

Isolation and Characterization of the Alkaline Protease Gene of *Aspergillus oryzae*

Kohji MURAKAMI, Yutaka ISHIDA, Atsushi MASAKI, Hiroki TATSUMI,*
Seiji MURAKAMI,* Eiichi NAKANO,* Hiroshi MOTAI,*
Haruhide KAWABE and Hirofumi ARIMURA

Research and Development Division, The Green Cross Corporation, 2-25-1 Shodai Ohtani,
Hirakata-shi, Osaka 573, Japan

*Research and Development Division, Kikkoman Corporation,
399 Noda, Noda-shi, Chiba 278, Japan

Received May 9, 1991

The genomic DNA for the alkaline protease (Alp) of the fungus *Aspergillus oryzae* was isolated using synthetic oligonucleotides as hybridization probes, and the complete nucleotide sequence was identified. The Alp gene is 1374 nucleotides long and contains three introns, one of which is in the pro region and two in the mature coding region. Sequences related to the TATA box (TATAAAT) and the CAAT box (CCAAAT) were found in the 5'-noncoding region. Primer extension analysis showed that three transcriptional start points are present.

The filamentous fungus, *Aspergillus oryzae*, is a widely used microorganism in the industrial manufacture of soy sauce and sake. The alkaline protease (Alp) of *A. oryzae*¹⁾ is considered to be most important in producing the delicious taste of soy sauce by hydrolyzing raw materials. We previously reported the cloning and sequencing of the cDNA of the *A. oryzae* Alp.^{2,3)} The Alp cDNA encodes a putative prepro-enzyme composed of 282 amino acids of the mature Alp and an additional N-terminal sequence of 121 amino acid residues. When the entire Alp cDNA was expressed in the yeasts *Saccharomyces cerevisiae*³⁾ or *Zygosaccharomyces rouxii*,⁴⁾ the enzymatically active Alp was secreted into the culture medium, with the same N-terminal amino acid sequence as the native Alp. However, the yield of the Alp secreted by these yeasts was not sufficiently high.

Industrial strains of *A. oryzae* secrete large amounts of proteins, such as amylases and proteases. *A. oryzae*, therefore, offers promise as a host for extracellular protein production.⁵⁾ Recently, several genes of *A. oryzae* have been cloned and characterized,⁶⁻⁹⁾ and gene-me-

diated transformation systems⁸⁻¹¹⁾ for *A. oryzae* have been developed.

The design of an improved Alp by protein-engineering techniques and its efficient production by *A. oryzae* are expected to be useful in soy sauce production. The cloning and structural analysis of the chromosomal Alp gene is the first step toward the development of a production system of native and mutant Alp in *A. oryzae*. This paper describes the isolation and structural analysis of the Alp gene.

Materials and Methods

Strains. *A. oryzae* ATCC 20386 was used as the DNA donor. *Escherichia coli* HB101 [*SupE44*, *hsdS20* (*r_B*, *m_B*), *recA13*, *ara-14*, *proA2*, *acY1*, *galK2*, *rpsL20*, *xyl-5*, *mtl-1*] was used as the host for DNA manipulation.

Genomic DNA preparation. *A. oryzae* ATCC 20386 was cultured in YPS medium (1.0% yeast extract, 2.0% bacto-peptone, and 2.0% soluble starch) with vigorous shaking at 30°C for 48 hr, and collected by filtering through a few sheets of gauze. The collected mycelia were washed with extraction buffer (0.1 M Tris-HCl, pH 8.0, 10 mM EDTA, 1.0% sodium dodecyl sulfate), and mixed with the same weight of sterilized sea sand (Nacalai Tesque Co., Ltd.). The mixture was homogenized in a mortar for 10 min.

The homogenate was centrifuged, and the supernatant treated with RNase A (Sigma Co.) at a final concentration of 5 µg/ml for 20 min at 60°C. After addition of 1/6-volume of 3 M potassium acetate, the mixture was centrifuged (10,000 × *g* 10 min) and the supernatant was transferred to a new tube. The supernatant was extracted twice with TE-saturated phenol and then 3 times with ether. Two volumes of ethanol was slowly added so as not to disturb the interface and the chromosomal DNA precipitated at the interface was spooled onto a Pasteur pipette. The DNA was rinsed with 70% ethanol and dissolved in TE buffer.

DNA oligonucleotide. DNA oligonucleotide probes and primers were synthesized on a DNA synthesizer (Applied Biosystems model 380A) and labelled with γ -³²P ATP using T4 polynucleotide kinase (Takara Shuzo Co., Ltd.). The oligonucleotide 5'-GCGCAAGAACAACCTC-AAGTCGGAGGATAGA-3' probe complementary to the 5'-noncoding region of the Alp cDNA (Probe 1) and 5'-CATGTACAGAGTATACTTATGTTAGTAGTC-3' probe corresponding to the 3'-noncoding region (Probe 2) were synthesized as hybridization probes. The synthetic oligonucleotide 5'-CGCGGGAAGGATAGCTCC-3' complementary to positions from +51 to +34 was synthesized for the primer extension experiments.

Hybridization. Hybridization was done at 42°C in 6 × SSC (0.9 M NaCl, 0.09 M trisodium citrate), 5 × Denhardt's solution (0.1% Ficoll, 0.1% polyvinyl-pyrrolidone, 0.1% bovine serum albumin), 0.1% SDS, and 100 µg/ml transfer RNA. Washing was done with 6 × SSC and 0.5% SDS.

Nucleotide sequencing analysis. Various restriction fragments were subcloned into pUC19. The nucleotides were sequenced by the dideoxy method using an M13 sequencing kit (Takara Shuzo Co., Ltd.).

Primer extension analysis. *A. oryzae* mRNA was prepared as described by Tatsumi *et al.*³⁾ and the primer extension¹²⁾ was performed with reverse transcriptase (Takara Shuzo Co., Ltd.).

Results and Discussion

Cloning of the chromosomal Alp gene

The genomic DNA of *Aspergillus oryzae* ATCC 20386 was digested with several restriction enzymes and then analyzed by Southern hybridization to clone the Alp gene. When the entire Alp cDNA was used as a probe, many positive bands were observed (data not shown). The *A. oryzae* ATCC 20386 genome may contain several genes similar to the Alp gene. Therefore, we synthesized two oligonucleotides, one complementary to the 5'-noncoding region (Probe 1) and the other to the 3'-noncoding region (Probe 2, see Materials and Methods). These synthetic oligonucleotides were labelled with γ -³²P ATP and used as hybridization probes. Southern hybridiza-

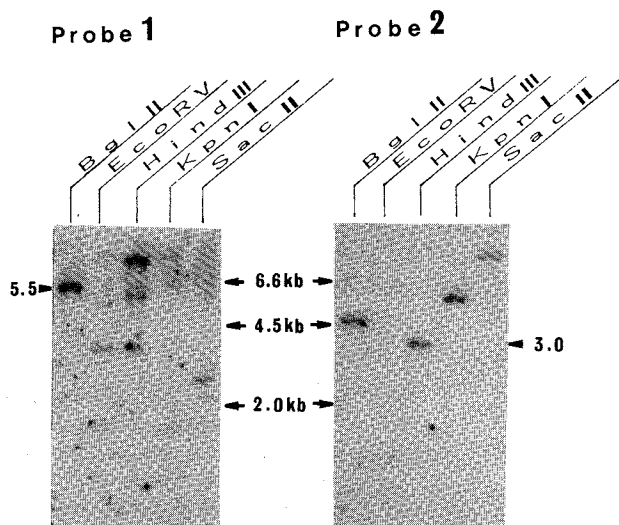


Fig. 1. Southern Blot Analysis of *A. oryzae* Genomic DNA.

Genomic DNA (5 µg) was digested with *Bgl*II, *Eco*RV, *Hind*III, *Kpn*I, or *Sac*II electrophoresed on an 0.8% agarose gel, transferred to a nitrocellulose filter, and then hybridized with probe 1 or probe 2.

tion using these probes against the digested genomic DNA of *A. oryzae* showed that probe 1 hybridized with a single 6.5-kb *Bgl*II fragment and probe 2 hybridized with a single 3.0-kb *Hind*III fragment (Fig. 1). Bands hybridized with both probes were not observed in this Southern blot analysis. The *Bgl*II site and the *Hind*III site are unique in the Alp cDNA sequence, the *Bgl*II site being downstream of the *Hind*III site. Since the 6.5-kb *Bgl*II fragment hybridized with the probe complementary to the 5'-noncoding region of the Alp cDNA and the 3.0-kb *Hind*III fragment hybridized with the probe complementary to the 3'-noncoding region, it was considered that the *Bgl*II fragment and the *Hind*III fragment would have an overlapping region and would thus cover the entire Alp chromosomal gene.

The 6.5-kb *Bgl*II fragment and the 3.0-kb *Hind*III fragment of the genomic DNA were purified by agarose gel electrophoresis, and inserted into *Bam*HI-digested pUC19 and *Hind*III -digested pUC19, respectively. *E. coli* HB101 cells were transformed with these plasmids. Each library was screened by colony hybridization with ^{32} P-labeled probes, and recombinant clones hybridizing with probe 1 or probe 2 were obtained. Plasmid DNA was

prepared from each clone and the restriction enzyme cleavage maps of the inserted fragments were examined (Fig. 2). The plasmids carrying the 6.5-kb *Bgl*II fragment hybridized with probe 1 and the 3.0-kb *Hind*III fragment hybridized with probe 2 were designated as pAP017 and pAP025, respectively (Fig. 2). The *Bgl*II fragment of pAP017 and the *Hind*III fragment of pAP025 shared a common 0.3-kb *Bgl*II-*Hind*III fragment. The nucleotides of this overlapping region (*Bgl*II-*Hind*III fragment) were sequenced. It was found that these sequences coincided with each other, and were identical to that of the *Bgl*II-*Hind*III fragment of the Alp cDNA. The *Bgl*II fragment of pAP017 was upstream of the overlap region, and the *Hind*III fragment of pAP025 was downstream of the overlap region. These results showed that these two overlapping fragments covered the entire Alp gene.

The nucleotide sequence of the Alp gene

We sequenced all of the nucleotides of the Alp gene. Various restriction fragments of pAP017 and pAP025 were subcloned into pUC19, and their nucleotides sequenced by the dideoxy method (Fig. 3). Comparison with the nucleotide sequence of Alp cDNA³⁾ showed that the sequence of the open reading frame of the Alp gene was in perfect agreement with that of Alp cDNA, and that the Alp gene contained three introns, one of which was in the pro region and two in the mature coding region. The size of all three introns is small (between 50 and 59 base pairs) relative to the introns of higher eukaryotes. The DNA sequence in these introns starts with GTNNGT and ends with PyAG. The PuCTPuAC sequence was observed near the 3' splice sites. The sequences in the introns of the Alp gene are similar to those of the genes of other filamentous fungi.^{13,14)}

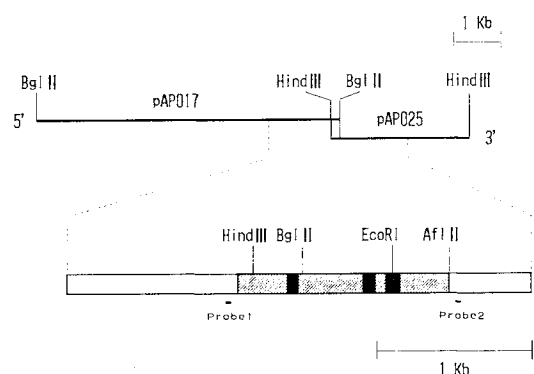


Fig. 2. Structure and Restriction Endonuclease Map of the Alkaline Protease Gene of *A. oryzae*.

The upper line represents the map of the *Bgl*II insert in pAP017 and of the *Hind*III insert in pAP025. The 5'- and 3'-noncoding region of the Alp gene are indicated by open boxes, the coding regions by striped boxes. The introns are indicated by solid boxes.

5'-Noncoding sequences

The transcription start point(s) of the Alp gene were identified by primer extension analysis (Fig. 4). Three possible transcription start points were found at -63 bp, -66 bp, and -69 bp, upstream of the translation start

tcagcaccacatcaaaaccccttgatccctgataaaaagcgtggaaaaagcgatcgcgtgaggatgccaaaggggaatgggtcacattgatct 625
 ctgtcgccgctgttaggatgatcttccactcctaaggcatcgccggcgattaggcccttccgtccaaagatcgggttactcctctcattatggcgag -526
 ctactttgtgaattaatgactgaggagataaccaccttcccttgaaggtagccgacactaccttgagcgttagttactttttcaggaaagacatcc -427
 tatgtctagctctcgcaatcactgcagcgtcgacaacttgcctagccttggcttccacggtctatcggaacaccggttcatgacigaaaggggtca -328
 ggcctcggtggtgtcaacatcatctcatcttccatcatgcccgcgtgattgatagagtaatttccggtggagcacaacgccgtccctcgatgcaatg -229

 tcaccctglaagtttacaclacaatcgtgtaglacagagcatccttgcattgcatgctglaagtgatccaaatccgtagaacttgctcgagaaacagg -130
 gaatatagaaactcctgaagggt[ataaaa]accacatgcatccctcgctccatccctcacttccatcaagccagcggttctatcctccgacttgagtt -31
 +1
 gttcttgcgcactctttacaactcttctcatcATGCAGTCCATCAAGCGTACCTTGCTCCTCCTCGGAGCTATCCTTCCCGCGGTCTCGGTGCCCTGTG 69
 M Q S I K R T L L L L G A I L P A V L G A P V
 CAGGAACCCGCCGGCCGCTGAGAAGCTTCTGGAAAGTACATTGTACACATTCAGCCCGGCATTGACGAGGCAAGATTGAGGAGCATACCCTGG 168
 Q E T R R A A E K L P G K Y I V T F K P G I D E A K I Q E H T T W
 GCTACCAACATTACCAGCGCAGTCTGGAGCGTCTGGCGCACTGGCGGTGATCTTCTGTGGTATTGAGCGCACTACAAGATCAACAAGTTCGCC 267
 A T N I H Q R S L E R R G A T G G D L P V G I E R N Y K I N K F A
 GCCTATGCAAGCTCTTTCGACGATGCTACCATGAGGAGATTGCCAAGAACGAGATgttllgtggtcatccgtcgcaatttgaatgacagctaactc 366
 A Y A G S F D D A T I E E I R K N E D Intron 1
 ggcggcagGTTGCTACGTCGAGGAGGACAGATCTACTCCTCGATGGCCTGACTACCCAGAAGAGTGCCCTGGGGTCTGGGCGACGATTCCCA 465
 V A Y V E E D Q I Y Y L D G L T T Q K S A P W G L G S I S H K
 AGGGCCAGCAGAGCACCAGTACATCTACGACACTAGTCCGGCGAGGGCACTATGCCTACGTTGGTGGATAGCGGTGTCAATGTGACCATGAGGAGT 564
 G Q Q S T D Y I Y D T S A G E G T Y A Y V V D S G V N V D H E E F
 TCGAGGCCCCCGCCAGCAAGGCTACAACGCTGCCGCTGGTCAAGATGTGGACAGCATTGGCCATGGCACCCACGTTCCGGCACCATTGCTGGCAAGA 663
 E G R A S K A Y N A A G G Q H V D S I G H G T H V S G T I A G K T
 CTTATGGTATCGCCAAGAAGGCCAGCATCCTTTCGGTCAAAGTTTTCAGGGTGAATCGAGCAGCACTTCCGTCTTCTGACGGCTTCAACTGGGCTG 762
 Y G I A K K A S I L S V K V F Q G E S S S T S V I L D G F N W A A
 CCAACGACATGTTAGCAAGAAGCGTACCAGCAAGGCTGCAATCAACATGAGCTTGGgtgagtttacattgttcttctacttgaagcgcgcagcgcc 861
 N D I V S K K R T S K A A I N M S I G Intron 2
 taatttcaaaaacacagCGGCTGGCTACTCTAAGGCTTCAACGATGCGGTGAGAACGCATTGAGCAGGGTGTTCTCTCGGTTGTCGCTGCCGTA 960
 G G Y S K A F N D A V E N A F E Q G V L S V V A A G N
 CGAGAAGctacgtctccccctccatcgccgaagacgaattcgtaactgacttgattttcttagTCTGATGCCGCCCAACACGCCCTGCTCTGCCCT 1059
 E N Intron 3 S D A G Q T S P A S A P
 GATGCCATCACTGTTGCCGCTATCCAGAAGAGCAACAACCGCCAGTTTCTCCAACCTTGGCAAGGTCGTTGACGTTCTTCTGCTCCCGGTCAAGATATC 1158
 D A I T V A A I Q K S N N R A S F S N F G K V V D V F A P G Q D I
 CTTTCTGCTGGATTGGCTCTTCTCTGCAACCAACACCATCTCTGGTACCTCCATGGCTACTCCCAACATTGTCGGCCTGCTCTACTCTGCTGCC 1257
 L S A W I G S S S A T N T I S G T S M A T P H I V G L S L Y L A A
 CTTGAGAACCTCGATGCCCCGCTGCCGTGACCAAGCGCATCAAGGAGTTGGCCACCAAGGACGTCGTAAGGATGTTAAGGCAGCCCTAACCTGCTT 1356
 L E N L D G P A A V T K R I K E L A T K D V V K D V K G S P N L L
 GCCTACAACGGTAACGCTtaagtaccaggagtagctcgaggaatttaccattgttacttggaalacaatgatgattagaaaacgaagagcgltatgatt 1455
 A Y N G N A Stop
 cggagcgatlatgcatggcaccatcacagcgtgatacatagggctgttctcagaatlaggattttatctgaatccatgtacagagatatacttattg 1554
 tagtagtcaataaatacttggcttctcaatttctccgactcacaaggggtcgctgatcacagaacgaactagatgtcaggggacgatgaicaccgt 1653
 tcttagcaagacctagtagttgtcgaccatagcctttagcgcgaatcagacctactacttctagattgcagaccaagtcgcatgacaatgtcctct 1752
 ttggattaggattagtagttgallagatccggaataatagggctggcttccaaactccgggagtgccgatttactgtacttattgaagaa 1851
 agtaagtctattgttacacagctgcag 1878

Fig. 3. Nucleotide Sequence of the Alkaline Protease Gene of *A. oryzae* and Its Deduced Amino Acid Sequence.

Coding regions are indicated in upper-case letters, all other sequences in lower-case letters. Below the coding regions, the predicted amino acid sequence is given using the standard one-letter code. Nucleotides are numbered with reference to the translation start codon (+1). The three major transcription start points are indicated by asterisks. The putative TATA box is indicated by the open box, the putative CAAT box by the dotted box. The consensus sequence for fungal introns,¹³⁾ GTPuNGT---PuCTPuAC---PyAG, is underlined. Positions corresponding to the hybridization probes and the primer for the analysis of the transcription start point(s) region are indicated by double underlining and by dashed underlining, respectively. The putative polyadenylation signal is indicated by dotted underlining.

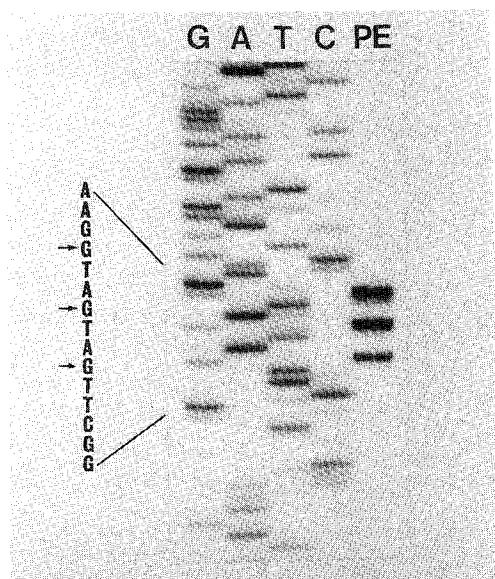


Fig. 4. Primer Extension Analysis to Identify the Transcription Start Points of the Alkaline Protease Gene.

Lane PE shows the extension products of poly(A)⁺ RNA from *A. oryzae* ATCC 20386 using the oligonucleotide primer (see Materials and Methods). Lanes G, A, T, C shows the sequencing ladders of pAP017 DNA by the dideoxy chain termination method using the same primer. The arrows indicate the positions of the major transcription start points.

codon (ATG) (Fig. 3). A sequence related to the TATA box¹⁵⁾ (TATAAAT) at nucleotides -102 to -108, and a sequence related to the CAAT box¹⁵⁾ (CCAAAT) at nucleotides -154 to -159 were found (Fig. 3). A number of filamentous fungal genes have isolated, and the 5'-noncoding regions of these genes are in an equivocal situation.¹⁶⁻²⁰⁾ Few fungal genes have sequences related to the TATA box or the CAAT box. It is interesting that the Alp gene has homology to the consensus sequences in the 5'-noncoding region of higher eukaryotes. A prominent C+T-rich region of about 30 bp was found between the TATA box and the transcription start points (Fig. 3). A similar feature is present in *A. nidulans* and *N. crassa* genes,^{13,19)} being particularly noticeable in highly expressed genes.

Acknowledgments. This work was supported by the Japanese R & D Association for the Improvement of Enzyme Function for the Food Industry.

References

- 1) T. Nakadaï, S. Nasuno and N. Iguchi, *Agric. Biol. Chem.*, **37**, 2685 (1973).
- 2) H. Tatsumi, H. Ohsawa, R. F. Tsuji, S. Murakami, E. Nakano, H. Motai, A. Masaki, Y. Ishida, K. Murakami, H. Kawabe and H. Arimura, *Agric. Biol. Chem.*, **52**, 1887 (1988).
- 3) H. Tatsumi, Y. Ogawa, S. Murakami, Y. Ishida, K. Murakami, A. Masaki, H. Kawabe, H. Arimura, E. Nakano and H. Motai, *Mol. Gen. Genet.*, **219**, 33 (1989).
- 4) Y. Ogawa, H. Tatsumi, S. Murakami, Y. Ishida, K. Murakami, A. Masaki, H. Kawabe, H. Arimura, E. Nakano, H. Motai and A. Toh-e, *Agric. Biol. Chem.*, **54**, 2521 (1990).
- 5) G. Saunders, T. M. Picknett, M. F. Tuite and M. Ward, *Trends Biotechnol.*, **7**, 283 (1989).
- 6) S. Tada, Y. Iimura, K. Gomi, K. Takahashi, S. Hara and K. Yoshizawa, *Agric. Biol. Chem.*, **53**, 593 (1989).
- 7) M. J. Gines, M. J. Dove and V. L. Seligy, *Gene*, **79**, 107 (1989).
- 8) Y. M. J. T. Ruiter-Jacobs, M. Broekhuijsen, S. E. Unkles, E. I. Campbell, J. R. Kinghorn, R. Contreras, P. H. Puuwels and C. A. M. J. J. van den Hondel, *Curr. Genet.*, **16**, 159 (1989).
- 9) S. E. Unkles, E. I. Campbell, Y. M. J. T. Ruiter-Jacobs, M. Broekhuijsen, J. A. Macro, D. Carrez, R. Contreras, C. A. M. J. J. van den Hondel and J. R. Kinghorn, *Mol. Gen. Genet.*, **218**, 99 (1989).
- 10) K. Gomi, Y. Iimura and S. Hara, *Agric. Biol. Chem.*, **51**, 2549 (1987).
- 11) T. Christensen, H. Woeldike, E. Boel, S. B. Mortensen, K. Hjortshoej, L. Thim and M. T. Hansen, *Bio/Technology*, **6**, 1419 (1988).
- 12) T. Maniatis, E. Fritsch and J. Sambrook, "Molecular Cloning," Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1982.
- 13) D. J. Ballance, *Yeast*, **2**, 229 (1986).
- 14) J. Rambosek and J. Leach, *CRC Critical Reviews in Biotechnology*, **6**, 357 (1987).
- 15) R. Nussinov, J. Owens and J. V. Maizel, Jr., *Biochim. Biophys. Acta*, **866**, 109 (1986).
- 16) E. Boel, M. T. Hansen, I. Hjort, I. Hoegh and N. P. Fiil, *EMBO J.*, **3**, 1581 (1984).
- 17) I. Shibuya, K. Gomi, Y. Iimura, K. Takahashi, G. Tamura and S. Hara, *Agric. Biol. Chem.*, **54**, 1905 (1990).
- 18) E. J. Mullaney, J. E. Hamer, K. A. Roberti, M. M. Yelton and W. E. Timberlake, *Mol. Gen. Genet.*, **199**, 37 (1985).
- 19) P. J. Punt, M. A. Dingemanse, A. Kuyvenhoven, R. D. M. Soede, P. H. Pouwels and C. A. M. J. J. van den Hondel, *Gene*, **93**, 101 (1990).
- 20) J. M. Clements and C. R. Roberts, *Gene*, **44**, 97 (1986).