

GENE 06163

Cloning and selective overexpression of an alkaline protease-encoding gene from *Aspergillus oryzae*

(Recombinant DNA; *alpA* gene; serine protease; transformation; gene expression; filamentous fungi; soy sauce fermentation)

S. Cheevadhanarak^b, D.V. Renno^a, G. Saunders^a and G. Holt^a

^a The International Institute of Biotechnology, Polytechnic of Central London, London W1M 8JS (U.K.); and ^b King Mongkut's Institute of Technology, Thonburi, Bangmod, Rasburana, Bangkok 10400 (Thailand)

Received by J.R. Kinghorn: 13 June 1991

Revised/Accepted: 17 July/18 July 1991

Received at publishers: 13 September 1991

SUMMARY

The gene *alpA* encoding *Aspergillus oryzae* alkaline protease (ALP) was isolated from a genomic library of an industrial strain used in Thailand by using oligodeoxyribonucleotide probes based on the published cDNA sequence [Tatsumi et al., Agric. Biol. Chem. 52 (1988) 1887–1888]. The entire nucleotide sequence of the genomic clone obtained was determined. By comparison with the published cDNA sequence, it was found that ALP is encoded by four exons of 314, 445, 89 and 351 bp. Three introns, which interrupt the coding sequence, are 50, 59 and 56 bp in length. The gene contains a typical TATA box 103 bp upstream from the start codon, and a consensus polyadenylation signal, AATAAA, 189 bp from the stop codon. The *alpA* gene, introduced into a protease deficient strain (*A. oryzae* U1638) by cotransformation, directed the secretion of enzymatically active ALP into the culture medium. Cotransformants of the high-level ALP-producing strain U212 containing multiple copies of the *alpA* gene were able to secrete up to five times more ALP than the parental strain.

INTRODUCTION

Soy sauce production is a solid state fermentation, in which many proteolytic and carbohydrase enzymes are involved. The alkaline protease of *A. oryzae* is a crucial enzyme in soy sauce manufacture. It is considered to play a most important role in producing the taste of soy sauce by hydrolysing the raw materials (Tatsumi et al., 1989). Im-

provement of *A. oryzae* in order to increase the amount of this enzyme has, in the past, been carried out successfully by mutation and selection (Kalayanamitr et al., 1987), and by protoplast fusion (Furuya et al., 1983). Besides these techniques, the development of transformation systems for *A. oryzae* (Cheevadhanarak et al., 1991) has made it possible to apply recombinant DNA techniques in efforts to obtain further improvements.

Increased enzyme activity as a result of gene amplification has already been observed in filamentous fungi. In the case of penicillin-producing strains it appears that the entire gene cluster encoding the penicillin biosynthetic genes has been amplified during strain improvement by mutation (Smith et al., 1989). Using recombinant DNA techniques specific amplification of the expandase gene from *Cephalosporium acremonium* was carried out and resulted in increased activity (Skatrud et al., 1989).

The genetic basis of alkaline protease overproduction in *A. oryzae* is of considerable interest if a new approach using

Correspondence to: Dr. G. Saunders, Polytechnic of Central London, 115 New Cavendish St., London W1M 8JS (U.K.)
Tel. (44) 71 911 5000; Fax (44) 71 911 5087.

Abbreviations: aa, amino acid(s); *A.*, *Aspergillus*; ALP, alkaline protease; *alpA*, gene encoding alkaline protease; bp, base pair(s); kb, kilobase(s) or 1000 bp; N, A or C or G or T; ND, not determined; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; Ph, phleomycin; PMSF, phenylmethylsulfonyl fluoride; ^R, resistance/resistant; SDS, sodium dodecyl sulfate; SSC, 0.15 M NaCl/0.015 M Na₃ citrate pH 7.6; u, unit(s); UV, ultraviolet; wt, wild type; Y, C or T.

molecular techniques is to be applied to future strain improvements. In order to understand the ALP structure and function at a molecular level, information on the nature and function of the enzyme, and the structure of its corresponding gene must be available. ALP from *A. oryzae* is known to be a serine protease (Nakagawa, 1970). Tatsumi et al. (1989) have reported the structure of the *alpA* cDNA and its expression in *Saccharomyces cerevisiae*.

The aim of the present study was to isolate and characterize a genomic clone of the ALP-encoding gene of *A. oryzae* U212. The expression of the *alpA* gene in an ALP-deficient and a high ALP-producing *A. oryzae* strain was also studied.

EXPERIMENTAL AND DISCUSSION

(a) Isolation of the *alpA* gene

The *A. oryzae alpA* gene was isolated from a λ EMBL3 gene library by hybridization with two synthetic oligos based on the published *alpA* gene cDNA sequence (Tatsumi et al., 1988). Two recombinant bacteriophages were isolated from the screening of 12 000 plaques. Restriction analysis of the two clones indicated that they contained the same insert. Southern blot analysis showed that the two probes hybridized to a single 8.5-kb *Bam*HI fragment which was subcloned into pUC18 generating pAP7 (Fig. 1). Subsequently a 4.0-kb *Xba*I fragment, containing sequences downstream from the *alpA* gene in pAP7, was removed and the resulting plasmid was designated pAP71.

(b) Structure of the *alpA* gene

The sequencing strategy used to determine the nt sequence of the *alpA* gene is given in Fig. 1B. Comparison of this sequence (Fig. 2) and the cDNA sequence published by Tatsumi et al. (1989) revealed a complete homology within the coding region. The only difference is the presence of a single triplet TCC located 43 nt upstream from the ATG start codon which is present in duplicate in the cDNA sequence published.

The comparison with the cDNA sequence also indicated the presence of three introns (designated 1, 2 and 3) within the *alpA* gene. These introns have the characteristic features of intron sequences characterised in other filamentous fungi (Gurr et al., 1987). First, they are small in size (50, 59 and 56 bp, respectively). Second, the boundaries are similar to those of *S. cerevisiae* and other filamentous fungi in possessing the nt sequences GTNNGT and YAG bordering the 5' and 3' extremities, respectively. Third, within each intron lies a sequence close to the consensus sequence YGCTAACN, thought to be necessary for intron splicing. These internal consensus sequences are located 9, 13 and 12 nt away from the 3' splicing site in the introns 1, 2 and

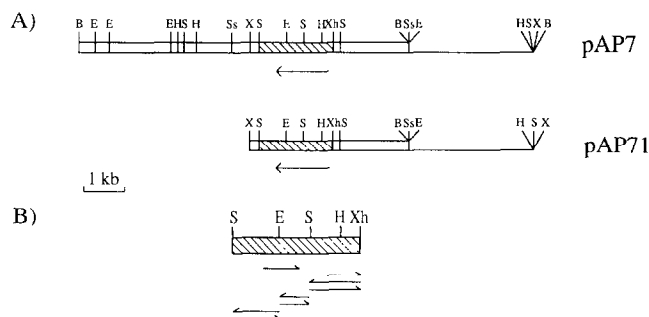


Fig. 1. Restriction maps of pAP7 and pAP71 and sequencing strategy. (A) Restriction map of pAP7 and pAP71. Single line, pUC18; box, *A. oryzae* DNA; hatched box, region sequenced. **Methods:** A genomic library of *A. oryzae* U212 was constructed by cloning *Bam*HI fragments of genomic DNA into a λ EMBL3 vector. After transfection into *E. coli* Q359, 12 000 resulting plaques were screened with two 21-mer oligo probes corresponding to the 5' and 3' ends of the published cDNA sequence (Tatsumi et al., 1988). The nylon membranes (Hybond N, Amersham) were incubated overnight at 65°C in the prehybridization solution (6 × SSC/0.1% SDS/5 × Denhardt's solution/100 µg per ml salmon sperm DNA). The prehybridization solution was then removed and filters were hybridized in 6 × SSC/0.1% SDS/5 × Denhardt's solution/1–10 ng per ml ³²P-labeled oligo for 2 h at 65°C. Following hybridization, the filters were washed four times in 6 × SSC/0.1% SDS, at room temperature for 5 min each, and were then washed once at 65°C. The membranes were exposed to X-ray film (Kodak X-AR) at –70°C. The positive plaques were isolated and characterized by restriction analysis. An 8.5-kb *Bam*HI fragment was subcloned into pUC18 resulting in the plasmid pAP7. The orientation of the *alpA* gene was inferred from Southern blot experiments using probes specific for either end of the gene. A 4.0-kb *Xba*I fragment containing sequences downstream from the *alpA* gene in pAP7 was removed and the resulting plasmid was designated pAP71. (B) Sequencing strategy. The nt sequence of *alpA* gene was obtained by sequencing the 1820-bp *Xho*I–*Sal*I fragment of pAP71. The 0.7-kb *Xho*I–*Sal*I, 0.3-kb *Sal*I–*Eco*RI and 0.8-kb *Eco*RI–*Sal*I fragments were subcloned into pBluescript (pAP72, pAP73 and pAP74, respectively). The sequencing was carried out according to Sanger et al. (1977) using Sequenase (U.S. Biochemical Corp., Cleveland, OH). The oligos used to isolate the gene were also used in sequencing pAP71 to determine the sequences overlapping the fragments. B, *Bam*HI; E, *Eco*RI; H, *Hind*III; S, *Sal*I; Ss, *Sst*I; X, *Xba*I; Xh, *Xho*I.

3, respectively. The four exon sequences are 314, 445, 89 and 351 bp, respectively.

Sequences thought to be involved in the control of gene expression are indicated in Fig. 2. The sequence TATAAA has been found 103 nt upstream from the ATG start codon. Translation of most fungal genes begins at the first ATG, and there is usually a consensus of DNA around the start codon. The nt located 3 nt upstream from the putative start codon is an adenosine. This is commonly observed for genes from filamentous fungi (Gurr et al., 1987).

In common with some other fungal genes (Ballance, 1986), the predicted polyadenylation signal AATAAA (Proudfoot and Brownlee, 1976) is present within the 3' flanking sequence of the *alpA* gene from *A. oryzae* U212 185 bp downstream from the stop codon.

```

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CTCGAGAACAGGGAATATAGAACTCCTGAAGGTATATAATACCATGCG 50
ATCCCTCGTCCATCCTCACTTCCATCATCAAGCCAGCGGTTTCTATCCGA 100
CTTGAGTTGTTCTTGGCATCTTTACAATCTTCTCATCATGCAGTCCATC 150
                                     m q s i
AAGCGTACCTTGCTCCTCGGAGTATCCTTCCCGGTCCTCGGTGC 200
k r t l l l l l g a i l p a v l g a
CCCTGTGCAGGAACCCGCGCGGCTGAGAAGCTTCTGGAAAGTACA 250
p v q e t r r a a e k l p g k y i
TTGTACATTCAAGCCCGGATTGACGAGGCAAGATTGAGGAGCATACC 300
v t f k p g i d e a k i q e h t
ACCTGGGCTACCAACATTACCAGCGCAGTCTGGAGCGTCGTGGCGCCAC 350
t w a t n i h q r s l e r r g a t
TGGCGGTGATCTTCTGTCGGTATTGAGCGCAACTACAAGATCAACAAGT 400
g g d l p v g i e r n y k i n k f
TCGCCGCTATGCAGGCTCTTTCAGCATGCTACCATGAGGAGATTCGC 450
a a y a g s f t d d a t i e e i r
AAGAACCAAGATGtttgggtcatcgctcgattttgaatgacagcta 500
k n e d
actcgcgccagGTTGCCTACGTCGAGGAGGACCAGATCTACTACCTCGA 550
v a y v e e d q i y y l d
TGGCCTGACTACCCAGAAGAGTGCCCTCGGGTCTGGGCAGCATTTCCC 600
g l t t q k s a p w g l g s i s h
ACAAGGGCCAGCAGAGCAGGACTACATCTACGACACTAGTCCGCGGAG 650
k g q q s t d y i y d t s a g e
GGCACCTATGCCTACGTGGTGGATAGCGGTGTCAATGTGACCATGAGGA 700
g t y a y v v d s g v n v d h e e
GTTTCGAGGGCCGCGCCAGCAAGGCTACAACGCTCCGGTGGTCAGCATG 750
f e g r a s k a y n a a g g q h v
TGGACAGCATTTGGCCATGGCACCACGTTTCGGGACCATGCTGGCAAG 800
d s i g h g t h v s g t i a g k
ACTTATGGTATGCCAAGAAGGCCAGCATCCTTTCGGTCAAAGTTTCCA 850
t y g i a k k a s i l s v k v f q
GGGTGAATCGAGCAGCACTTCCGTATTCTTGACGGCTTCAACTGGGCTG 900
g e s s s t s v i l d g f n w a a
CCAACGACATTGTTAGCAAGAAGCTACCAGCAAGGCTGCAATCAACATG 950
n d i v s k k r t s k a a i n m
AGCTTGGgtgagtttaccattgttcttctctacttggaaatcgagagtcgct 1000
s l g
aatttcaaaaacacagGCGGTGGCTACTCTAAGGCTTTCAACGATGCGGT 1050
g g y s k a f n d a v
CGAGAACGCATTGAGCAGGGTGTCTCTCGGTTGTCTGCTGCCGTAACG 1100
e n a f e q g v l s v v a a g n e
AGAACGtagctctccctccatcgcgcaaaagacgaattcgtaactgactt 1150
n
gattttcttagTCTGATGCCGGCCAAACAGCCCTGCCTCTGCCCTGAT 1200
s d a g q t s p a s a p d
GCCATCACTGTTGCCGCTATCCAGAAGAGCAACAACCGCGCAGTTTCTC 1250
a i t v a a i q k s n n r a s f s
CAACTTTGGCAAGGTCGTTGACGCTTTCGCTCCCGGTCAAGATATCCTTT 1300
n f g k v v d v f a p g q d i l s
CTGCCTGGATTGGCTCTTCTCTGCCACCAACACCATCTCTGGTACCTCC 1350
a w i g s s s a t n t i s g t s
ATGGCTACTCCCAACATTGTGCGGCTGTCCCTCTACCTCGCTGCCCTTGA 1400
m a t p h i v g l s l y l a a l e
GAACCTCGATGGCCCGCTGCCGTGACCAAGCGCATCAAGGAGTTGGCCA 1450
n l d g p a a v t k r i k e l a t
CCAAGGACGTCGTCAAGGATGTTAAGGGCAGCCCTAACCTGCTTGCCCTAC 1500
k d v v k d v k g s p n l l a y
AACGGTAACGCTTAAGTACCAGGAGTACGTCGAGGATTCTACCATGTT 1550
n g n a
ACTGGAATACAATGATGATTAGAAAACGAAGAGCGTTATGATTCGGACGG 1600
ATATATGCATGGCACCACATACAGCGTGATACATAGGCTGTTTGCTCAAGA 1650
ATTAGGATTTTATCTGAATCCATGTACAGAGTATCTTATGTTAGTAGTC 1700
AATAAAATCTTGGCTTTCTAATTTTGTCCGATCTACAAGGGGTCGTCGAT 1750
CACAGAACGAAGTATGTCAGGGGACGATGATCACCCGTTCTTAGCAA 1800
GACCTCTAGTAGTTGTGCGAC 1820

```

Fig. 2. The nt sequence of the *alpA* gene of *A. oryzae* U212 and deduced aa sequence. The putative TATA box and polyadenylation sequences are underlined. Lower-case letters in the nt sequence are used to indicate the position of the introns.

(c) Expression of the *alpA* gene in *Aspergillus oryzae* U1638

The *alpA* gene was further characterized by expression of the isolated gene in *A. oryzae* strain U1638, a protease

nonproducing mutant (Kalayanamitr et al., 1987). Transformation was conducted as described by Cheevadhanarak et al. (1991). As pAP71 contains no selectable marker, which can be expressed in *A. oryzae* U1638, the pOBT, containing a Ph^R gene was used together with pAP71 to allow the screening of the transformants. The Ph^R transformants obtained were screened for protease production on skim milk plates (Kalayanamitr et al., 1987). After incubation at 30°C for three days, two of the transformants out of 36 obtained were producing protease, as assessed by the clearing zone around the colonies.

To investigate whether the clearing zone was due to the ALP, the transformants were grown in liquid medium (Kalayanamitr et al., 1987). The supernatant collected after 72 h incubation at 30°C was tested for alkaline protease activity using the modified Anson's method (1938). In both cases, the transformants secreted enzymatically active ALP into the medium. This ALP activity was completely inhibited by the presence of 0.1 mM PMSF thus further indicating that the enzyme produced was a serine protease.

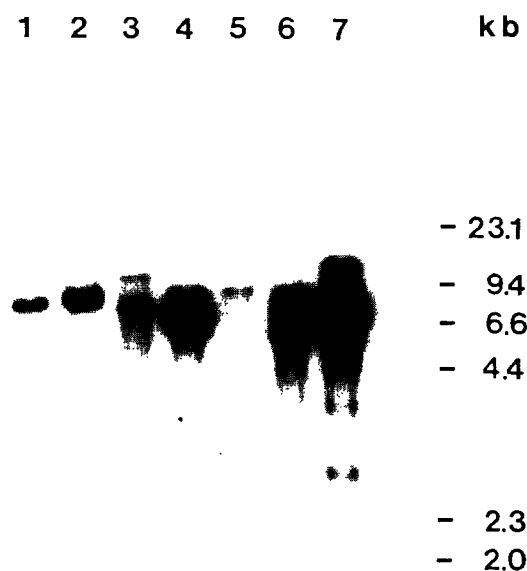


Fig. 3. Southern-blot analysis of transformants of the *A. oryzae* strain U212. Total DNA was isolated following the technique described by Picknett et al. (1987). The DNAs (5 µg) of the transformants were digested to completion with *Bam*HI. The fragments were separated in a 0.8% agarose gel, transferred to a nylon membrane (Hybond N, Amersham Int.) which was then hybridized with ^{32}P -labeled *Bam*HI-*Xba*I fragment of pAP71 (see Fig. 1A). Hybridization and post-hybridization washes were conducted as recommended by the manufacturer. The positions of the size markers are indicated on the side of the figure. Lanes 1–7 contain, respectively, U212 (a Ph^R transformant from U212, which does not produce higher ALP levels), U212, U271, U821, U921, U1421 and U1521 DNA (see Table I).

In addition, Southern-blot analysis was performed to confirm that the increase in the ALP level was due to the presence of additional copies of the *alpA* gene in the genome of the transformants. In both ALP-producing transformants the hybridization signal obtained, using the radio-labeled 4.5-kb *Bam*HI-*Xba*I fragment from pAP71 as a probe, indicated the presence of additional copies of the gene within the transformants (data not shown). In these two transformants, the presence of an intense hybridization signal at a molecular weight corresponding to that for the linear plasmid indicates that the plasmid was integrated in tandem arrays. Interestingly, one transformant isolated after cotransformation contained additional copies of the *alpA* gene but did not produce ALP. In this case it is possible that one of the sites of integration was involved in the regulation of ALP production.

(d) Expression of the *alpA* gene in *Aspergillus oryzae* U212

In order to investigate whether or not multiple copies of the *alpA* gene would affect the level of ALP produced, transformants of the strain U212 were obtained. Four transformants out of five selected were able to produce ALP at a higher level than the parental strain U212 (Table I).

Southern-blot analysis showed that the strain U921

contained no extra copies of the *alpA* gene (Fig. 3). The number of copies of the integrated pAP71 was determined by slot-blot analysis in the transformants U821, U1421 and U1521 (Table I). No linear relation between copy number and ALP production was observed but the strain U1521, which contained an estimated twelve copies of the plasmid, produced five times the amount of ALP, when compared with the parental strain U212.

(e) Conclusions

(1) The *alpA* gene from *A. oryzae* contains three introns of 50, 59 and 56 bp, respectively.

(2) Amplification of the *alpA* gene by transformation can result in elevated levels of ALP in transformants. High-level ALP-producing strains such as U212 obtained by UV mutagenesis can be improved further using this cloning/amplification method.

ACKNOWLEDGEMENTS

This project was conducted under an agreement between the International Institute of Biotechnology (U.K.) and The National Centre for Genetic Engineering and Biotechnology (Thailand) and made possible by the generous support of British Council and the Thai Government.

TABLE I

The ALP level in U212 transformants

Strain ^a	ALP activity ^b (u/g wheat-bran)	Inhibition by PMSF ^c (%)	Number of copies of the <i>alpA</i> gene ^d
U212	8143	80	1 (Control)
U721	17100	93	ND
U821	32816	99	3
U921	8224	90	1
U1421	18240	92	3
U1521	41529	99	12

^a The strain U212 is a high protease producing mutant obtained by UV mutagenesis (Kalayanamitr et al., 1987) which produces twice the ALP level of the wt strain. Cotransformation was achieved as described in section c with pOBT and pAP71. Six Ph^R transformants out of 40 were selected for their ability to produce more protease on skim milk plates. Five of these transformants (U721, U821, U921, U1421 and U1521) were then tested for their ability to produce ALP in solid state fermentation on wheat-bran (Kalayanamitr et al., 1987).

^b The wheat-bran extract was assayed for ALP activity according to Impoolsup et al. (1981). One unit of protease activity was defined as the amount of enzyme that liberated 1 µg of tyrosine in a reaction mixture per min at 40 °C.

^c The PMSF (0.1 mM) was used in the inhibition test of ALP activity.

^d The *alpA* gene copy number was estimated by slot blot hybridization. Serial dilutions of genomic DNA from each strain were blotted onto a nitrocellulose filter. Hybridization was conducted as described in the legend to Fig. 3. The intensity of the signal on the X-ray film was measured using a densitometer. The strain U212 was used as a control.

REFERENCES

- Anson, M.L.: Estimation of pepsin, trypsin, papain and cathepsin with haemoglobin. *J. Gen. Physiol.* 22 (1938) 79–89.
- Ballance, D.J.: Sequences important for gene expression in filamentous fungi. *Yeast* 2 (1986) 229–236.
- Cheevadhanarak, S., Saunders, G., Renno, D.V., Holt, G. and Flegel, T.W.: Transformation of *Aspergillus oryzae* with a dominant selectable marker. *J. Bio/Technology* 19 (1991) 117–122.
- Furuya, T., Ishige, M., Uchida, K. and Yoshino, H.: Koji mold breeding by protoplast fusion for soy sauce production. *J. Agric. Chem. Soc. Japan* 57 (1983) 1–8.
- Gurr, S.J., Unkles, S.E. and Kinghorn, J.R.: The structure and organization of nuclear genes of filamentous fungi. In: Kinghorn, J.R. (Ed.), *Gene Structure in Eukaryotic Microbes*. IRL Press, Oxford, 1987, pp. 93–139.
- Impoolsup, A., Bhumiratana, A., Flegel, T.W.: Isolation of alkaline and neutral proteases from *Aspergillus var. columnaris*, a soy sauce koji mold. *Appl. Environ. Microbiol.* 42 (1981) 619–628.
- Kalayanamitr, A., Bhumiratana, A., Flegel, T.W., Glinsukon, T. and Shinmyo, A.: Occurrence of toxicity among protease, amylase, and color mutants of a nontoxic soy sauce koji mold. *Appl. Environ. Microbiol.* 53 (1987) 1980–1982.
- Nakagawa, Y.: Alkaline proteinases from *Aspergillus*. *Methods Enzymol.* 19 (1970) 581–591.
- Picknett, T.M., Saunders, G., Ford, P. and Holt, G.: Development of a gene transfer system for *Penicillium chrysogenum*. *Curr. Genet.* 12 (1987) 449–455.

- Proudfoot, N.K. and Brownlee, G.G.: 3' Non-coding region sequences in eukaryotic messenger RNA. *Nature* 263 (1976) 211–214.
- Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74 (1977) 5463–5467.
- Skatrud, P.L., Tietz, A.J., Ingolia, T.D., Cantwell, C.A., Fisher, D.L., Chapman, J.L. and Queener, S.W.: Use of recombinant DNA to improve production of cephalosporin C by *Cephalosporium acremonium*. *Bio/Technology* 7 (1989) 477–485.
- Smith, D.J., Bull, J.H., Edwards, J. and Turner, G.: Amplification of the isopenicillin N synthetase gene in a strain of *Penicillium chrysogenum* producing high level of penicillin. *Mol. Gen. Genet.* 216 (1989) 492–497.
- Tatsumi, H., Ohsawa, M., Tsuji, R.F., Murakami, S., Nakano, E., Motai, H., Masaki, A., Ishida, Y., Murakami, K., Kawabe, H. and Arimura, H.: Cloning and sequencing of the alkaline protease cDNA from *Aspergillus oryzae*. *Agric. Biol. Chem.* 52 (1988) 1887–1888.
- Tatsumi, H., Ogawa, Y., Murakami, S., Ishida, Y., Murakami, K., Masaki, A., Kawabe, H., Arimura, H., Nakano, E. and Motai, H.: A full length cDNA clone for the alkaline protease from *Aspergillus oryzae*: structural analysis and expression in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 219 (1989) 33–38.