results showed that members of this group shared the same 3' DNA sequences. This result was expected because an oligo(dT) primer was used to synthesize the first strand cDNA. The primer bound at 3' end of the mRNA molecule and extended the cDNA chain to the 5'-part of every mRNA molecules.

The obtained DNA sequences were analyzed using web-based computer analysis programs as indicated in Materials and Methods (Table 2). The summaries of the sequence analysis and extracted biological information are shown in Figures 19 to 36 and Table 5. The restriction map analysis of clones in groups 1-5 are shown in Appendix B.

5.9.1) Analysis of group 1 (Figure 19)

Group 1 is composed of clones P1, P2, P4, P8, P16, P19, and P20. The insert sizes were 2.5, 1.25, 1.3, 1.0, 1.6, 3.0, and 1.7 kb, respectively. Whole clone DNA sequence of P1 is shown in Figure 19. The first nucleotide of clones P2, P4, and P8 were at nucleotide number 1,155, 1,242, and 1,392 in the P1 sequence, respectively. All clones in this group shared the 3' DNA sequence from the sequencing results (data not shown). Similarity search found that P1 DNA was a homologue of the gene encoding a catalase-peroxidase, a growing class of hydrogen peroxide detoxifying enzyme found in pathogenic bateria, archaebacteria, fungi and plants. Multiple sequence alignment showed a match of the complete length to the subject sequences at both nucleotide and polypeptide levels. The result from motif scan search confirmed finding of the peroxidase-encoding gene.

The P1 DNA sequence was subjected to six-frame translation to determine the coding frame and encoded polypeptide sequence. The forward frame 3 revealed an open reading frame of 748 amino acids without inappropriate stop codon. In addition, the result from BLASTX search showed that the polypeptide from the forward frame 3 was functionally identical with the catalase-peroxidase. This could give authentic conclusion that the forward frame 3 was the correct open reading frame. The P1 clone contained 68 nucleotides of 5'-untranslated region (UTR), 2,247 nucleotides of an open reading frame, and 176 nucleotides of 3'-UTR.

5.9.2) Analysis of group 2 (Figure 20)

Group 2 contained only P3 clone. BLASTN analysis of the whole-clone DNA sequence showed a significant hit to a gene for heparan sulfate-6-O-sulfotransferase of a zebra fish. However, when considered from the local alignment, only a short segment (46 of 3,162 bases) query was aligned to the subject sequence. Moreover, the 46-nt is in the 3' UTR (nucleotide 3117-3162). This suggested a high possibility that this 46-nt identity was spurious. The other hits to the sequence were for genes encoding a papatin-like protein, an unknown protein, and heparanase 2. However, these are also spurious for the same reason. Thus, the function of P3 clone could not be determined from the similarity found in BLASTN analysis. The BLASTX result showed a significant hit of forward frame 3 polypeptide to a hypothetical protein of *Shizosaccharomyces pombe* (fission yeast). The identity was 151 in 545 amino acids (27%) and the positive alignment was 270 in 545 (49%) from 669 amino acids of query sequence. These numbers were high enough to be considered this match as a true positive. Yet, the function of P3 could not be identified since it was similar to a hypothetical protein of unknown activity. An attempt to find its function by the motif scan found weak matches to profiles of proteasome protease and bipartite nuclear localization signal. In the context of no additional experimental information supported, these weak matches could only serve for prediction. Therefore, P3 clone contains a functionally unidentified gene.

The coding frame of P3 was readily identified from the six-frame translation. The open reading frame started at nucleotide 846 and stop at nucleotide 2,855. The encoded polypeptide is 669 amino acids in length. The predicted molecular weight of the monomeric protein should be 75.86 kDa. However, the potential N-glycosylation sites N-X-S/T were detected from PROSITE scanning analysis at the amino acid positions 377- 380 (NSTS), 513-516 (NASY), 642-645 (NLTI), 865-868 (NGTF), and 879-882 (NGTT). The molecular weight of P3 protein that is naturally synthesized from *P. marneffei* may be higher than the predicted weight.

5.9.3) Analysis of group 3 (Figure 21)

The members of group 3 were P5 and P23, with the insert sizes of 0.9 and 1.1 kb, respectively. Clone P23 was whole-clone sequenced. Six-frame translation revealed that P23 was the full-length clone by observing an open reading frame at nucleotide numbers 100 to 662. This group was functionally identified as the heat shock protein 30 (Hsp30). Notable hits were observed to the genes encoding the Hsp30 of *Aspergillus nidulans* and *Aspergillus oryzae* (both are fungi). BLASTX search results also showed high similarity to Hsp30. When the encoded polypeptide was used as a subject for similarity search by BLASTP program, a conserved domain for the small heat shock protein 20 family was detected. The fungal Hsp30s are belonging to this protein family. Result from motif scan search also matched with the Hsp20 family. These evidences indicate that P23 was the clone containing a gene encoding the small heat shock protein. The predicted molecular mass of this protein was 20.76 kDa.

5.9.4) Analysis of group 4 (Figure 22)

The entire insert (1.2-kb) of P6 clone was sequenced. The obtained whole-clone DNA sequence contained no poly (A) tail. The *Not*I site that was observed at the junction between the vector and the insert sequence was an internal *Not*I site of P6 DNA, not from the primer adapter like the former clones. This result indicates that the insert in P6 clone contains *Not*I. Therefore *Not*I digestion, which is a process in preparing of directional cDNAs in the cDNA library construction, purged the 3' sequence from the gene. Six-frame translation of the obtained sequence could identify

the open reading frame. The first methionine (Met) was recognized. However, the encoded polypeptide is truncated.

Functional analysis of the P6 sequence by BLASTN found no similarity to any sequences in the DDBJ/EMBL/Genbank databases as of March 10, 2004. The BLASTX result gave a significant match with a hypothetical protein of *Shizosaccharomyces pombe* and *Saccharomyces cerevisiae*. Even though the motif scan searching found a conserved SAND domain, the function of this gene could not be identified by a hit to a domain of unknown function. Thus, the function of clone P6 could not be identified.

5.9.5) Analysis of group 5 (Figure 23)

The member P7 contained a 1.6-kb insert. Whole clone sequencing was performed on this clone. The result sequence data showed that this clone was also trimmed at the 3'-end because the poly(A) sequence was missing. The encoded protein from the forward frame 2 started its first Met at the nucleotide 410, but no termination was recognized.

BLAST search could not establish the function of the gene carried by clone P7. However, an interesting observation was made based on the results of motif scan search. The encoded polypeptide contained the threonine (Thr)-, serine (Ser)-, proline (Pro)- and alanine (Ala)-rich regions. Thr and Ser are polar amino acids. The Thr- and Ala- rich regions are located at N-terminal part of the encoded polypeptide. Regularly, polar amino acids provide hydrophilic property to protein molecules (Branden & Tooze, 1999). Ala and Pro are uncharged amino acids. The regions that are rich in these amino acids are frequently found in transmembrane proteins. These findings inferred that the polypeptide encoded by P7 could be an integral membrane protein. A domain of unknown function contains two transmembrane regions was found at the amino acid position 179-279. This domain is found in a small set of bacterial integral proteins (InterPro description entry IPR007140).

5.9.6) Analysis of group 6 (Figure 24)

Group 6 contained clone P9, with an insert size of 1.8-kb. This clone was 5-prime sequencing. The sequence obtained was analyzed by BLAST. BLASTN found a similarity to an *fbpA* gene for fructose-1,6-bisphosphatase of *Aspergillus oryzae* and two other genes with unknown function from fungi. P9 DNA sequence aligned with the *fbpA* DNA sequence as shown in Figure 24. Further support for the BLASTN result came by significant hits to fructose-1,6-bisphosphatase proteins of fungi and plants shown by BLASTX analysis. Therefore, a possible function could be assigned to the protein of clone P9. Motif scan result found no match to the deduced amino acids. However, it is possible that the 21-amino acid sequence of P9 did not contain the conserve domain of this protein. If a longer DNA sequence or longer polypeptide chain of P9 could be achieved, the search should find some conserved residues.

5.9.7) Analysis of group 7 (Figure 25)

Group 7 contained a member clone P10. Its insert size was 1.0-kb. Five prime DNA sequence was obtained and analyzed for possible function of the gene embedded in this clone. Similarity search by BLASTN found three best hits to unknown genes of *Botrytis cinerea* and *Shizosaccharomyces pombe* (fungi). However, significant matches were found to the ribosomal genes at the forth and fifth hits. Alignment of the P10 DNA and the gene for ribosomal protein of *S. pombe* is shown in Figure 25. BLASTX analysis found significant similarity of the polypeptide from forward frame 1 to the ribosomal proteins of yeast, fish, and worm. Multiple sequence alignment indicated that they were homologues. A conserved domain for ribosomal protein was found. Additionally, a weak match to a death domain was found from the motif scan. This domain has been described as a region in the cytoplasmic tail of several proteins that are involved in cell-death signaling (Prosite document PDOC50017). The domain has been found mostly in the C-terminal region of proteins. However, this domain has never been reported in the ribosomal protein. Therefore, it is possible that finding of this domain by weak match scoring could be an artifact.

Determination of the coding frame by six-frame translation found several frames that could be recognized as the correct frames. However, only the forward frames were taken in account since directional cloning permits only the forward frames to be cloned. Thus, only two frames (forward frame 1 and 2) were heeded. When taken together with the result from BLASTX analysis, the forward frame 1 was chosen because it was functionally defined.

5.9.8) Analysis of group 8 (Figure 26)

The member of this group was clone P11. The insert size was 2.0-kb. Functional analysis of the five-prime, single pass DNA sequence by BLAST search and motif scan could not establish the biological function of this clone.

The correct coding frame of this clone could either be forward frame 2 or 3. Without the information of protein function from BLASTX analysis, a definite conclusion could not been made on the correct frame. Thus, the amino acid composition was not shown in the analysis of this clone.

5.9.9) Analysis of group 9 (Figure 27)

Group 9 contained P12, P25, and P27. The insert sizes were 1.4-, 1.2-, and 1.3-kb, respectively. BLASTN analysis of the DNA sequence of the 5'-end of P12 showed a hit to the *snaD* mRNA for spindle pole body associated protein of *Emericella nidulans* (fungus). However, the similarity score was only 52, with the identity to a small portion of this gene. The *snaD* gene is 3,213-nt long, and the open reading frame spans over the positions 964-2,985. BLASTN result showed the local alignment of P12 at nucleotide 259-316 to the nucleotide 711-654 of the *snaD* gene. Part of the alignment was out of the range of the *snaD* open reading frame. Hence, the hit from BLASTN could not be taken into account. Functional analysis of the BLASTX reveals that P12 protein could be a cytochrome C oxidase polypeptide. Only a single BLASTX hit was found with this protein from *Thermus thermophilus* (bacterium). Although the conserved domain could not be identified from the sequence, the polypeptide alignment contributed a clue to the functionality of these similar proteins.

Six-frame translation was readily recognized the forward frame 2 as a correct coding frame. Additional consideration from the BLASTX result also supported the result from translation.

5.9.10) Analysis of group 10 (Figure 28)

P13, which contained the insert size of 0.9-kb, was the member of this group. This is another clone whose function could not be identified by BLASTN search. It showed no significant similarity to any genes in the nucleotide databases. However, it found similarity hit to the NADH-ubiquinone oxidoreductase of *Neurospora crassa* (fungus) when searched by BLASTX. The similarity was significant as shown in the polypeptide sequence alignment. The translation frame, forward frame 1, could be unanimously identified from the six-frame translation analysis and from BLASTX matching to the NADH-ubiquinone oxidoreductase protein.

5.9.11) Analysis of group 11 (Figure 29)

The member of group 11 was the clone P14, which contained a 1.6-kb of insert DNA. Similarity search could not identify the function of this clone. Even though there were significant hits by BLASTX analysis, the biological function of the subject sequences have not been investigated. The motifs scan searching found weak matches to the zinc finger C_2H_2- and AT hook-motif, which could help in defining the functional configuration. Zinc finger domains are nucleic acid-binding protein structures that have been found in numerous DNA-binding proteins. AT-hook is also a domain that functions in DNA binding with the presence for AT-rich regions. Since both of them are the DNA-binding domains, it could be assumed that P14 protein functions as a transcription factor. The coding frame of P14 is forward frame 1. The polypeptide encoded from this clone matched with hypothetical proteins in *Aspergillus fumigatus* (fungus) and *Magnoporthe grisea* (plant).

5.9.12) Analysis of group 12 (Figure 30)

Clones P15 and P18 were the members of group 12. They contained the same size of insert length of 1.0-kb. The five-prime sequence of P15 is similar to the *MP1* gene encoding a cell-wall mannoprotein of *P. marneffei* when determined by BLASTN. The P15 DNA could be aligned to the *MP1* sequence as shown in Figure 29. Although significant matches were not found by BLASTX and motif scan, the possible function could be determine considering from match to the *MP1* gene. The *MP1* gene was previously identified from *P. marneffei* as an antigenic protein-encoding gene (Cao *et al.*,1998). In this study, a related gene was isolated. The correct open reading frame is forward frame 3 identified from the result of six-frame translation.

5.9.13) Analysis of group 13 (Figure 31)

Clone P17, member of this group, had an insert size of 0.9-kb. Its five-prime DNA sequence had no similarity to any genes in the DNA databank. However, BLASTX found significant similarities when search through the SWISSPROT database. The encoded polypeptide from forward frame 1 was similar to the glutathione peroxidase protein as shown in the multiple alignment result (Figure 31). The motif scan search found the conserved domain for glutathione peroxidase in the deduced amino acid sequence of clone P17. This match thus confirmed the possible glutathione peroxidase function of this clone.

P17 could be a full-length clone. The translation of P17 DNA sequence found the first Met at the nucleotide position 118. The number of amino acids of the reported glutathione peroxidases ranged from 157 to 163, therefore the gene transcript should have the size of at least 500 base pairs. The size of this clone corresponds to the expected size of a full-length transcript.

5.9.14) Analysis of group 14 (Figure 32)

Group 14 contained only one member, P21. The six-frame translation revealed two possibilities of the correct coding frames: forward frame 1 and frame 3. Since no significant similarity was made using BLASTX analysis. There was no supporting evidence to predict the correct reading frame. However, considering the size of this clone, which was only 0.8 kb, the more likely possibility was considered that forward frame 3 was the correct open reading frame given its longer uninterrupted sequence.

Functional analysis showed no similarity by BLAST analysis, and no match was found with the motif scan searching. Thus, the possible function could not be assigned to this clone.

5.9.15) Analysis of group 15 (Figure 33)

This group was discarded as an artifact. It contained P22 clone as a member. Its insert DNA was only 17-nt long, which was very unlikely to be a gene encoding protein.

5.9.16) Analysis of group 16 (Figure 34)

This group was functionally identified. The group contained P24, which contained a 1.5 kb of insert DNA. The 440-bases of 5' DNA sequence was demonstrated to be a part of thiamine synthesis enzyme-coding gene. This clone did not contain a full-length gene. The deduced amino acid showed a reading frame of 146-amino acids; no conserved motif was detected. However, the similarity search found significant matches at both nucleotide and polypeptide levels, as shown in the Figure 34. The genes for thymine synthase in *Aspergillus parasiticus*, *Neurospora crassa*, *Botrytinia fuckeliana*, and *Candida tropicalis* (all are fungi) were matched to the queried P24 DNA sequence. Also, BLASTX result found similarity hits to the thymine synthase. Multiple alignments at both nucleotide and polypeptide levels are shown (Figure 34).

5.9.17) Analysis of group 17 (Figure 35)

The 1.0-kb-clone P26 was five-prime sequenced. The DNA sequence of 444-bp was analyzed for the possible function. Similarity search could not establish the function of this clone. Motif scan search found the significant match to Thr-rich region. However, this match could not determine the function of protein encoded from P24. The coding frame was readily identified as the forward frame 2 from the six-frame translation.

5.9.18) Analysis of group 18 (Figure 36)

Clone P28 of this group contained a 1.6-kb insert. A five-prime DNA sequence of 466 nucleotides was subjected to similarity searching. The BLASTN analysis demonstrated homology to the *sdeA* gene, which encodes for stearic acid desaturase of the fungus *Emericella nidulans*. BLASTX analysis also gave significant hits to the acyl-CoA-desaturase. Multiple sequence alignment on the polypeptide level is shown. These results indicated that the P28 contained a gene encoding an enzyme in fatty acid biosynthesis even though the match from the motif scan searching was not found. The coding forward frame 3 of this gene was identified from the results of six-frame translation and the BLASTX.

5.10. Characterization of the *cpeA*

Group 1 was chosen for detailed characterization. The gene in this group encoded catalase-peroxidase, thus it was designated *cpeA*. This gene was the most striking since the similar gene was reported as a virulence factor in *Mycobacterium tuberculosis*, an intracellular pathogenic bacterium. Moreover, it was isolated in a high proportion (7 of 28 clones) among the positive clones.

The *cpeA* gene contained 2,490 nucleotides with an open reading frame of 2,247 bp encoding a protein of 748 amino acids with a calculated molecular mass of 82.4 kDa. BLAST sequence analysis showed homology to the catalase-peroxidase protein-encoding genes and their encoding amino acid sequence from several fungi and pathogenic bacteria (Figure 19). The nucleotide sequence of *cpeA* of *P. marneffei* was about 80 % identical to the sequence of *cpeA* from *Aspergillus nidulans* (Scherer *et al.*,

2002), 81 % with the *cat2* from *Aspergillus fumigatus* (Paris *et al.*, 2003), and 83 % with the *cat2* from *Neurospora crassa* (Peraza & Hansberg, 2002). The inferred amino acid sequence of *P. marneffei* CpeA had a similarity ranging from 45 to 69 % to the catalase-peroxidases in the SWISS-PROT database. The DNA sequence and deduced amino acid compositions, including putative heme ligands and active site residues, are shown in Figure 37. Residues, forming the peroxidase active site on the distal side of the heme and those that bind the proximal side of the heme in yeast cytochrome C peroxidase (CCP) (Finzel, Poulos & Kraut, 1984), are conserved in *CpeA*. Using comparative alignment of catalase-peroxidases and the sequence of *Hm*CP, the catalase-peroxidase of *Haloarcula marismortui* which the crystal stucture has been characterized (Yamada *et al.*, 2002), predictions are made that the N-terminal half of the *P. marneffei* CpeA, contrary to the C-terminal one, binds one heme group. The residues that play a role in catalysis, folding and structural stability, i.e. W95, Y218, M244, R409 in *Hm*CP, are highly conserved in CpeA (Figure 37).

5.11. Southern blot analysis of *cpeA*

Southern blot analysis was carried out to investigate the possibility that multiple *cpeA* copies are present in the genome of *P. marneffei*. Relatively low stringency washes were used in the hybridization procedures. In accordance with the restriction map of *cpeA* (Appendix B), the results indicated that the *P. marneffei* genome contained only a single copy of the *cpeA* gene (Figure 38). From the picture, the uncut genomic DNA of *P. marneffei*, F4 strain gave a faint signal on the film. This might be due to stearic hindrance of intact DNA accession of the probe. The enzymes that have no restriction site in the *cpeA* gene (*Xho*I, *Xba*I, *Eco*RI, *Hin*dIII, *Pvu*II, and *Eco*RV) resulted in only one band, whereas those that have one restriction site (*Pst*I and *Bam*HI) and two restriction sites (*Sal*I) displayed 2 and 3 bands, respectively. These results indicated that *P. marneffei* genome contained a single copy of *cpeA* gene.

5.12. Expression of *cpeA* transcript

The expression of *cpeA* during development of both the yeast and mold phases of *P. marneffei* was examined by Northern blot analysis. The results showed a single transcript of approximately 2.5-kb *cpeA* mRNA, consistent with the size of obtained full-length cDNA clone. The *CpeA* mRNA transcript was nearly absent at 25°C except for some low level of expression at 48 h of incubation. However, this expression was no longer present at 72 h. Expression of *cpeA* seemed to be induced by incubation at 37°C with the high level observed at 72 h after temperature shift (Figure 39A). Reverse transcription PCR was used to confirm the expression of *cpeA*. The results generally supported the Northern blot analysis by showing the presence of the *cpeA* transcript at every time point of incubation at 37°C, while the transcript was present only at 48 h of incubation at 25°C (Figure 39B). Direct PCR of the DNase-treated RNA samples failed to amplify the *cpeA* gene fragment. This result indicated no DNA contamination in the RNA samples used in RT-PCR (data not shown).

5.13. Cloning of the *cpeA* gene fragment into pRSET B expression vector

Cloning strategy is shown in Figure 40. Initially, the insert DNA was amplified from clone P2 using M13R and P1-*Eco*RI primers. An amplified fragment contained 120 nucleotides of the pZL1 vector and 1,200 nucleotides of the insert *cpeA* gene, resulting in the expected size of 1.32 kb. *Bam*HI and *Eco*RI digestion of the 1.32-kb amplified product would generate a 0.95-kb fragment, which corresponded to the nucleotides 1,383-2,315 of *cpeA* gene or amino acid number 461-748 of CpeA protein. However, the PCR result found two bands of amplified products: a 1.4-kb, which is larger than the expected size and a non-specific band of 800-bp. (lane 2, Figure 41). The PCR product was *Eco*RI and *Bam*HI digested (lanes 3 and 4, Figure 41). The desired fragment of about 1-kb was gel purified (QIAquick gel purification kit, QIAGEN) and cloned into the *Eco*RI/*Bam*HI-cut pRSET B vector.

Only sixty-six colonies were grown on 25 μg/ml ampicillin-LB agar plate, whereas the 10 ng of uncut pRSET B plasmid that served as a transformation control gave more than 10^5 transformants on the ampicillin-LB plate. Plasmids from all 66 transformants were isolated using alkaline lysis method in the small-scale plasmid

isolation (Sambrook & Russell, 2001). Analysis of the isolated plasmid found only 6 in 66 transformants containing inserts (colony number 6, 28, 35, 42, 45, and 51). Digestion of all six recombinant plasmids with *Bam*HI and *Eco*RI revealed unexpected results. Digestion of the recombinant plasmid should release a 1-kb fragment from the 2.9-kb pRSET B vector. However, they gave 3 fragments of 0.25, 1.1, and 2.9 kb (Figure 42). Thus the recombinant plasmid number 28 was subjected to DNA sequencing. Result from sequence analysis found that the recombinant plasmid indeed contained the *cpeA* fragment, but it was in the inverse direction. Moreover, the *cpeA* fragment that was embedded in the recombinant plasmid contained two *Bam*HI sites. One was the site inside the *cpeA* gene, and the other one, which served as a cloning site was derived from the pZL1 multiple cloning sites. The *Eco*RI site that served as a cloning site was also the site in the pZL1 plasmid instead of that previously designed to generate from the P1-*Eco*RI primer (Figure 43). This eventual wrong direction recombinant plasmid resulted from unexpected disposition of primer binding. The P1-*Eco*RI primer that was primarily designed to bind at the stop codon of the *cpeA* gene bound at unexpected site, approximately 100-bp beyond to the designed binding site. The extended 100-bp fragment contained an additional *Bam*HI site of the pZL1 vector, which is located outside the insert part of P2 clone (Figure 44). This additional hundred base pairs fragment resulting in an amplified product of 1,400 bp is shown in the Figure 41, lane 2. Therefore, the obvious very low transformation efficiency and the evidence of 3 unexpected fragments after *Eco*RI/*Bam*HI digestions (Figure 42) could be explained. Small portion of incomplete restriction digestion fragments that could be cloned in the pRSET B plasmid with *Bam*HI/*Eco*RI arms led to low efficiency of ligation and transformation (Figure 44). Thus, the recombinant plasmids containing two *Bam*HI sites and single *Eco*RI site resulted in three bands following the digestion.

Inversion of the insert was then performed on the recombinant plasmid #28. *Bam*HI digestion of the clone #28 released a 1.1-kb fragment from the pRSET B. This fragment contains the *cpeA* gene fragment from nucleotide 1,383 to polyA tail (1.1-kb) and 15 nucleotides of pZL1. The 1.1-kb fragment was re-ligated to a dephosphorylated *Bam*HI-cut pRSET B. The resulting recombinants were composed of 2 populations, one with the insert in the right orientation and the other insert with the wrong orientation (Figure 45). Digestion with *Not*I and *Hind*III should give fragments of 3.8-kb and 0.056-kb for the recombinant plasmid with the insert in the right orientation, while the wrong orientation recombinant should give 2.8-kb and 1.07-kb fragments (Figure 46A). Fourteen out of approximately 10^5 transformants were digested with *Not*I and *Hind*III to check their orientations. Figure 46B shows that 7 transformants in lanes 3, 5, 8, 9, 12, 13, and 14 contain recombinant plasmids with the proper direction for the expression of CpeA-His tag fusion protein.

5.14. Expression of the recombinant CpeA-His_6 tag fusion protein

The CpeA fusion protein is composed of 31-residues of N-terminal tag polypeptide and 311 residues of the CpeA. Thus, an expected size of the fusion protein should be 34 kDa. Two recombinant plasmids with proper direction (#5 and #8), and one with an improper direction (#1) of the insert DNA were used to transform the BL21(DE3)pLysS *E. coli* expression strain. Initially, the conditions for optimal expression were determined. It was found that a condition of 1 mM IPTG and 37 °C incubation could induce a high expression from both transformant #5 and #8 (contain insert in the right direction) as the inclusion bodies. The transformant #1 (contain insert in the wrong direction) showed no expression of the fusion protein (Figure 47). Variation of induction time when using 1 mM IPTG is shown in Figure 48. The induction time of 1.5 to 6 h gave similar quantity of the fusion protein from the transformant #8. The result was the same in transformant #5 (data not shown). Thus an optimal induction time could be at any time point between 1.5 to 6 h. Additional studies used the time point at 2 h for further experiments. An optimization of the IPTG concentration and induction temperature when using 2 h of induction time was compared as shown in Figure 49. Soluble and insoluble fractions were analyzed. Even

though the IPTG concentration was decreased to 0 mM and the induction temperature was lower to 25 °C, the fusion protein was aggregated in the inclusion bodies. Since the denaturing purification method would not affect the application of this fusion protein, thus the denaturing purification method could be used to purify the insoluble fusion protein. Therefore, the induction temperature at 37 °C was chosen since larger amount of the fusion protein was produced, and the IPTG concentration at 0.05 M was chosen for economically reasons. In conclusion, the suitable conditions for the CpeA-His_6 tag fusion protein expression were the addition of 0.05 M of IPTG to the bacterial culture for 2 h at 37 °C.

5.15. Purification of the CpeA-His_6 tag fusion protein

The fusion protein was purified from an induced culture of transformant #5. Twelve fractions were collected. Analysis of 10-µl from each fraction by SDS-PAGE and Coomassie blue staining found that the 34-kDa fusion protein was purified from the *E. coli* lysate in fraction number 4-9 (Figure 50). Small amount of fusion protein was eluted in the first 3 fractions, where some *E. coli* proteins were eluted. Beginning from fraction no. 4, a large amount of fusion protein was eluted. The peak of protein were seen in the fractions no. 6 and 7. The amount of eluted protein was decreased until it could not be seen using the Coomassie blue staining at fraction number 10. Amount of the proteins in fractions 4 to 9 determined by Bradford dye binding assay were 0.35, 0.81, 0.97, 0.93, 0.55, and 0.06 mg, respectively. Total amount of the eluted protein from fractions 4-9 was 3.67 mg from 50-ml culture volume.

5.16. Immunoblot analysis of the CpeA-His_6 tag fusion protein

Western immunoblot analysis was performed on 20 serum samples: 15 individual serum from *P. marneffei*-infected AIDS patients, a serum from non-*P. marneffei* infected AIDS patient, a pooled serum (n = 3) from fungal laboratory personnels, a pooled sera (n = 3) from normal healthy people in Chiang Mai (endemic regions). A pooled serum (n = 3), which possessed positive reactivities to *P. marneffei* antigen (previous study, Poolsri, 1999) was used as positive control. Normal serum of people in non-endemic region (Sigma) was used as a negative control. Each serum was used

in immunoblot assay with both nitrocellulose strips containing crude protein antigens of 3-day-old yeast culture of *P. marneffei* and the purified fusion protein. Result is shown in Figure 51. Panel A shows the immunoreactivities of individual serum to *P. marneffei* crude antigen. Panel B shows the reactivity of the same individual serum samples to the CpeA fusion protein at the molecular weight of 34-kDa. In the panel B, positive control serum (lane 18), which was a pool of *P. marneffei*-infected AIDS patients' sera, shows strong positive signal at the fusion protein band. Negative control serum (lane 19), which was the pool of normal healthy individuals in non-endemic region, did not react to the blotted protein. Pooled sera obtained from normal healthy people (n = 3) in Chiang Mai (lane 20), also show negative signal to the fusion protein. However, a pooled sera from fungal laboratory personnels showed weak positive reactivity to the fusion protein. This level of reactivity was used as a baseline to consider positive results when tested with culture confirmed-*P. marneffei*-infected sera. Eight in fifteen samples were considered positive, which were serum number 2, 3, 4, 7, 8, 11, 13 and 15. They were accounted to 53.3% of all tested infected sera. Immunoblot results in the panel B correspond with those in the panel A, but more intensity bands were seen in panel B. This indicated that there were specific antibodies to both the CpeA-fusion protein and its cognate native protein in the tested sera. Serum from an AIDS patient without *P. marneffei* infection (lane 17) was negative to the CpeA-fusion protein. The positive signals to the bands beside the 34-kDa fusion protein could be the reactivities to remnant *E. coli* protein. This result suggested that the fusion protein was not purified to homogeneity, and there were anti-*E. coli* antibodies in the patient's sera.

In addition, the fusion protein was tested by immunoblot assay with 10 serum samples from several mycotic patients: one from candidiasis patient, one from aspergillosis patient, one from histoplasmosis patient, and seven cryptococcosis patients. All sera gave negative results to the fusion protein (Figure 52). Moreover, testing with two sera from *Mycobacterium tuberculosis*-infected AIDS sera also showed no reactivity at the 34-kDa band (Figure 52). These results show that this CpeA fusion protein could be a potential diagnostic marker for *P. marneffei* infection in HIV-infected patients.

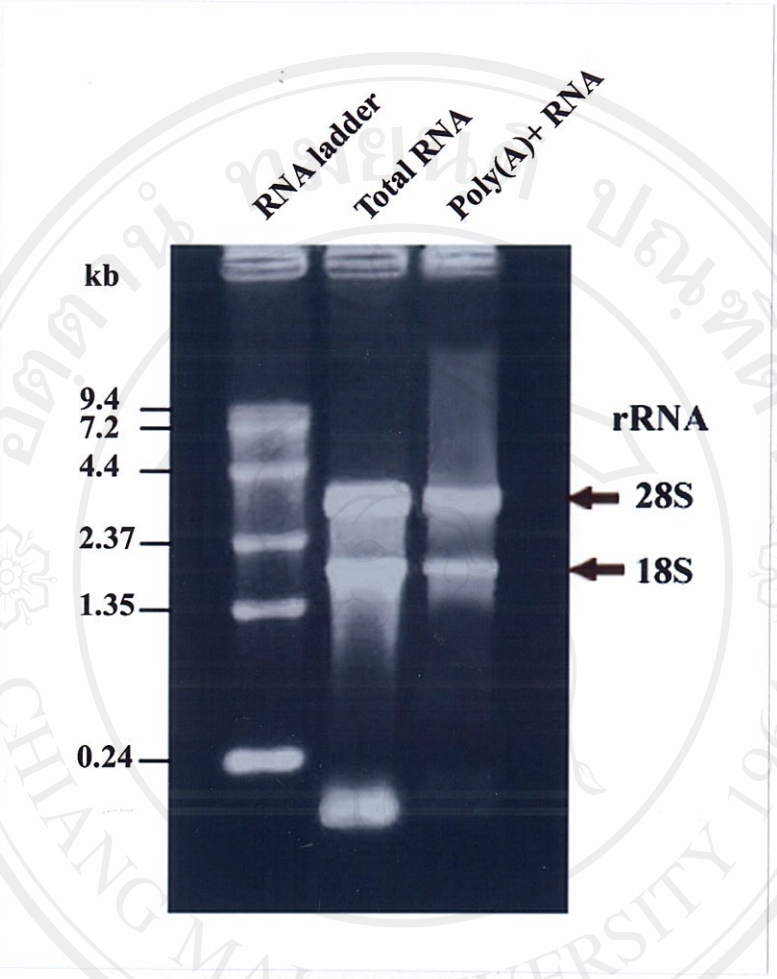


Figure 11. Formaldehyde gel analysis for integrity of isolated total and poly(A)$^+$ RNA. Four micrograms of RNA from total RNA (lane 2) and poly (A)$^+$ purification (lane 3) samples were electrophoresed through a 1.5% denaturing formaldehyde gel. Both RNAs showed integrity. The poly(A)$^+$ RNA was enriched as visualized by the decreased rRNA and tRNA population. The size of RNA ladder (lane 1, Gibco BRL) is shown on the left.

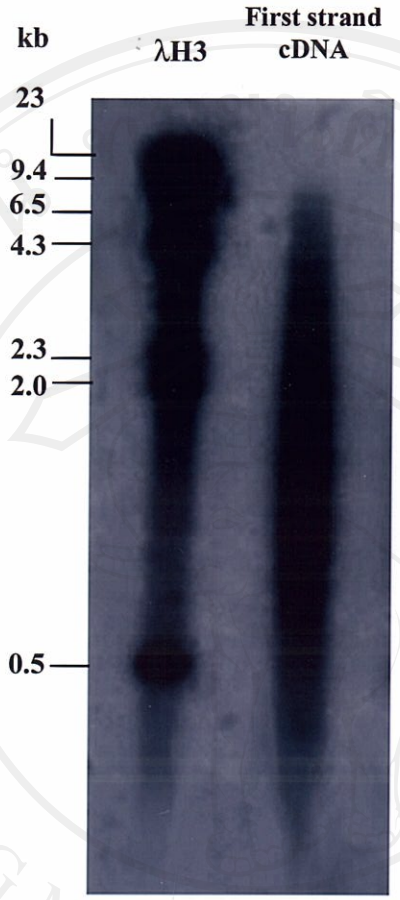


Figure 12. Analysis of the first strand cDNA. Two microliters aliquot of first strand reaction was analyzed on 1.4% alkaline agarose gel to determine first strand cDNA size distribution. The figure shows the sizes of first strand cDNAs varied from less than 0.5 to 9 kb. ^{32}P-labeled lambda *Hin*dIII (λH3) was used. The size in kb is shown on the left.

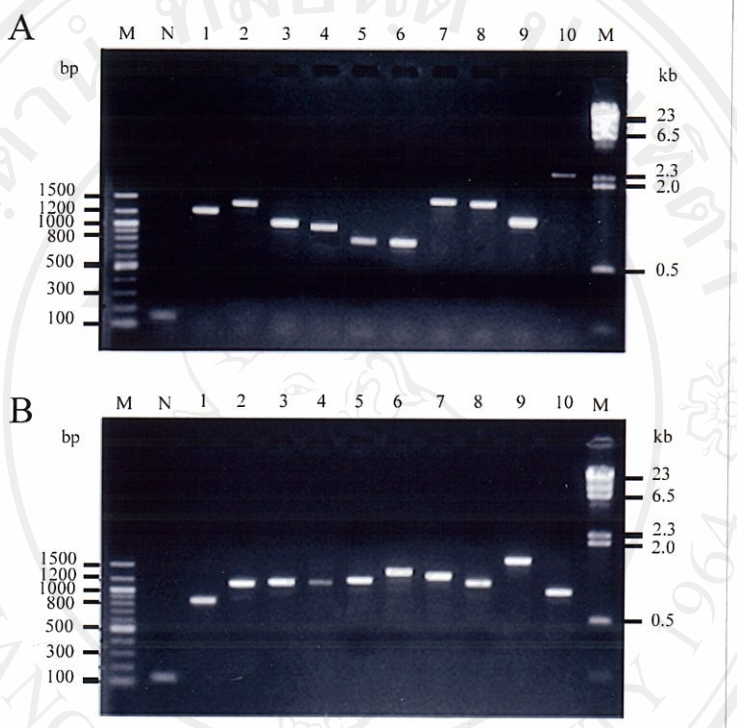
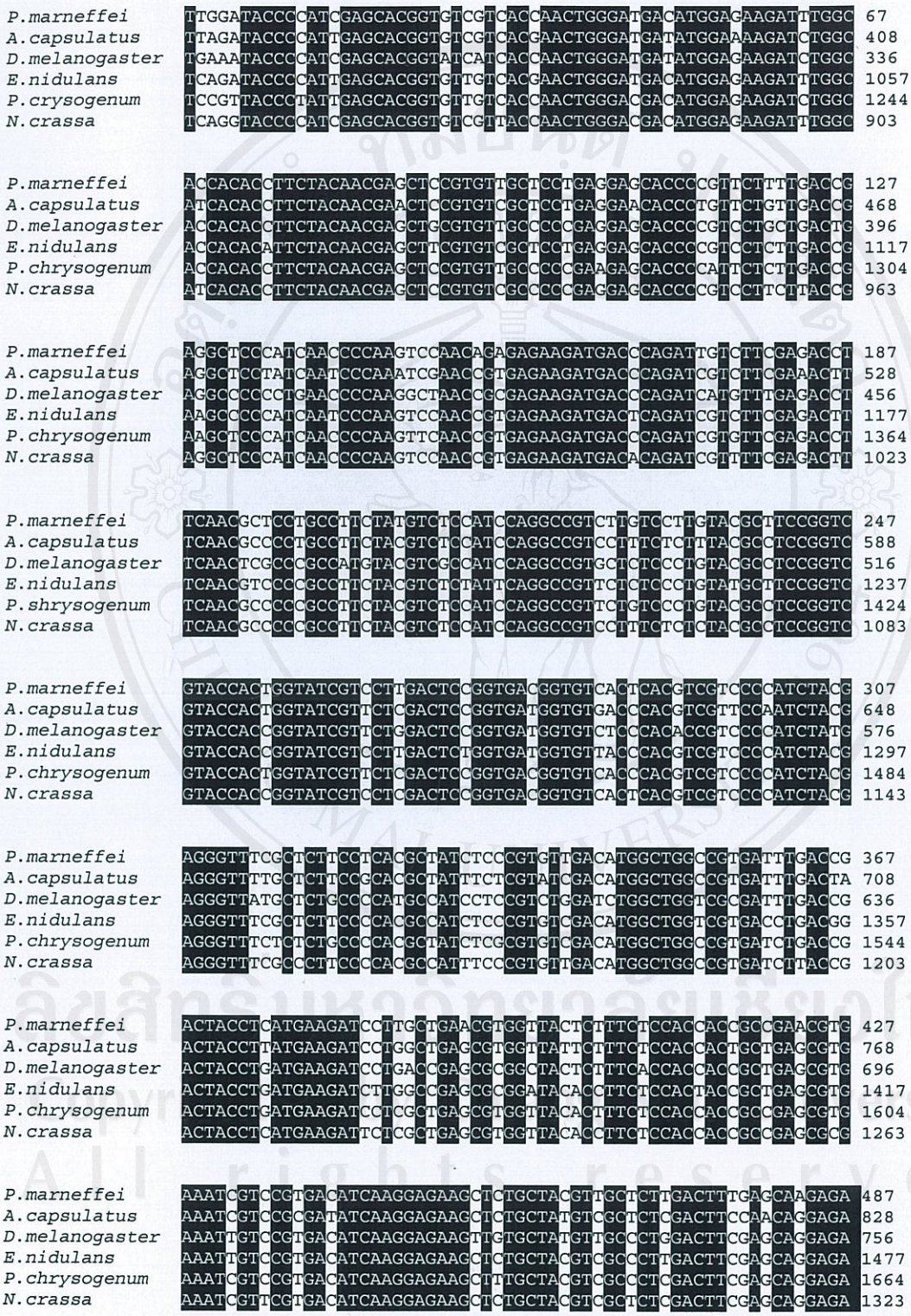


Figure 13. Inserted size determination by PCR. Ten plaques from the primary and amplified libraries were randomly picked and the insert portions were amplified. The PCR products were then analyzed on 1 % agarose gel. A, shows the insert size distribution from the primary library and B, from the amplified library. M = 100-bp marker (New England Biolabs) and N = product from clone with no insert.

A.



B.

```
P. marneffei        ---------YPIEHGVVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPINPKSNREKM  55
N. crassa           SKRGILTLRYPIEHGVVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPINPKSNREKM 119
A. crysogenum       SKRGILTLRYPIEHGVVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPINPKSNREKM 119
A. nidulans         SKRGILTLRYPIEHGVVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPINPKSNREKM 120
E. dermatitidis     SKRGILTLRYPIEHGVVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPINPKSNREKM 119

P. marneffei        TQIVFETFNAPAFYVSIQAVLSLYASGRTTGIVLDSGDGVTHVVPIYEGFALPHAISRVD 115
N. crassa           TQIVFETFNAPAFYVSIQAVLSLYASGRTTGIVLDSGDGVTHVVPIYEGFALPHAISRVD 179
A. crysogenum       TQIVFETFNAPAFYVSIQAVLSLYASGRTTGIVLDSGDGVTHVVPIYEGFALPHAIARVD 179
A. nidulans         TQIVFETFNAPAFYVSIQAVLSLYASGRTTGIVLDSGDGVTHVVPIYEGFALPHAISRVD 180
E. dermatitidis     TQIVFETFNAPAFYVSIQAVLSLYASGRTTGIVLDSGDGVTHVVPIYEGFALPHAISRVD 179
                                                                              :

P. marneffei        MAGRDLTDYLMKILAERGYSFSTTAEREIVRDIKEKLCYVALDFEQEIQT---------- 165
N. crassa           MAGRDLTDYLMKILAERGYTFSTTAEREIVRDIKEKLCYVALDFEQEIQTAAQSSSLEKS 239
A. crysogenum       MAGRDLTDYLMKILAERGYTFSTTAEREIVRDIKEKLCYVALDFEQEIQTAAQSSSLEKS 239
A. nidulans         MAGRDLTDYLMKILAERGYTFSTTAEREIVRDIKEKLCYVALDFEQEIQTASQSSSLEKS 240
E. dermatitidis     MAGRDLTDYLMKILAERGYSFSTTAEREIVRDIKEKLCYVALDFEQEIQTASQSSRLEQS 239
                                      :
```

C.

```
P. marneffei        ------CACACACACACACATCCTTCATCATGGAAGAGGAAGTCGCTGCTCTCGTTATTG  54
H. capsulatum       TCCTTTGCCACCTCGTTAAGTAGCCCACAATGGAGGAAGAAGTCGCTGCCCTCGTTATCG 180
                           ***     *  *    **  ***** ** ********** ******** *

P. marneffei        ACAATGGTTCGGGCATGTGCAAGGCCGGTTTCGCCGGTGATGACGCACCCCGAGCCGTGT 114
H. capsulatum       ACAATGGATCCGGTATGTGCAAGGCCGGTTTCGCTGGCGATGATGCTCCGCGGAGTGTCT 240
                    ******* ** ** ******************** ** ***** ** ** **    ** *

P. marneffei        TCCCTTCGATCGTTGGTCGTCCCCGTCACCATGGTATCATGATTGGTATGGGCCAAAAGG 174
H. capsulatum       TCCCTTCGATTGTGGGCCGTCCTCGCCATCATGGTATTATGATTGGTATGGGCCAGAAGG 300
                    ********** ** ** ***** ** ** ******** ***************** ****

P. marneffei        ACTCCTATGTCGGTGATGAGGCACAGTCTAAGCGTGGTATCCTCACCTTGAGATACCCCA 234
H. capsulatum       ACTCCTATGTTGGTGATGAAGCACAATCCAAGCGTGGTATCCTTACCCTTAGATACCCCA 360
                    ********** ******** ***** ** ************** *** * **********

P. marneffei        TCGAGCACGGTAGTCGTCACCAACTGGGATGACATGGA---------------------- 272
H. capsulatum       TTGAGCACGGT-GTCGTCACGAACTGGGATGATATGGAAAAGATCTGGCATCACACCTTC 419
                    * ********* ******** ********** *****
```

Figure 14. Sequence alignment results to confirm the isolation of a fragment from an actin-encoding gene. The alignment compares a 494 bases of a 530-bp fragment that was amplified from genomic DNA of *P. marneffei* to actin-encoding genes from other fungi (A). An alignment of translated amino acids from the 494 bases to actin sequences of several fungi in SWISS-PROT database (B). Alignment of 5' sequence of the actin-encoding clone obtained from screening of the cDNA library to an actin mRNA of *Histoplasma capsulatum* (C).

A.

```
P. marneffei        CGTTAACAACGCTGTCATCACTGTCCCCGCCTACTTCAACGACTCCCAGCGTCAGGCCAC 152
T. rubrum           CGTCAACAACGCTGTCATCACTGTCCCTGCCTACTTCAACGACTCCCAGCGTCAAGCTAC 162
C. elegans          ------------GTTGTCACTGTCCCGACTTACTTCAACGACTCGCAGCGTCAAGCCAC 47
P. brasiliensis     TGTTAACAACGCTGTCGTTACTGTCCCTGCCTATTTCAACGATTCCCAGCGTCAAGCTAC 167
D. melanogaster     GGTCACCCATGCCGTCGTTACCGTGCCAGCCTACTTCAACGATGCCCAGCGTCAGGCCAC 81
A. capsulatus       TGTCAACAATGCCGTCGTCACCGTCCCCGCCTACTTCAACGACTCCCAGCGTCAAGCTAC 180

P. marneffei        CAAGGATGCTGGTCTCATTGCCGGTTTGAACGTCCTCCGTATCATCAACGAACCTACTGC 212
T. rubrum           TAAGGACGCCGGTCTCATTGCCGGTCTCAACGTCCTCCGTATCATCAACGAGCCCACTGC 222
C. elegans          CAAGGATGCCGGAGCCATCGCTGGACTCAACGTTCTCCGTATCATCAACGAGCCAACCGC 107
P. brasiliensis     CAAGGATGCTGGTCTCATCGCTGGCCTGAATGTCCTTCGGATCATCAACGAACCCACTGC 227
D. melanogaster     CAAGGATGCTGGTGTTATTGCCGGTCTGCAGGTGATGCGCATCATCAACGAACCCACTGC 141
A. capsulatus       CAAGGATGCTGGTCTTATTGCCGGCCTCAATGTCCTCCGTATCATTAACGAGCCCACTGC 240

P. marneffei        TGCCGCCATTGCCTACGGTCTCGACAAGAAGGTTGAGGGTGAGCGCAACGTTCTCATCTT 272
T. rubrum           TGGTGGTATTGCCTACGGTCTCGACAAGAAGGCCGACGGTGAGCGCAACGTTCTCATTTT 282
C. elegans          TGCAGCTATCGCTTACGGACTTGACAAGAAGGGACACGGAGAACGCAACGTTCTTATCTT 167
P. brasiliensis     TGCCGCCATTGCCTATGGTCTTGACAAGAAGGCCGAGGGTGAACGCAACGTTCTCATCTT 287
D. melanogaster     TGCGGCTATTGCCTACGGTCTGGACAAGAAGG---AGGGGGAGAAGAACGTGCTTGTGTT 198
A. capsulatus       GGCTGCTATTGCTTACGGTCTCGACAAGAAAGCTGACGGTGAACGCAATGTTCTGATCTT 300

P. marneffei        CGATCTTGGTGGTGGTACCTTCGATGTTTCTCTCCTCACCATCGAGGACGG--------- 323
T. rubrum           CGATTTCGGTGGTGGTACCTTCGATGTCTCCCTCCTGACCATCGAAGAGGGTATTTTCGA 342
C. elegans          CGATCTTGGAGGTGGTACCTTCGATGTCTCCATTCTTACCATTGAGGACGGAATCTTCGA 227
P. brasiliensis     CGACTTGGGTGGGGGCACCTTTGATGTCTCCCTCCTGACCATTGAAGAGGACATCTTCGA 347
D. melanogaster     CGATTTGGGCGGCGGCACCTTCGATGTCTCCCTGCTGACCAT----------------- 240
A. capsulatus       CGACTTGGGCGGCGGTACTTTCGATGTGTCTCTCCTCACCATCGAAGAGGGTATCTTCGA 360
```

B.

```
P. marneffei        ---------------------------------------------DPRVRQFTPEEISSM 15
P. brasiliensis     PSNTVFDAKRLIGRKFADPEVQSDMKHFPFKVIDKAGKPVISVEFKGEEKQFTPEEISSM 120
H. capsulatum       PANTVFDAKRLIGRKFADPEVQADMKHFPFKITDKGGKPVIQVEFKGETKEFTPEEISSM 120
T. verrucosum       PINTVFDAKRLIGRKFNDAEVQADMKHFPFKVVEKSGKPIVQVEFKGEEKQFTPEEISSM 120
A. nidulans         PHNTVFDAKRLIGRRFGDAEVQADMKHWPFKVVDKSGKPIIEVEFKGETKQFTPEEISSM 120
                                                      . . ::

P. marnefei         VLIKMRETAEAYLGGTVNNAVITVPAYFNDSQRQATKDAGLIAGLNVLRIINEPTAAAIA 75
P. brasiliensis     VLTKMRETAESYLGGTVNNAVVTVPAYFNDSQRQATKDAGLIAGLNVLRIINEPTAAAIA 180
H. capsulatum       VLTKMRETAEAYLGGTVNNAVVTVPAYFNDSQRQATKDAGLIAGLNVLRIINEPTAAAIA 180
T. verrucosum       VLTKMRETAEAYLGGTVNNAVITVPAYFNDSQRQATKDAGLIAGLNVLRIINEPTAAAIA 180
A. nidulans         VLTKMRETAEAFLGGTVNNAVITVPAYFNDSQRQATKDAGLIAGLNVLRIINEPTAAAIA 180
                          ::         :

P. marneffei        YGLDKKVEGERNVLIFDLGGGTFDVSLLTIEDG------------------------- 108
P. brasiliensis     YGLDKKAEGERNVLIFDLGGGTFDVSLLTIEEDIFEVKSTAGDTHLGGEDFDNRLVNHFV 240
H. capsulatum       YGLDKKADGERNVLIFDLGGGTFDVSLLTIEEGIFEVKSTAGDTHLGGEDFDNRLVNHFV 240
T. verrucosum       YGLDKKAECERNVLIFDLGGGTFDVSLLTIEEGIFEVKSTAGDTHLGGEDFDNRLVNHFV 240
A. nidulans         YGLDKKVEGERNVLIFDLGGGTFDVSLLTIEEGIFEVKATAGDTHLGGEDFDNRLVNHFV 240
                       .:         :.
```

Figure 15. Homology of *P. marneffei* Hsp70 to that of other species. DNA sequence alignment of the obtained sequence from Hsp70-encoding clone of *P. marneffei* to the sequences *hsp70* from other organisms (A). The deduced amino acids were highly conserved among fungal Hsp70 (B).

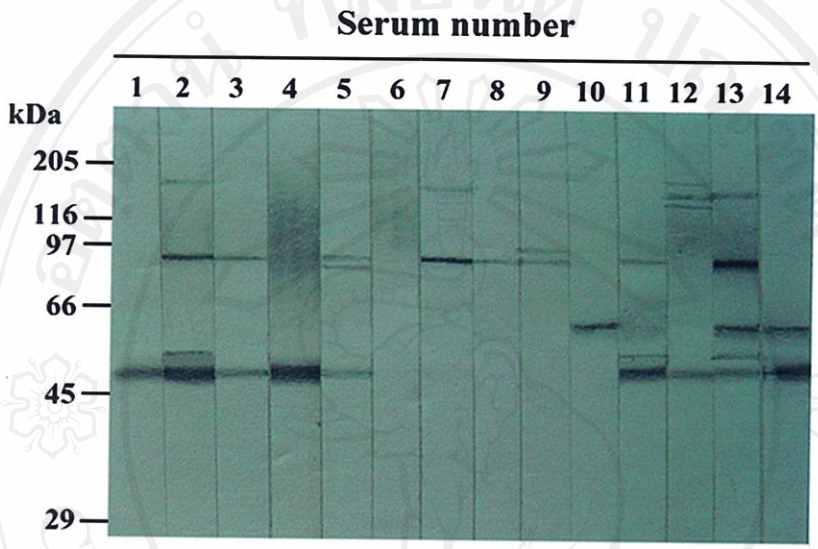


Figure 16. Antibody responses of *P. marneffei*-infected patients' sera to antigens of *P. marneffei*. Fourteen sera obtained from *P. marneffei*-infected AIDS patients were reacted to the 8-day-old yeast antigen of *P. marneffei*. Individual sera gave different pattern of immunoreactivities. The sera with distinct reactivities to *P. marneffei* antigen (number 2, 5, 9, 12, and 13) were chosen.

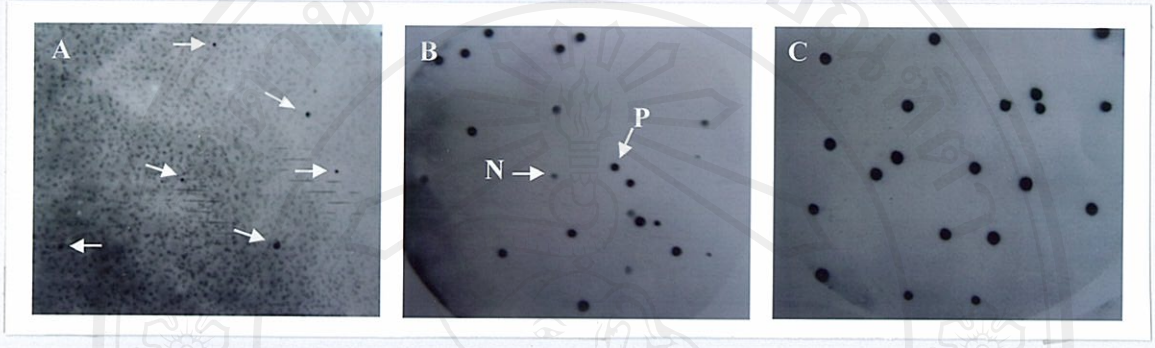


Figure 17. Positive signals from the screening of the cDNA library of *P. marneffei* with the pooled patients' sera. A) Primary screening. Arrows pointed at the positive signal position prominently seen on the negative background clones. B) Secondary screening. Bacteriophages from the primary screening were plated with low pfu (200-400 pfu). The positive plaques (P) were distinguished from those negative (N) clones. C) Tertiary screening. The positive plaque from well-isolated secondary screening was repeatedly screened. Homogeneous positive signals were generated, indicating that the positive clone was purified.

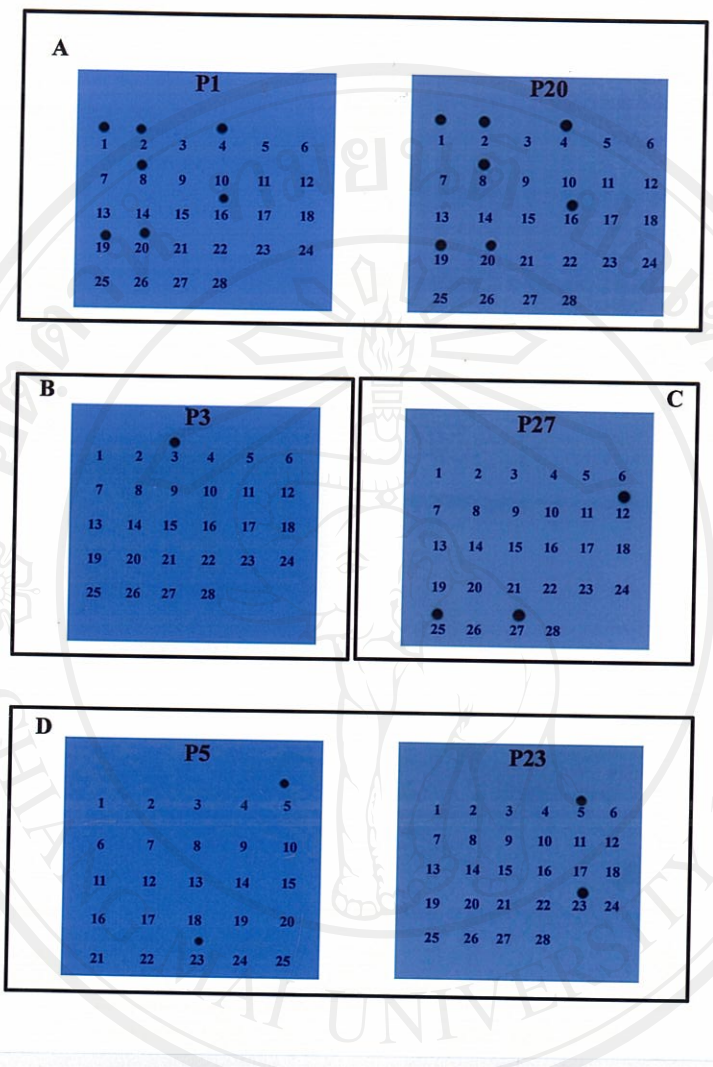


Figure 18. Dot blot hybridization assay for grouping of the antigenic protein-encoding clones. Positive signals generated on each probe were categorized into the same group. A) Representatives from group 1, positive signals generated were the same in clones of group 1 (consisting of clones P1, 2, 4, 8, 16, 19, and 20). B) group 2, the unique clone P3. C) group 9, consisting of clones P12, 25 and 27. D) group 3 had P5 and P23 as members.

GROUP 1

1.1) Members: P1, P2, P4, P8, P16, P19, P20

1.2) DNA sequence of P1
GC Content of the ORF: 56%

```
1       CGGACGCGTG GGCGTACTCG AATCCCCCAC ACTGTCTGAA CTATCTTATT     50
51      TGAAGCAAAC CAAACAAGAT GGCTGAGAGC AAGTGTCCCG CTCACCAGCA    100
101     TGTGTTGAAG GCCAACGTCG GCGGTGCTGG TACCAGCAAC CAAGATTGGT    150
151     GGCCAGACCG CTTGAAGCTT AACATCCTCC GCCAAAACAA CCCCGTCTCC    200
201     AACCCTCTGG GCGAGGAATT TGACTATGCC GCCGCCTTCA ACAGCCTAGA    250
251     TTACTTTGCG CTCAAGAAGG ATATTCAAGA TCTGATGACT GACTCCCAGG    300
301     ACTGGTGGCC GGCTGACTTT GGCCACTATG GTGGTCTCTT TATTCGTATG    350
351     GCCTGGCATA GTGCCGGTAC CTACCGAGTC GCCGACGGTC GAGGCGGCGG    400
401     TGGCGGCGGC CAACAGCGCT TTGCTCCTCT CAACAGCTGG CCCGACAATG    450
451     TCGGTCTCGA CAAGGCCCGC CGTTTGTTGT GGCCCATCAA GCAGAAATAC    500
501     GGAAACAAGA TCTCGTGGGC GGATCTCCTA TTGCTCACTG GTAACGTCGC    550
551     CCTTGAGTCC ATGGGTTTCA AGACCTTTGG TTTCTCTGGC GGTCGTGCCG    600
601     ACACATGGGA AGTGGATGAG TCAGCCAACT GGGGAGGGGA AACCACCTGG    650
651     CTAGGCAATG ACGTCCGCTA CTCCGGCGGT AAGGCTGATC ACAAGGATAT    700
701     CCACAACCGT GACTTGGACA AGCCACTGGC CGCTGCCCAC ATGGGTTTGA    750
751     TCTATGTCAA CCCCGAAGGT CCTGATGGAA ACCCCGACCC CATCGCCGCT    800
801     GCCAAAGATA TTCGCACCAC CTTCGGTCGT ATGGCCATGA ACGACGAGGA    850
851     GACGGTTGCC CTTATTGCCG GCGGTCACAC CTTCGGTAAG ACACACGGTG    900
901     CTGGCCCAGC AGACAAGCTC GGCCCGGAAC CAGAGGCTGC AGACATGGCA    950
951     CAACAGGGTT TAGGCTGGAC CAATAGCTTC AAAAGCGGCA AGGGTCCTGA   1000
1001    TACCACAACC AGCGGTCTCG AAGTTACCTG GACCAAGACT CCTACTAAAT   1050
1051    GGAGTAACCA ATTCTTGGAG TACCTCTTCC GCTACGACTG GGAACTCACT   1100
1101    AAGAGTCCTG CCGGCGCCCA CCAGTGGGTC GCCAAAAATG CAGAGGCTTT   1150
1151    CATCCCCGAT GCATTTGACC CATCCAAGAA GCGCAAGCCA ATGATGCTCA   1200
1201    CGACCGATCT TTCCCTTCGC TATGACCCTA TCTACGAGAA GATCTCTCGT   1250
1251    CGCTTCTTGG AGCACCCTGA CCAGTTTGCT GATGCGTTTG CCCGTGCCTG   1300
1301    GTTCAAGTTA CTTCACCGTG ACCTTGGCCC ACGAGCTCTC TACATTGGTC   1350
1351    CCGAAGTGCC TGCAGAGGTT CTACCCTGGC AGGATCCCGT TCCCGCTGTC   1400
1401    GACCACCCCC TCATTAGCAA TGAAGACGCG TCGGCTTTGA AACAGCGCAT   1450
1451    TTTGGCCTCG GGTGTCAAAC CATCCAGCTT GATTTCCACT GCTTGGGCAT   1500
1501    CCGCTTCTAC GTTCAGAGGT AGCGACAAGC GCGGCGGTGC CAATGGTGCT   1550
1551    CGCATCCGCC TGTCTCCTCA GCGTGAGTGG GCAGTTAACA ACCAACCCTG   1600
1601    GTTGCGCGAG ACCCTTTCTG TGCTTGAAGC CATACAGAAA CAATTCAATA   1650
1651    CCTCCCAGTC TGGAGGCAAG AAGGTGTCTA TTGCAGACTT GATTGTTCTC   1700
1701    GCTGGTGTCG CCGCTGTTGA GAAGGCTGCT CGCGACGCCG GATACGCCGT   1750
1751    CACAGTACCC TTCACTCCCG GTCGCACAGA TGCTTCCCAA GAGCAGACTG   1800
1801    ACGTCCAATC CTTCAGCGAC ATGGAGCCCA TTGCTGATGG TTTCCGTAAC   1850
1851    TACGGCTCAT CCACCTCTCG CGTTCGTGCT GAGGAGTGGC TCATCGATAA   1900
1901    GGCACAGCTT TTGACCCTCA GTGCACCCGA GTTGGCCGTT CTCATCGGCG   1950
1951    GTCTCCGTGT CCTCAACACA AACTACGACG GCTCTGCTCA CGGTGTCTTC   2000
2001    ACCCAGCGCC CAGGCAAGTT GACCAACGAC TTCTTCGTCA ACCTCTTGGA   2050
2051    CATGAACACC GCATGGAAAT CAATTGGTGG TGTCGACCTC TACGAGGGCA   2100
2101    CAGATCGCAA GACTGGCGCC AAGAAGTGGA CTGCTACTCG TAACGATCTC   2150
2151    GTCTTTGGCT CCAACGCTGA GTTGCGTGCT ATTGCTGAGG TGTACGGTAG   2200
2201    CTCTGATGGC CAGGAGAAGT TCGTCAAGGA CTTTGTGGCT GCTTGGGACA   2250
2251    AGGTCATGAA CTTGGATCGA TTCGACTTGA AGAAGAAGCA ATCCACTTCC   2300
2301    AGTCACCGCC TTTAAATGTG AATAGTGGAC AATTTGACGC AAACTATATA   2350
2351    ATAATTCTGA TGAGATTATG CCAGTAATGA GAAAGTTTGT TGTTTTGCTG   2400
```

```
2401    TTTCGAACTT GGTGGTAGTT GAATGTAACT TAAGCAGAAC ATGAAACAAT   2450
2451    ATCAGGACAC ATATCCCAGC AAG
```

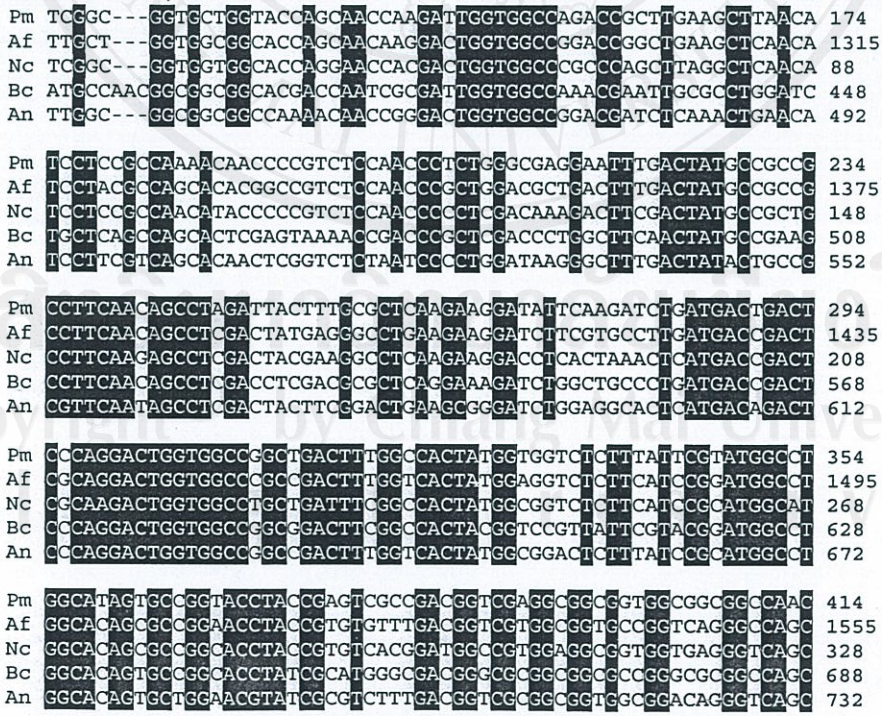
1.3) Functional analysis

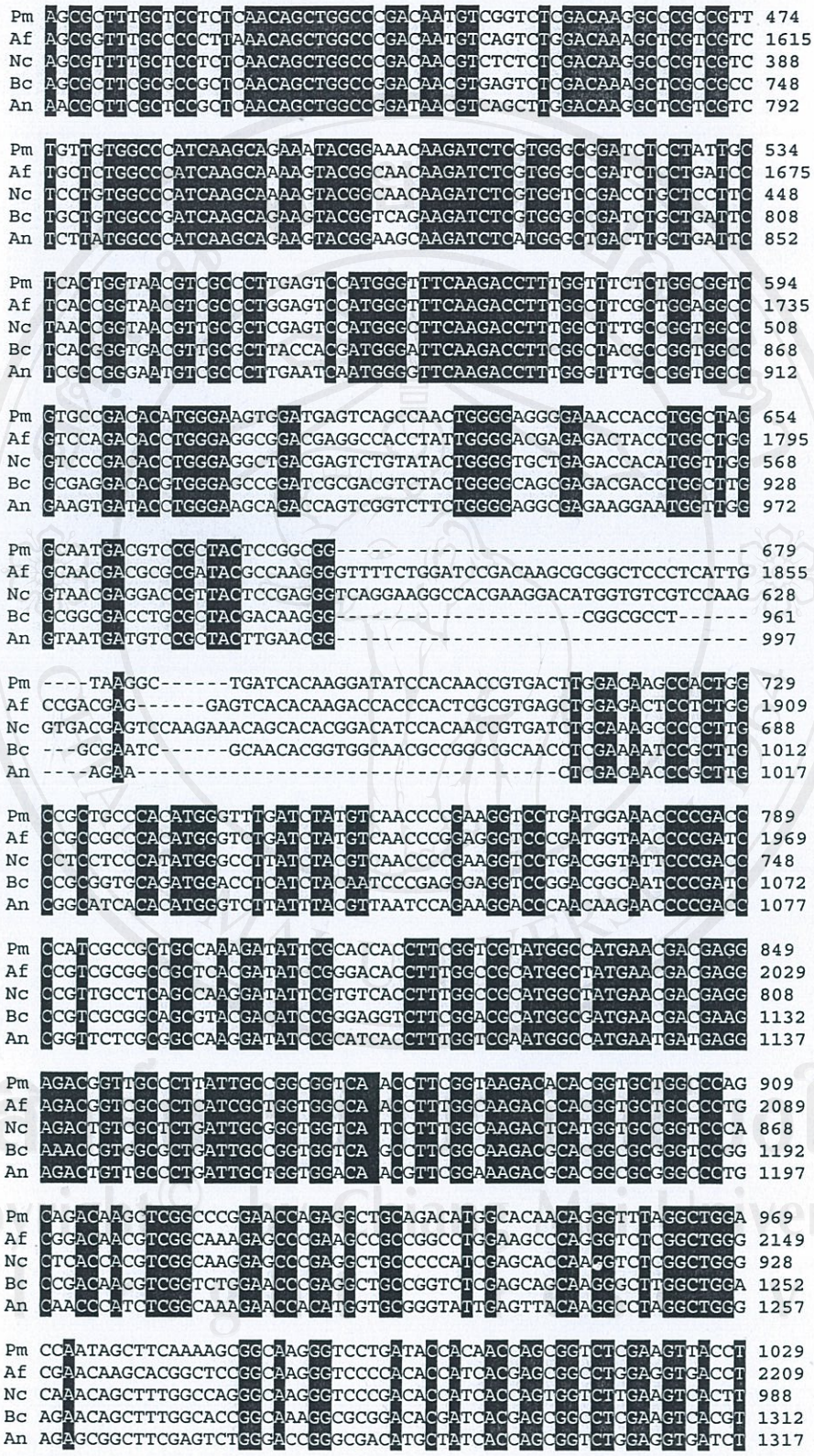
A. Similarity search

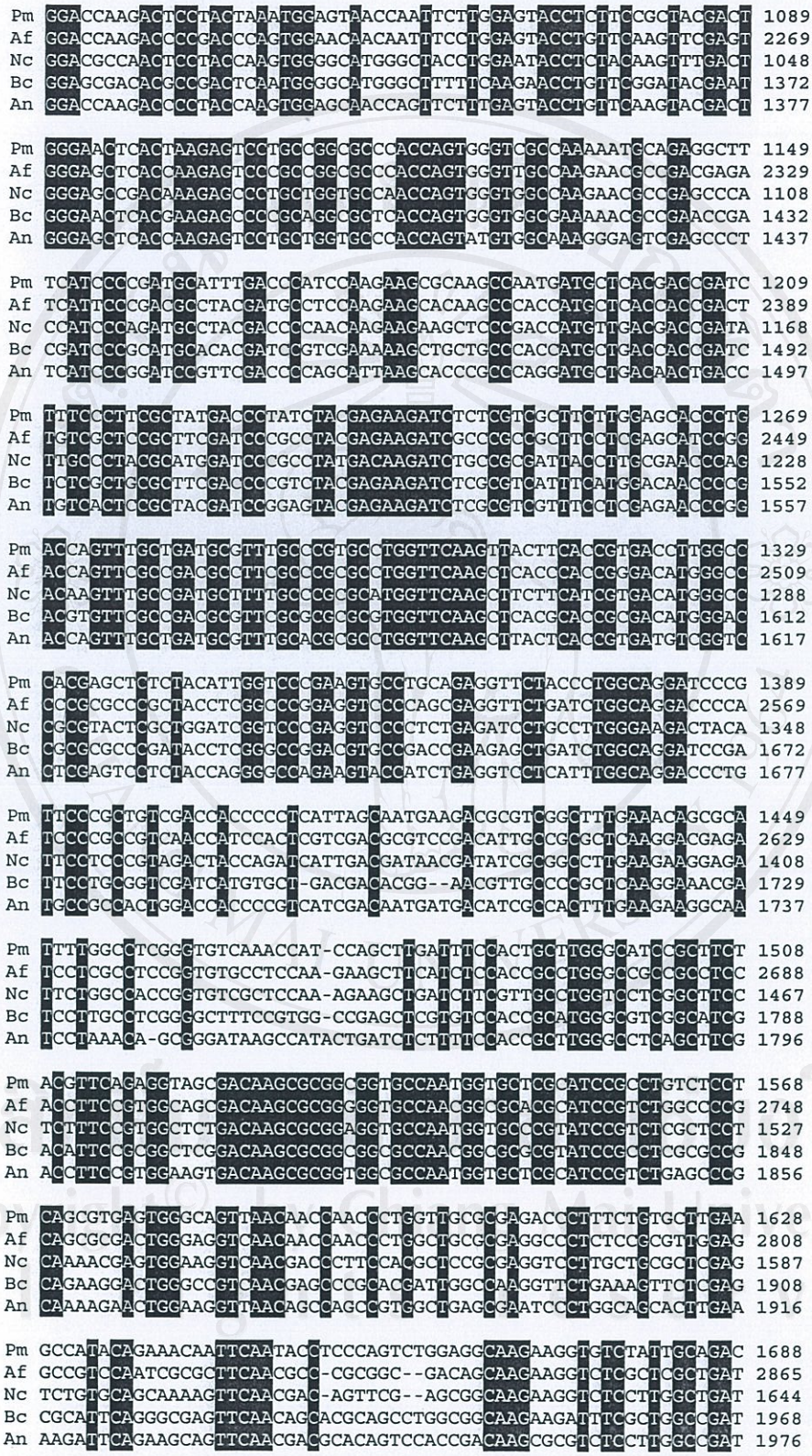
Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLASTN	gi|31074944|gb|AY125354.1|	*CAT2* for catalase-peroxidase	*Aspergillus fumigatus* (fungus)	252	3e-63	391/479 (81%)
	gi|32410288|ref|XM_325624.1	*Cat-2* for catalase-peroxidase	*Neurospora crassa* (fungus)	190	9e-45	246/296 (83%)
	gi|36958823|gb|AY316747.1|	*katG* for catalase-peroxidase	*Rhizobium* sp. (bacterium)	137	1e-28	258/321 (80%)
	gi|15384982|emb|AJ305225.1|ANI305225	*cpeA* for catalase-peroxidase	*Aspergillus nidulans* (fungus)	135	5e-28	239/296 (80%)
	gi|12620401|gb|AF317697.1|AF317697	*katG* for catalase-peroxidase	*Burkholderia cepacia* (bacterium)	129	3e-26	200/245 (81%)
BLASTX	gi|30580356|sp|Q8X182|CAT2 NEUCR	Catalase-peroxidase	*Neurospora crassa* (fungus)	1083	0.0	525/753 (69%)
	gi|9972752|sp|O87864|CATB_STRRE	Catalase-peroxidase	*Streptomyces reticuli* (bacterium)	976	0.0	477/734 (64%)
	gi|1345687|sp|P14412|CATA_BACST	Catalase-peroxidase	*Geobacillus stearothermophilus* (bacterium)	924	0.0	462/748 (61%)
	gi|6226842|sp|Q08129|CATA_MYCTU	Catalase-peroxidase	*Mycobacterium tuberculosis* (bacterium)	913	0.0	460/741 (62%)
	gi|9972736|sp|O59651|CATA_HALMA	Catalase-peroxidase	*Haloarcula marismortui* (Archaeobacterium)	817	0.0	406/720 (56%)

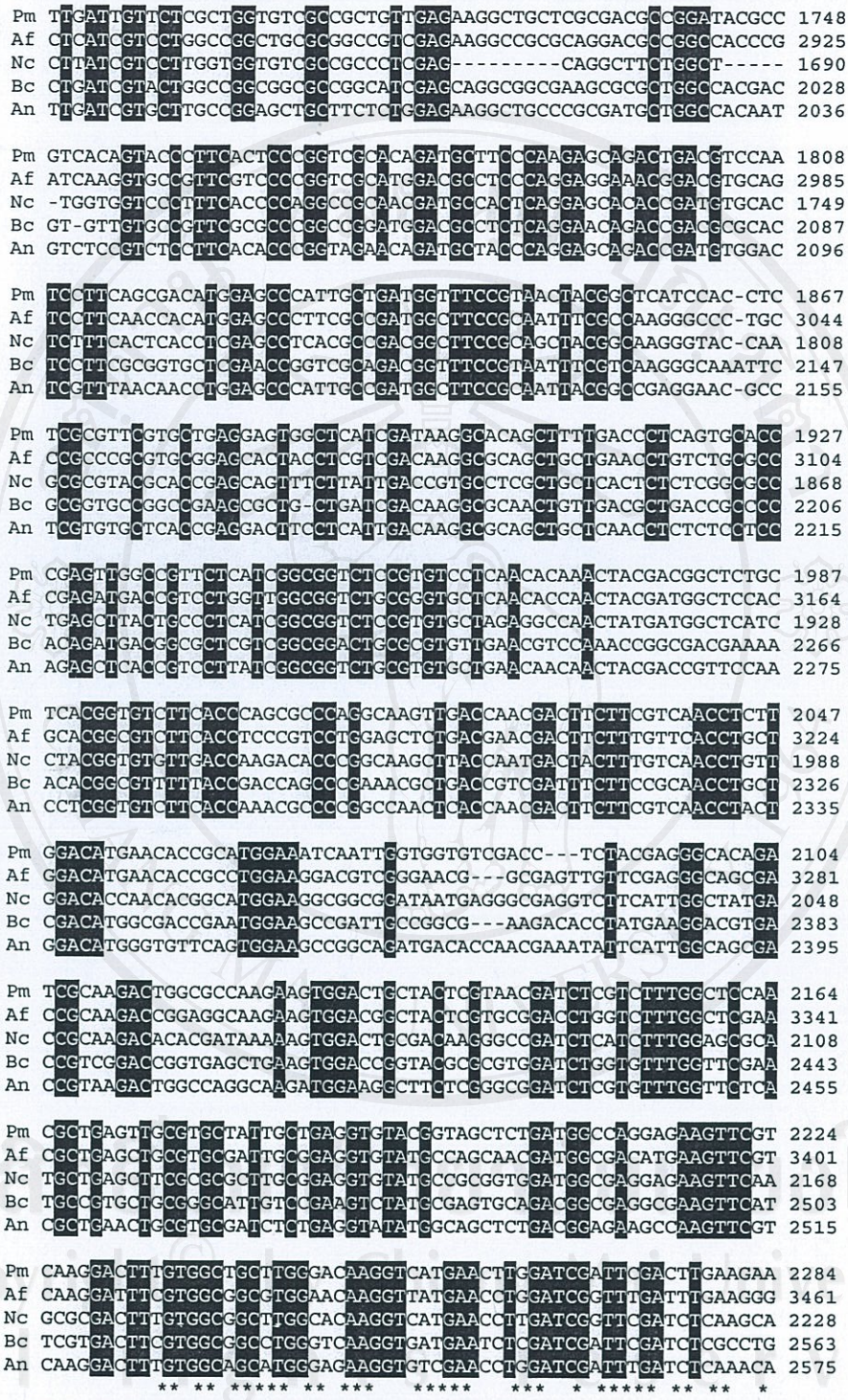
B. Multiple sequence alignment

Nucleotide level;









Note Pm = *Penicillium marneffei*, Af = *Aspergillus fumigatus*,
Nc = *Neurospora crassa*, Bc = *Burkholderia cepacia*, An = *Aspergillus nidulans*

Polypeptide level;

```
Pm --------------MAESKCPAHQHVLKANVGGAGTSNQDWWPDRLKLNILRQNNPVSNP 46
Nc ----------------MSECP----VRKSNVGGGGTRNHDWWPAQLRLNILRQHTPVSNP 40
Sr MTENHDAIVTDAKSEGSGGCPVAHDRALHPTQGGG--NRQWWPERLNLKILAKNPAVANP 58
Mt MPEQHPPITETTTGAASNGCPVVG-HMKYPVEGGG--NQDWWPNRLNLKVLHQNPAVADP 57
Gs --------MENQNRQNAAQCPFHG--SVTNQSSNRTTNKDWWPNQLNLSILHQHDRKTNP 50
Hm -------------MAETPNSDMSG--ATGGRSKRPKSNQDWWPSKLNLEILDQNARDVGP 45
                         .             *::*** :*.*.:* ::    .*

Pm LGEEFDYAAAFNSLDYFALKKDIQDLMTDSQDWWPADFGHYGGLFIRMAWHSAGTYRVAD 106
Nc LDKDFDYAAAFKSLDYEGLKKDLTKLMTDSQDWWPADFGHYGGLFIRMAWHSAGTYRVTD 100
Sr LDEDFDYAEAFKALDLAAVKRDIAEVLTTSQDWWPADFGNYGPLMIRMAWHSAGTYRISD 118
Mt MGAAFDYAAEVATIDVDALTRDIEEVMTTSQPWWPADYGHYGPLFIRMAWHAAGTYRIHD 117
Gs HDEEFNYAEEFQKLDYWALKEDLRKLMTESQDWWPADYGHYGPLFIRMAWHSAGTYRIGD 110
Hm VEDDFDYAEEFQKLDLEAVKSDLEELMTSSQDWWPADYGHYGPLFIRMAWHSAGTYRTAD 105
      *:**   . :*  .:. *: .::* ** *****:*:** *:******:*****  *

Pm GRGGGGGGQQRFAPLNSWPDNVGLDKARRLLWPIKQKYGNKISWADLLLLTGNVALESMG 166
Nc GRGGGGEGQQRFAPLNSWPDNVSLDKARRLLWPIKQKYGNKISWSDLLLLTGNVALESMG 160
Sr GRGGAGAGQQRFAPLNSWPDNGNLDKARRLLWPVKKKYGQSISWADLLILTGNVALETMG 178
Mt GRGGAGGGMQRFAPLNSWPDNASLDKARRLLWPVKKKYGKKLSWADLIVFAGNCALESMG 177
Gs GRGGASTGTQRFAPLNSWPDNANLDKARRLLWPIKKKYGNKISWADLFILAGNVAIESMG 170
Hm GRGGAAGGRQRFAPINSWPDNANLDKARRLLLPIKQKYGQKISWADLMILAGNVAIESMG 165
   ****.. * *****:****** .******** *:*:***:.:**:**::::** *:*:**

Pm FKTFGFSGGRADTWEVDESANWGGETTWLGNDVRYSGG--------------KADHKDI 211
Nc FKTFGFAGGRPDTWEADESVYWGAETTWLGNEDRYSEGQEGHEGHGVVQGDESKKQHTDI 220
Sr FKTFGFGGGRADVWEAEEDVYWGPETTWLDDR-RYTGD---------------------- 215
Mt FKTFGFGFGRVDQWEPDE-VYWGKEATWLGDE-RYSGK---------------------- 213
Gs GKTIGFGGGRVDVWHPEEDVYWGSEKEWLASE-RYSGD---------------------- 207
Hm FKTFGYAGGREDAFEEDKAVNWGPEDEFETQE-RFDEP---------------------- 202
    **:*:. ** * :. :: . ** *  :  .  *:

Pm HNRDLDKPLAAAHMGLIYVNPEGPDGNPDPIAAAKDIRTTFGRMAMNDEETVALIAGGHT 271
Nc HNRDLQSPLASSHMGLIYVNPEGPDGIPDPVASAKDIRVTFGRMAMNDEETVALIAGGHS 280
Sr --RELENPLGAVQMGLIYVNPEGPNGNPDPIAAARDIRETFRRMAMNDEETVALIAGGHT 273
Mt --RDLENPLAAVQMGLIYVNPEGPNGNPDPMAAAVDIRETFRRMAMNDVETAALIVGGHT 271
Gs --RELENPLAAVQMGLIYVNPEGPDGKPDPKAAARDIRETFRRMGMNDEETVALIAGGHT 265
Hm --GEIQEGLGASVMGLIYVNPEGPDGNPDPEASAKNIRQTFDRMAMNDKETAALIAGGHT 260
      :::. *.:  ***********:* *** *:* :** ** **.*** **.***.***:

Pm FGKTHGAG-PADKLGPEPEAADMAQQGLGWTNSFKSGKGPDTTTSGLEVTWTKTPTKWSN 330
Nc FGKTHGAG-PTHHVGKEPEAAPIEHQGLGWANSFGQGKGPDTITSGLEVTWTPTPTKWGM 339
Sr FGKTHGAG-PADHVGADPEAASLEEQGLGWRSTYGTGKGADAITSGLEVTWTSTPTQWSN 332
Mt FGKTHGAG-PADLVGPEPEAAPLEQMGLGWKSSYGTGTGKDAITSGIEVVWTNTPTKWDN 330
Gs FGKAHGAG-PATHVGPEPEAAPIEAQGLGWISSYGKGKGSDTITSGIEGAWTPTPTQWDT 324
Hm FGKVHGADDPEENLGPEPEAAPIEQQGLGWQNKNGNSKGGEMITSGIEGPWTQSPTEWDM 320
   ***.***. *   :* :**** :  **** ..   ..* :  ***:*  ** :**:*.

Pm QFLEYLFRYDWELTKSPAGAHQWVAKN--AEAFIPDAFDPSKKRKPMMLTTDLSLRYDPI 388
Nc GYLEYLYKFDWEPTKSPAGANQWVAKN--AEPTIPDAYDPNKKKLPTMLTTDIALRMDPA 397
Sr GFFKNLFEYEYELEQSPAGAHQWVAKN--APEIIPDAHDPSKKHRPRMLTTDLSLRFDPI 390
Mt SFLEILYGYEWELTKSPAGAWQYTAKDGAGAGTIPDPFGGPGRS-PTMLATDLSLRVDPI 389
Gs SYFDMLFGYDWWLTKSPAGAWQWMAVDPDEKDLAPDAEDPSKKVPTMMMTTDLALRFDPE 384
Hm GYINNLLDYEWEPEKGPGGAWQWAPKSEELKNSVPDAHDPDEKQTPMMLTTDIALKRDPD 380
    ::. *  :::   :.*.** *: . .      **. .   :  . *::**::*: **

Pm YEKISRRFLEHPDQFADAFARAWFKLLHRDLGPRALYIGPEVPAEVLPWQDPVPAVDHPL 448
Nc YDKICRDYLANPDKFADAFARAWFKLLHRDMGPRTRWIGPEVPSEILPWEDYIPPVDYQI 457
Sr YEPISRRFYENPEEFADAFARAWYKLTHRDMGPKSLYLGPEVPEETLLWQDPLPEREGEL 450
Mt YERITRRWLEHPEELADEFAKAWYKLIHRDMGPVARYLGPLVPKQTLLWQDPVPAVSHDL 449
Gs YEKIARRFHQNPEEFAEAFARAWFKLTHRDMGPKTRYLGPEVPKEDFIWQDPIPEVDYEL 444
Hm YREVMETFQENPMEFGMNFAKAWYKLTHRDMGPPERFLGPEVPDEEMIWQDPLPDADYDL 440
   *  : . :  :* ::.  **:**:** ***:**   ::** ** : : *:* :*  .  :
```

```
Pm ISNEDASALKQRILASGVKPSSLISTAWASASTFRGSDKRGGANGARIRLSPQREWAVNN 508
Nc IDDNDIAALKKEILATGVAPKKLIFVAWSSASSFRGSDKRGGANGARIRLAPQNEWKVND 517
Sr IDDADIAILKTKLLESGLSVSQLVTTAWASASTFRASDKRGGANGARIRLAPQRGWEVND 510
Mt VGEAEIASLKSQIRASGLTVSQLVSTAWAAASSFRGSDKRGGANGGRIRLQPQVGWEVND 509
Gs T-EAEIEEIKAKILNSGLTVSELVKTAWASASTFRNSDKRGGANGARIRLAPQKDWEVNE 503
Hm IGDEEIAELKEEILDSDLSVSQLVKTAWASASTYRDSDKRGGANGARLRLEPQKNWEVNE 500
     : :   :* .:  :.: ..*:  .**::**::* *********.*:** **  * **:

Pm QP-WLRETLSVLEAIQKQFNTSQSGGKKVSIADLIVLAGVAAVEKAARDAGYAVTVPFTP 567
Nc PS-TLREVLAALESVQQKFNDSSSG-KKVSLADLIVLGGVAALEQAS-----GLVVPFTP 570
Sr PD-QLAQVLRTLENVQQEFN-ASSGAKKVSLADLIVLGGAAGVEKAAKEAGFEIQVPFTP 568
Mt PDGDLRKVIRTLEEIQESFNSAAPGNIKVSFADLVVLGGCAAIEKAAKAAGHNITVPFTP 569
Gs PE-RLAKVLSVYEDIQRELP------KKVSIADLIVLGGSAAVEKAARDAGFDVKVPFFP 556
Hm PE-QLETVLGTLENIQTEFNDSRSDGTQVSLADLIVLGGNAAVEQAAANAGYDVEIPFEP 559
      *  .: . * :* .:        :**:***:**.* *.:*:*:      : :** *

Pm GRTDASQEQTDVQSFSDMEPIADGFRNYGSSTSRVRAEEWLIDKAQLLTLSAPELAVLIG 627
Nc GRNDATQEHTDVHSFTHLEPHADGFRSYGKGTKRVRTEQFLIDRASLLTLSAPELTALIG 630
Sr GRVDATEEHTDVESFEALEPTADGFRNYLGKGNRLPAEYLLLDKANLLNLSAPEMTVLVG 628
Mt GRTDASQEQTDVESFAVLEPKADGFRNYLGKGNPLPAEYMLLDKANLLTLSAPEMTVLVG 629
Gs GRGDATQEQTDVESFAVLEPFADGFRNYQKQEYSVPPEELLVDKAQLLGLTAPEMTVLVG 616
Hm GRVDAGPEHTDAPSFDALKPKVDGVRNYIQDDITRPAEEVLVDNADLLNLTASELTALIG 619
   ** **  *:**. **  ::* .**.*.*        .*  *:*.*.** *:*.*::.*:*

Pm GLRVLNTNYDGSAHGVFTQRPGKLTNDFFVNLLDMNTAWKSIGG-VDLYEGTDRKTGAKK 686
Nc GLRVLEANYDGSSYGVLTKTPGKLTNDYFVNLLDTNTAWKAADNEGEVFIGYDRKTHDKK 690
Sr GLRVLGANHQQSQLGVFTKTPGVLTNDFFVNLLDMGTTWKATSEDQTTFEGRDAATGEVK 688
Mt GLRVLGANYKRLPLGVFTEASESLTNDFFVNLLDMGITWEPSPADDGTYQGKDGS-GKVK 688
Gs GLRVLGANYRDLPHGVFTDRIGVLTNDFFVNLLDMNYEWVPTDS--GIYEIRDRKTGEVR 674
Hm GMRSIGANYQDTDLGVFTDEPETLTNDFFVNLLDMGTEWEPAADSEHRYKGLDRDTGEVK 679
   *:* : :*:     **:*.    ****:****** .  * .      :   *      :

Pm WTATRNDLVFGSNAELRAIAEVYGSSDGQEKFVKDFVAAWDKVMNLDRFDLKKK-QSTSS 745
Nc WTATRADLIFGAHAELRALAEVYAAVDGEEKFKRDFVAAWHKVMNLDRFDLKQEGRGQNA 750
Sr WAGSRADLVFGSNSELRALAEVYASDDAKEKFVKDFVAAWHKVMDADRFDLV-------- 740
Mt WTGSRVDLVFGSNSELRALVEVYGADDAQPKFVQDFVAAWDKVMNLDRFDVR-------- 740
Gs WTATRVDLIFGSNSILRSYAEFYAQDDNQEKFVRDFINAWVKVMNADRFDLVKKARESVT 734
Hm WEATRIDLIFGSNDRLRAISEVYGSADAEKKLVHDFVDTWSKVMKLDRFDLE-------- 731
   * .:* **:**::  **:  *.*.  * : *: :**: :* ***. ****:
```

Note Pm = *Penicillium marneffei*, Nc = *Neurospora crassa*, Sr = *Streptomyces reticuli*, Mt = *Mycobacterium tuberculosis*, Gs = *Geobacillus stearothermophilus*, Hm = *Haloarcula marismortui*

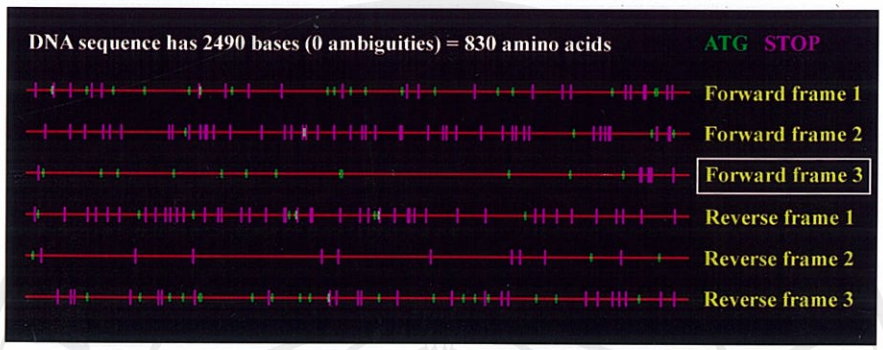
C. Motif scan search

Hits	Status	Amino acid position	Database
Proximal heme ligand signature (peroxidase 1)	Weak match	88-99	PROSITE
Peroxidase active site signature (peroxidase 2)	Weak match	262-272	PROSITE
Peroxidase	Significant match	65-399	Pfam

D. Conclusion of functional analysis:
Possible function of encoded protein = Catalase-peroxidase

1.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 3 was functionally identified as catalase-peroxidase

C. Deduced amino acid composition (748 amino acids)
Predicted molecular weight (kDa): 82.4
Theoretical pI: 6.35

```
      Forward Frame 3:

  3  GACGCGTGGGCGTACTCGAATCCCCCACACTGTCTGAACTATCTTATTTGAAGCAAACCA

  1        M  A  E  S  K  C  P  A  H  Q  H  V  L  K  A  N  V  G
 63  AACAAGATGGCTGAGAGCAAGTGTCCCGCTCACCAGCATGTGTTGAAGGCCAACGTCGGC

 19   G  A  G  T  S  N  Q  D  W  W  P  D  R  L  K  L  N  I  L  R
123  GGTGCTGGTACCAGCAACCAAGATTGGTGGCCAGACCGCTTGAAGCTTAACATCCTCCGC

 39   Q  N  N  P  V  S  N  P  L  G  E  E  F  D  Y  A  A  A  F  N
183  CAAAACAACCCCGTCTCCAACCCTCTGGGCGAGGAATTTGACTATGCCGCCGCCTTCAAC

 59   S  L  D  Y  F  A  L  K  K  D  I  Q  D  L  M  T  D  S  Q  D
243  AGCCTAGATTACTTTGCGCTCAAGAAGGATATTCAAGATCTGATGACTGACTCCCAGGAC

 79   W  W  P  A  D  F  G  H  Y  G  G  L  F  I  R  M  A  W  H  S
303  TGGTGGCCGGCTGACTTTGGCCACTATGGTGGTCTCTTTATTCGTATGGCCTGGCATAGT

 99   A  G  T  Y  R  V  A  D  G  R  G  G  G  G  G  G  Q  Q  R  F
363  GCCGGTACCTACCGAGTCGCCGACGGTCGAGGCGGCGGTGGCGGCGGCCAACAGCGCTTT

119   A  P  L  N  S  W  P  D  N  V  G  L  D  K  A  R  R  L  L  W
423  GCTCCTCTCAACAGCTGGCCCGACAATGTCGGTCTCGACAAGGCCCGCCGTTTGTTGTGG

139   P  I  K  Q  K  Y  G  N  K  I  S  W  A  D  L  L  L  L  T  G
483  CCCATCAAGCAGAAATACGGAAACAAGATCTCGTGGGCGGATCTCCTATTGCTCACTGGT

159   N  V  A  L  E  S  M  G  F  K  T  F  G  F  S  G  G  R  A  D
543  AACGTCGCCCTTGAGTCCATGGGTTTCAAGACCTTTGGTTTCTCTGGCGGTCGTGCCGAC

179   T  W  E  V  D  E  S  A  N  W  G  G  E  T  T  W  L  G  N  D
603  ACATGGGAAGTGGATGAGTCAGCCAACTGGGGAGGGGAAACCACCTGGCTAGGCAATGAC

199   V  R  Y  S  G  G  K  A  D  H  K  D  I  H  N  R  D  L  D  K
663  GTCCGCTACTCCGGCGGTAAGGCTGATCACAAGGATATCCACAACCGTGACTTGGACAAG

219   P  L  A  A  A  H  M  G  L  I  Y  V  N  P  E  G  P  D  G  N
723  CCACTGGCCGCTGCCCACATGGGTTTGATCTATGTCAACCCCGAAGGTCCTGATGGAAAC
```

```
239   P  D  P  I  A  A  A  K  D  I  R  T  T  F  G  R  M  A  M  N
783   CCCGACCCCATCGCCGCTGCCAAAGATATTCGCACCACCTTCGGTCGTATGGCCATGAAC

259   D  E  E  T  V  A  L  I  A  G  G  H  T  F  G  K  T  H  G  A
843   GACGAGGAGACGGTTGCCCTTATTGCCGGCGGTCACACCTTCGGTAAGACACACGGTGCT

279   G  P  A  D  K  L  G  P  E  P  E  A  A  D  M  A  Q  Q  G  L
903   GGCCCAGCAGACAAGCTCGGCCCGGAACCAGAGGCTGCAGACATGGCACAACAGGGTTTA

299   G  W  T  N  S  F  K  S  G  K  G  P  D  T  T  T  S  G  L  E
963   GGCTGGACCAATAGCTTCAAAAGCGGCAAGGGTCCTGATACCACAACCAGCGGTCTCGAA

319   V  T  W  T  K  T  P  T  K  W  S  N  Q  F  L  E  Y  L  F  R
1023  GTTACCTGGACCAAGACTCCTACTAAATGGAGTAACCAATTCTTGGAGTACCTCTTCCGC

339   Y  D  W  E  L  T  K  S  P  A  G  A  H  Q  W  V  A  K  N  A
1083  TACGACTGGGAACTCACTAAGAGTCCTGCCGGCGCCCACCAGTGGGTCGCCAAAAATGCA

359   E  A  F  I  P  D  A  F  D  P  S  K  K  R  K  P  M  M  L  T
1143  GAGGCTTTCATCCCCGATGCATTTGACCCATCCAAGAAGCGCAAGCCAATGATGCTCACG

379   T  D  L  S  L  R  Y  D  P  I  Y  E  K  I  S  R  R  F  L  E
1203  ACCGATCTTTCCCTTCGCTATGACCCTATCTACGAGAAGATCTCTCGTCGCTTCTTGGAG

399   H  P  D  Q  F  A  D  A  F  A  R  A  W  F  K  L  L  H  R  D
1263  CACCCTGACCAGTTTGCTGATGCGTTTGCCCGTGCCTGGTTCAAGTTACTTCACCGTGAC

419   L  G  P  R  A  L  Y  I  G  P  E  V  P  A  E  V  L  P  W  Q
1323  CTTGGCCCACGAGCTCTCTACATTGGTCCCGAAGTGCCTGCAGAGGTTCTACCCTGGCAG

439   D  P  V  P  A  V  D  H  P  L  I  S  N  E  D  A  S  A  L  K
1383  GATCCCGTTCCCGCTGTCGACCACCCCCTCATTAGCAATGAAGACGCGTCGGCTTTGAAA

459   Q  R  I  L  A  S  G  V  K  P  S  S  L  I  S  T  A  W  A  S
1443  CAGCGCATTTTGGCCTCGGGTGTCAAACCATCCAGCTTGATTTCCACTGCTTGGGCATCC

479   A  S  T  F  R  G  S  D  K  R  G  G  A  N  G  A  R  I  R  L
1503  GCTTCTACGTTCAGAGGTAGCGACAAGCGCGGCGGTGCCAATGGTGCTCGCATCCGCCTG

499   S  P  Q  R  E  W  A  V  N  N  Q  P  W  L  R  E  T  L  S  V
1563  TCTCCTCAGCGTGAGTGGGCAGTTAACAACCAACCCTGGTTGCGCGAGACCCTTTCTGTG

519   L  E  A  I  Q  K  Q  F  N  T  S  Q  S  G  G  K  K  V  S  I
1623  CTTGAAGCCATACAGAAACAATTCAATACCTCCCAGTCTGGAGGCAAGAAGGTGTCTATT

539   A  D  L  I  V  L  A  G  V  A  A  V  E  K  A  A  R  D  A  G
1683  GCAGACTTGATTGTTCTCGCTGGTGTCGCCGCTGTTGAGAAGGCTGCTCGCGACGCCGGA

559   Y  A  V  T  V  P  F  T  P  G  R  T  D  A  S  Q  E  Q  T  D
1743  TACGCCGTCACAGTACCCTTCACTCCCGGTCGCACAGATGCTTCCCAAGAGCAGACTGAC

579   V  Q  S  F  S  D  M  E  P  I  A  D  G  F  R  N  Y  G  S  S
1803  GTCCAATCCTTCAGCGACATGGAGCCCATTGCTGATGGTTTCCGTAACTACGGCTCATCC

599   T  S  R  V  R  A  E  E  W  L  I  D  K  A  Q  L  L  T  L  S
1863  ACCTCTCGCGTTCGTGCTGAGGAGTGGCTCATCGATAAGGCACAGCTTTTGACCCTCAGT

619   A  P  E  L  A  V  L  I  G  G  L  R  V  L  N  T  N  Y  D  G
1923  GCACCCGAGTTGGCCGTTCTCATCGGCGGTCTCCGTGTCCTCAACACAAACTACGACGGC

639   S  A  H  G  V  F  T  Q  R  P  G  K  L  T  N  D  F  F  V  N
1983  TCTGCTCACGGTGTCTTCACCCAGCGCCCAGGCAAGTTGACCAACGACTTCTTCGTCAAC

659   L  L  D  M  N  T  A  W  K  S  I  G  G  V  D  L  Y  E  G  T
2043  CTCTTGGACATGAACACCGCATGGAAATCAATTGGTGGTGTCGACCTCTACGAGGGCACA

679   D  R  K  T  G  A  K  K  W  T  A  T  R  N  D  L  V  F  G  S
2103  GATCGCAAGACTGGCGCCAAGAAGTGGACTGCTACTCGTAACGATCTCGTCTTTGGCTCC
```

```
 699   N A E L R A I A E V Y G S S D G Q E K F
2163   AACGCTGAGTTGCGTGCTATTGCTGAGGTGTACGGTAGCTCTGATGGCCAGGAGAAGTTC

 719   V K D F V A A W D K V M N L D R F D L K
2223   GTCAAGGACTTTGTGGCTGCTTGGGACAAGGTCATGAACTTGGATCGATTCGACTTGAAG

 739   K K Q S T S S H R L -
2283   AAGAAGCAATCCACTTCCAGTCACCGCCTTTAAATGTGAATAGTGGACAATTTGACGCAA

2343   ACTATATAATAATTCTGATGAGATTATGCCAGTAATGAGAAAGTTTGTTGTTTTGCTGTT

2403   TCGAAACTTGGTGGTAGTTGAATGTAACTTAAGCAGAACATGAAACAATATCAGGACACA

2463   TATCCCAGCAAGAAAAAAAAAAAAAAAA
```

Figure 19. Sequence analysis of clones in group 1 (P1, P2, P4, P8, P16, P19, P20)

GROUP 2

2.1) Member: P3

2.2) DNA sequence of P3:
GC content of the ORF: 50%

```
1     CCCCAGTTAC GCACACGCGA CGTTTGACAT CATTCTGTCC ATTGTACGCC   50
51    TAATAACAAC ATCACTATCG TTTGGAATTG ATCACTATCG AATCCGTTGC  100
101   GACATATTTG ACCTCATCAT CACTTAGTTT TGGCGTCTGC TATATATCTT  150
151   GAATCATAGT TCCACCCCAT TGACAACGTG TATACACCGC CATCATGGGT  200
201   GCGAAGAAGT CCGCTGTGAA CCGTCCAATG AATGCCCAGC AGCGGGACGC  250
251   TGATATTAGC ACAAAGTTAC AAATATACGG CATCTATAGT GCATTTGCCA  300
301   ACGGGAAACT ACCATCAAAC AAGCAATGTG ATGTCGCCTT GAACAGCGCC  350
351   ATCAAGTCAA AATGGCTCTC GTCTCCACCA AAAGAACTCT CTGAAGATGG  400
401   CCGAACTCTG GTCAAGGACC TTCGTGATGT TATTGACAAG ACAAAGCTAC  450
451   TCTTCCTTAC CAAGAACGAG GGCGAGCTTA TACAAGAGTT CATCTGGGAT  500
501   GCCCAACAGA TTACTGGTGA AGAATTTCAG AGAATTACCG GCCCTGTCAG  550
551   CAAGGAGTCT GCTCGCGGAG ATGCCGATCA GGCCGCCGAG GGTTTCAAAA  600
601   CCCTGGGCAC GCTTTTGATC ACTAATGGAG AATTCCGCAA GTTGTTGAGT  650
651   GATGCTGTCA CCCTCCTTCG AGATATTGCT GGCGACACCG CGAGCAAGGC  700
701   AGCCTCCAAG CTTCGACCCG ACGAGGACGC TCTTGCGCAA ATCGATCAAC  750
751   CTGCCGAAGA GAATGTCTGG CATGACAAGC CTGACCTCAA CAAGGAGTCA  800
801   TTGAAAGCAC AATTCAAGGA ACAGACCGAC CGATTTAAAC CTGCTAGCAA  850
851   GCAAGATGTT CAAGAAGCAG CTAATGCTGC TACTACTGCT GCAACTGGTG  900
901   GTCAGCAAGA TGCTTCAGTC TCTCAGATTG ACGCACGTGC TGGTGTCAAT  950
951   GCTGCCAAGG GAGACCCTGC AACAGCGCGC CGAGCAGAAT GTTGCACCTG 1000
1001  AAGACCGGGA CCAAGTTCGA CAGGTTACTG AACAAGCACA AGCGGTATCT 1050
1051  TCTGAATATA ATCGGCGTAT TAAGGATTTC TTGGCCTCCA AGATGCCCAA 1100
1101  AGAACGTCGT GAACAGATCG TTTGGCGTTT GAAGAAGATG ATTGTTGAGA 1150
1151  TCCAGGGTCA CTCTGATTAC CAACAAGCCA TCGAGACTCT CTTGAGTTTG 1200
1201  GCTGAGTCAT ATGCTGGACA TGGAAGAGAT ATCTCTTCGC AAGGAACGAC 1250
1251  CGCTACTAAA GGGTTCATGG ATAGCAAGAA GGACATGCTG ATGCGACTCA 1300
1301  AGATTTTAAT CGAGAGATTT GCGAATAGCA CATCGACCGA TGACTTCTTT 1350
1351  GATTCGCTCA ACACGATCTA TCGCGATGCC GACCAGGACC CGCGGTTGAA 1400
1401  GGAATGGTTT AGGGGTGTCG ACACTTACAT TCGAAAATGC CTGCGCGAGC 1450
1451  AAGGGTTCAT CATGCAAGAT GAAGCTAATG ACCAGTGGAA CAAGCTGTAC 1500
1501  GATGAGGGTC GCTTCTTGCT TCGAGACCGG TACAGAAGTC ACACAGATCG 1550
1551  TATTGCCGAT GAAGCCAAGT TTTTGGCCAC TCAATTTGAC GAAGATCCAC 1600
1601  AGAACCGAGC TTTCCGTCAA TCTTTGGAGA GACTCTTCAA GGATCTGGGT 1650
1651  CAGGACCAGT ATGGAAAGCC GACGTTCAAG CCCCATTTGA TCAAAGACAT 1700
1701  TACCAATGTC ATCGTTCCTG AGATATTTGA GAATGCCAGC TACATTCCTA 1750
1751  TTCCTCGAAT TGAGGTTTCC GACCCAGCTG TTGACATGGT TATCGAAAAC 1800
1801  CTTATCATCG AGAGCGACAA CTTGATGCCC AATGTTTTAG AATTCGGCAC 1850
1851  TGACAACTAC TGGCGGTGGG GTCGTAAGAA GATCAGCAGC TTCGACGACC 1900
1901  ACAAGGTTAT GATTTCTGCA TCTGGCATCC AAGCTGATCT CCGTGATGTG 1950
1951  AGCTACTATT TCAAGAAGAA GCAAGGATTC CCTTCCCTCA CCGACATCGG 2000
2001  TGTCATGGAC ATTCTTCTTG GTGGTTCAGG CTTCGGCTTC AAGATTGCTG 2050
2051  CCTCCAAAGC ACAGAAGAAT GACCACAACG CAGTTTTTAA GCTTGATAGT 2100
2101  GTCAAGGTCA ACGTGAAGAA CCTTACCATC AAGCTGAAGA AGTCCAAGCA 2150
2151  CAAGATGCTT TTCATCATCT TCCGACCCAT GCTGTTAAAT GTTGTCCGAC 2200
2201  CTGTTCTTGA AAAGTGTTGG AGGCCCACAT TCCTGAGGCA TTCCAAAAGG 2250
2251  CAGACGCATT TGCCTACCAG GTACAGACCG AGGCTCAGCG TGCCCAGGAG 2300
2301  ACCATGCGTG AGGATCCTGA AAATGCAAAG AATATCTTCG CCCGCTACGC 2350
2351  CGATGCTACT CGTCACGTCA TCACTGGAGA AGAAGAAGCA GGCCGAAGCT 2400
```

```
2401    ATCGCCGAGC GCGGAACCAA GGTTCATATG GCAATGACTC ACCAGGATGC   2450
2451    CATGTTCAAG GACATCAAAT TGCCTGGTGG TGTCACTAAC AAAGCTACTG   2500
2501    AGTTCAAGGA GCTTTCTGCT AAGGGTGATC GCTGGCAATC ACCAGTTTTC   2550
2551    AACTGGGGCG GTGCTTCACC CACTAGCAAC TTTCCCAAGG CTTCCGCAGT   2600
2601    GTCTCGCAAG CCTCATACTG CCGCTGAAAG TCATCTGCAG GAGAAGTCAA   2650
2651    CCGATGGTGT CAATGGAGTC GGAGCTAGTG CAGTCAATGG ACTCCACGGA   2700
2701    GTCTCGACGA CTGGAGCCAC AGCTAGGGGA TCCAGTGGCA AGGCAATGTT   2750
2751    GCCAACCAAT GGTGTCACCA ACGGCCAGTC TGCCACCAAC GGTACTTTCC   2800
2801    AGAAGGAAGT CGAGCGTGCT TTTGACGCAA ACGGCACCAC TCTTCCAACC   2850
2851    CTAGGTGGTG TTTAATCGAC ATAAATTGCG AGTTTGACAT GGTAAACGAA   2900
2901    TGTCTGGAGA ATTCTTTTTT GATACCAAGA ACCAACCTAT CTCAATAATG   2950
2951    ACTAGTTTGG ATGGAATGAT GTGATTTGAG TTTGACTTTT GAAGACGAGG   3000
3001    TATGATTCAC CAATTATTAT AGCGAATCTG TTGGATACTT GTACAATGTT   3050
3051    GTCTAGTCTC TAGTTTCCTT TTGTCAGATA CTGATACTTA ATAAAATAAT   3100
3101    TACTATTTTC CAAAAAAAAA AAAAAAGGGC GGCCGCTCTA GAGGATCCAA   3150
3151    GCTTACGTAC GCGTGCATGC GA
```

2.3) Functional analysis

A. Similarity search

Type of Search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLAST N	gi|29150000|emb|AJ505990.1|DRE505990	*hs6st* mRNA for heparan sulfate-6-O sulfotransferase	*Danio rerio* (zebra fish)	91.7	8e-15	46/46 (100%)
	gi|2462264|emb|Y12793.1|CSPATATLP	mRNA for patatin-like protein	*Cucumis sativus* (cucumber)	91.7	8e-15	46/46 (100%)
	gi|28465790|dbj|AU301063.1|	cDNA clone	*Cyprinus carpio* (carp)	91.7	8e-15	46/46 (100%)
	gi|18073436|emb|AJ299719.1|HSA299719	mRNA for heparanase 2	*Homo sapiens* (human)	89.7	3e-14	45/45 (100%)
BLAST X	gi|1723487|sp|Q10327|YD72_SCHPO	Hypothetical protein	*Shizosaccharomyces pombe*(fungus)	235	4e-61	151/545 (27%)
	gi|29337193|sp|P18459|TY3H_DROME	Tyrosine-3-monooxygenase	*Drosophila melanogaster* (fly)	35	1.0	30/125 (24%)
	gi|27734598|sp|O34538|YCDA_BACSU	Hypothetical lipoprotein ycdA precursor	*Bacillus subtilis*	34.7	1.3	37/146 (27%)

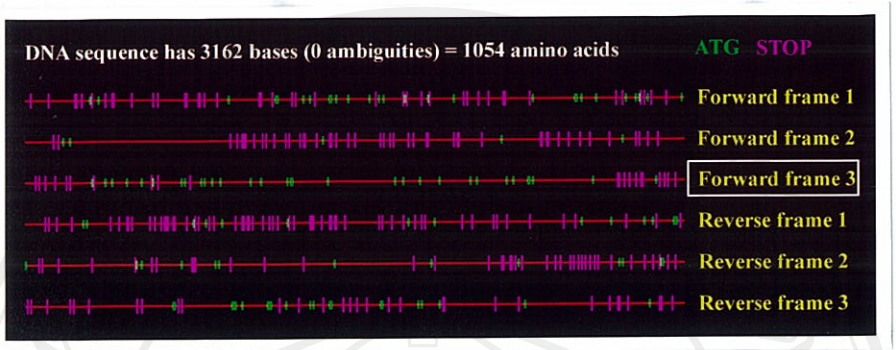
B. Motif scan search

Hits	Status	Amino acid position	Database
Proteasome protease	Weak match	341-372	PROSITE
Bipartite nuclear localization signal	Weak match	70-87	PROSITE

C. Conclusion of functional analysis:
Possible function of encoded protein = Unknown

2.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 3 matched with the hypothetical protein

C. Deduced amino acid composition (669 amino acids)
Predicted molecular weight (kDa): 75.86
Theoretical pI: 9.6

```
Forward Frame 3:

   3  CCAGTTACGCACACGCGACGTTTGACATCATTCTGTCCATTGTACGCCTAATAACAACTT

  63  TGGAATTGATCACTATCGAATCCGTTGCGACATATTTGACCTCATCATCACTTAGTTTTG

 123  GCGTCTGCTATATATCTTGAATCATAGTTCCACCCCATTGACAACGTGTATACACCGCCA

 183  TCATGGGTGCGAAGAAGTCCGCTGTGAACCGTCCAATGAATGCCCAGCAGCGGGACGCTG

 243  ATATTAGCACAAAGTTACAAATATACGGCATCTATAGTGCATTTGCCAACGGGAAACTAC

 303  CATCAAACAAGCAATGTGATGTCGCCTTGAACAGCGCCATCAAGTCAAAATGGCTCTCGT

 363  CTCCACCAAAAGAACTCTCTGAAGATGGCCGAACTCTGGTCAAGGACCTTCGTGATGTTA

 423  TTGACAAGACAAAGCTACTCTTCCTTACCAAGAACGAGGGCGAGCTTATACAAGAGTTCA

 483  TCTGGGATGCCCAACAGATTACTGGTGAAGAATTTCAGAGAATTACCGGCCCTGTCAGCA

 543  AGGAGTCTGCTCGCGGAGATGCCGATCAGGCCGCCGAGGGTTTCAAAACCCTGGGCACGC

 603  TTTTGATCACTAATGGAGAATTCCGCAAGTTGTTGAGTGATGCTGTCACCCTCCTTCGAG

 663  ATATTGCTGGCGACACCGCGAGCAAGGCAGCCTCCAAGCTTCGACCCGACGAGGACGCTC

 723  TTGCGCAAATCGATCAACCTGCCGAAGAGAATGTCTGGCATGACAAGCCTGACCTCAACA

 783  AGGAGTCATTGAAAGCACAATTCAAGGAACAGACCGACCGATTTAAACCTGCTAGCAAGC

          M  F  K  K  Q  L  M  L  L  L  L  L  Q  L  V  V  S  K  M
 843  AAGATGTTCAAGAAGCAGCTAATGCTGCTACTACTGCTGCAACTGGTGGTCAGCAAGATG

  20   L  Q  S  L  R  L  T  H  V  L  V  S  M  L  P  R  E  T  L  Q
 903  CTTCAGTCTCTCAGATTGACGCACGTGCTGGTGTCAATGCTGCCAAGGGAGACCCTGCAA

  40   Q  R  A  E  Q  N  V  A  P  E  D  R  D  Q  V  R  Q  V  T  E
 963  CAGCGCGCCGAGCAGAATGTTGCACCTGAAGACCGGGACCAAGTTCGACAGGTTACTGAA

  60   Q  A  Q  A  V  S  S  E  Y  N  R  R  I  K  D  F  L  A  S  K
1023  CAAGCACAAGCGGTATCTTCTGAATATAATCGGCGTATTAAGGATTTCTTGGCCTCCAAG
```

```
  80  M P K E R R E Q I V W R L K K M I V E I
1083  ATGCCCAAAGAACGTCGTGAACAGATCGTTTGGCGTTTGAAGAAGATGATTGTTGAGATC

 100  Q G H S D Y Q Q A I E T L L S L A E S Y
1143  CAGGGTCACTCTGATTACCAACAAGCCATCGAGACTCTCTTGAGTTTGGCTGAGTCATAT

 120  A G H G R D I S S Q G T T A T K G F M D
1203  GCTGGACATGGAAGAGATATCTCTTCGCAAGGAACGACCGCTACTAAAGGGTTCATGGAT

 140  S K K D M L M R L K I L I E R F A N S T
1263  AGCAAGAAGGACATGCTGATGCGACTCAAGATTTTAATCGAGAGATTTGCGAATAGCACA

 160  S T D D F F D S L N T I Y R D A D Q D P
1323  TCGACCGATGACTTCTTTGATTCGCTCAACACGATCTATCGCGATGCCGACCAGGACCCG

 180  R L K E W F R G V D T Y I R K C L R E Q
1383  CGGTTGAAGGAATGGTTTAGGGGTGTCGACACTTACATTCGAAAATGCCTGCGCGAGCAA

 200  G F I M Q D E A N D Q W N K L Y D E G R
1443  GGGTTCATCATGCAAGATGAAGCTAATGACCAGTGGAACAAGCTGTACGATGAGGGTCGC

 220  F L L R D R Y R S H T D R I A D E A K F
1503  TTCTTGCTTCGAGACCGGTACAGAAGTCACACAGATCGTATTGCCGATGAAGCCAAGTTT

 240  L A T Q F D E D P Q N R A F R Q S L E R
1563  TTGGCCACTCAATTTGACGAAGATCCACAGAACCGAGCTTTCCGTCAATCTTTGGAGAGA

 260  L F K D L G Q D Q Y G K P T F K P H L I
1623  CTCTTCAAGGATCTGGGTCAGGACCAGTATGGAAAGCCGACGTTCAAGCCCCATTTGATC

 280  K D I T N V I V P E I F E N A S Y I P I
1683  AAAGACATTACCAATGTCATCGTTCCTGAGATATTTGAGAATGCCAGCTACATTCCTATT

 300  P R I E V S D P A V D M V I E N L I I E
1743  CCTCGAATTGAGGTTTCCGACCCAGCTGTTGACATGGTTATCGAAAACCTTATCATCGAG

 320  S D N L M P N V L E F G T D N Y W R W G
1803  AGCGACAACTTGATGCCCAATGTTTTAGAATTCGGCACTGACAACTACTGGCGGTGGGGT

 340  R K K I S S F D D H K V M I S A S G I Q
1863  CGTAAGAAGATCAGCAGCTTCGACGACCACAAGGTTATGATTTCTGCATCTGGCATCCAA

 360  A D L R D V S Y Y F K K K Q G F P S L T
1923  GCTGATCTCCGTGATGTGAGCTACTATTTCAAGAAGAAGCAAGGATTCCCTTCCCTCACC

 380  D I G V M D I L L G G S G F G F K I A A
1983  GACATCGGTGTCATGGACATTCTTCTTGGTGGTTCAGGCTTCGGCTTCAAGATTGCTGCC

 400  S K A Q K N D H N A V F K L D S V K V N
2043  TCCAAAGCACAGAAGAATGACCACAACGCAGTTTTTAAGCTTGATAGTGTCAAGGTCAAC

 420  V K N L T I K L K K S K H K M L F I I F
2103  GTGAAGAACCTTACCATCAAGCTGAAGAAGTCCAAGCACAAGATGCTTTTCATCATCTTC

 440  R P M L L N V V R P V L E K C W R P T F
2163  CGACCCATGCTGTTAAATGTTGTCCGACCTGTTCTTGAAAAGTGTTGGAGGCCCACATTC

 460  L R H S K R Q T H L P T R Y R P R L S V
2223  CTGAGGCATTCCAAAAGGCAGACGCATTTGCCTACCAGGTACAGACCGAGGCTCAGCGTG

 480  P R R P C V R I L K M Q R I S S P A T P
2283  CCCAGGAGACCATGCGTGAGGATCCTGAAAATGCAAAGAATATCTTCGCCCGCTACGCCG

 500  M L L V T S S L E K K K Q A E A I A E R
2343  ATGCTACTCGTCACGTCATCACTGGAGAAGAAGAAGCAGGCCGAAGCTATCGCCGAGCGC

 520  G T K V H M A M T H Q D A M F K D I K L
2403  GGAACCAAGGTTCATATGGCAATGACTCACCAGGATGCCATGTTCAAGGACATCAAATTG
```

```
 540   P  G  G  V  T  N  K  A  T  E  F  K  E  L  S  A  K  G  D  R
2463   CCTGGTGGTGTCACTAACAAAGCTACTGAGTTCAAGGAGCTTTCTGCTAAGGGTGATCGC

 560   W  Q  S  P  V  F  N  W  G  G  A  S  P  T  S  N  F  P  K  A
2523   TGGCAATCACCAGTTTTCAACTGGGGCGGTGCTTCACCCACTAGCAACTTTCCCAAGGCT

 580   S  A  V  S  R  K  P  H  T  A  A  E  S  H  L  Q  E  K  S  T
2583   TCCGCAGTGTCTCGCAAGCCTCATACTGCCGCTGAAAGTCATCTGCAGGAGAAGTCAACC

 600   D  G  V  N  G  V  G  A  S  A  V  N  G  L  H  G  V  S  T  T
2643   GATGGTGTCAATGGAGTCGGAGCTAGTGCAGTCAATGGACTCCACGGAGTCTCGACGACT

 620   G  A  T  A  R  G  S  S  G  K  A  M  L  P  T  N  G  V  T  N
2703   GGAGCCACAGCTAGGGGATCCAGTGGCAAGGCAATGTTGCCAACCAATGGTGTCACCAAC

 640   G  Q  S  A  T  N  G  T  F  Q  K  E  V  E  R  A  F  D  A  N
2763   GGCCAGTCTGCCACCAACGGTACTTTCCAGAAGGAAGTCGAGCGTGCTTTTGACGCAAAC

 660   G  T  T  L  P  T  L  G  G  V  -
2823   GGCACCACTCTTCCAACCCTAGGTGGTGTTTAATCGACATAAATTGCGAGTTTGACATGG

2883   TAAACGAATGTCTGGAGAATTCTTTTTTGATACCAAGAACCAACCTATCTCAATAATGAC

2943   TAGTTTGGATGGAATGATGTGATTTGAGTTTGACTTTTGAAGACGAGGTATGATTCACCA

3003   ATTATTATAGCGAATCTGTTGGATACTTGTACAATGTTGTCTAGTCTCTAGTTTCCTTTT

3063   GTCAGATACTGATACTTAATAAAATAATTACTATTTTCCAAAAAAAAAAAAAAAGGGCGG

3123   CCGCTCTAGAGGATCCAAGCTTACGTACGCGTGCATGCGA
```

Figure 20. Sequence analysis of clone in group 2 (P3)

GROUP 3

3.1) Member: P5, P23

3.2) DNA sequence of P23:
GC content of ORF: 57%

```
1     CCAGCTGACT ACCATTGGAA CCCAACTACC CAACAACTTA AAACATCATC   50
51    ACTTCTTCAC CCTTCTTAAA TCCTTTACTA CCAAATCAAC CAATTAAAAT  100
101   GTCTCTCTTC CACCGCAGCG GCGACTTTGC TCCCCTCTTC CGTCTCCTCG  150
151   ACGACTATGA CCTCCACCGC TCCGGACGCG ACGGTCAGAC TCCTGCCTCC  200
201   AGTAGCATCT CAAGCTTCGC GCCACGCTTC GACGTCCGCG AGTCCAAGGA  250
251   TGCTTACCAT CTGGACGGCG AACTCCCCGG CATTGCTCAA AAGGACGTTG  300
301   AAATCGAATT CTCCGACCCG CAGACATTGA CCATCAAGGG TCGCTCGGTT  350
351   CGGGAATACC ACACCCTTCC CGAGAACGAG AACCCTCATG CTCCTAAGCC  400
401   CGCTTCTGTC GAAGACGCAC CCGAGTCCAG TGACGAGACA GCCGTCCAAA  450
451   AGTCTTCCGA CAAAAAAGAG GTCTCGAAGG CTCAGGGTAA CGGCTACAAG  500
501   TACTGGGTCA GCGAGCGCTC AGTCGGCGAG TTCCATCGCT CATTCAACTT  550
551   CCCTAGCCGC GTTGATCACA ATGGCGTCAA GGCTAGTTTG AAGAATGGTG  600
601   TTCTTACGGT GACGGTGCCC AAGGCTGCCC CTCCTACTAG TCGCAAGATC  650
651   ACAATTGAGT AAATTCTTCA ACCTCTCATC AATCGATTTG GAGAGATAAA  700
701   ATCCTCAACC TTCATCATGA ACCCGGTTAC AAAAATTTGA CTCTTTTTTT  750
751   TTCTTTCTCA ATTGATTCAT CCGACTCTTT TTCGCTCATT GCGAATGTCT  800
801   GAATGATTCT GCATTATCCT CTGCATTGCG GCGGTGCTTA GTTGGTTCCG  850
851   GTTACTTTGT TTCGTATAGC ACGATCCCGA TGTGTAATAG TTAGCTGATA  900
901   ATAATAATCA ACATCATTGA TT
```

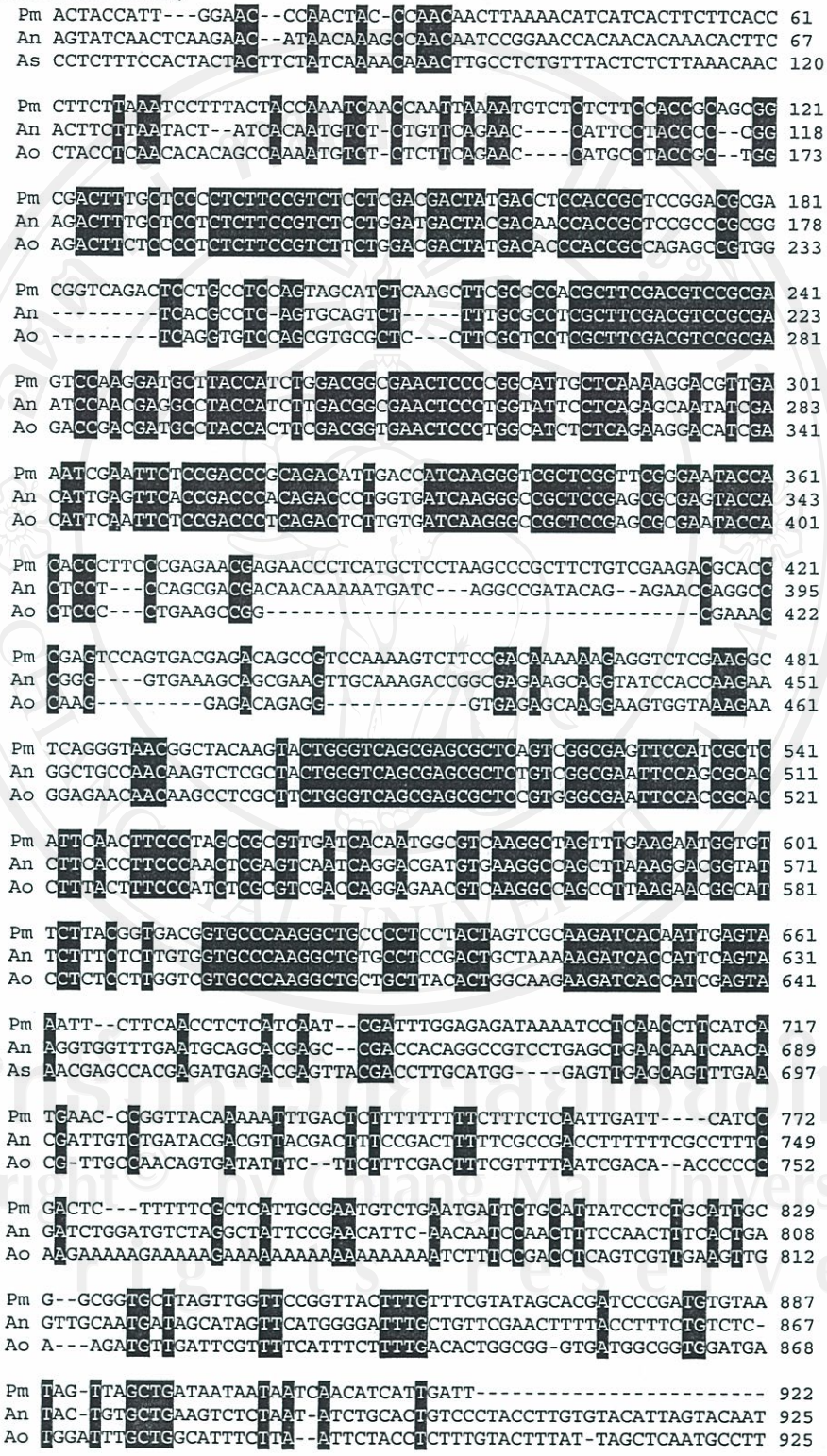
3.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLAST N	gi|577277|dbj|D32070.1| ASNG034ARA	mRNA for heat-shock protein 30	*Aspergillus nidulans* (fungus)	69.9	8e-09	53/59 (89%)
	gi|38564190|dbj|AB126868.1|	mRNA for heat shock protein 30	*Aspergillus oryzae* (fungus)	61.9	2e-06	91/111 (81%)
	gi|32492418|emb|AL731623.3|OSJN00263	Genomic DNA clone	*Oryza sativa* (rice)	42.1	1.8	24/25 (96%)
BLAST X	gi|38564191|dbj|BAD02411.1|	Hsp30	*Aspergillus oryzae* (fungus)	204	1e-51	111/190 (58%)
	gi|729763|sp|P40920|HS30_EMENI	Hsp30	*Aspergillus nidulans* (fungus)	204	2e-51	108/190 (56%)
	gi|32423237|ref|XP_332056.1|	Hsp30	*Neurospora crassa* (fungus)	151	2e-35	88/210 (41%)
	gi|19073387|gb|AAL84791.1|	Hsp30	*Exophiala dermatitidis* (fungus)	129	7e-29	75/168 (44%)

B. Multiple sequence alignment

Nucleotide level;



Note Pm = *Penicillium marneffei*, An = *Aspergillus nidulans*,
Ao = *Aspergillus oryzae*

Polypeptide level;

```
Pm MSLFHRS--------GDFAPLFRLLDDYDLH-RSGRDGQ-----TPASSSISSFAPRFDV 46
Ao MSLFRTMPT-----AGDFSPLFRLLDDYDTH-RQSRG---------QVSSVRSFAPRFDV 45
An MSLFRTIPT-----PGDFAPLFRLLDDYDNH-RSARG---------HAS-VQSFAPRFDV 44
Nc MALFPRGFYGSYGSDPSFTNLFRLLDDFDTYTREVQGSAPETGSRRHTQPTRTFSPKFDV 60
Ed MALFPRLSTSFGPSRQELGPFFNLFNDTFSELQKLSE----------SASRTFAPKFDV 49
   *:**           .:  :*.*::*   :.  .          .. :*:*:***

Pm RESKDAYHLDGELPGIAQKDVEIEFSDPQTLTIKGRSVREY------------------- 87
Ao RETDDAYHFDGELPGISQKDIDIQFSDPQTLVIKGRSEREY------------------- 86
An RESNEAYHLDGELPGIPQSNIDIEFTDPQTLVIKGRSEREY------------------- 85
Nc RETEQTYELHGELPGIDRDNVQIEFTDPQTIVIRGRVERNYTAGTPPAQVAGVLTEKGEP 120
Ed KEAQDKYILEGELPGIDQKNVTIQFEDDQTLTIKGRTEHHR------------------- 90
  :*:.: * :.****** :::: *:* * **:.*:**  :.

Pm ------HTLPENENPHAPKPASVEDAPESSDET--AVQKSSDKKEVSKAQGNGYKYWVSE 139
Ao ------HSP--------EAGETKETEGESKE-----VVK---------KENNKPRFWVSE 118
An ------HSSSD-DNKNDQADTENQARGESSE-----VAKTGEKQVSTKKAANKSRYWVSE 133
Nc HSPAAHHATVEDDVDEDNRSVATTATGANNQNNQQVAQRASAPTTEEKPKAPAEKYWVSE 180
Ed ---------EESQRPDQTSGEQGQQQGTSSS-----KEVATTGSKEVARSEPKHTYWVSE 136
          :  ...              .....      :              :****

Pm RSVGEFHRSFNFPSRVDHNGVKASLKNGVLTVTVPKAAPPT-SRKITIE-- 187
Ao RSVGEFHRTFTFPSRVDQENVKASLKNGILSLVVPKAAAYT-GKKITIE-- 166
An RSVGEFQRTFTFPTRVNQDDVKASLKDGILSLVVPKAVPPT-AKKITIQ-- 181
Nc RSIGEFSRTFNFPGRVDQNAVSASLNNGILTITVPKAKKHE-TIRIAIN-- 228
Ed RQVGEFARSFAFPNPVDQDNVKASLKNGILSVIVPKLEKSKGSKQIQITSE 187
   *.:*** *:* **  *::: *.***::*:*:: ***       :* *
```

Note Pm = *Penicillium marneffei*, Ao = *Aspergillus oryzae*, An = *Aspergillus nidulans*, Nc = *Neurospora crassa*, Ed = *Exophiala dermatitidis*

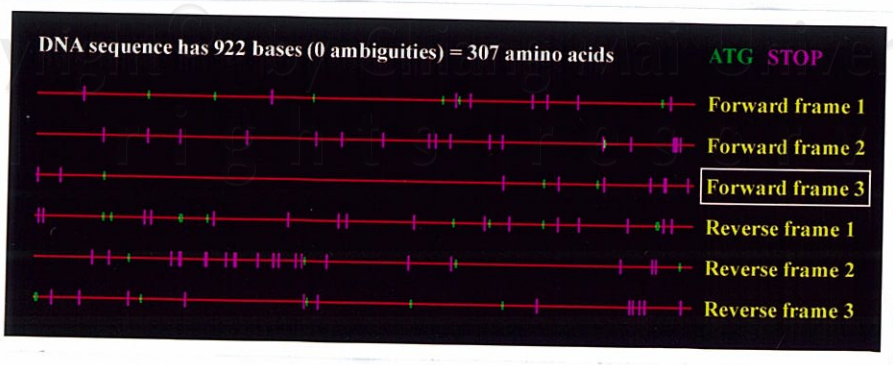
C. Motif scan search

Hits	Status	Amino acid position	Database
Heat shock protein 20 family	Significant match	45-175	PROSITE
Heat shock protein 20 family	Significant match	45-187	Pfam

D. Conclusion of functional analysis:
Possible function of encoded protein: Heat Shock Protein 30

3.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 3 was functionally identified as heat shock protein 30

C. Deduced amino acid composition (187 amino acids)
Predicted molecular weight (kDa): 20.76
Theoretical pI: 6.17

```
Forward Frame 3:

  3  AGCTGACTACCATTGGAACCCAACTACCCAACAACTTAAAACATCATCACTTCTTCACCC

                                          M  S  L  F  H  R  S  G
 63  TTCTTAAATCCTTTACTACCAAATCAACCAATTAAAATGTCTCTCTTCCACCGCAGCGGC

  9   D  F  A  P  L  F  R  L  L  D  D  Y  D  L  H  R  S  G  R  D
123  GACTTTGCTCCCCTCTTCCGTCTCCTCGACGACTATGACCTCCACCGCTCCGGACGCGAC

 29   G  Q  T  P  A  S  S  S  I  S  S  F  A  P  R  F  D  V  R  E
183  GGTCAGACTCCTGCCTCCAGTAGCATCTCAAGCTTCGCGCCACGCTTCGACGTCCGCGAG

 49   S  K  D  A  Y  H  L  D  G  E  L  P  G  I  A  Q  K  D  V  E
243  TCCAAGGATGCTTACCATCTGGACGGCGAACTCCCCGGCATTGCTCAAAAGGACGTTGAA

 69   I  E  F  S  D  P  Q  T  L  T  I  K  G  R  S  V  R  E  Y  H
303  ATCGAATTCTCCGACCCGCAGACATTGACCATCAAGGGTCGCTCGGTTCGGGAATACCAC

 99   T  L  P  E  N  E  N  P  H  A  P  K  P  A  S  V  E  D  A  P
363  ACCCTTCCCGAGAACGAGAACCCTCATGCTCCTAAGCCCGCTTCTGTCGAAGACGCACCC

119   E  S  S  D  E  T  A  V  Q  K  S  S  D  K  K  E  V  S  K  A
423  GAGTCCAGTGACGAGACAGCCGTCCAAAAGTCTTCCGACAAAAAAGAGGTCTCGAAGGCT

139   Q  G  N  G  Y  K  Y  W  V  S  E  R  S  V  G  E  F  H  R  S
483  CAGGGTAACGGCTACAAGTACTGGGTCAGCGAGCGCTCAGTCGGCGAGTTCCATCGCTCA

159   F  N  F  P  S  R  V  D  H  N  G  V  K  A  S  L  K  N  G  V
543  TTCAACTTCCCTAGCCGCGTTGATCACAATGGCGTCAAGGCTAGTTTGAAGAATGGTGTT

179   L  T  V  T  V  P  K  A  A  P  P  T  S  R  K  I  T  I  E  -
603  CTTACGGTGACGGTGCCCAAGGCTGCCCCTCCTACTAGTCGCAAGATCACAATTGAGTAA

663  ATTCTTCAACCTCTCATCAATCGATTTGGAGAGATAAAATCCTCAACCTTCATCATGAAC

723  CCGGTTACAAAAATTTGACTCTTTTTTTTCTTTCTCAATTGATTCATCCGACTCTTTTT

783  CGCTCATTGCGAATGTCTGAATGATTCTGCATTATCCTCTGCATTGCGGCGGTGCTTAGT

843  TGGTTCCGGTTACTTTGTTTCGTATAGCACGATCCCGATGTGTAATAGTTAGCTGATAAT

903  AATAATCAACATCATTGATT
```

Figure 21. Sequence analysis of clones in group 3 (P5, P23)

Group 4

4.1) Member: P6

4.2) DNA sequence of P6:

```
1     GACCGTTCCC GTCTTTTCTT GCAACATGAA AGAAGATGGA GACTGCGAAC   50
51    GACGCACCCT CAACTACACT GGAGGGACAA CATGAAGCAG ACGCTCAGAA  100
101   CAGCCGAACT CAACAACCCC AACAATCCCC CGACAGTCAA GAAGACGACG  150
151   CAGAAGAAGT ACGACCACCA TTGCCACCAC GTCCTGAGAC CATCGATCTG  200
201   CTAAATGAAG GCATTGCCTT TCGTACCTCC ACGGCGAGAC CAAACCTGCA  250
251   GTCGCACGCG ACAACGGCAC TGTCATTGAC AGACATTACC GGCCAAACGA  300
301   ATGCAGATGG AAGAGACGGT TTCGTTGCAG GCTTTGGGCG TACCTTGCTG  350
351   GGGCGGGGAC TACGAGCAAA GGCTAGTTTG AGCCAACTGA ATAGTGCTCG  400
401   TGGTAGTGAA GCCGGGGACA CCGCGAGCGT GTTAAGCTTT GCGCCGAATT  450
451   CCGAGGAAGG CCAGGATGAG AGTCTGTTTG GGGAGTTTGC AAATGAAACC  500
501   AACGCGCAGG ATATTTCTGG GAATATTGAA GTCCTGGGCT ATGATGAATA  550
551   TCCCCAGGAC GGCAATGAGT ACGAATTTGT AGAGGAATTC GAGCCGATAG  600
601   GGGAACTGGA TGAGGATGGC CAAAATGAAG AATCACTGCT CCAGAAATGG  650
651   TTCAGTGCAG GAGATACGAA GATTGTGATA CTCTCAAAGA CTCCGCTATA  700
701   TTTGGTTGCT ATCAGTCGAT TATTGGAAAG CGAAAGCCAT CTGAGACTTC  750
751   AACTTGATGC GCTGTACATG CAGATATTAT CGACCTTGAC CCTTCCGGCA  800
801   TTGAACCATC TATTTTCGAT CCGGCCATCT ACAGATCTAA AAAGACCCCT  850
851   ACAAGGTACC GAGACGTTAC TTTCCTCGCT CGCCGATTCT TTTACCAAGG  900
901   GCTCTCCCAC GACTCTCTTA TCGGCATTGG AGTGCTTGAA ACTACGCAAA  950
951   CATCAACAAT ATCTTACTAA GAAACAGAGC GG
```

4.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLAST N	gi|32978556|dbj|AK068539.1|	cDNA clone	*Oryza sativa* (rice)	48.1	0.038	27/28 (96%)
	gi|32972352|dbj|AK062334.1|	cDNA clone	*Oryza sativa* (rice)	48.1	0.038	27/28 (96%)
	gi|29150082|emb|BX295539.1|NCB1D14	BAC clone	*Neurospora crassa* (fungus)	46.1	0.15	32/35 (91%)
BLAST X	gi|1723226|sp|Q10150|YAT3_SCHPO	Hypothetical protein	*Shizosaccharomyces pombe* (fission yeast)	90.9	5e-18	52/162 (32%)
	gi|1723912|sp|P53129|YGM4_YEAST	Hypothetical protein	*Saccharomyces cerevisiae* (yeast)	77.8	4e-14	50/163 (30%)
	gi|6226165|sp|O67323|SYA_AQUAE	Alanyl-tRNA synthetase	Aquifex aeolicus (bacterium)	34.7	0.41	28/126 (22%)

B. Motif scan search

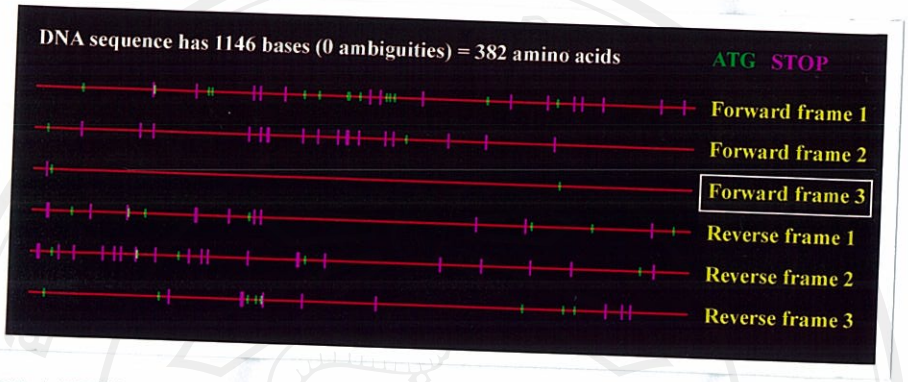
Hits	Status	Amino acid position	Database
SAND domain family (Domain of unknown function)	Weak match	199-370	Pfam

C. Conclusion of functional analysis:

Possible function of encoded protein: Unknown

4.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 3 matched to the hypothetical protein

C. Deduced amino acid composition (370 amino acids)

```
Forward Frame 3:

                                                M  E  T  A  N  D  A  P  S
  3  CCGTTCCCGTCTTTTCTTGCAACATGAAAGAAGATGGAGACTGCGAACGACGCACCCTCA

 10   T  T  L  E  G  Q  H  E  A  D  A  Q  N  S  R  T  Q  Q  P  Q
 63  ACTACACTGGAGGGACAACATGAAGCAGACGCTCAGAACAGCCGAACTCAACAACCCCAA

 30   Q  S  P  D  S  Q  E  D  D  A  E  E  V  R  P  P  L  P  P  R
123  CAATCCCCCGACAGTCAAGAAGACGACGCAGAAGAAGTACGACCACCATTGCCACCACGT

 50   P  E  T  I  D  L  L  N  E  G  I  A  F  R  T  S  T  A  R  P
183  CCTGAGACCATCGATCTGCTAAATGAAGGCATTGCCTTTCGTACCTCCACGGCGAGACCA

 70   N  L  Q  S  H  A  T  T  A  L  S  L  T  D  I  T  G  Q  T  N
243  AACCTGCAGTCGCACGCGACAACGGCACTGTCATTGACAGACATTACCGGCCAAACGAAT

 90   A  D  G  R  D  G  F  V  A  G  F  G  R  T  L  L  G  R  G  L
303  GCAGATGGAAGAGACGGTTTCGTTGCAGGCTTTGGGCGTACCTTGCTGGGGCGGGGACTA

110   R  A  K  A  S  L  S  Q  L  N  S  A  R  G  S  E  A  G  D  T
363  CGAGCAAAGGCTAGTTTGAGCCAACTGAATAGTGCTCGTGGTAGTGAAGCCGGGGACACC

130   A  S  V  L  S  F  A  P  N  S  E  E  G  Q  D  E  S  L  F  G
423  GCGAGCGTGTTAAGCTTTGCGCCGAATTCCGAGGAAGGCCAGGATGAGAGTCTGTTTGGG

150   E  F  A  N  E  T  N  A  Q  D  I  S  G  N  I  E  V  L  G  Y
483  GAGTTTGCAAATGAAACCAACGCGCAGGATATTTCTGGGAATATTGAAGTCCTGGGCTAT

170   D  E  Y  P  Q  D  G  N  E  Y  E  F  V  E  E  F  E  P  I  G
543  GATGAATATCCCCAGGACGGCAATGAGTACGAATTTGTAGAGGAATTCGAGCCGATAGGG

190   E  L  D  E  D  G  Q  N  E  E  S  L  L  Q  K  W  K  E  K  R
603  GAACTGGATGAGGATGGCCAAAATGAAGAATCACTGCTCCAGAAATGGAAGGAAAAGCGA

210   K  H  Y  L  I  L  S  A  A  G  K  P  I  Y  T  R  H  G  D  S
663  AAGCATTACTTGATTTTATCAGCTGCTGGAAAGCCAATATATACTCGACACGGCGATAGT

230   G  L  V  S  G  Y  I  G  I  I  Q  T  I  I  S  F  Y  Q  D  A
723  GGTTTGGTTTCGGGCTACATTGGTATCATACAAACAATCATCTCATTCTATCAGGACGCA

250   D  D  T  L  R  S  F  S  A  G  D  T  K  I  V  I  L  S  K  T
783  GATGACACTTTGCGGAGTTTCAGTGCAGGAGATACGAAGATTGTGATACTCTCAAAGACT
```

```
 270   P  L  Y  L  V  A  I  S  R  L  L  E  S  E  S  H  L  R  L  Q
 843   CCGCTATATTTGGTTGCTATCAGTCGATTATTGGAAAGCGAAAGCCATCTGAGACTTCAA

 290   L  D  A  L  Y  M  Q  I  L  S  T  L  T  L  P  A  L  N  H  L
 903   CTTGATGCGCTGTACATGCAGATATTATCGACCTTGACCCTTCCGGCATTGAACCATCTA

 310   F  S  I  R  P  S  T  D  L  K  R  P  L  Q  G  T  E  T  L  L
 963   TTTTCGATCCGGCCATCTACAGATCTAAAAAGACCCCTACAAGGTACCGAGACGTTACTT

 330   S  S  L  A  D  S  F  T  K  G  S  P  T  T  L  L  S  A  L  E
1023   TCCTCGCTCGCCGATTCTTTTACCAAGGGCTCTCCCACGACTCTCTTATCGGCATTGGAG

 350   C  L  K  L  R  K  A  H  R  Q  V  I  N  N  I  L  L  R  N  R
1083   TGCTTGAAACTACGCAAAGCACATCGACAAGTCATCAACAATATCTTACTAAGAAACAGA

 370   A
1143   GCGG
```

Figure 22. Sequence analysis of clone in group 4 (P6)

Group 5

5.1) Member: P7

5.2) DNA sequence of P7:

```
1       CCTCTTCCTC TCTTCACGCT CCTCTCCCT CGTGGAATCT ACATTCTTTC     50
51      ATGTGGTGAT CTCTCCACACT CCCATCTTT CTGCTGTAGC CTGGCTGCTC    100
101     GTCTCATTCA TTATATGTTCG TGATTACGA TTTGACCTGC ACAGGCTGTA    150
151     GTAGTCTGCC TTTCGTTCTCT ACCCCACGT TCTCTTTCAA AGTATTTGTG    200
201     TATCTTCCTC TGGCGCTGGCT ATCGACCTC CAGCGCTGTC CTTCCCTACC    250
251     AGAGTACTGC CTGGTTCGTAC TCGTTTCGC ATTTTCGATC CTCGGATCAC    300
301     CATTCGTCTC GATATCGTTAT CTTTCGCAT ATGACAGTTT GAATCGCCAT    350
351     AATACCGCGC CAATTACTTGT TGAACTTGT AAACAGCCAT ATTATCCAAG    400
401     AGCTTCGTCA TGTCGCACGCC CATTTCCAC CGTCATGAAA GACGGAATTT    450
451     CATTTCGGAC GTGGAGGATTT CTTTGGGGT CAATTACGCA AAGGGTCCAC    500
501     AGAGTACTGT AACTGAAACTG CGTCGATTG TATACGTGAC TGCGTCGCCA    550
551     ACCTTCACCG GCCCAATCGGT GGTTATATC ACGGGCACAG ATACTGCCAG    600
601     CCCTGCGGAT CCAACTACAAC CGCAGGAAA GGGTGCTCCT GTAGCTCAGT    650
651     CATCAACAAA GACAACCGCCG CGAGTTCAA AATCAGACAC ATCTACCACC    700
701     ACTAAGGTTC CTGCTCCCACG ACTCATACT ACTTCAACTC TCTCGACAAC    750
751     GTCAACCTCG CTGTCCGACTT GTTGACCAC CTCGAGCGAT TCTCCTGCGA    800
801     CAACTTTCTT GACGTCTTCTT CAACGCAAT CTGCATCTTC CACTTCAACC    850
851     GATACAGCAC TAGACCAGCTC TCGTCTTCT CCCACTGCTT CTCCCACTAC    900
901     TTCTGCAGCC GCATCCACTTC CAATGGTCT TACAGGAGGT GCCAAAGCAG    950
951     GAATCGCCAT TGGTGTCTTGT TCGGCGTTG GTCTTATTGC TGGATTGATT   1000
1001    CTGTTCTGGC TACATAAGCAG AAGAAGAAC AGAGAAGAAG CGGCTGCTGC   1050
1051    TGCCGCTGCT GCAGAAAATGA GAAATTTAC TCCAAGTCAA CCCCCACCAA   1100
1101    TGACTAGCGC TCAGTCACCAC AGTCAATGG CAGCTTACTC TTCTGCGCCG   1150
1151    TCAACTCCAG CGACTGCCCCT CAAGTCAGC TTGCGTCCGA TCACTCAGTT   1200
1201    CAACCCTTTG TTGAGCCAGCC TGGTGGTGC CAATCCTTAC GCCGCTGGTG   1250
1251    CTGTAGGTGC TGCCGTCGGTG GTGCTGCTG CTGCTGCCGG TGGCCTACAA   1300
1301    GTAAATCGGT CAGCTGAACGC CCTTACAGC GGCTCAGCAC ATGTTCCTCC   1350
1351    TCAGTCTCCT CGTCAAGATCC ATTCACTGA TCCTGTTAAT CCTTTCGATA   1400
1401    ATGGAGCGCA AACTGCATCTC CTCCTATGC CCCCTGCTAA AGATGCATCG   1450
1451    TCTCCAGTCA GGGATTTGACG CCTTCGCCT ACCGGTAGTG CCCACAACTT   1500
1501    GGCCAGCCCT ATTGCTGAAGA ACCGTCCGC TGGGTCAGTA GAAGCTGCCG   1550
1551    CTGCCGGTGC TGTCGCAGGTG CTGCTGTTG GCGCTGCTGC TGTTGC
```

5.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organisms			
BLASTN	gi|34850232|dbj|AP005611.2|	BAC clone	*Oryza sativa* (rice)	50.1	0.014	28/29 (96%)
	gi|38199254|emb|BX248355.1|	Genomic sequence	*Corynebacterium diphtheriae* (bacterium)	50.1	0.014	34/37 (91%)
	gi|34787436|gb|AC091776.19|	clone cr-4i21	*Chlamydomonas reinhardtii* (bacterium)	48.1	0.054	24/24 (100%)
BLASTX	gi|465861|sp|P34735|YLU2_PICAN	Hypothetical protein	*Pichia angusta* (yeast)	38.9	0.034	19/39 (48%)
	gi|418612|sp|Q04584|ZYX_CHICK	Zyxin	*Gallus gallus* (plant)	36.6	0.17	20/65 (30%)
	gi|729114|sp|P40198|CEA3_HUMAN	Carcinoembryonic antigen CGM1	*Homo sapiens* (human)	36.6	0.17	26/74 (35%)

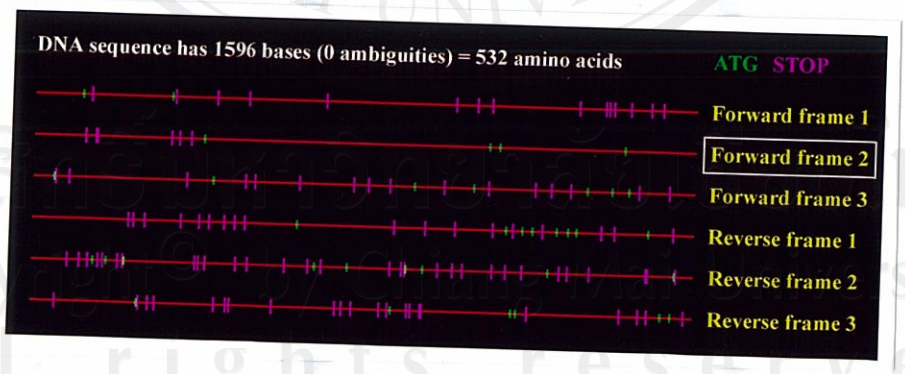
B. Motif scan search

Hits	Status	Amino acid position	Database
Threonine rich region	Significant match	33-175	PROSITE
Serine rich region	Significant match	81-171	PROSITE
Proline rich region	Weak match	313-357	PROSITE
Alanine rich region	Weak match	211-293	PROSITE
Domain of unknown function	Weak match	179-279	Pfam

C. Conclusion of functional analysis:
Possible function of encoded protein: Unknown

5.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 2 matched to the hypothetical protein

C. Deduced Amino Acid Composition (406 amino acids)

Forward Frame 2

```
   2  CTCTTCCTCTCTTCACGCTTCCTCTCCCTCGTGGAATCTACATTCTTTCATGTGGTGATC

  62  TCTCCACACTCCCATCTTTCTGCTGTAGCCTGGCTGCTCGTCTCATTCATTATATGTTCG

 122  TGATTACGATTTGACCTGCACAGGCTGTAGTAGTCTGCCTTTCGTTCTCTACCCCACGTT

 182  CTCTTTCAAAGTATTTGTGTATCTTCCTCTGGCGCTGGCTATCGACCTCCAGCGCTGTCC

 242  TTCCCTACCAGAGTACTGCCTGGTTCGTACTCGTTTCGCATTTTCGATCCTCGGATCACC

 302  ATTCGTCTCGATATCGTTATCTTTCGCATATGACAGTTTGAATCGCCATAATACCGCGCC

                                                      M  S  H  A
 362  AATTACTTGTTGAACTTGTAAACAGCCATATTATCCAAGAGCTTCGTCATGTCGCACGCC

   5   H  F  H  R  H  E  R  R  N  F  I  S  D  V  E  D  F  F  G  V
 422  CATTTCCACCGTCATGAAAGACGGAATTTCATTTCGGACGTGGAGGATTTCTTTGGGGTC

  25   N  Y  A  K  G  P  Q  S  T  V  T  E  T  A  S  I  V  Y  V  T
 482  AATTACGCAAAGGGTCCACAGAGTACTGTAACTGAAACTGCGTCGATTGTATACGTGACT

  45   A  S  P  T  F  T  G  P  I  G  G  Y  I  T  G  T  D  T  A  S
 542  GCGTCGCCAACCTTCACCGGCCCAATCGGTGGTTATATCACGGGCACAGATACTGCCAGC

  65   P  A  D  P  T  T  T  A  G  K  G  A  P  V  A  Q  S  S  T  K
 602  CCTGCGGATCCAACTACAACCGCAGGAAAGGGTGCTCCTGTAGCTCAGTCATCAACAAAG

  95   T  T  A  A  S  S  K  S  D  T  S  T  T  T  K  V  P  A  P  T
 662  ACAACCGCCGCGAGTTCAAAATCAGACACATCTACCACCACTAAGGTTCCTGCTCCCACG

 115   T  H  T  T  S  T  L  S  T  T  S  T  S  L  S  D  L  L  T  T
 722  ACTCATACTACTTCAACTCTCTCGACAACGTCAACCTCGCTGTCCGACTTGTTGACCACC

 135   S  S  D  S  P  A  T  T  F  L  T  S  S  S  T  Q  S  A  S  S
 782  TCGAGCGATTCTCCTGCGACAACTTTCTTGACGTCTTCTTCAACGCAATCTGCATCTTCC

 155   T  S  T  D  T  A  L  D  Q  L  S  S  S  P  T  A  S  P  T  T
 842  ACTTCAACCGATACAGCACTAGACCAGCTCTCGTCTTCTCCCACTGCTTCTCCCACTACT

 175   S  A  A  A  S  T  S  N  G  L  T  G  G  A  K  A  G  I  A  I
 902  TCTGCAGCCGCATCCACTTCCAATGGTCTTACAGGAGGTGCCAAAGCAGGAATCGCCATT

 195   G  V  L  F  G  V  G  L  I  A  G  L  I  L  F  W  L  H  K  Q
 962  GGTGTCTTGTTCGGCGTTGGTCTTATTGCTGGATTGATTCTGTTCTGGCTACATAAGCAG

 215   K  K  N  R  E  E  A  A  A  A  A  A  A  E  N  E  K  F  T
1022  AAGAAGAACAGAGAAGAAGCGGCTGCTGCTGCCGCTGCTGCAGAAAATGAGAAATTTACT

 235   P  S  Q  P  P  P  M  T  S  A  Q  S  P  Q  S  M  A  A  Y  S
1082  CCAAGTCAACCCCCACCAATGACTAGCGCTCAGTCACCACAGTCAATGGCAGCTTACTCT

 255   S  A  P  S  T  P  A  T  A  P  Q  V  S  L  R  P  I  T  Q  F
1142  TCTGCGCCGTCAACTCCAGCGACTGCCCCTCAAGTCAGCTTGCGTCCGATCACTCAGTTC

 275   N  P  L  L  S  Q  P  G  G  A  N  P  Y  A  A  G  A  V  G  A
1202  AACCCTTTGTTGAGCCAGCCTGGTGGTGCCAATCCTTACGCCGCTGGTGCTGTAGGTGCT

 295   A  V  G  G  A  A  A  A  A  G  G  L  Q  V  N  R  S  A  E  R
1262  GCCGTCGGTGGTGCTGCTGCTGCTGCCGGTGGCCTACAAGTAAATCGGTCAGCTGAACGC

 315   P  Y  S  G  S  A  H  V  P  P  Q  S  P  R  Q  D  P  F  T  D
1322  CCTTACAGCGGCTCAGCACATGTTCCTCCTCAGTCTCCTCGTCAAGATCCATTCACTGAT

 335   P  V  N  P  F  D  N  G  A  Q  T  A  S  P  P  M  P  P  A  K
1382  CCTGTTAATCCTTTCGATAATGGAGCGCAAACTGCATCTCCTCCTATGCCCCCTGCTAAA
```

```
 355   D  A  S  S  P  V  R  D  L  T  P  S  P  T  G  S  A  H  N  L
1442   GATGCATCGTCTCCAGTCAGGGATTTGACGCCTTCGCCTACCGGTAGTGCCCACAACTTG

 375   A  S  P  I  A  E  E  P  S  A  G  S  V  E  A  A  A  A  G  A
1502   GCCAGCCCTATTGCTGAAGAACCGTCCGCTGGGTCAGTAGAAGCTGCCGCTGCCGGTGCT

 395   V  A  G  A  A  V  G  A  A  A  V  A
1562   GTCGCAGGTGCTGCTGTTGGCGCTGCTGCTGTTGC
```

Figure 23. Sequence analysis of clone in group 5 (P7)

Group 6

6.1) Member: P9

6.2) Five-prime sequence of P9:

```
1       CGTTGTTCGT GTAACGAATAC CTCCATACC TATACTTGTA TATCCGACGT     50
51      ACGTACAGCT ACGGCTCTTCT CCGCTGAAC CGAGTTCCTC ACATCGTCGA    100
101     TATTCTTGTA TCTCAGCAAAA ATAGATCAC CAATATCCTA AACGATCATC    150
151     ACCACAATGG CAGCCAACGGC AATGGCGGT TCCGGCGTTG GCCAGGAGAA    200
201     CATCAACACA GATATCATCAC TCTCACACG ATTCCTGACG GAAGAGCAGC    250
251     TTCGAGTCCC AGAAGCTACCG GTGATTTCA CTCTCCTCTG CCACGCCCTG    300
301     CAATTCGCTT TCAAGTAATTG CCTACTACA TTCGACGAGC CAGCCTCATC    350
351     AACCTCACCG GTCTAGCCGGC TCCTCCAAC ATAACCGGCG ACGACCAAAA    400
401     GAAACTCGAC GTAATCGGCAA CGATGTCTT CATCGCCGCC ATGCGCGGCT    450
451     CAGGCAAGGT GCGCCTGCTCG TCTCCGAAG AAGAAGAAGA GGCAATCATC    500
501     TTCG
```

6.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organisms			
BLASTN	gi|9955388|dbj|AB030248.1|	*fbpA* gene for fructose-1,6-bisphosphatase	*Aspergillus oryzae* (fungus)	100	5e-18	101/118 (85%)
	gi|32407116|ref|XM_324153.1|	mRNA	Neurospora crassa (fungus)	82	1e-12	135/165 (81%)
	gi|5831147|emb|AL115931.1|CNS01CSJ	cDNA clone under nitrogen deprivation conditions	*Botrytis cineria* (fungus)	64	3e-07	83/100 (83%)
	gi|21104554|dbj|AP003206.4|	BAC clone	*Oryza sativa* (rice)	44	0.24	22/22 (100%)
BLASTX	gi|119747|sp|P09201|F16P_YEAST	Fructose-1,6-bisphosphatase	*Saccharomyces cerevisiae* (yeast)	81	3e-27	39/65 (60%),
	gi|462044|sp|Q05079|F16P_KLULA	Fructose-1,6-bisphosphatase	*Kluyveromyces lactis* (fungus)	86	8e-26	40/65 (61%)
	gi|1169585|sp|P46267|F16Q_BRANA	Fructose-1,6-bisphosphatase	*Brassica napus* (plant)	67	2e-16	31/55 (56%)
	gi|119748|sp|P14766|F16Q_SPIOL	Fructose-1,6-bisphosphatase	*Spinacia oleracea* (plant)	66	4e-16	30/55 (54%)
	gi|3913640|sp|O64421|F16Q_ORYSA	Fructose-1,6-bisphosphatase	*Orysa sativa* (rice)	66	7e-16	30/59 (50%)

B. Multiple sequence alignment

Nucleotide level;

```
Pm ------------------CGTTGTTCGTGTAACGAATACCTCCATACCTATACTTGTATA  42
Ao ACCCGGCCATTATCACCTCACATCCATTGGTGCCCTAGCCTTCCTGCTTTCTCCGCTGAA 780

Pm TCCGACGTACGTACAGCT----ACGGCTCTTCTCCGCT---GAACC----GAGTTCCTCA  91
Ao CCGAGGTTGCGCATTACTTTAAACCGGTCCATTGCGCCCCAGAACCACTTGATCTCCCCT 840

Pm CATCGTCGATATTCTTGTATCTCAGCAAAAATAGATCA-CCAATATCCTAAACGATCATC 150
Ao CCTGAACAATCTGCGCATATCCTCTCTGTGCTGAAGCATCCCAAATCAACATCAATCATC 900

Pm ACCACAATGGCAGCCAACGGCAATGGCGGTTCCGGCGTTGGCCAGGAGAACATCAACACA 210
Ao ATCATGTCTGGACACCAGAACAACGGTGCTCCCCCCGTGGGAAAGGAGAAGATCAACACC 960
```

```
Pm GATATCATCACTCTCACACGATTCCTGACGGAAGAGCAGCTTCGAGTCCCAGAAGCTACC 270
Ao GACATCGTCACCCTCACAAGGTTTCTGACAGAAGAACAAACCAAGGTGCCAGAAGCCACT 1020

Pm GGTGATTTCACTCTCCTCTGCCACGGCCTGCAATTCGCTTTCAAGTA-ATTGCCTACTAC 329
Ao GGTGACTTCACACTCCTCTGCCACGCTCTTCAGTTCTCCTTCAAGTCCATCGCCTACTAT 1080

Pm ATTCGACGAGCCAGCCTCATCAACCTCACCGGTCTAGCCGGCTCCTCCAACATAACCGGC 389
Ao ATCCGTTGCGCATTCTTAATCAACCTCACAGGACTCGCGGGTTCCTCAAACACCACAGGC 1140

Pm GACGACCAAAAGAAACTCGACGTAATCGGCAACGATGTCTTCATCGCCGCCATGCGCGGC 449
Ao GATGACCAGAAGAAGCTCGACGTAATCGGCAATGATATCTTCATCTCCGCCATGCGCGGC 1200

Pm TCAGGCAAGGTGCGCCTGCTCGTCTCCGAAGAAGAAGAAGAGGCAATCATCTTCG----- 504
Ao TCAGGTAAATGCCGTATCCTCGTCTCGGAAGAAGAAGAAGAAGCCATCATCTTCGACGAG 1260
```

Note Pm = *Penicillium marneffei*, Ao = *Aspergillus oryzae*

Polypeptide level;

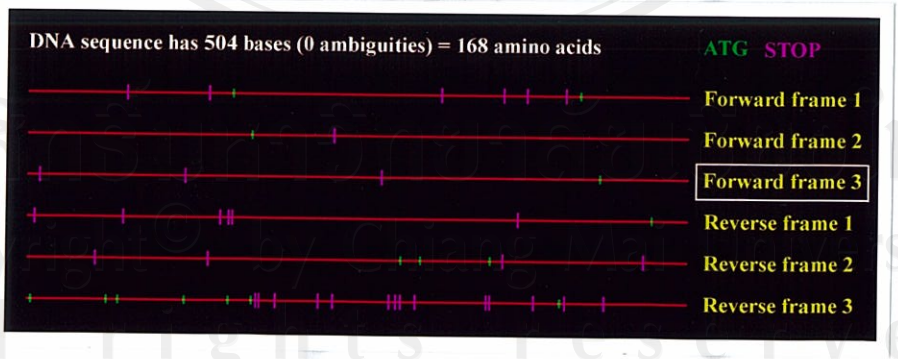
```
Pm ----------------------------------MRGSGKVRLLVSEEEEEAIIF--- 21
Sc RRAELVNLVGLAGASNFTGDQQKKLDVLGDEIFINAMRASGIIKVLVSEEQEDLIVFPTN 119
Kl RRAELVNLIGLAGASNSTGDQQKKLDVLGDEIFINAMKASGNVKVLVSEEQEDLIVFRNS 118
Bn NKAGLAKLIGLAGDTNIQGEEQKKLDVLSNDVFVKALVSSGRTSVLVSEEDEEATFVESS 110
So NKAGLAKLIGLAGETNIQGEEQKKLDVLSNEVFVKALTSSGRTCILVSEEDEEATFIEPS 111
Os NKAGLAKLIGLAGETNVQGEEQKKLDVLSNEVFVKALVSSGRTCVLVSEEDEEATFVDPA 111
                                    : .**   :*****:*:  ..
```

Note Pm = *Penicillium marneffei*, Sc = *Saccharomyces cerevisiae*,
Kl = *Kluyveromyces lactis* (fungus), Bn = *Brassica napus* (plant),
So = *Spinacia oleracea* (plant), Os = *Orysa sativa* (rice)

C. Motif scan search: No match found
D. Conclusion of functional analysis:
Possible function of encoded protein: Fructose-1,6-bisphosphatase

6.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 3 match to the fructose-1,6-bisphosphatase

C. Deduced amino acid composition

```
Forward Frame 3:

  3  TTGTTCGTGTAACGAATACCTCCATACCTATACTTGTATATCCGACGTACGTACAGCTAC

 63  GGCTCTTCTCCGCTGAACCGAGTTCCTCACATCGTCGATATTCTTGTATCTCAGCAAAAA

123  TAGATCACCAATATCCTAAACGATCATCACCACAATGGCAGCCAACGGCAATGGCGGTTC

183  CGGCGTTGGCCAGGAGAACATCAACACAGATATCATCACTCTCACACGATTCCTGACGGA

243  AGAGCAGCTTCGAGTCCCAGAAGCTACCGGTGATTTCACTCTCCTCTGCCACGCCCTGCA

303  ATTCGCTTTCAAGTAATTGCCTACTACATTCGACGAGCCAGCCTCATCAACCTCACCGGT

363  CTAGCCGGCTCCTCCAACATAACCGGCGACGACCAAAAGAAACTCGACGTAATCGGCAAC

                        M  R  G  S  G  K  V  R  L  L  V  S  E  E
423  GATGTCTTCATCGCCGCCATGCGCGGCTCAGGCAAGGTGCGCCTGCTCGTCTCCGAAGAA

 15   E  E  E  A  I  I  F
483  GAAGAAGAGGCAATCATCTTCG
```

Figure 24. Sequence analysis of clone in group 6 (P9)

Group 7

7.1) Member: P10

7.2) Five-prime sequence of P10:

```
1     AAACGACACA AAATGGGCCG TCTCAACGAG TACCAGGTCA TCGGGCGCCA    50
51    CTTGCCCACC GAGGCAAACC CGACCCCCAA GTTGTACCGC ATGCGCATCT   100
101   TTGCTCCCAA CACAGTTGTT GCCAAGTCGC GCTTCTGGTA CTTCCTCACC   150
151   AAGCTCAAGA AGGTCAAGAA GGCCAACGGT GAGATCGTCA GCCTCAACGT   200
201   GATCTCCGAG AAGCGTCCCA CCAAGGTCAA GAACTTCGGT ATCTGGTTGC   250
251   GTTACGACTC TCGCTCCGGC ACCCACAACA TGTACAAGGA GTTCCGTGAG   300
301   CTCAGCCGAA CTGATGCCGT CGACTCTTTG TACCAGGACA TGGCTGCCCG   350
351   TCACCGTGCC CGCTTCGGCT CCATTCACAT TCTCCGCGTC ATCGAGATCG   400
401   TTACATCAAG C
```

7.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLASTN	gi|5829175|emb|AL114556.1|CNS01BQC	cDNA clone under conditions of nitrogen deprivation	*Botrytis cinerea* (fungus)	147	2e-32	221/270 (81%)
	gi|5826574|emb|AL111955.1|CNS019Q3	cDNA clone under conditions of nitrogen deprivation	*Botrytis cinerea* (fungus)	98	1e-17	118/141 (83%)
	gi|1177347|emb|Z69240.1|SPAC26A3	Cosmid clone	*Shizosaccharomyces pombe* (fission yeast)	92	9e-16	61/66 (92%)
	gi|4927949|gb|AF127913.1|AF127913	ribosomal protein L20A (rpl20-1) mRNA	*Shizosaccharomyces pombe* (fission yeast)	92	9e-16	61/66 (92%)
	gi|2647409|dbj|AB009012.1|	mRNA for ribosomal protein L18	*Schizosaccharomyces pombe* (fission yeast)	92	9e-16	61/66 (92%)
BLASTX	gi|1350679|sp|P47913|RL20_YEAST	60S ribosomal protein L20 (L18A)	*Saccharomyces cerevisiae* (yeast)	210	9e-55	99/141 (70%)
	gi|1710497|sp|P05732|RL20_SCHPO	60S ribosomal protein L20 (YL17)	*Schizosaccharomyces pombe* (fission yeast)	199	1e-51	92/138 (66%)
	gi|3914712|sp|O57561|RL1X_SALSA	60S ribosomal protein L18a	*Salmo salar* (Atlantic salmon)	186	2e-47	91/140 (65%)
	gi|21362864|sp|Q90YU9|RL1X_ICTPU	60S ribosomal protein L18a	*Ictalurus punctatus* (channel catfish)	179	2e-45	88/140 (62%)
	gi|21362850|sp|Q8WQI7|RL1X_SPOFR	60S ribosomal protein L18a	*Spodoptera frugiperda* (fall armyworm)	178	4e-45	86/140 (61%)

B. Multiple sequence alignment

Nucleotide level;

```
Pm ------------------AAACGACACAAAATGGCCGT------CTCAACGAGTACCAG 36
Sp GCAAAGTGAAACTTTTACGAACAACACCAATTGGGGAAAATGGCACTCAAGGAGTATCAA 60

Pm GTCATCGGGCGCCACTTGCCCACCGAGGCAAACCCGACCCCCAAGTTGTACCGCATGCGC 96
Sp GTCGTTGGACGCAAGGTTCCTACGGAACATGAGCCTGTTCCCAAGCTATTCCGTATGCGT 120

Pm ATCTTTGCTCCCAACACAGTTGTTGCCAAGTCGCGCTTCTGGTACTTCCTCACCAAGCTC 156
Sp TTGTTCGCACCTAATGAATCTGTTGCCAAGTCTCGTTATTGGTACTTCTTGAAGATGATC 180

Pm AAGAAGGTCAAGAAGGCCAACGGTGAGATCGTCAGCCTCAACGTGATCTCCGAGAAGCGT 216
Sp AACAAAGTCAAGAAGGCCACTGGTGAGATCGTTGCCATCAATGAGATTTCCGAGCCTAAG 240

Pm CCCACCAAGGTCAAGAACTTCGGTATCTGGTTGCGTTACGACTCTCGCTCCGGCACCCAC 276
Sp CCGTTGAAGGCCAAGGTCTTCGGTATCTGGATTCGTTATGACTCTCGCTCTGGTACCCAC 300

Pm AACATGTACAAGGAGTTCCGTGAGCTCAGCCCAACTGATGCCGTCGACTCTTTGTACCAG 336
Sp AACATGTACAAGGAGTTCCGTGACACTACTCGTGTTGGTGCCGTTGAGGCTATGTATGCC 360

Pm GACATGGCTGCCCGTCACCGTGCCCGCTTCGGCTCCATTCACATTCTCCGCGTCATCGAG 396
Sp GATATGGCTGCCCGTCATCGTGCTCGTTTCCGCAGCATCCGTATCCTTAAGGTTGTTGAG 420

Pm ATCGAGGACAACGAGAGCATCCGCCGCCCTTACATCAAGC-------------------- 436
Sp GTTGAGAAGAAGGAGGACGTTCGCCGTAACTATGTCAAGCAACTCCTTAACCCTCACTTG 480
```

Note Pm = *Penicillium marneffei*, Sp = *Shizosaccharomyces pombe*

Polypeptide level;

```
Pm KRHKMGRLNEYQVIGRHLPTEANPTPKLYRMRIFAPNTVVAKSRFWYFLTKLKKVKKANG 60
Sc --MYLAHFKEYQVIGRRLPTESVPEPKLFRMRIFASNEVIAKSRYWYFLQKLHKVKKASG 58
Sp -----MALKEYQVVGRKVPTEHEPVPKLFRMRLFAPNESVAKSRYWYFLKMINKVKKATG 55
Ss -MKASGTLREYKVVGRLLPSVKNPTPPLYRMRIFAPNHVVAKSRFWYFVSQLRKMKKANG 59
Ip -MKASGTLREYKVVGRLLPSVKNPTPPLYRMRIFAPNHVVAKSRFWYFVSQLRKMKKANG 59
      :.**:*:** :*:   * * *:***:**.*  :****:***:  :.*:***.*

Pm EIVSLNVISEKRPTKVKNFGIWLRYDSRSGTHNMYKEFRELSRTDAVDSLYQDMAARHRA 120
Sc EIVSINQINEAHPTKVKNFGVWVRYDSRSGTHNMYKEIRDVSRVAAVETLYQDMAARHRA 118
Sp EIVAINEISEPKPLKAKVFGIWIRYDSRSGTHNMYKEFRDTTRVGAVEAMYADMAARHRA 115
Ss ETVYCGLVHEKTPLKVKNFGVWLRNDSRSGTHNMYREYRDLTTSAAVTQCYRDMGARHRA 119
Ip ETVYCGLVHEKTPLKVKNFGVWLRNDSRSGTHNMYREYRDLTTSAAVTQCYRDMGARHRA 119
   * *  . : *  * *.* **:*:* **********:* *: :   **   * **.*****

Pm RFGSIHILRVIEIEDNESIRRPYIK---------------------------------- 145
Sc RFRSIHILKVAEIEKTADVKRQYVKQFLTKDLKFPLPHR----VQKS-TKTFSYKRPSTF 173
Sp RFRSIRILKVVEVEKKEDVRRNYVKQLLNPHLKFPLPHRRTGVVGLAGKKVFAPHRPSTF 175
Ss RAHSIHIMKVQEIAAN-KCRRPAIKQFHDSKIKFPLPHR---VLRRQHNPRFTTKRPNTF 175
Ip RAHSIHIMKVQEIAAN-KCRRPAIKQFHDSKIKFPLPHR---VLRRQHNPRFTTKRPNTF 175
   *  *:*::* *:   . . :*  :*
```

Note Pm = *Penicillium marneffei*, Sc = *Saccharomyces cerevisiae*,
Sp = *Shizosaccharomyces pombe*, Ss = *Salmo salar* (salmon),
Ip = *Ictalurus punctatus* (channel catfish)

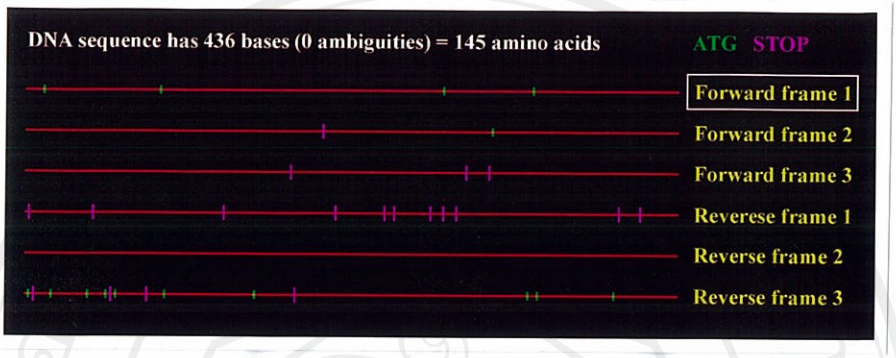
C. Motif scan search

Hits	Status	Amino acid position	Database
Death domain	Weak match	46-114	PROSITE
Ribosomal L18AE family	Significant match	7-145	Pfam

D. Conclusion of functional analysis:
Possible function of encoded protein: 60S ribosomal protein

7.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 1 match with the 60S ribosomal protein

C. Deduced amino acid composition

```
Forward Frame 1:
  1    K  R  H  K  M  G  R  L  N  E  Y  Q  V  I  G  R  H  L  P  T
  1    AAACGACACAAAATGGGCCGTCTCAACGAGTACCAGGTCATCGGGCGCCACTTGCCCACC

 21    E  A  N  P  T  P  K  L  Y  R  M  R  I  F  A  P  N  T  V  V
 61    GAGGCAAACCCGACCCCCAAGTTGTACCGCATGCGCATCTTTGCTCCCAACACAGTTGTT

 41    A  K  S  R  F  W  Y  F  L  T  K  L  K  K  V  K  K  A  N  G
121    GCCAAGTCGCGCTTCTGGTACTTCCTCACCAAGCTCAAGAAGGTCAAGAAGGCCAACGGT

 61    E  I  V  S  L  N  V  I  S  E  K  R  P  T  K  V  K  N  F  G
181    GAGATCGTCAGCCTCAACGTGATCTCCGAGAAGCGTCCCACCAAGGTCAAGAACTTCGGT

 81    I  W  L  R  Y  D  S  R  S  G  T  H  N  M  Y  K  E  F  R  E
241    ATCTGGTTGCGTTACGACTCTCGCTCCGGCACCCACAACATGTACAAGGAGTTCCGTGAG

101    L  S  R  T  D  A  V  D  S  L  Y  Q  D  M  A  A  R  H  R  A
301    CTCAGCCGAACTGATGCCGTCGACTCTTTGTACCAGGACATGGCTGCCCGTCACCGTGCC

121    R  F  G  S  I  H  I  L  R  V  I  E  I  E  D  N  E  S  I  R
361    CGCTTCGGCTCCATTCACATTCTCCGCGTCATCGAGATCGAGGACAACGAGAGCATCCGC

141    R  P  Y  I  K
421    CGCCCTTACATCAAGC
```

Figure 25. Sequence analysis of clone in group 7 (P10)

Group 8

8.1) Member: P11

8.2) Five-prime sequence of P11:

```
1     GACTCCTCAA AAGAAATGGA CGTAGTTTCA AACGAAAGAC TGGACGTGAG   50
51    TTGATCTCGT TGACTCCCAA ACTTAACGGA AAGACAAAAT GTTCTTCGAG  100
101   GAAGGAGGAC ACTTTGCCCG TTGCAGGAAA GCAGTCACGT GAGCCCACTA  150
151   TCAGTGGCAA CGGCGGAGCA CCGCGGACAG AGCGAACCAA TAAACAGCCG  200
201   CAAACCGGGG AAGCTAATCT AATCCATAGA ACGATTTGGA AGCTCTACGA  250
251   TGGCGCCACA GATACCAGCA AACTAGTCAT TTTGATACAT AACAGAATCC  300
301   GTGGTATGAG TGAATGGTCT TACCAGCTGA ATAAATCCAC AATATACCCA  350
351   GCCCATAACG AGTCAGCAAC TAGTAGTATA TGGCGGATAA GATACAGCTT  400
401   CCCCCACATT TCAGCCACGG CTTGTGCACG TCA
```

8.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLASTN	gi|15706113|gb|AC022176.6|	clone RP11-149C3	*Homo sapiens* (human)	44	0.22	22/22 (100%)
	gi|40789195|emb|AL773518.3|	clone RP24-231F12 on chromosome 4	*Mus musculus* (mouse)	40	3.4	20/20 (100%)
BLASTX	gi|26327013|dbj|BAC27250.1|	unnamed protein product	*Mus musculus* (mouse)	39	0.022	32/110 (29%)
	gi|45191043|ref|NP_985297.1|	AER442Wp	*Eremothecium gossypii* (fungus)	35	0.32	20/61 (32%)

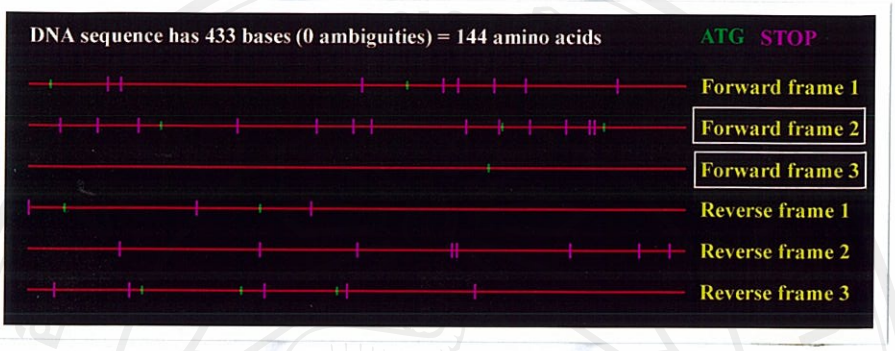
B. Motif scan search: No match

C. Conclusion of functional analysis:

Possible function of encoded protein: Unknown

8.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: No match

Figure 26. Sequence analysis of clone in group 8 (P11)

Group 9

9.1) Member: P12, P25, P27

9.2) Five-prime sequence of P12:

```
1     CAAAAACGCC AATCACACTC CGCCACAGCA AGAGTCAATC ATCAGCCGCA   50
51    TACATACAAT AACTCTAGCC TAGCTTAATC AGTACTCAAT CCCACGTTCA  100
101   ACTTCAACCG ACATAATGGC CATGCTCTCG TCATCATTGT CCCGGCCCAT  150
151   CCTCCTGGGC ATCCGAACCA TGCAATGGGC CAGCTCCGTC ATCGCCTTGG  200
201   GTATATACGC CTACTTTGTG CATCACCAGC GCAGTGGCAC AAACCCCATC  250
251   TTCAATCTAG TCATTTCTGT CCTGTCCGTT GTCTTCTTCA TCCCTGCTTT  300
301   CGTGTCGCCG TTCATGACCG TGCTCAGCAA GTGGGTTGCT TTGATTGATA  350
351   TGGTGTTTTC GTACCTATGG TTGACGGCAT TCGTCCTCGC CGCACAGAGC  400
401   ACAACTACGG TGATGTCTAC TTGAAGGCCC CCTCCGGCC
```

9.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLAST N	gi|3236453|gb|AF070480.1|	*snaD* mRNA for spindle pole body associated protein	*Emericella nidulans* (fungus)	52	9e-04	50/58 (86%)
	gi|22657551|gb|AC026898.8|	clone RP11-767C4 chromosome 18	*Homo sapiens* (human)	40	3.3	23/24 (95%)
BLAST X	gi|1352145|sp|P98005|CO13_THETH	Cytochrome c oxidase polypeptide I+III	*Thermus thermophilus* (bacterium)	42	5e-04	22/92 (23%)
	gi|1352177|sp|P98000|COXN_BRAJA	Alternative cytochrome c oxidase polypeptide I	*Bradyrhizobium japonicum* (bacterium)	41	0.001	26/95 (27%)
	gi|464512|sp|P34956|QOX1_BACSU	Quinol oxidase polypeptide I	*Bacillus subtilis* (bacterium)	37	0.019	29/120 (24%)

B. Multiple sequence alignment

Polypeptide level;

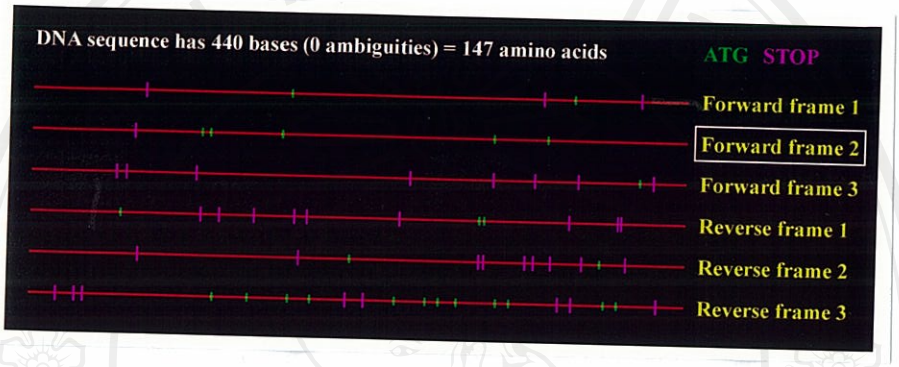
```
Pm --------MAMLSSSLSRPILLGIRTMQWASSVIALGIYAYFVHHQ-RSGTNPIFNLVIS 51
Tt MLLPYLGILAEVASTFARKPLFGYRQMVWAQMGIVVLGTMVWAHHMFTVGESTLFQIAFA 315
Bj VALPAFGIVSDLISTHARKNIFGYRMMVWAIVAIGALSFVVWAHHMYVSGMYPYFGFFFA 345
Bs VILPAFGIFSEIISSFARKQLFGYKAMVGSIIAISVLSFLVWTHHFFTMGNSASVNSFFS 345
            .: : *: :*  ::* : *  :   *       :.**    *  . .   ::

Pm VLSVVFFIPAFVSPFMTVLSKWVALIDMVFSYLWLTAFVL--------------AAQSYN 97
Tt FFTALIAVPTGVKLFNIIGTLWGGKLQMKTPLYWVLGFIFNFLLGGITGVMLSMTPLDYQ 375
Bj TTTLIIAIPTAIKVYNWVLTLWHGDIHLTVPMLFALGFIITFVNGGLTGLFLGNVVVDVP 405
Bs ITTMAISIPTGVKIFNWLFTMYKGRISFTTPMLWALAFIPNFVIGGVTGVMLAMAAADYQ 405
     :  : :*: :. : : : : . : :  .  :  .*:           .  .

Pm YGDVYLKAPSG---------------------------------------------------- 108
Tt FHDSYFVVAHFHNVLMAGSGFGAFAGLYYWWPKMTGRMYDERLGRLHFWLFLVGYLLTFL 435
Bj LSDTMFVVAHFHMVMGVAPIMVVLGAIYHWYPKVTGRMLNDVLGKFHFWVTFLGAYLIFF 465
Bs YHNTYFLVSHFHYVLIAGTVFACFAGFIFWYPKMFGHKLNERIGKWFFWIFMIGFNICFF 465
     :  : ..
```

Note Pm = *Penicillium marneffei*, Tt = *Thermus thermophilus*,
Bj = *Bradyrhizobium japonicum*, Bs = *Bacillus subtilis*

C. Motif scan search: No match found
D. Conclusion of functional analysis:
Possible function of encoded protein: Cytochrome C oxidase

9.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 2 match with Cytochrome C oxidase polypeptide

C. Deduced amino aid composition

```
Forward Frame 2:

   2  AAAAACGCCAATCACACTCCGCCACAGCAAGAGTCAATCATCAGCCGCATACATACAATA

                                                              M  A
  62  ACTCTAGCCTAGCTTAATCAGTACTCAATCCCACGTTCAACTTCAACCGACATAATGGCC

   3   M  L  S  S  S  L  S  R  P  I  L  L  G  I  R  T  M  Q  W  A
 122  ATGCTCTCGTCATCATTGTCCCGGCCCATCCTCCTGGGCATCCGAACCATGCAATGGGCC

  23   S  S  V  I  A  L  G  I  Y  A  Y  F  V  H  H  Q  R  S  G  T
 182  AGCTCCGTCATCGCCTTGGGTATATACGCCTACTTTGTGCATCACCAGCGCAGTGGCACA

  43   N  P  I  F  N  L  V  I  S  V  L  S  V  V  F  F  I  P  A  F
 242  AACCCCATCTTCAATCTAGTCATTTCTGTCCTGTCCGTTGTCTTCTTCATCCCTGCTTTC

  63   V  S  P  F  M  T  V  L  S  K  W  V  A  L  I  D  M  V  F  S
 302  GTGTCGCCGTTCATGACCGTGCTCAGCAAGTGGGTTGCTTTGATTGATATGGTGTTTTCG

  83   Y  L  W  L  T  A  F  V  L  A  A  Q  S  Y  N  Y  G  D  V  Y
 362  TACCTATGGTTGACGGCATTCGTCCTCGCCGCACAGAGCTACAACTACGGTGATGTCTAC

 103   L  K  A  P  S  G
 422  TTGAAGGCCCCCTCCGGCC
```


Figure 27. Sequence analysis of clones in group 9 (P12, P25, P27)

Group 10

10.1) Member: P13

10.2) Five-prime sequence of P13:

```
1       GACCTCACAG TCAGCAGACT CAAAATGTCT CGAGAAGTGG CGAAAGCGGC     50
51      CAAGTCCGCC TCGAACGCCA TCGCTGTTTC CAAGAAATAC ACCGTTCAGT    100
101     CTACAGGCAT CTGGGAAGTG ATCCGACGCA CACTGGCCGT CGACCCAACC    150
151     CGATCCACCG GTGTCCCATT GAACTCTCAA TTCCGCAATC CAGCACCAGG    200
201     CGCCCTACCA CCTCAATCAT ACGATGAACC CGTCACTCTC CCCGCCGCCG    250
251     ACCTCGCAGA TAACCCTTAC TGGAAACGCG ATGTCCGCCG CAAATACCCT    300
301     CAACTCAGCG TTTTCAGCCA AGGCGACGTT GCCGGACTTT TGACCTTTGG    350
351     AAATAAGCAG GCACCCAAGG AAGACGCTCT TCAGCTTGGA GAGGCTGGTG    400
401     AGAAGCAATT GATTGCTGCT AAGCAGGAAG GAGATGAAAA A
```

10.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLAST N	gi|28193406|emb|AL6701 14.17|	clone RP23-7C19 on chromosome X	*Mus musculus* (mouse)	42	0.74	21/21 (100%)
	gi|24580615|ref|NM_057 503.3|	mRNA for phospholipase C	*Drosophila melanogaster* (fly)	40	2.9	23/24 (95%)
BLAST X	gi|128868|sp|P19968|NU ZM_NEUCR	NADH-ubiquinone oxidoreductase 21.3 kDa subunit	*Neurospora crassa* (fungus)	140	1e-33	65/115 (56%)

B. Multiple sequence alignment

Polypeptide level;

```
Pm DLTVSRLKMSREVAKAAKSASN-AIAVSKKYTVQSTGIWEVIRRTLAVDPTRSTGVPLNS 59
Nc -------MASKVVTGVVKTTAGGVVPVSQKYTVQSVGVWERIRRAFAIDPNRSNGVPLVP 53
           : *: *: ..*:::. .:.**:******.*:** ***::*:**.**.**** .

Pm QFRNPAPGALPPQSYDEPVTLPAADLADNPYWKRDVRRKYPQLSVFSQGDVAGLLTFGNK 119
Nc YNRNPSPGSLDPLAYDDPVTIPAGDIADNPYWKRDARRNYPRLSVVGQAEAVALLSVGSA 113
     ***:**:* * :**:***:**.*:*********.**:**:***..*.:...**:.*.

Pm QAPKEDALQLGEAGEKQLIAAKQEGD------------------------------EK 147
Nc THPR--VELVGENGSKQLVAAQEAGKTGGLAKYFEGTGVEAGKLVLAETGGLPPLPSGEK 171
       *:  ...:** *.***:**:: *.                             **

Pm ------------------------------
Nc LGEGGKWDVYKYQLAEEPSYSEAYPCRSFS 201
```

Note Pm = *Penicillium marneffei*, Nc = *Neurospora crassa*

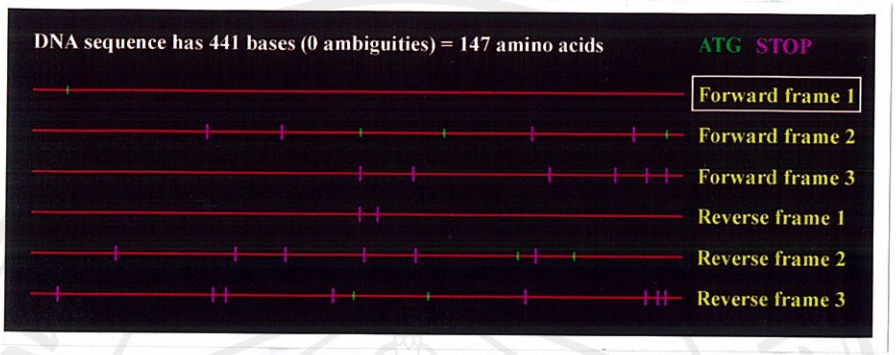
C. Motif scan search: No match found

D. Conclusion of functional analysis:

Possible function of encoded protein: NADH-ubiquinone oxidoreductase

10.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 1 matched with NADH-ubiquinone oxidoreductase of *Neurospora crassa*

C. Deduced amino acid composition

```
Forward Frame 1:

  1   D  L  T  V  S  R  L  K  M  S  R  E  V  A  K  A  A  K  S  A
  1   GACCTCACAGTCAGCAGACTCAAAATGTCTCGAGAAGTGGCGAAAGCGGCCAAGTCCGCC

 21   S  N  A  I  A  V  S  K  K  Y  T  V  Q  S  T  G  I  W  E  V
 61   TCGAACGCCATCGCTGTTTCCAAGAAATACACCGTTCAGTCTACAGGCATCTGGGAAGTG

 41   I  R  R  T  L  A  V  D  P  T  R  S  T  G  V  P  L  N  S  Q
121   ATCCGACGCACACTGGCCGTCGACCCAACCCGATCCACCGGTGTCCCATTGAACTCTCAA

 61   F  R  N  P  A  P  G  A  L  P  P  Q  S  Y  D  E  P  V  T  L
181   TTCCGCAATCCAGCACCAGGCGCCCTACCACCTCAATCATACGATGAACCCGTCACTCTC

 81   P  A  A  D  L  A  D  N  P  Y  W  K  R  D  V  R  R  K  Y  P
241   CCCGCCGCCGACCTCGCAGATAACCCTTACTGGAAACGCGATGTCCGCCGCAAATACCCT

101   Q  L  S  V  F  S  Q  G  D  V  A  G  L  L  T  F  G  N  K  Q
301   CAACTCAGCGTTTTCAGCCAAGGCGACGTTGCCGGACTTTTGACCTTTGGAAATAAGCAG

121   A  P  K  E  D  A  L  Q  L  G  E  A  G  E  K  Q  L  I  A  A
361   GCACCCAAGGAAGACGCTCTTCAGCTTGGAGAGGCTGGTGAGAAGCAATTGATTGCTGCT

141   K  Q  E  G  D  E  K
421   AAGCAGGAAGGAGATGAAAAA
```

Figure 28. Sequence analysis of clone in group 10 (P13)

Group 11

11.1) Member: P14

11.2) Five-Prime sequence of P14:

```
1       ACGCCTCACG GTGTACGTGA TTCAATGGTT GGCACACCGA CCACTGGTCA     50
51      GAGGACCGGT CGTCGCGGTC GGCCGCCTGG CGCCAAAAAC AAAAACCCTA    100
101     CGAAAGCTAC CTTGAAAGCT CTTGTCAAGA CTACCGCTGG GAGCTCTACG    150
151     AGAGACGCGC AACCCGCCGT ACCTGCACCT GCACCTGCTA CCTCCGTCCC    200
201     TGAGCTTTCC TATCCTATGT TCACGTGTGA ATGGGCGTCT TGTCCTGCGC    250
251     AACTCCACGA CGTGCATACG TTGGAGCGCC ATGTTGTCAA AAACCACATA    300
301     TCTGGTCAAA CAACATGTCT ATGGCAGAAC TGTCCCAATC TTGCAACAGA    350
351     GTATAGCGGC GAAGGTTTGA AGGAGCATTT AGCACAAGCG CATATCCAAC    400
401     CTCTGGCATG GAAGTACGGA GATGGGGCCT CCGTTAATGG AAATGGTGAG    450
451     AAAGAGATCA CCTCACAAGG TTGCTGCTAT TGTGTCATGA ATAAATTATT    500
501     GAGGAAGAAC TC
```

11.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLAST N	gi|28630144|gb|AC122268.5|	BAC clone from chromosome 14	*Mus musculus* (mouse)	42	0.98	21/21 (100%)
	gi|27923689|gb|AC104866.10|	BAC clone from chromosome 17	*Mus musculus* (mouse)	42	0.98	21/21 (100%)
	gi|35763020|emb|AL929245.13|	BAC clone from chromosome 2	*Mus musculus* (mouse)	42	0.98	24/25 (96%)
BLAST X	gi|21627826|emb|CAD37158.1|	hypothetical protein	*Aspergillus fumigatus* (fungus)	69	3e-11	28/84 (33%)
	gi|38109490|gb|EAA55354.1|	hypothetical protein	*Magnaporthe grisea* (plant)	54	1e-06	30/80 (37%)
	gi|25144535|ref|NP_492104.2|	Polybromodomain protein	*Caenorhabditis elegans* (worm)	38	0.09	27/98 (27%)

B. Motif scan search

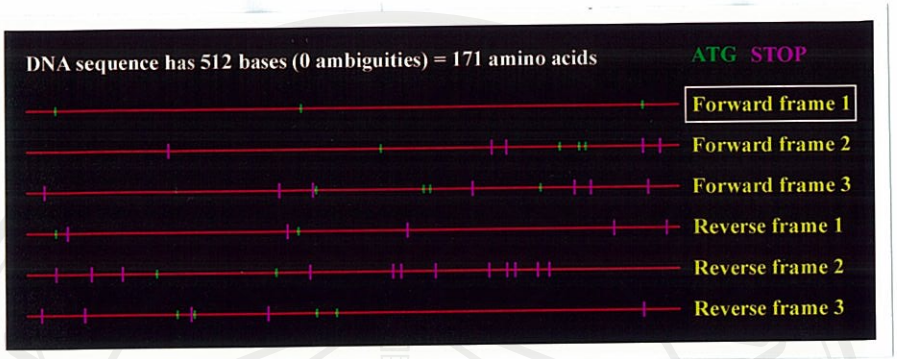
Hits	Status	Amino acid position	Database
Zinc finger C2H2 type domain signature	Weak match	76-99	PROSITE
AT hook motif	Weak match	19-31	Pfam
Zince finger C2H2 type	Weak match	74-99	Pfam

C. Conclusion of functional analysis:

Possible function of encoded protein= Unknown

11.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 1 match with hypothetical protein

C. Deduced Amino Acid Composition

```
Forward Frame 1:

  1   T  P  H  G  V  R  D  S  M  V  G  T  P  T  T  G  Q  R  T  G
  1   ACGCCTCACGGTGTACGTGATTCAATGGTTGGCACACCGACCACTGGTCAGAGGACCGGT

 21   R  R  G  R  P  P  G  A  K  N  K  N  P  T  K  A  T  L  K  A
 61   CGTCGCGGTCGGCCGCCTGGCGCCAAAAACAAAAACCCTACGAAAGCTACCTTGAAAGCT

 41   L  V  K  T  T  A  G  S  S  T  R  D  A  Q  P  A  V  P  A  P
121   CTTGTCAAGACTACCGCTGGGAGCTCTACGAGAGACGCGCAACCCGCCGTACCTGCACCT

 61   A  P  A  T  S  V  P  E  L  S  Y  P  M  F  T  C  E  W  A  S
181   GCACCTGCTACCTCCGTCCCTGAGCTTTCCTATCCTATGTTCACGTGTGAATGGGCGTCT

 81   C  P  A  Q  L  H  D  V  H  T  L  E  R  H  V  V  K  N  H  I
241   TGTCCTGCGCAACTCCACGACGTGCATACGTTGGAGCGCCATGTTGTCAAAAACCACATA

101   S  G  Q  T  T  C  L  W  Q  N  C  P  N  L  A  T  E  Y  S  G
301   TCTGGTCAAACAACATGTCTATGGCAGAACTGTCCCAATCTTGCAACAGAGTATAGCGGC

121   E  G  L  K  E  H  L  A  Q  A  H  I  Q  P  L  A  W  K  Y  G
361   GAAGGTTTGAAGGAGCATTTAGCACAAGCGCATATCCAACCTCTGGCATGGAAGTACGGA

141   D  G  A  S  V  N  G  N  G  E  K  E  I  T  S  Q  G  C  C  Y
421   GATGGGGCCTCCGTTAATGGAAATGGTGAGAAAGAGATCACCTCACAAGGTTGCTGCTAT

161   C  V  M  N  K  L  L  R  K  N  S
481   TGTGTCATGAATAAATTATTGAGGAAGAACTC
```

Figure 29. Sequence analysis of clone in group 11 (P14)

Group 12

12.1) Member: P15, P18

12.2) Five-prime sequence of P15:

```
1       TGAAGTTCTT ACCCTCTCTC GTCGTCCTCG GTCTTTCTAC CCAGGCTCTT     50
51      GCGAGCTCTT ACGTCGATTA TGTTACTAAG GACCAGCATG GTCTTACTGT    100
101     TTATGAGATG GTCATAAATA TCATTAACAC CACTACCTCA GACTTCAATA    150
151     CCCGGATTCA GAGTTACCAG GGTGGTGATC TCAGCAGCAT TCTCGAAGGC    200
201     TGTAATCAAG TTACCCAAAT AATGAAGCTT GGGGCTACGA TACTCGATCA    250
251     GCAAACAACG AAACCCCTTA CCAATAATGA GTGGCTTAGC CTCCTCTCAC    300
301     ATATGGAGAA AAAGGGTGGT TTAGAAGATA TGTTGAAAAT GGCTATAAAT    350
351     ACTCTTATCC TGAAGAAGTC ACTTATTCTT GATTCTGGTT TGGGTTCTAA    400
401     GCTTGGGTTA GCGCTTTACA GTCAGCAGAT GGCTTCTATA GACCTCGGTG    450
451     CTAAGTTCTT TGAGAAGACC CCCCCAGGAA AGGTCGAAGA TTTCAGAGAA    500
501     CACTGGTTTA AGAACATCAT GTGGGTCATC GGGCGAGGTG TTGATACCTT    550
551     TGATAATCTA CCCATCACGC GACAATA
```

12.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organisms			
BLAST N	gi|2921741|gb|AF009957.1|AF009957	cell wall antigen (MP1) mRNA	*Penicillium marneffei*	88	2e-14	59/64 (92%)
	gi|21539135|gb|AC023683.4|	BAC clone RP98-29G2	*Drosophila melanogaster (fly)*	42	1.2	21/21 (100%)
	gi|21397195|gb|AC023723.4|	BAC clone RP98-48E6	*Drosophila melanogaster (fly)*	42	1.2	21/21 (100%)
BLAST X	gi|2494526|sp|P75936|FLGD_ECOLI	Basal-body rod modification protein flgD	*Escherichia coli (bacterium)*	33	0.57	20/73 (27%)
	gi|6686041|sp|Q9YEW8|THII_AERPE	Probable thiamine biosynthesis protein thiI	*Aeropyrum pernix* (archaebacterium)	30	4..9	23/82 (28%)
	gi|232167|sp|P30537|GLGB_BACCL	1,4-alpha-glucan branching enzyme (Glycogen branching enzyme)	*Bacillus caldolyticus* (bacterium)	30	4.9	18/64 (28%)

B. Multiple sequence alignment

Nucleotide level;

```
MP1     GGCACGAGCGTTAATCAACATGAAGTTCTTATCCTCCCTCGTCGTCCTCGGTCTTTCTGC  60
P15     --------------------TGAAGTTCTTACCCTCTCTCGTCGTCCTCGGTCTTTCTAC  40

MP1     CCAGGCTCTTGCAAGCCCTTACGTTGATCACCAGGCTACCAAGGACCAGCGTCATGTAAA  120
P15     CCAGGCTCTTGCGAGCTCTTACGTCGATTAT---GTTACTAAGGACCAGCATGGTCTTAC  97

MP1     TGTTTTCAAGCAGGTCCTCCAAGATATTAACCTCGATGTGCAGAAATTCGACCAGGATAT  180
P15     TGTTTATGAGATGGTCATAAATATCATTAACACCACTACCTCAGACTTCAATACCCGGAT  157

MP1     CACTCAATACCAGGGCGGTGATCCCACAGTCCTTCTCGCTCACTCTGATGCTATTATCAA  240
P15     TCAGAGTTACCAGGGTGGTGATCTCAGCAGCATTCTCGAAGGCTGTAATCAAGTTACCCA  217
```

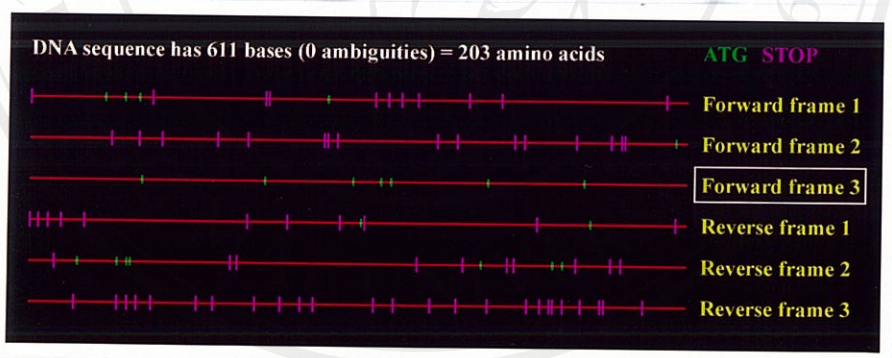
```
MP1   AACCACTGAGGAAGGCATTCAGACAATCGGACCTCAGCCT------CCCCTTAGTGTCAC 294
P15   AATAATGAAGCTTGGGGCTACGATACTCGATCAGCAAACAACGAAACCCCTTACCAATAA 277

MP1   TGAGGCCCTTGCCCTTGTTGGCCCTGTTCAG---------GGTGTAAACAAGTTGATTAT 345
P15   TGAGTGGCTTAGCCTCCTCTCACATATGGAGAAAAAGGGTGGTTTACAAGATATGTTGAA 337

MP1   GAAGGCTGTCCATCACCTTATTGAAAGAAGGGTCCTCTTGTTGGTGGGGGTTATGGTCC 405
P15   AATGGCTATAAATACTCTTATCCTGAAGAAGTCACTTATTCTTGATTCTGGTTTGGGTTC 397

MP1   TCAAGTCAAGGATAGT-CTTGAGAGGCAGGCCCATGCTGCGAGTAAACTCAGCGAGTTAG 464
P15   T-AAGCTTGGGTTAGCGCTTTACAGTCAGCAGATGGCTTCTATACACCTCGGTGCTAAGT 456

MP1   TCTCCTCAAAGGTCCCTAGTCCACTCGCCCCAATTTCCAAACAGCTCT---CCGATCAGG 521
P15   TCTTTGAGAAGACCCCCCCAGGAAAGGTCGAAGATTTCAGAGAACACTGGTTTAAGAACA 516

MP1   TCGCCCAAGCCCTCCAGAAAGGTATCCAAGCCTTCTCCATTAGCGCTCGCCAGGCCACCA 581
P15   TCATGTGGGTCATCGGGCGAGGTGTTGATACCTTTGATAATC-TACCCATCACGCGAC-A 574

MP1   AGGTAAACGCGTGAGGCCACCAAGGTCCAGCGTGATATTTCTGCTTTCAAGAAGGTCATCC 641
P15   ATACCTACCTTACCCCCTAGAGGTGCTATGGCGACTA---------------------- 611
```

C. Motif scan search: No match found

D. Conclusion of functional analysis:
Possible function of encoded protein: Mp1p-like protein

12.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: No significant homology

C. Deduced amino acid composition

```
Forward frame 3:

  1   K  F  L  P  S  L  V  V  L  G  L  S  T  Q  A  L  A  S  S  Y
  3   AAGTTCTTACCCTCTCTCGTCGTCCTCGGTCTTTCTACCCAGGCTCTTGCGAGCTCTTAC

 21   V  D  Y  V  T  K  D  Q  H  G  L  T  V  Y  E  M  V  I  N  I
 63   GTCGATTATGTTACTAAGGACCAGCATGGTCTTACTGTTTATGAGATGGTCATAAATATC

 41   I  N  T  T  T  S  D  F  N  T  R  I  Q  S  Y  Q  G  G  D  L
123   ATTAACACCACTACCTCAGACTTCAATACCCGGATTCAGAGTTACCAGGGTGGTGATCTC

 61   S  S  I  L  E  G  C  N  Q  V  T  Q  I  M  K  L  G  A  T  I
183   AGCAGCATTCTCGAAGGCTGTAATCAAGTTACCCAAATAATGAAGCTTGGGGCTACGATA

 81   L  D  Q  Q  T  T  K  P  L  T  N  N  E  W  L  S  L  L  S  H
243   CTCGATCAGCAAACAACGAAACCCCTTACCAATAATGAGTGGCTTAGCCTCCTCTCACAT
```

```
101   M  E  K  K  G  G  L  E  D  M  L  K  M  A  I  N  T  L  I  L
303   ATGGAGAAAAAGGGTGGTTTAGAAGATATGTTGAAAATGGCTATAAATACTCTTATCCTG

121   K  K  S  L  I  L  D  S  G  L  G  S  K  L  G  L  A  L  Y  S
363   AAGAAGTCACTTATTCTTGATTCTGGTTTGGGTTCTAAGCTTGGGTTAGCGCTTTACAGT

141   Q  Q  M  A  S  I  D  L  G  A  K  F  F  E  K  T  P  P  G  K
423   CAGCAGATGGCTTCTATAGACCTCGGTGCTAAGTTCTTTGAGAAGACCCCCCCAGGAAAG

161   V  E  D  F  R  E  H  W  F  K  N  I  M  W  V  I  G  R  G  V
483   GTCGAAGATTTCAGAGAACACTGGTTTAAGAACATCATGTGGGTCATCGGGCGAGGTGTT

181   D  T  F  D  N  L  P  I  T  R  Q  Y  L  P  Y  P  L  E  V  L
543   GATACCTTTGATAATCTACCCATCACGCGACAATACCTACCTTACCCCCTAGAGGTGCTA

201   W  R  L
603   TGGCGACTA
```

Figure 30. Sequence analysis of clones in group 12 (P15, P18)

Group 13

13.1) Member: P17

13.2) Five-prime sequence of P17:

```
1       ATCCAACTCA ATCCGACAGG ATTTCGTTCT TACGCTATTC GCATCATCCT     50
51      AGGCTGTTGT TCGTGACGAC CCCTCATTTG TATACACAAC CGCCTCAGAT    100
101     TAAATCACAA CAGCACAATG GCCTCCGCGA CCACGTTCTA CGACTTTTCT    150
151     CCTCCTGACA AAAAAGGAAA CCCTTACCCC TTGACAGACT ACAAGGGCAA    200
201     AGTCGTCCTC GTCGTCAACA CAGCATCCAA ATGCGGCTTC ACGCCCCAAT    250
251     TCGCCGGCCT TGAAAAACTC TACAAATCTA TCGAAGCCAA GCATCCCGGC    300
301     GCCTTCACCA TCCTGGGTTT TCCCTGCAAT CA
```

13.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLAST N	gi|5826556|emb|AL111937.1|CNS019PL	cDNA library under conditions of nitrogen deprivation	*Botrytis cinerea* (fungus)	52	6e-04	56/66 (84%)
	gi|5832390|emb|AL117174.1|CNS01DR2	cDNA library under conditions of nitrogen deprivation	*Botrytis cinerea* (fungus)	46	0.035	35/39 (89%)
BLAST X	gi|6225487|sp|O59858|GSHJ_SCHPO	Glutathione peroxidase	*Shizosaccharomyces pombe* (fission yeast)	91	3e-19	42/65 (64%)
	gi|729640|sp|P40581|GSHJ_YEAST	Glutathione peroxidase	*Saccharomyces cerevisiae* (yeast)	84	5e-17	39/65 (60%)
	gi|6225467|sp|O32770|GPO_LACLC	Glutathione peroxidase	*Lactococcus lactis subsp. cremoris* (bacterium)	83	1e-16	40/65 (61%)
	gi|585222|sp|P38143|GSHI_YEAST	Glutathione peroxidase GPX2	*Saccharomyces cerevisiae* (yeast)	81	4e-16	39/67 (58%)
	gi|22653728|sp|Q9CFV1|GPO_LACLA	Glutathione peroxidase	*Lactococcus lactis subsp. lactis* (bacterium)	81	6e-16	39/65 (60%)

B. Multiple sequence alignment

Polypeptide level;

```
Pm MASATTFYDFSPPDKKGNPYPLTDYKGKVVLVVNTASKCGFTPQFAGLEKLYKSIEAKHP 60
Ll ----MNFYDFSAVKMNGETVSMSDYKGKVVIVVNTASKCGFTPQFEGLEKLYETYKDQ-- 54
Sp ---MSHFYDLAPKDKDGNPFPFSNLKGKVVLVVNTASKCGFTPQYKGLEALYQKYKDR-- 55
Sc ---MSEFYKLAPVDKKGQPFPFDQLKGKVVLIVNVASKCGFTPQYKELEALYKRYKDE-- 55
          **.::. . .*:. .: : *****::**.*********:  ** **: : .

Pm GAFTILGFPCN---------------------------------------------------- 71
Ll -GLEILGFPCNQFANQDAGENTEINEFCQLNYGVTFTMFQKIKVNGKEAHPLYQFLKKEA 113
Sp -GFIILGFPCNQFGNQEPGSDEEIAQFCQKNYGVTFPVLAKINVNGDNVDPVYQFLKSQK 114
Sc -GFTIIGFPCNQFGHQEPGSDEEIAQFCQLNYGVTFPIMKKIDVNGGNEDPVYKFLKSQK 114
     .: *:*****

Pm ----------------------------------------------
Ll KGALSGT-IKWNFTKFLIDRDGQVIERFAPKTEPEEMEEEIKKLL---- 157
Sp K-QLGLERIKWNFEKFLVNRQGQVIERYSSISKPEHLENDIESVL---- 158
Sc SGMLGLRGIKWNFEKFLVDKKGKVYERYSSLTKPSSLSETIEELLKEVE 163
```

Note Pm = *Penicillium marneffei*, Ll = *Lactococcus lactis*,
Sp = *Shizosaccharomyces pombe*, Sc = *Saccharomyces cerecisiae*

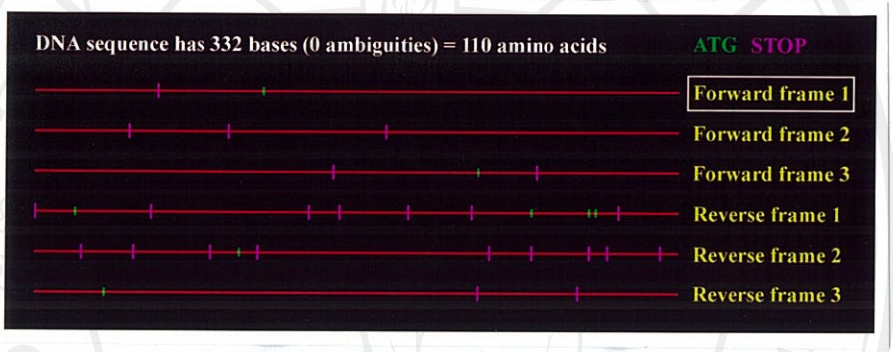
C. Motif scan search

Hits	Status	Amino acid position	Database
Glutathione peroxidases active site	Significant match	27-42	PROSITE
Glutathione peroxidase	Significant match	6-71	Pfam

D. Possible function of encoded protein: Glutathione peroxidase

13.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 1 matched with Glutathione peroxidase

C. Deduced amino acid composition

```
Forward Frame 1:

  1  ATCCAACTCAATCCGACAGGATTTCGTTCTTACGCTATTCGCATCATCCTAGGCTGTTGT
                                                               M
 61  TCGTGACGACCCCTCATTTGTATACACAACCGCCTCAGATTAAATCACAACAGCACAATG

  2   A  S  A  T  T  F  Y  D  F  S  P  P  D  K  K  G  N  P  Y  P
121  GCCTCCGCGACCACGTTCTACGACTTTTCTCCTCCTGACAAAAAAGGAAACCCTTACCCC

 22   L  T  D  Y  K  G  K  V  V  L  V  V  N  T  A  S  K  C  G  F
181  TTGACAGACTACAAGGGCAAAGTCGTCCTCGTCGTCAACACAGCATCCAAATGCGGCTTC

 42   T  P  Q  F  A  G  L  E  K  L  Y  K  S  I  E  A  K  H  P  G
241  ACGCCCCAATTCGCCGGCCTTGAAAAACTCTACAAATCTATCGAAGCCAAGCATCCCGGC

 62   A  F  T  I  L  G  F  P  C  N
301  GCCTTCACCATCCTGGGTTTTCCCTGCAATCA
```

Figure 31. Sequence analysis of clone in group 13 (P17)

Group 14

14.1) Member: P21

14.2) Five-prime sequence of P21:

```
1       TCTCCCTCAC CACCACCACA CAGTATCTTT GTTTTATCAT CTTCATCCAT     50
51      CGGTTTTGAC CCAGAATAAT TCTGTCATTT GATTTCCCAT CAAAATTCCC    100
101     GTCGTCATTC ATCGAAATCC ATAATTTTTC TGGCCCCCAG CATTCTCCCG    150
151     TCAGCCTAAG CTTGGTTGAG CATTAGCAGT GCCAAATCAG AACCACTAAC    200
201     TATCATCGAT TTTTCATCAC ACTGAAAAAA AACTACATTG ATCACAACGC    250
251     CTGATTCAAA ATGAGTGATC TGGTGCCTTT ATCAAAATCG TCTGCGATTT    300
301     TGTCGACCTC CATGGCGACT CCCTCCCCCT CAGCACCGGC CGTCGGTCCA    350
351     GACACATTGA TCACTATCAA GATTCTTCAC AATGACTCGG TTAACCGTCG    400
401     CTTC
```

14.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLASTN	gi|20198260|gb|AC007584.4|	chromosome 2 clone MJB20	*Arabidopsis thaliana* (plant)	42	0.67	21/21 (100%)
	gi|31335548|emb|AL928793.12|	clone RP23-89M14 on chromosome 2	*Mus musculus* (mouse)	42	0.67	24/25 (96%)
	gi|1729582|emb|Z83228.1|CEF52F12	cosmid F52F12	*Caenorhabditis elegans* (worm)	42	0.67	21/21 (100%)
BLASTX	gi|1352832|sp|P34012|VC04_VARV	PROTEIN C4	Variola virus	27	8.6	12/36 (33%)

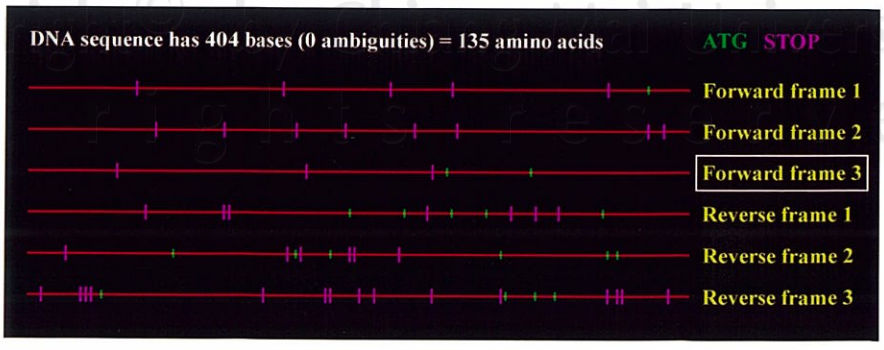
C. Motif scan search: No match found

D. Conclusion of functional analysis:

Possible function of encoded protein: Unknown

14.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: No significant homology
C. Deduced amino acid composition

```
Forward Frame 3:

  3  TCCCTCACCACCACCACACAGTATCTTTGTTTTATCATCTTCATCCATCGGTTTTGACCC

 63  AGAATAATTCTGTCATTTGATTTCCCATCAAAATTCCCGTCGTCATTCATCGAAATCCAT

123  AATTTTTCTGGCCCCCAGCATTCTCCCGTCAGCCTAAGCTTGGTTGAGCATTAGCAGTGC

183  CAAATCAGAACCACTAACTATCATCGATTTTTCATCACACTGAAAAAAAACTACATTGAT

                         M  S  D  L  V  P  L  S  K  S  S  A  I  L
243  CACAACGCCTGATTCAAAATGAGTGATCTGGTGCCTTTATCAAAATCGTCTGCGATTTTG

 15   S  T  S  M  A  T  P  S  P  S  A  P  A  V  G  P  D  T  L  I
303  TCGACCTCCATGGCGACTCCCTCCCCCTCAGCACCGGCCGTCGGTCCAGACACATTGATC

 36   T  I  K  I  L  H  N  D  S  V  N  R  R  F
363  ACTATCAAGATTCTTCACAATGACTCGGTTAACCGTCGCTTC
```

Figure 32. Sequence analysis of clone in group 14 (P21)

Group 15

15.1) Member: P22

15.2) DNA sequence of P22: ACTAGTTCTAGATCGCGA

15.3) Possible function: Artifact

Figure 33. Sequence analysis of clone in group 15 (P22)

Group 16

16.1) Member: P24

16.2) Five-prime sequence of P24:

```
1       ACACGCTAAT ATTTACAAGGC ACGCAACCCC GTACCACGC ACCTCTTTAT      50
51      CTTGCTCAAA GCAAGGGTTAC TTCAAGGATG AGGGACTTA AAGTTGCCTT     100
101     GCTAGAACCT AATGACCCTTC GGATGTGACC GAGATCATT GGCAGCGGCA     150
151     AGGTTGACAT GGGTTTCAAAG CCATGATCCA CACGTTGGC TGCCAAAGCC     200
201     CGTAACTTCC CGGTTACCTCC TTCGGTTCCC TTCTGGACG AGCCGTTTAC     250
251     TGGAGTTGTC TACCTCAAGGA CAGCGGTATT ACCACCGAC TTCAAGACCC     300
301     TCAAGGGCAA ACGTATTGGTT ACGTTGGAGA GTTTGGCAA AATCCAAATC     350
351     GATGAGCTTA CCAAGTACTAC GGTATGACTC CCGATGACT ACACTGCTGT     400
401     ATGAACGTCA CCAAGGCCATT ATCGAGGGG
```

16.3) Functional analysis

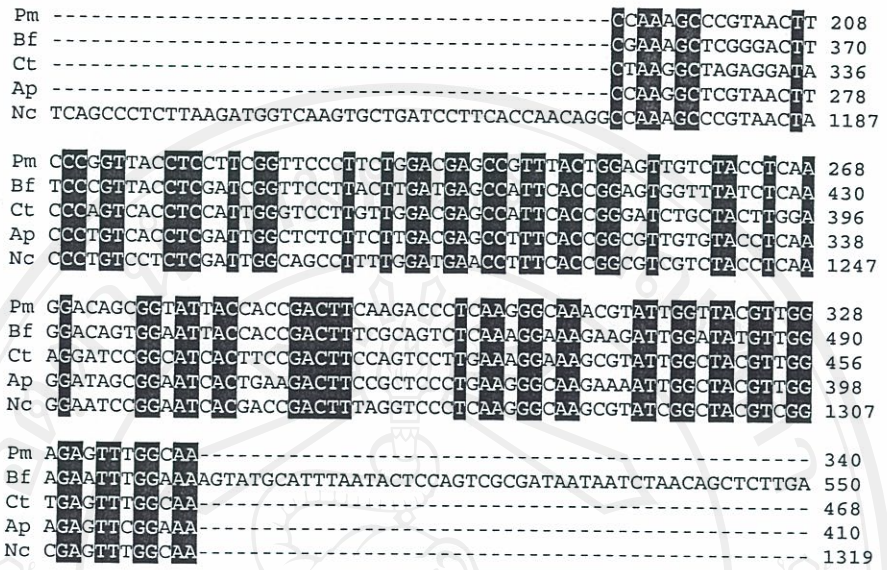
A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLAST N	gi|557049|gb|U15196.1|APU15196	*nmt1* mRNA for thymine synthase	*Aspergillus parasiticus* (fungus)	206	2e-50	335/412 (81%)
	gi|15822512|gb|AY007661.1|	thiamine biosynthesis protein NMT-1 (nmt-1) gene	*Neurospora crassa* (fungus)	109	4e-21	151/183 (82%)
	gi|3282219|gb|U68718.1|BFU68718	thiamine synthase homolog (*BcNMT1*) gene	*Botryotinia fuckeliana* (fungus)	84	2e-13	138/170 (81%)
	gi|5777325|dbj|AB031293.1|	gene for pyrimidine precursor biosynthesis	*Candida tropicalis* (yeast)	84	2e-13	96/114 (84%)
BLAST X	gi|1171741|sp|P42882|NMT1_ASPPA	NMT1 (thymine synthase) protein homolog	*Aspergillus parasiticus* (fungus)	272	1e-13	133/146 (91%)
	gi|25453135|sp|O00057|NMT1_UROFA	NMT1 protein homolog	*Uromyces viciae-fabae* (fungus)	240	8e-64	111/146 (76%)
	gi|1174672|sp|P43534|THI5_YEAST	Pyrimidine precursor biosynthesis enzyme THI5	*Saccharomyces cerevisiae* (yeast)	236	2e-62	112/146 (76%)

B. Multiple sequence alignment

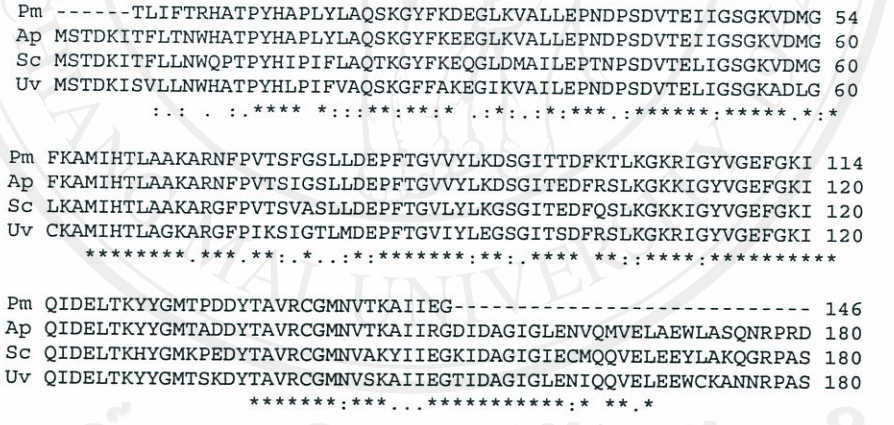
Nucleotide level;

```
Pm CCCGTACCACGCACCTCTTTATCTTGCTCAAAGCAAGGGTTACTTCAAGGATGAGGGACT 88
Bf G--GTAGCTCTTCTCGAGCCTAATGA--TCCCAGTGGTTAGCAATCAAATACTCAAATTA 251
Ct TCCATACCACATCCCAGTCTACTTGGCCTCCCAGAAGGGATACTTCAAGGAGGAAGGCAT 216
Ap G--GTTGCTCTGCTGGAGCCCAATGA--CCCC---------------------------- 189
Nc G--GTGGCTCTCCTGGAGCCCAACGA--CCCCAGCGTACGTAATTCATGCC-----GTTT 1007

Pm TAAAGTTGCCTTGCTAGAACCTAATGACCCTTCG-GATGTGACCGAGATCATTGGCAGCG 147
Bf TTTCGCTCGGGTGGT-GACTGATTGAGTGAACTA-GATGTTACCGAGATCATCGGCTCCG 309
Ct CGACGTCGCTATCTTAGAGCCATCCAACCCATCT-GATGTCACCGAGTTGATCGGGTCCG 275
Ap ------------TCC-------------------GATGTCACTGAGATAATTGGTAGCG 217
Nc CAATCTGCCAGTTCCCGACTAACAGAGCATGGCAGGACGTAACCGAAATCATCGGCAGCG 1067

Pm GCAAGGTTGACATGGTTTCAAAGCCATGATCCACACGTTGGCTG--------------- 192
Bf GAAAAGTCGATATGGTTTCAAAGCCATGATCCACACTCTTGCCG--------------- 354
Ct GCAAGGTCGACATGGTTTGAAAGCCATGGTCCACACCTTGGCTG--------------- 320
Ap GTAAGGTTGACATGGCTTCAAGGCCATGATCCATACTCTGGCTG--------------- 262
Nc GCAAAGTCGACCTCGGTTTCAAGGCCATGATTCACACTCTTGCTGTAAGATTCGCCATTC 1127
```



Note Pm = *Penicillium marneffei*, Bf = *Botryotinia fuckeliana,* Ct = Candida tropicalis, Ap = *Aspergillus parasiticus*, Nc = *Neurospora crassa*

Polypeptide level;



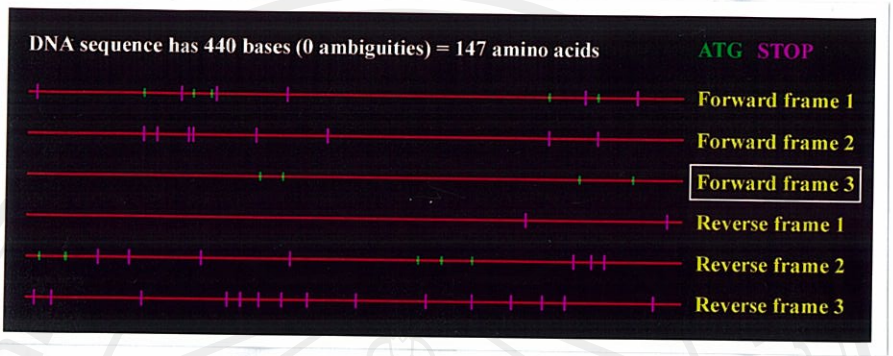
Note Pm = *Penicillium marneffei*, Ap = *Aspergillus parasiticus,* Sc = *Sc = Saccharomyces cerevisiae,* Uv = *Uromyces viciae-fabae*

C. Motif scan search: No match found

D. Conclusion of functional analysis:
Possible function of encoded protein: Thymine synthase

16.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 3 matched with thymine synthase

C. Deduced amino acid composition

```
Forward Frame 3:

  1   T  L  I  F  T  R  H  A  T  P  Y  H  A  P  L  Y  L  A  Q  S
  3  ACGCTAATATTTACAAGGCACGCAACCCCGTACCACGCACCTCTTTATCTTGCTCAAAGC

 21   K  G  Y  F  K  D  E  G  L  K  V  A  L  L  E  P  N  D  P  S
 63  AAGGGTTACTTCAAGGATGAGGGACTTAAAGTTGCCTTGCTAGAACCTAATGACCCTTCG

 41   D  V  T  E  I  I  G  S  G  K  V  D  M  G  F  K  A  M  I  H
123  GATGTGACCGAGATCATTGGCAGCGGCAAGGTTGACATGGGTTTCAAAGCCATGATCCAC

 61   T  L  A  A  K  A  R  N  F  P  V  T  S  F  G  S  L  L  D  E
183  ACGTTGGCTGCCAAAGCCCGTAACTTCCCGGTTACCTCCTTCGGTTCCCTTCTGGACGAG

 81   P  F  T  G  V  V  Y  L  K  D  S  G  I  T  T  D  F  K  T  L
243  CCGTTTACTGGAGTTGTCTACCTCAAGGACAGCGGTATTACCACCGACTTCAAGACCCTC

101   K  G  K  R  I  G  Y  V  G  E  F  G  K  I  Q  I  D  E  L  T
303  AAGGGCAAACGTATTGGTTACGTTGGAGAGTTTGGCAAAATCCAAATCGATGAGCTTACC

121   K  Y  Y  G  M  T  P  D  D  Y  T  A  V  R  C  G  M  N  V  T
363  AAGTACTACGGTATGACTCCCGATGACTACACTGCTGTGCGCTGCGGCATGAACGTCACC

141   K  A  I  I  E  G
423  AAGGCCATTATCGAGGGG
```

Figure 34. Sequence analysis of clone in group 16 (P24)

Group 17

17.1) Member: P26

17.2) Five-prime sequence of P26:

```
1       CTACTGGGCT GAACGCAGAC ACTATGGCAC CGGCATCGGC ACCGGCACTC     50
51      AACCTCATCC ACCGCCACAG ACCAGTACCC TCACCACTAC CTCCATGCCT    100
101     CCTGCTAGCA CCACCACTAC CACCACCAGC GTAAGTGAAT GGAGCACCGT    150
151     AACCGAGACG GAGACGAACA CAAAGACAGT TTTTGTCCCT TGCTCGACTT    200
201     CTGTGGGGAC ACGTGGGTCG TCGACTGTCT ACTCAACTTG GCTTACGACT    250
251     ACTACCTCCG TGTGGACTAC GACTCATACT ACCACCGTTC CTGTCTACGG    300
301     CACCACGACG ACACCTGTTG GTGGTGTTGG TCCCGTTGGA TCTGAGACGG    350
351     GGCTCGCGTG TCCCTTGCCG GTTACGACTA CTCATACGGA GACTAAGACT    400
401     ACTACTGCAA CTGTGACTGT GACTGTGGCC GTGACGGCGC CTGG
```

17.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLAST N	gi|28416271|gb|AC121946.3|	BAC clone RP24-212F15 from chromosome 10	*Mus musculus* (mouse)	48	0.014	27/28 (96%)
	gi|27452955|gb|AC122417.4|	BAC clone RP24-159C12 from chromosome 7	*Mus musculus* (mouse)	48	0.014	24/24 (100%)
BLAST X	gi|3024622|sp|Q92154|SMP_COTJA	Schwann cell myelin protein precursor (Siglec-4b)	*Coturnix japonica* (Japanese quail)	30	1.8	21/65 (32%)

B. Motif scan search

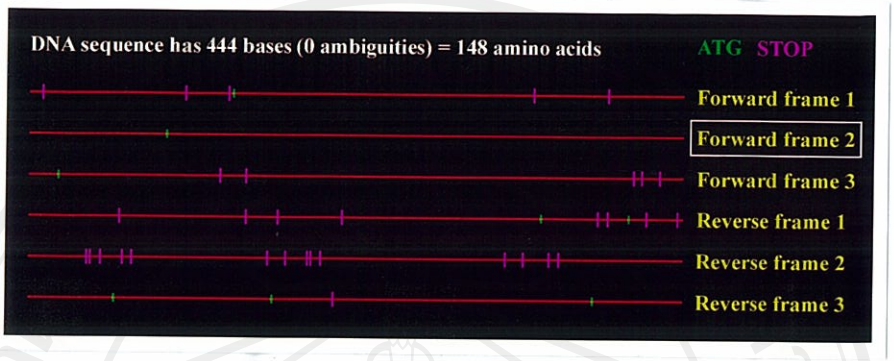
Hits	Status	Amino acid position	Database
Threonine-rich region	Significant match	10-104	PROSITE

C. Conclusion of functional analysis:

Possible function of encoded protein: Unknown

17.4) Recognition of the coding frame

A. Six-frame translation Note



B. BLASTX result: No significant homology

C. Deduced amino acid composition

```
Forward Frame 2:

  1   Y  W  A  E  R  R  H  Y  G  T  G  I  G  T  G  T  Q  P  H  P
  2   TACTGGGCTGAACGCAGACACTATGGCACCGGCATCGGCACCGGCACTCAACCTCATCCA

 21   P  P  Q  T  S  T  L  T  T  T  S  M  P  P  A  S  T  T  T  T
 62   CCGCCACAGACCAGTACCCTCACCACTACCTCCATGCCTCCTGCTAGCACCACCACTACC

 41   T  T  S  V  S  E  W  S  T  V  T  E  T  E  T  N  T  K  T  V
122   ACCACCAGCGTAAGTGAATGGAGCACCGTAACCGAGACGGAGACGAACACAAAGACAGTT

 61   F  V  P  C  S  T  S  V  G  T  R  G  S  S  T  V  Y  S  T  W
182   TTTGTCCCTTGCTCGACTTCTGTGGGGACACGTGGGTCGTCGACTGTCTACTCAACTTGG

 81   L  T  T  T  T  S  V  W  T  T  T  H  T  T  T  V  P  V  Y  G
242   CTTACGACTACTACCTCCGTGTGGACTACGACTCATACTACCACCGTTCCTGTCTACGGC

101   T  T  T  T  P  V  G  G  V  G  P  V  G  S  E  T  G  L  A  C
302   ACCACGACGACACCTGTTGGTGGTGTTGGTCCCGTTGGATCTGAGACGGGGCTCGCGTGT

121   P  L  P  V  T  T  T  H  T  E  T  K  T  T  T  A  T  V  T  V
362   CCCTTGCCGGTTACGACTACTCATACGGAGACTAAGACTACTACTGCAACTGTGACTGTG

141   T  V  A  V  T  A  P  G
422   ACTGTGGCCGTGACGGCGCCTGG
```

Figure 35. Sequence analysis of clone in group 17 (P26)

Group 18

18.1) Member: P28

18.2) Five-prime sequence of P28

```
1     CACTTTGTCA TTGTCATCCA TTGTGTGTGT ATCTCTGTGA CTGACGACTG   50
51    ACAACCAACC ACAGCAGCGC CATGTCTGCT ACCACTGCAG ATGCCAGGCC  100
101   TCGGCCTGCT GCGGACACCA AGAAGGTGCA CATCGCCGAC ACAAAAATGA  150
151   CCCTTAAGAA TTGGTATAAG CATGTCGACT GGTTGAATGT GTACTTCATC  200
201   ATCGGTATTC CACTCTATGG ATGCATCCAG TCACTCTGGG TTCCCTTGCA  250
251   GCTCAAGACA GCTGTCTGGG CTGTTCTCTA CTACTTCTAT ACGGGCCTGG  300
351   GAATCACAGC TGGATACCAC CGTCTTTGGG CTCACTGCTC CTACTCTGCC  350
351   ACTCTTCCTC TCCAGATCTT TCTCGCTGCC GCTGGCGGTG GTGCCGTCGA  400
401   GGGCTCCATC CGCTGGTGGG CTCGTGGCCA CCGTGCTCAT CACCGTTACA  450
451   CCGACACCGA CAAAGG
```

18.3) Functional analysis

A. Similarity search

Type of search	Sequence producing significant alignment			Score (bits)	E-Value	Identities (%)
	Accession No.	Description	Organism			
BLASTN	gi|21262989|gb|AF51086 1.1|	stearic acid desaturase (*sdeA*) gene	*Emericella nidulans* (fungus)	133	3e-28	130/151 (86%)
	gi|6473656|dbj|AB02785 9.1|	gene for Hypothetical protein	*Schizosaccharomyces pombe* (fission yeast)	74	2e-10	58/65 (89%)
	gi|5725456|emb|AJ00797 4.1|MSP007974	mRNA for delta-9 fatty acid desaturase	*Mortierella* sp. (fungus)	56	5e-05	61/72 (84%)
BLASTX	gi|1703084|sp|P21147|ACO1_YEAST	Acyl-CoA desaturase 1	*Saccharomyces cerevisiae* (yeast)	102	4e-22	50/114 (43%)
	gi|21431735|sp|P13516|ACO1_MOUSE	Acyl-CoA desaturase	*Mus musculus* (mouse)	62	5e-10	31/78 (39%)
	gi|13431274|sp|O62849|ACOD_SHEEP	Acyl-CoA desaturase	*Ovis aries* (sheep)	61	1e-09	30/96 (31%),
	gi|21431730|sp|O00767|ACOD_HUMAN	Acyl-CoA desaturase	*Homo sapiens* (human)	61	1e-09	27/51 (52%)
	gi|3023238|sp|O02858|ACOD_PIG	Acyl-CoA desaturase	*Sus scrofa* (pig)	61	1e-09	30/83 (36%)

B. Multiple sequence alignment

Nucleotide level;

```
Pm ----------------------------------------------CACTTTGTCAT 11
An CCCTGTGTCCTGCCGTGGTGGACCCTTCGAAGTCGACCTCGCCCAGGCCGACCTTTTTTT 540

Pm TGTCATCCATT-GTGTGTGTA-TCTCTGTGACTGACGACTCAC--AACCAACCACAGC-- 65
An TTTCTTCTGTTTAAATACTCGCTTTCTTTCATTTTCACTTATCAAAACAACCGTCACTT 600

Pm ----AGCGCCATGTCTGCTACCACTGCAGATGCCAGGCCTCGGCCTGCTGCGGACACCAA 121
An AGAGATAGCCATGTCTGCACCAACGGCGGACATCAGGGCTCG---CGCCCCGGAGGCCAA 657

Pm GAAGGTGCACATCGCCGACACAAAAATCACCCTTAAGAATTGGTATAAGCATGTCGACTG 181
An AAAGGTTCACATCGCTGACACTGCTATCAACCGCCATAACTGGTACAAGCATGTGAACTG 717

Pm GTTGAATGTGTACTTCATCATCGGTATTCCACTCTATGGATGCATCCAGTCACTCTGGGT 241
An GCTGAACGTTTTCCTGATCATCGGTATCCCGCTTTATGGGTGCATTCAGCGCTTCTGGGT 777
```

```
Pm TCCCTTGCAGCTCAAGACAGCTGTCTGGGCTGTTCTCTACTACTTCTATACGGGCCTGGG 301
An GCCACTGCAGCTGAAGACTGCCATCTGGGCCGTCATCTACTACTTTTTCACCGGTCTCGG 837

Pm AATCACAGCTGGA---------------------------------------------- 314
An TATCACAGCAGGTAAGAAGCTTCGAGCTCCAGGATGGTTGATGCCGAATGCTGACCTCCT 897
```

Note Pm = *Penicillium marneffei*, An = *Aspergillus nidulans*

Polypeptide level;

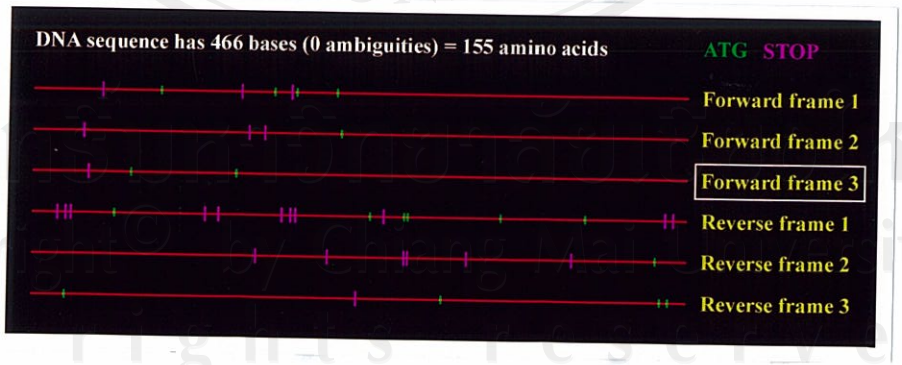
```
Pm     MSATTADAR----------------PRPAADTKKVHIADTKMTLKNWYKHVDWLNVYFII 44
Sheep  ITAPPSRVLQNGGGKLEKTPLYLEEDIRPEMRDDIYDPNYQDKEGPKPKLEYVWRNIILM 79
Pig    ITAPSSRVLQNGGGKSEKTPQYVEEDIRPEMKDDIYDPTYQDKEGPQGKLEYVWRNIILM 68
Human  ITAPPSRVLQNGGDKLETMPLYLEDDIRPDIKDDIYDPTYKDKEGPSPKVEYVWRNIILM 79
Mouse  ITAPPS---GNEREKVKTVPLHLEEDIRPEMKEDIHDPTYQDEEGPPPKLEYVWRNIILM 75
Yeast  MVSVEFDKKGNEKKSNLDRLLEKDNQEKEEAKTKIHISEQPWTLNNWHQHLNWLNMVLVC 120
         : :                                .:: .        :    .  ::

Pm     GIPLYGCIQSLW--VPLQLKTAVWAVLYYFYTGLGITAGYHRLWAHCSYSATLPLQIFLA 102
Sheep  GLLHLGALYGITLIPTCKIYTFLWVLFYYVISALGITAGVHRLWSHRTYKARLPLRVFLI 139
Pig    SLLHLGALYGIILIPTCKIYTLLWAFAYYLLSAVGVTAGAHRLWSHRTYKARLPLRVFLI 128
Human  SLLHLGALYGITLIPTCKFYTWLWGVFYYFVSALGITAGAHRLWSHRSYKARLPLRLFLI 139
Mouse  VLLHLGGLYGIILVPSCKLYTCLFGIFYYMTSALGITAGAHRLWSHRTYKARLPLRIFLI 135
Yeast  GMPMIGWYFALSGKVPLHLNVFLFSVFYYAVGGVSITAGYHRLWSHRSYSAHWPLRLFYA 180
          :   *   .:    . :: . :: . **   .:.:*** ****:* :*.*  **::*

Pm     AAGGGAVEGSIRWWARGHRAHHRYTDTDK------------------------------ 131
Sheep  IANTMAFQNDVFEWSRDHRAHHKFSETDADPHNSRRGFFFSHVGWLLVRKHPAVREKGAT 199
Pig    IANTMAFQNDVYEWARDHRAHHKFSETDADPHNSRRGFFFSHVGWLLVRKHPAVKEKGGL 188
Human  IANTMAFQNDVYEWARDHRAHHKFSETHADPHNSRRGFFFSHVGWLLVRKHPAVKEKGST 199
Mouse  IANTMAFQNDVYEWARDHRAHHKFSETHADPHNSRRGFFFSHVGWLLVRKHPAVKEKGGK 195
Yeast  IFGCASVEGSAKWWGHSHRIHHRYTDTLRDPYDARRGLWYSHMGWMLLKPNPKYKAR--- 237
                          .  :.:..   *.:.** **::::*
```

C. Motif scan search: No match found

D. Conclusion of functional analysis:
Possible function of encoded protein: Acyl-CoA desaturase

18.4) Recognition of the coding frame

A. Six-frame translation



B. BLASTX result: Forward frame 3 matched with Acyl-CoA desaturase
C. Deduced Amino Acid Composition:

```
Forward Frame 3:

  3  CTTTGTCATTGTCATCCATTGTGTGTGTATCTCTGTGACTGACGACTGACAACCAACCAC

                M  S  A  T  T  A  D  A  R  P  R  P  A  A  D  T  K
 63  AGCAGCGCCATGTCTGCTACCACTGCAGATGCCAGGCCTCGGCCTGCTGCGGACACCAAG

 41   K  V  H  I  A  D  T  K  M  T  L  K  N  W  Y  K  H  V  D  W
123  AAGGTGCACATCGCCGACACAAAAATGACCCTTAAGAATTGGTATAAGCATGTCGACTGG

 61   L  N  V  Y  F  I  I  G  I  P  L  Y  G  C  I  Q  S  L  W  V
183  TTGAATGTGTACTTCATCATCGGTATTCCACTCTATGGATGCATCCAGTCACTCTGGGTT

 81   P  L  Q  L  K  T  A  V  W  A  V  L  Y  Y  F  Y  T  G  L  G
243  CCCTTGCAGCTCAAGACAGCTGTCTGGGCTGTTCTCTACTACTTCTATACGGGCCTGGGA

101   I  T  A  G  Y  H  R  L  W  A  H  C  S  Y  S  A  T  L  P  L
303  ATCACAGCTGGATACCACCGTCTTTGGGCTCACTGCTCCTACTCTGCCACTCTTCCTCTC

121   Q  I  F  L  A  A  A  G  G  G  A  V  E  G  S  I  R  W  W  A
363  CAGATCTTTCTCGCTGCCGCTGGCGGTGGTGCCGTCGAGGGCTCCATCCGCTGGTGGGCT

141   R  G  H  R  A  H  H  R  Y  T  D  T  D  K  G
423  CGTGGCCACCGTGCTCATCACCGTTACACCGACACCGACAAAGG
```

Figure 36. Sequence analysis of clone in group 18 (P28)

Table 5. Conclusion of similarity searching of antigenic protein-encoding clones.

Group	Clone (size in kb)	Result of BLAST searching					Possible function
		Nucleotide BLAST		Translated BLAST			
		Similar to	Bit(identity)	Frame	Similar to	Bit (Positives)	
1	**1**(2.5), **2**(1.25), **4**(1.3), **8**(1.0), **16**(1.6), **19** (3.0), **20**(1.7)	*Neurospora crassa* catalase-peroxidase (*cat-2*) mRNA, Length = 2705	190 (246/296)	+3	Peroxidase/catalase of *Neurospora crassa* Length = 740	976 (477/734)	Catalase-peroxidase
2	**3**(3.2)	*Danio rerio* (Zebra fish) mRNA for heparan sulfate 6-O-sulfotransferase Length = 2658	92 (46/46)	+3	Hypothetical protein C32A11.02c in chromosome I of *Shizosaccharomyces pombe* Length = 851	235 (151/545)	Unknown
3	**5**(0.9), **23**(1.1)	*Aspergillus nidulans* mRNA for heat-shock protein 30 (HSP30), Length = 935	70 (53/59)	+3	heat shock protein 30 [*Aspergillus oryzae*] Length = 166	204 (111/190)	Heat shock protein 30

Table 5. (Continued)

Group	Clone (size in kb)	Result of BLAST searching					Possible function
		Nucleotide BLAST		Translated BLAST			
		Similar to	Bit(identity)	Frame	Similar to	Bit (Positives)	
4	6(1.2)	*Oryza sativa* chromosome 5 clone P0431G05 Length = 145012	44 (22/22)	+3	Hypothetical protein in *Shizosaccharo-myces pombe* Length = 513	91 (84/162)	Unknown
5	7(1.6)	*Mus musculus* androgen receptor mRNA Length = 2700	48 (27/28)	+2	Hypothetical protein in LEU2 3'region Length = 373	39 (27/39)	Unknown
6	9(1.8)	*Aspergillus oryzae fbpA* gene for fructose-1,6-bisphosphatase Length = 2179	100 (101/118)	+3	Fructose-1,6-bisphosphatase of *S. cerevisiae* Length = 348	87 (55/65)	Fructose-1,6-bisphosphatase

Table 5. (Continued)

Group	Clone (size in kb)	Result of BLAST searching					Possible function
		Nucleotide BLAST		Translated BLAST			
		Similar to	Bit(identity)	Frame	Similar to	Bit (Positives)	
7	**10**(1.0)	*Botrytis cinerea* strain T4 cDNA library under conditions of nitrogen deprivation Length = 600	147 (221/270)	+1	*S. cerevisiae* 60S ribosomal protein L20 (L18A) of *S. cerevisiae* Length = 174	210 (99/171)	60S ribosomal protein
8	**11**(2.0)	*Homo sapiens* chromosome , clone RP11-149C3, complete sequence Length = 184026	44 (22/22)	+3	Unnamed protein product of *Mus musculus*	39 (48/110)	Unknown
9	**12**(1.4), **25** (1.2), **27**(1.3)	*Emericella nidulans SNAD* gene for spindle pole body associated protein Length = 3213	52 (50/58)	+2	Cytochrome C aa(3) subunit 1 (*Thermus thermophilus*) Length = 791	42 (22/92)	Cytochome C oxidase polypeptide

Table 5. (Continued)

Group	Clone (size in kb)	Result of BLAST searching					Possible function
		Nucleotide BLAST		Translated BLAST			
		Similar to	Bit(identity)	Frame	Similar to	Bit (Positives)	
10	**13**(0.9)	Mouse DNA sequence from clone RP23-7C19 on chromosome X Length = 128978	42 (21/21)	+1	NADH-ubiquinone oxidoreductase 21.3 kDa subunit (*Neurospora crassa*) Length = 201	140 (65/115)	NADH-ubiquinone oxidoreductase
11	**14**(1.6)	*Mus musculus* BAC clone RP23-216B15 Length = 207136	42 (21/21)	+1	Hypothetical protein [*Aspergillus fumigatus*] Length = 525	69 (28/84)	Unknown
12	**15**(1.0), **18** (1.0)	*Penicillium marneffei* cell wall antigen (*MP1*) mRNA, complete cds Length = 1549	88 (59/64)	+3	Basal-body rod modification protein flgD (*E. coli*) Length = 231	33 (20/73)	Mp1-like protein

Table 5. (Continued)

Group	Clone (size in kb)	Result of BLAST searching					Possible function
		Nucleotide BLAST		Translated BLAST			
		Similar to	Bit(identity)	Frame	Similar to	Bit (Positives)	
13	**17**(0.9)	*Botrytis cinerea* strain T4 cDNA library under conditions of nitrogen deprivation Length = 720	52 (56/66)	+1	Glutathione peroxidase of *Shizosaccharomyces pombe* Length = 158	93 (42/65)	Glutathione peroxidase
14	**21**(0.8)	*Arabidopsis thaliana* chromosome 2 clone MJB20 sequence Length = 84592	42 (21/21)	+3	PROTEIN C4	27 (12/36)	Unknown
15	**22**(0.017)	-	-	-	-	-	Artifact
16	**24**(1.5)	*Aspergillus parasiticus* (*nmt1*) mRNA, complete cds Length = 1177	206 (335/412)	+3	NMT1 protein homolog of *spergillus. parasiticus* Length = 342	272 (133/146)	Thiamine biosynthesis enzyme NMT1 (Thymine synthase)

Table 5 (Continued)

Group	Clone (size in kb)	Result of BLAST searching					Possible function
		Nucleotide BLAST		Translated BLAST			
		Similar to	Bit(identity)	Frame	Similar to	Bit (Positives)	
17	**26**(1.0)	*Mus musculus* BAC clone RP24-212F15 from chromosome 10, complete sequence Length = 181828	48 (27/28)	+2	Schwann cell myelin protein precursor of Japanese quail (Siglec-4b) Length = 620	30 (21/65)	Unknown
18	**28**(1.6)	*Emericella nidulans* stearic acid desaturase (*sdeA*) gene, complete cds Length = 2219	133 (130/151)	+3	Acyl-CoA desaturase 1 (Stearoyl-CoA desaturase 1) (Fatty acid desaturase 1) from *S. cerevisiae* Length = 510	102 (50/114)	Stearic acid desaturase

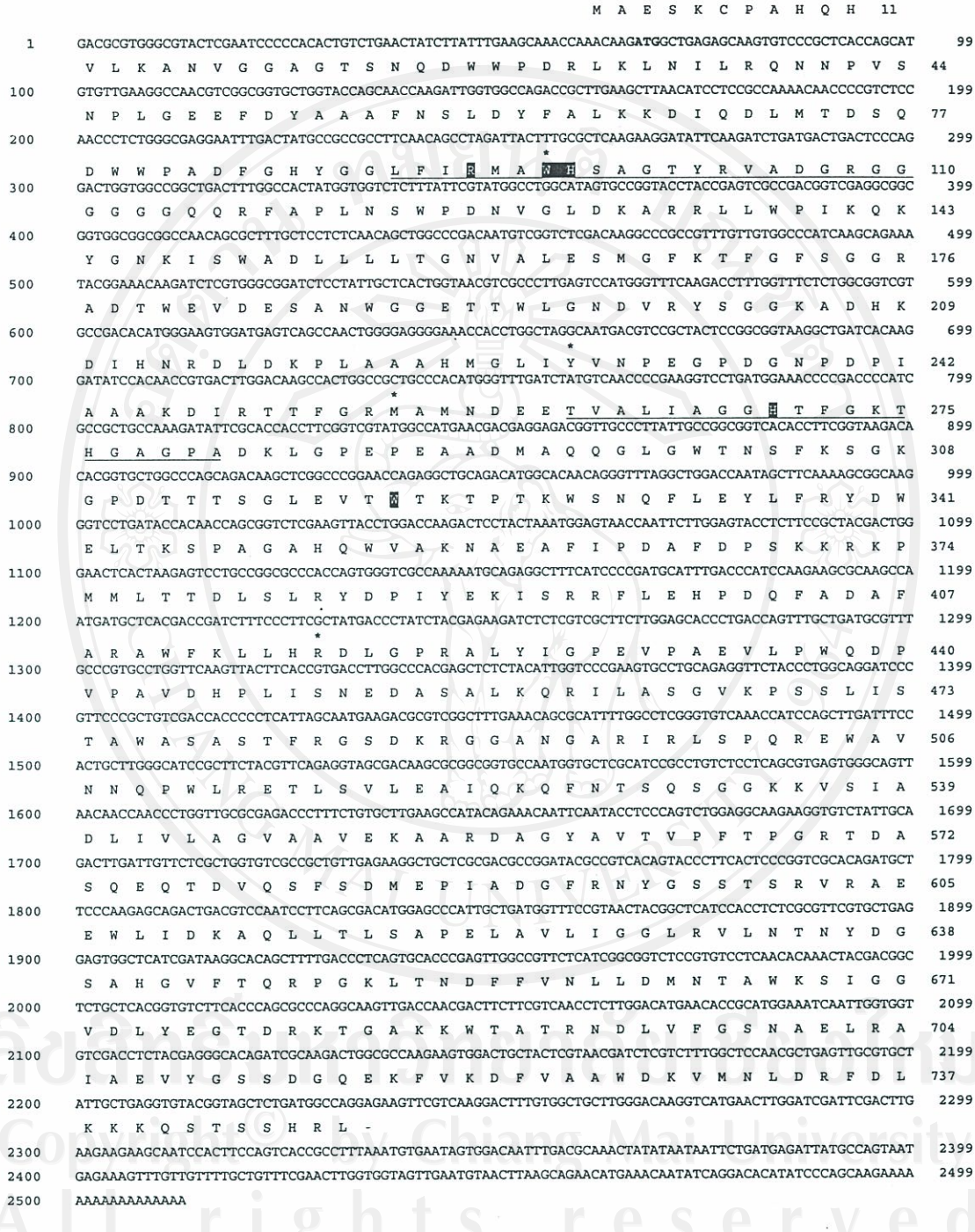


Figure 37. DNA and deduced amino acid sequences of *cpeA* gene. The *cpeA* cDNA encodes 748 amino acids with a predicted molecular mass of 82.4 kDa. Conserved heme associated residues in bacterial catalase-peroxidases and yeast cytochrome C peroxidase are marked by white letters on a black background. Proximal (amino acid residues 262-282) and distal (residues 90-110) sides of coordination to the heme molecule are underlined. Asterisks mark the conserved covalent bonds that are important for the catalase-peroxidase activity.

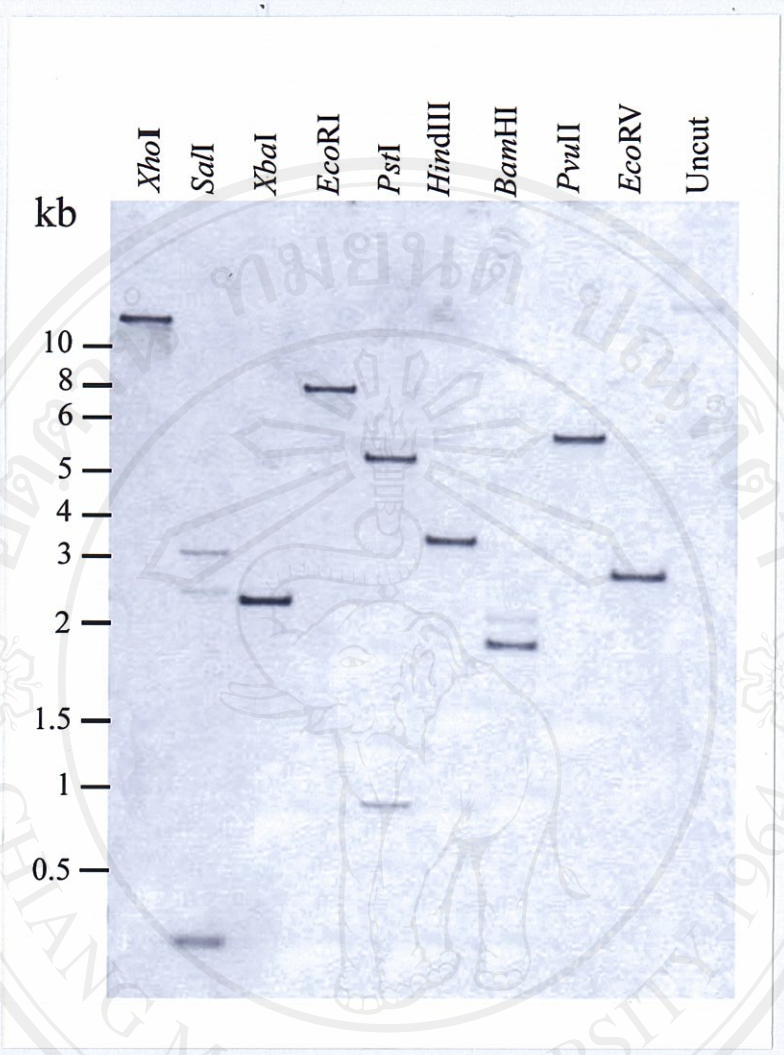


Figure 38. Southern blot analysis of *cpeA*. In accordance with the restriction map of c*peA*, *Bam*HI and *Pst*I cut once inside the targeted hybridization region, thereby resulting in 2 distinct hybridization signals. In contrast, *Sal*I cuts twice within the probe region, thus generating 3 hybridization bands. The enzymes that do not cut at any point in the known sequence (*Xho*I, *Xba*I, *Eco*RI, *Hin*dIII, *Pvu*II, and *Eco*RV), yield a single band as expected. The absence of additional signals under low stringent hybridization indicates that the genome of *P. marneffei* contains a single copy of *cpeA*.

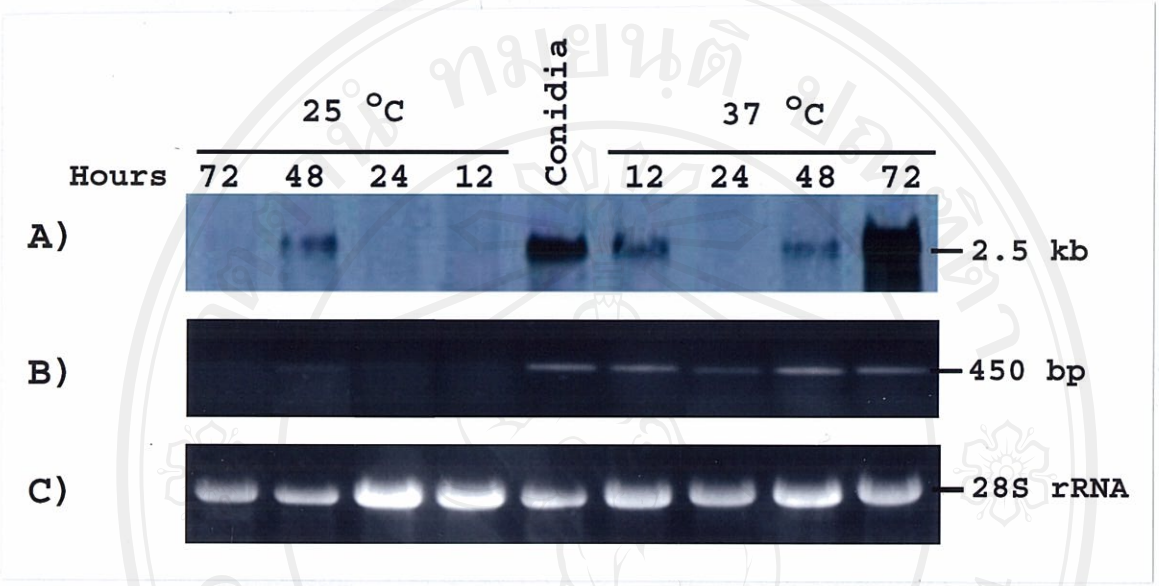


Figure 39. Differential expression of *cpeA*. Total RNA (4 μg) isolated from *P. marneffei* cells at different time points of cultivation at 25°C and 37°C was examined for the c*peA* gene expression. A) Northern blot analysis. Total RNA was probed with a *cpeA* DNA fragment. A 2.5-kb transcript visualized in the analysis was consistent with the size of isolated full-length clone. This gene displayed a high level of expression, being induced when the temperature was shifted to 37°C (yeast phase), especially at 72 h. *CpeA* was nearly absent at 25°C (mold phase). B) RT-PCR result generally supported the Northern blot result. C) RNA loading control for Northern blot analysis. (Result were the same from duplicated experiments).

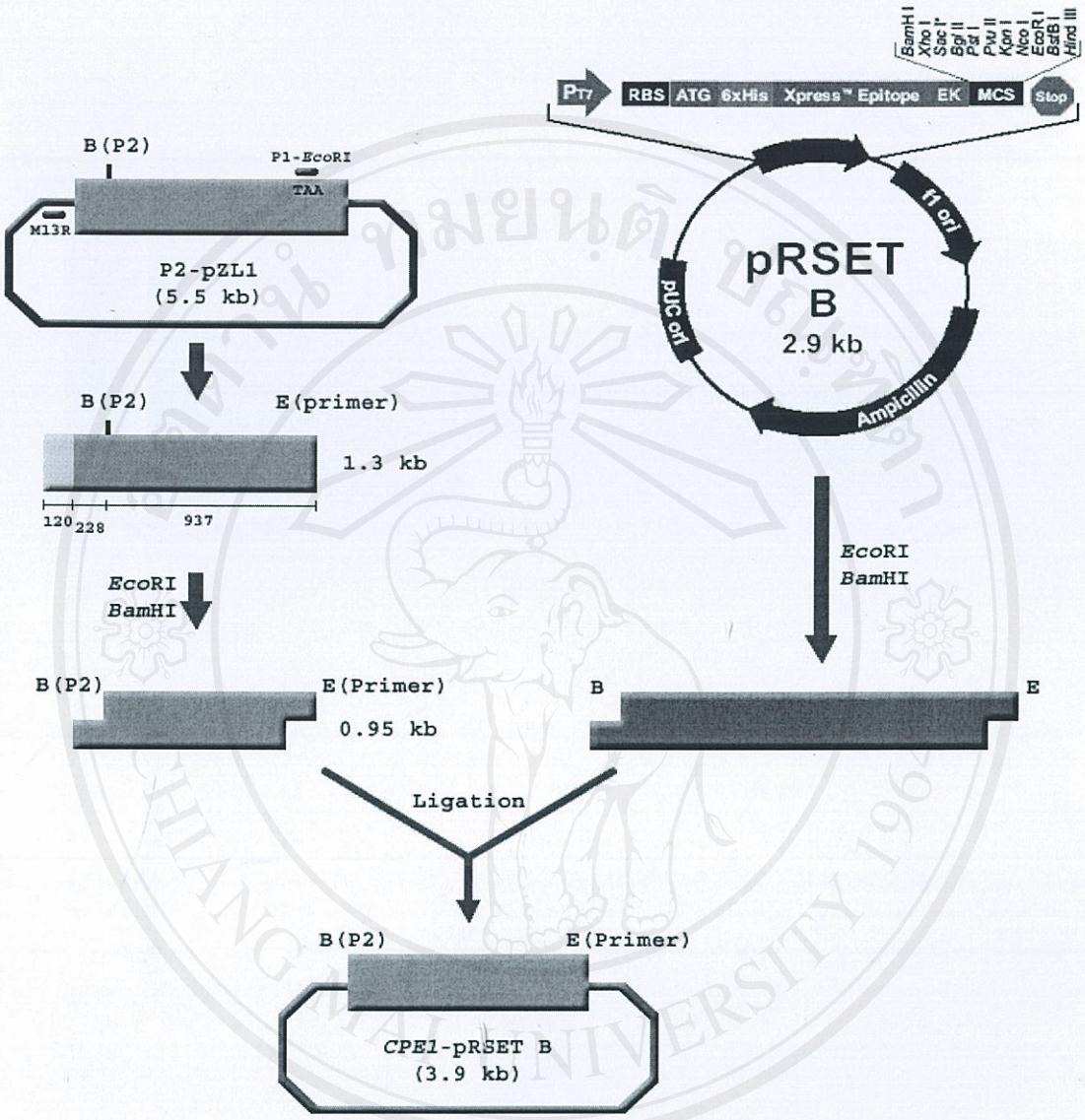


Figure 40. Strategy to clone the *cpeA* gene fragment into the pRSET B vector. The c*peA* gene fragment was PCR amplified from P2 clone using M13R and P1-*Eco*RI primers. The 1.3-kb amplified product contained a *Bam*HI site inside the P2 sequence and *Eco*RI site at the end of P1-primer. Both serve as cloning sites. The PCR product was digested with *Bam*HI and *Eco*RI to produce a directional cloning DNA fragment. The 2.9-kb pRSET B plasmid was cut with the same enzymes. The prepared insert and vector were ligated to produce a recombinant plasmid.

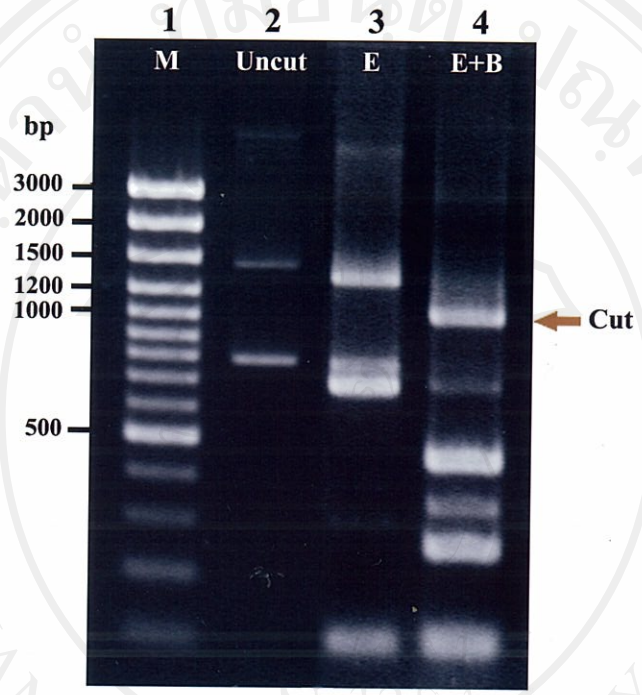


Figure 41. Preparation of a *cpeA* gene fragment for cloning into the pRSET B vector, *Bam*HI/*Eco*RI arms. The *cpeA* gene fragment was amplified from the clone P2 containing 3' sequence of the *cpeA*. The amplified product is shown in lane 2 (uncut). The products after digestion with *Eco*RI is shown in lane 3 (E). The *Eco*RI-digested PCR product was then digested with *Bam*HI (lane 4, E+B). The resulting 1-kb from double enzyme digestion, indicated by an arrow, was cut and purified from the gel and further used in a ligation reaction. The DNA standard sizes (100-bp ladder, New England Biolabs) are labeled on the left.

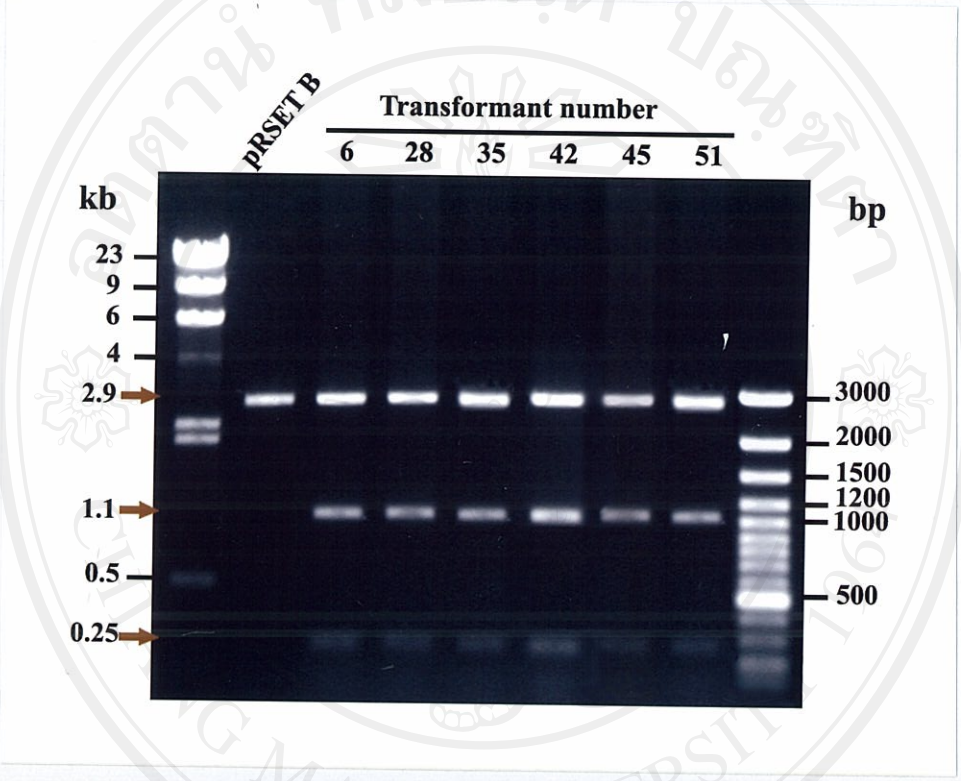


Figure 42. *Eco*RI and *Bam*HI double digestion of the recombinant *cpeA*-pRSET B plasmid. The recombinant plasmid was digested with *Eco*RI and *Bam*HI, the cloning enzymes. The fragments of 1.1 and 0.25 were released from the 2.9-kb pRSET B instead of the single 1.3-kb fragment.

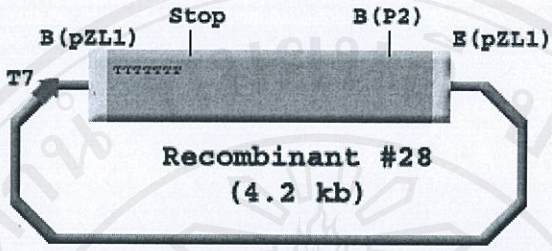


Figure 43. The recombinant plasmid from the cloning of *cpeA* gene fragment into the pRSET B. Cloning of the *cpeA* fragment into the pRSET B vector using *Eco*RI and *Bam*HI as the cloning sites resulting in the recombinant plasmid with the incorrect direction. The inversion of the cloned gene occurred from the disposition of the P1-*Eco*RI primer that was used to amplify the insert fragment, resulting in amplification of additional *Bam*HI sequence marked as B(pZL1). This *Bam*HI site came from the pZL1 plasmid which serves as an recipient vector for the clone P2.

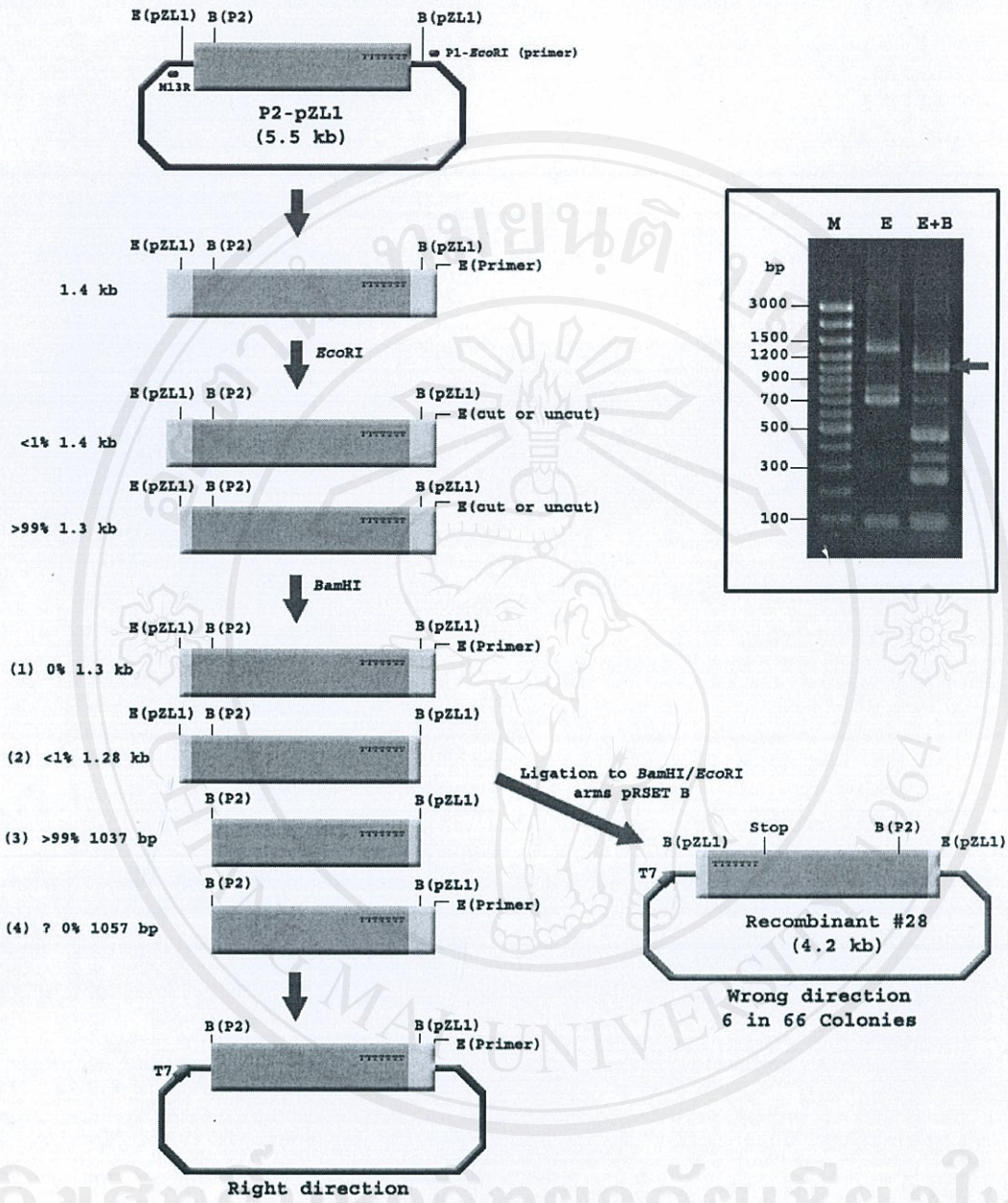


Figure 44. The actual event in cloning of the *cpeA* gene fragment into the pRSET B vector using *Bam*HI and *Eco*RI cloning sites. The 1.4-kp amplified fragment resulted from the dispositioning of P1-*Eco*RI primer contains two *Eco*RI and two *Bam*HI sites. One of the *Eco*RI was from the pZL1 [E(pZL1)] and the other one was from P1-*Eco*RI primer [E(primer)]. Two *Bam*HI sites came from the internal site inside the *cpeA* sequence [B(P2)] and pZL1 plasmid [B(pZL1)]. In case of incomplete cutting, four types of fragment could be generated. The fragments

containing *Eco*RI and *Bam*HI sites at each end could be cloned into the corresponding sites of the cut pRSET B plasmid. After excising of the 1-kb DNA band from the gel (shown in palette), all 4 types of fragment could be purified. The fragments that contained 2 *Eco*RI sites [type (1)] and 2 *Bam*HI sites [type (3)] could not be cloned. The fragments of type (2) and (4) that occurred in low proportion (<1 %) could be cloned into the cut pRSET. However, a recombinant plasmid with the insert fragment type 4 was not found in any of the transformants.

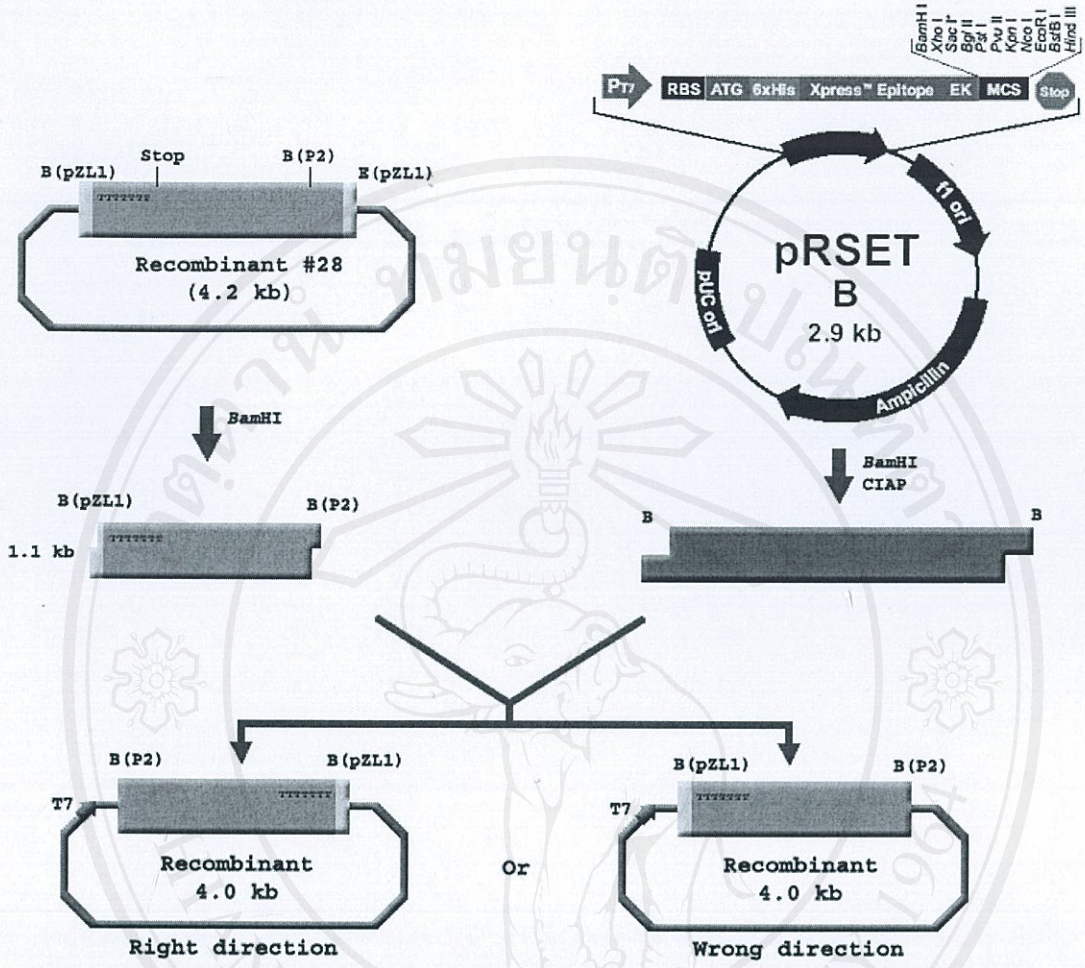


Figure 45. Strategy to invert the direction of the insert *cpeA* gene fragment cloned in the pRSET B vector. The recombinant # 28 was re-digested with *Bam*HI. The 1.1-kb fragment was released. One end of the 1.1-kb fragment was the *Bam*HI site generated from *cpeA* sequence [B(P2)], and the other *Bam*HI end was from pZL1 sequence [B(pZL1)]. This fragment was re-ligated to the dephosphorylated *Bam*HI-digested pRSET B. The regenerated recombinant plasmids contain either insert with proper or improper direction for the expression.

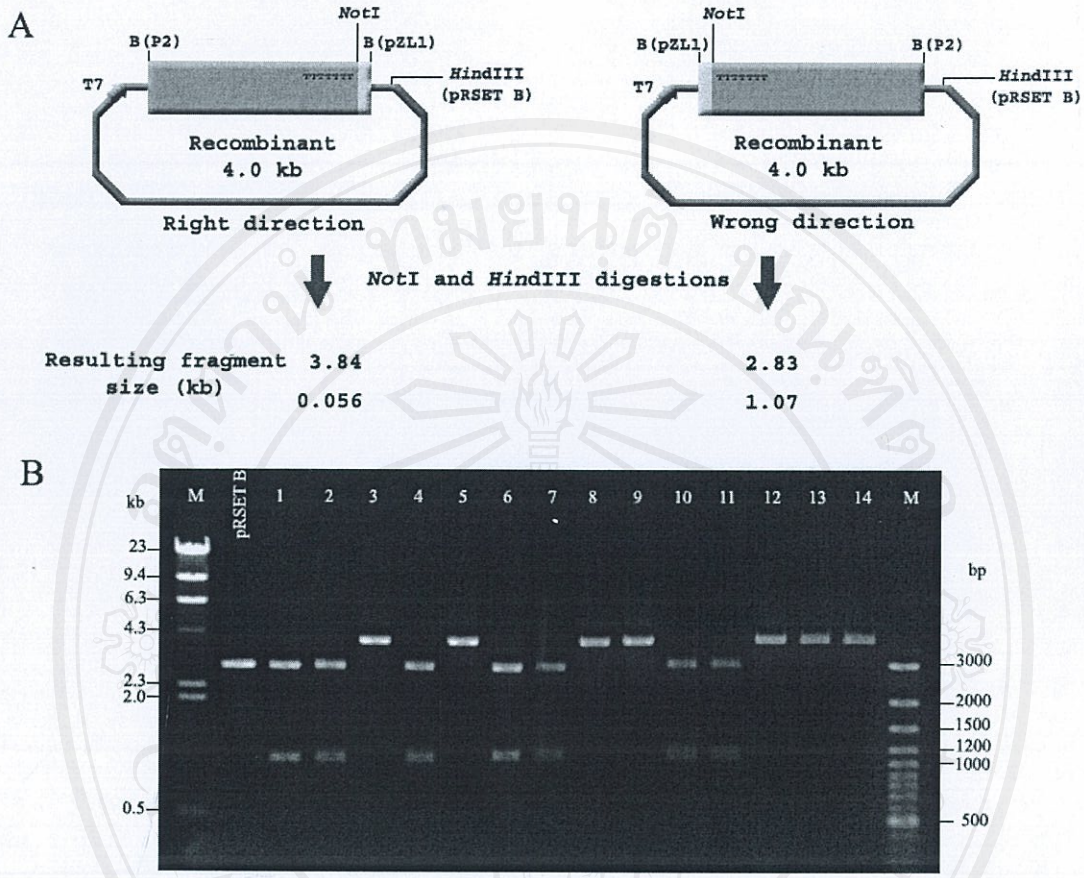


Figure 46. Determining for the orientation of the re-generated recombinant plasmids. *Not*I and *Hin*dIII digestions could distinguish the recombinant plasmids. *Not*I site is in the insert fragment, and *Hin*dIII is a unique site inside the pRSET B vector. Digestion of the recombinant plasmid that contained insert in the right direction gave the fragments of 3.84-kb and 56-bp, whereas those containing insert in the wrong direction gave the fragments of 2.83- and 1.07-kb (A). The result from digestion of the plasmids isolated from 14 transformants found that 7 of 14 (number 3, 5, 8, 9, 12, 13, and 14) contained an insert with the right direction (B).

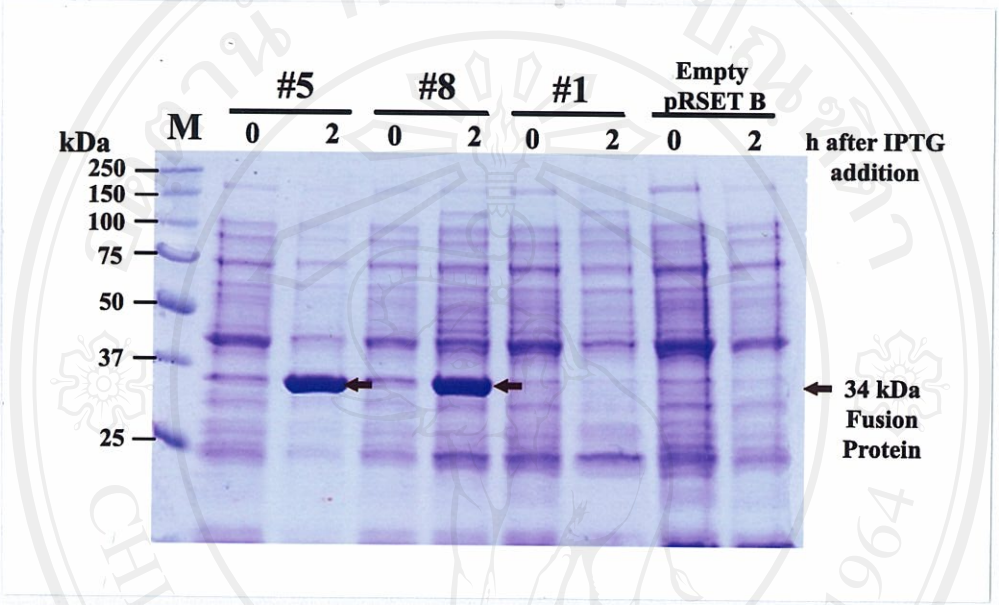


Figure 47. Expression of the CpeA-His$_6$ tag fusion protein. Two transformants containing recombinant plasmid with the proper direction for expression (#5 and #8) produced the fusion protein of predicted molecular weight (34-kDa) under an induction of 1 mM IPTG for 2 h at 37 ^{o}C (lane 2 of #5 and #8). The transformant that received the recombinant plasmid with improper direction could not produce the fusion protein (lane 2 of #1). The fusion protein was not seen in the induction of transformant containing an empty pRSET B, which serve as a negative control.

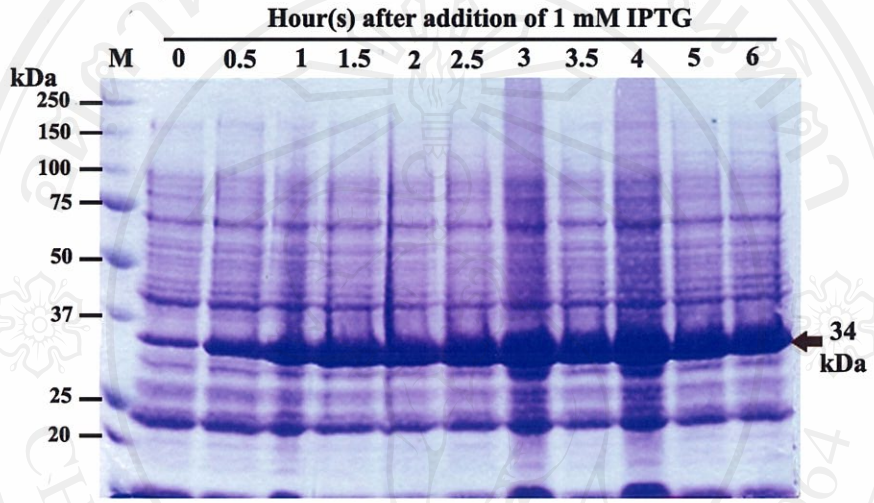


Figure 48. Time optimization of the CpeA-His$_6$ tag fusion protein induction. An optimal time of induction was determined under the expression condition at 37 °C and 1 mM IPTG. The *E. coli* cells were collected at various time points and analyzed by Coomassie blue stained 10 % SDS-PAGE gel. At time 0 (before addition of IPTG), there was some levels of expression. After 0.5 h of induction, the amount of fusion protein has gradually increased until it reached a peak at 1.5 h of induction. Since then, the amount of expressed protein did not significantly increase. Therefore, optimal time of induction ranged from 1.5 to 6 h.

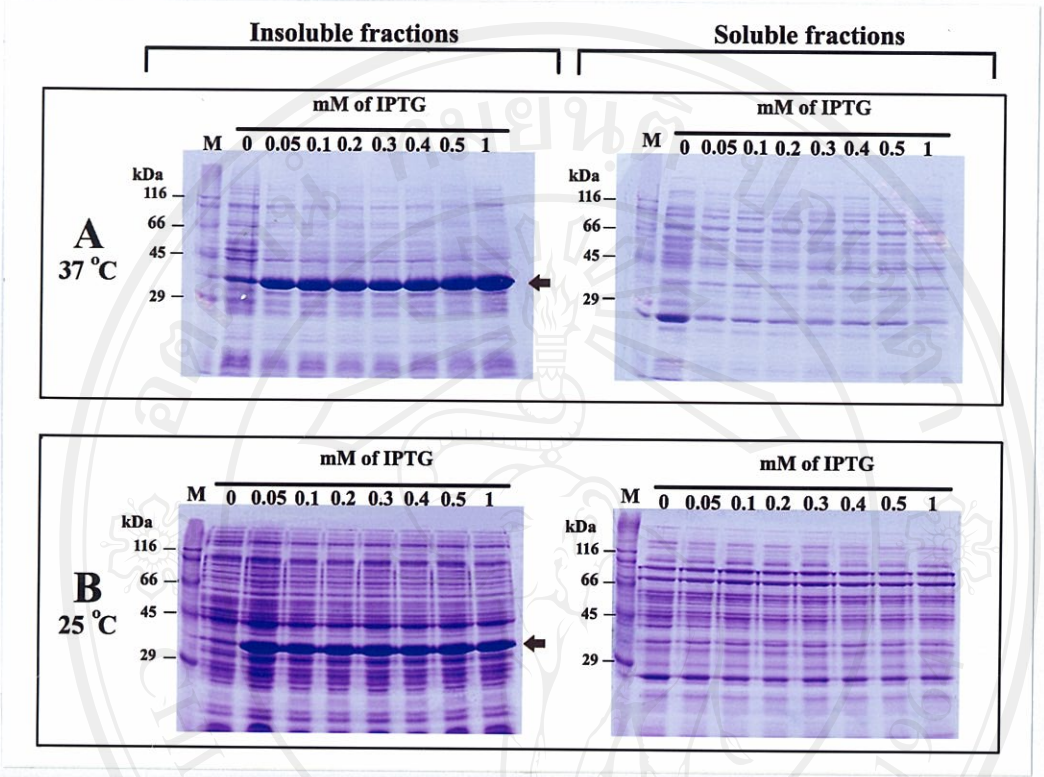


Figure 49. Effects of temperature and various IPTG concentrations on expression of the CpeA-His$_6$ tag fusion protein. Two induction temperatures and various concentrations of IPTG (0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 and 1 mM) were tested to optimize the expression conditions. The cells were collected at 2 h after addition of IPTG and the soluble and insoluble parts were fractionated. Each part was analyzed by 10 % SDS-PAGE. The fusion protein was found only in the insoluble fraction at both induction temperatures. Lowering the IPTG concentration could not improve the solubility of the expressed protein.

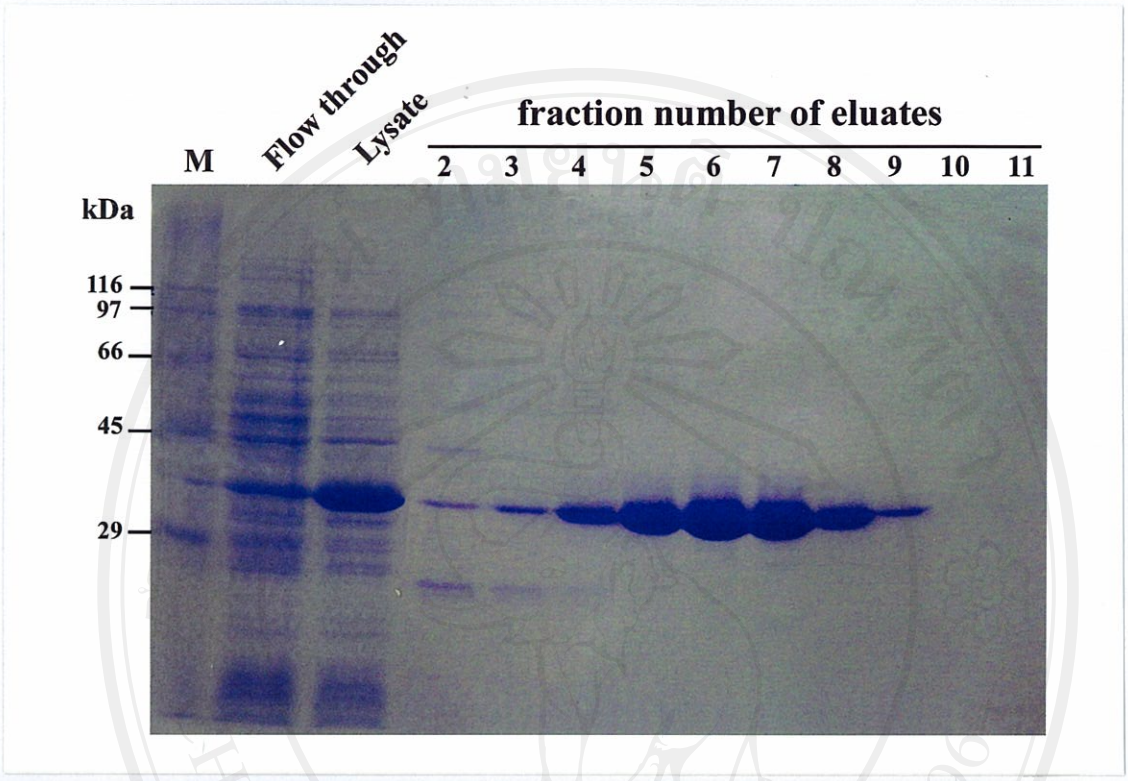


Figure 50. Purification of the CpeA-His$_6$ tag fusion protein. The fusion protein was prepared from the transformant #5. Twelve fractions were collected. Ten microliters of each fraction were analyzed by SDS-PAGE, including the *E. coli* lysate before (Lysate) and flow through after (Flow through) binding to the Probond™ resin column (Gibco BRL). The fusion protein was eluted in large amount in fractions no. 4-9. The first three fractions contained co-purified *E. coli* protein. The amount of fusion protein that exceed the capacity of column was retained in the lysate after binding to the column.

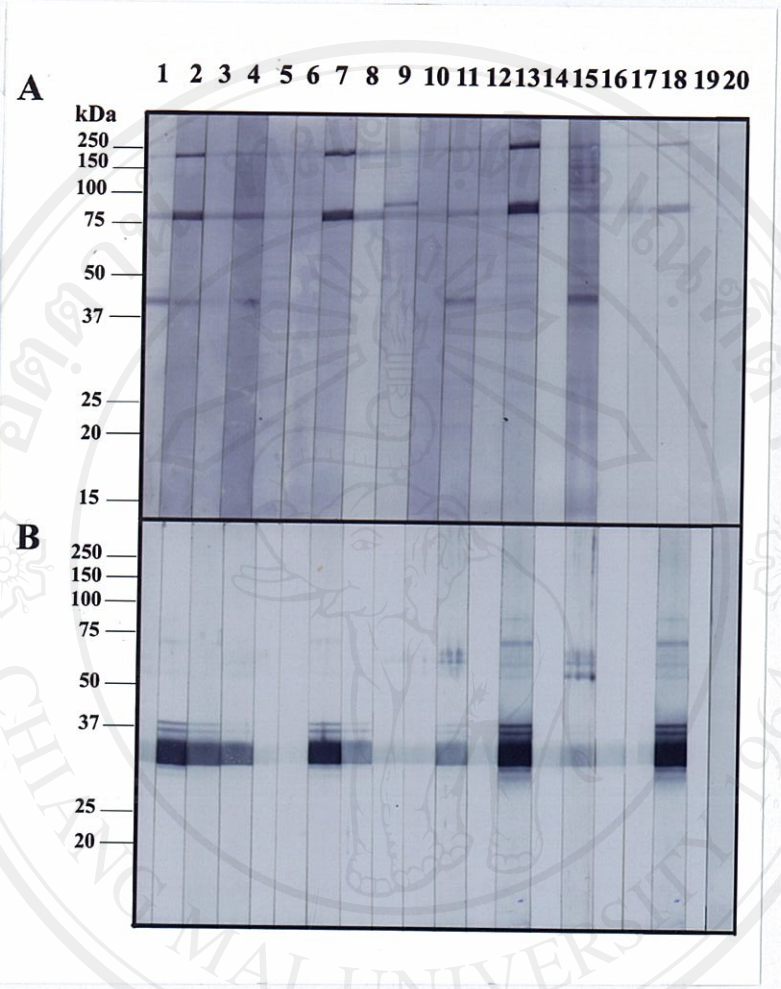


Figure 51. Immunoblot reactivities of *P. marneffei*-infected AIDS patients' sera to the CpeA-His$_6$ tag fusion protein. Panel A shows the reactivities of tested sera to *P. marneffei* crude protein antigens collected from 3-day-old yeast cells. Panel B shows the reactivities to the CpeA-His$_6$ tag fusion protein. Twenty serum samples were tested: fifteen individual serum from *P. marneffei*-infected AIDS patients (lane 1-15), a pooled fungal laboratory personnels (lane 16), a non-*P. marneffei* infected AIDS patient (lane 17), a pooled serum of positive control (lane 18), a pooled serum from people in non-endemic area (purchased from Sigma) as a negative control (lane 19), and a pooled serum from healthy people in Chiang Mai (lane 20). Protein standard (Precision Plus; BIO-RAD) size is shown on the left.

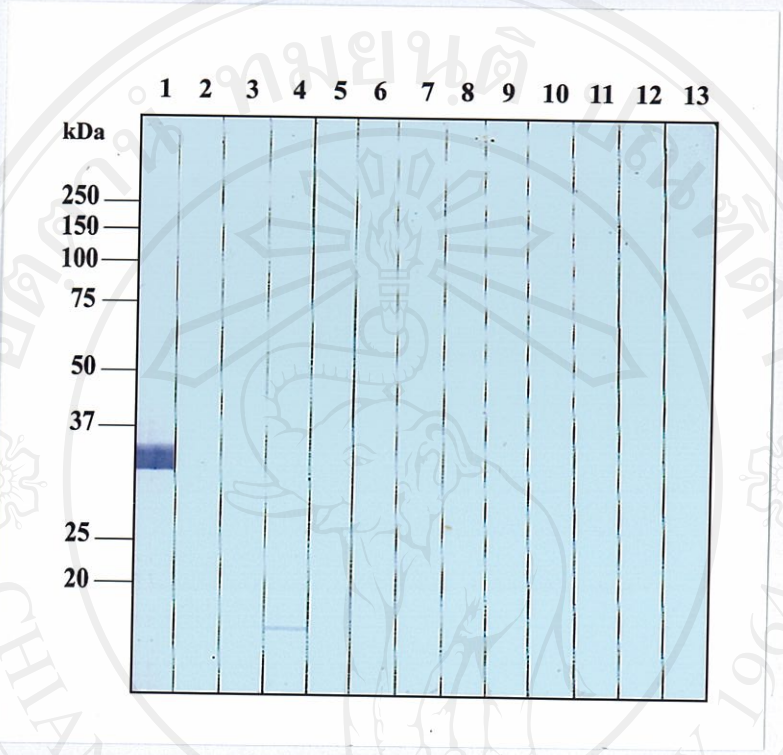


Figure 52. Immunoblot reactivities of fungal- and *M. tubercusis*-infected AIDS patients' sera to the CpeA-His$_6$ tag fusion protein. Sera obtained from AIDS patients with penicilliosis marneffei (lane 1), candidiasis (lane 2), aspergillosis (lane 3), histoplasmosis (lane 4), cryptococcosis (lane 5-11), and tuberculosis (lane 12-13) were Western immunoblot analysis with the CpeA-His$_6$ tag fusion protein. Protein standard size (Precision Plus; BIO-RAD) is shown on the left.