

Short Communication

Cloning and Sequencing of the
Alkaline Protease cDNA from
Aspergillus oryzae

Hiroki TATSUMI, Manabu OHSAWA,
Ryohei F. TSUJI, Seiji MURAKAMI,
Eiichi NAKANO, Hiroshi MOTAI,
Atsushi MASAKI,* Yutaka ISHIDA,*
Kohji MURAKAMI,* Haruhide KAWABE*
and Hirofumi ARIMURA*

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

*Research Division, The Green Cross Corporation,
2-1180-1 Shodai Ohtani, Hirakata-shi,
Osaka 573, Japan

Received April 14, 1988

The alkaline protease of *Aspergillus oryzae* is one of the most important enzymes in the manufacturing process for soy sauce. In order to change some properties of this enzyme by means of protein engineering and to apply mutant enzymes in food industries, first of all, it is necessary to determine the primary structure of the enzyme. In this paper, we report the cloning and sequencing of the mature region of the alkaline protease cDNA from *Aspergillus oryzae* (ATCC20386).

Total RNA was isolated from *Aspergillus oryzae* mycelia by the guanidium isothiocyanate procedure and the alkaline protease mRNA was partially purified as described by Tatsumi *et al.*¹⁾ cDNA was synthesized using an Amersham cDNA synthesis kit. After the addition of *Eco*RI linkers to both ends, the cDNA was inserted into the *Eco*RI site of the vector, pUC119 (Takara Shuzo), and then transformed into *E. coli*, DH1.

Eighty clones harboring inserted cDNA of more than 300 bp long were subjected to hybridization selection.²⁾ One clone (Fig. 1, lane 1) was found to be positive on this selection and the plasmid of this clone was designated as

pOAP3. On restriction enzyme analysis, we found that the cDNA cloned in pOAP3 was 750 bp long. As the size of the alkaline protease mRNA is about 1,700 bp, as determined on Northern blot analysis (data not shown), we consider that the 750 bp cDNA fragment is not enough to code for the mature alkaline protease. By means of the colony hybridization procedure with the 750 bp cDNA fragment as a probe, we have isolated plasmid pOAP5, which has a 1,100 bp cDNA insert, and sequenced this 1,100 bp fragment by the dideoxy chain termination method.

Figure 2 shows the nucleotide sequence of the cDNA, which corresponds to the coding region of the mature alkaline protease and the 3'-noncoding region. As the N-terminal and internal amino acid sequences of the alkaline protease coincide with the largest open reading frame of this cDNA fragment, it became clear that the alkaline protease cDNA had been cloned. From these nucleotide sequence data, it is deduced that the mature alkaline protease from *Aspergillus oryzae* consists of 282 amino acids and has a molecular weight of 29,010. The molecular weight calculated here is about 30% larger than that estimated by the gel filtration method.³⁾

As shown in Fig. 2, the 3'-noncoding region

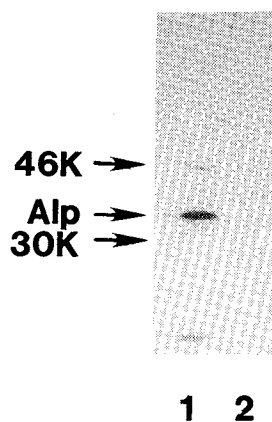


FIG. 1. Fluorography after Hybridization Selection.

Lane 1, the positive clone harboring pOAP3; lane 2, a negative clone. Alp indicates the position of the alkaline protease produced on *in vitro* translation. The molecular weight markers are [¹⁴C] methylated ovalbumin (46K) and carbonic anhydrase (30K).

```

GGCGTACTACCCAGAAGACTGCCCGCTGGGGTCTGGGAGCATTCCCAAGGGCCAGAGCAGCAGCTACATCTACGACACTAGT 90
G L T T Q K S A P W G L G S I S H K G Q Q S T D Y I Y D T S
GCCGCCAGGGCACCTATGCTACGTGGTGGATAGCGGTCAATGTCCAGCCATGAGGAGTTCCAGGGCCGCCAGCAAGGCTACAA 180
A G E G T Y A Y V V D S G V N V D H E E F E G R A S K A Y N
GCTGCCGGTGCTCAGCATGTGACAGCATTGGCCATGGCACCACGTTTCGGGACCATTTGCTGGCAAGACTTATGGTATCGGCAAGA 270
A A G G Q H V D S I G H G T H V S G T I A G K T Y G I A K K
GCCAGCATCTTTTGGTCAAAGTTTTCAGGGTGAATCGAGCAGCACTTCGCTCATTCTTGACGGCTTCACTGGGCTGCCAAGCAGCAT 360
A S I L S V K V F Q G E S S S T S L I L D G F N W A A N D I
GTTAGCAAGAAGCTACCAAGAGCTGCAATCAACATGAGCTTGGGCGGTGGCTACTCTAAGGCTTCAACGATGGCGTCGGAACCGCA 450
V S K K R T S K A A I N M S L G G G Y S K A F N D A V E N A
TTCGAGCAGGGTGTCTCTCGGTTGTCGCTGCGGTAAACGAGAACTCTGATGCCGCCAAACAGCCCTGCCTCTGCCCTGATGCCATC 540
F E Q G V L S V V A A G N E N S D A G Q T S P A S A P D A I
ACTGTGGCGCTATCCAGAAAGAGCAACAACCGCGCCAGTTTCTCAACTTTGGCAAGGTCGTTGACGCTCTTCGCTCCCGGTCAAGATATC 630
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
CTTCTGCGCTGGATTGCTCTTCTGTCGCCACCAACCATCTCTGTTACCTCCATGGCTACTCCCAACTTGTGGCGCTGTCCCTCTAC 720
L S A W I G S S A T N T I S G T S M A T P H I V A L S L Y
CTCGTCCCTTGAGAACCTCGATGCCCGCGTGGCTGACCAAGCGCATCAAGGAGTTGGCCACCAAGGAGCGTGTCAAGGATGTTAAG 810
L A A 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
GGCAGCCCTAACCTGCTTGCCTACAACGGTAACGCTTAAGTACCAGGAGTACGTCGCAGGATTCTACCATTTGTACTGGAATACAATGAT 900
G S F N L L A Y N G N A
GATTAGAAAACGAAGCGTTATGATTGCGACGATATATGCATGGCACCATACAGCGTGATACATAGGCTGTTTGCTCAAGAATTAGG 990
ATTTTACTCTGAATCCATGTACAGAGTATACTTATGTTAGTAGTCAATAAAATCTTGGCTTTCTAAAAA

```

FIG. 2. Nucleotide Sequence of the Alkaline Protease cDNA of *A. oryzae*

The nucleotide sequence which codes for the mature alkaline protease and its 3'-noncoding region is shown. The amino acid sequence deduced from the nucleotide sequence of the largest open reading frame is also shown. The thick bar indicates the amino acid sequence which is coincident with the N-terminal amino acid sequence determined on Edman degradation of the purified alkaline protease. Thin bars indicate the amino acid sequences of the alkaline protease fragments obtained on lysylendopeptidase digestion or autolysis.

consists of 226 bp including a 19 bp long poly(A) tail. A consensus polyadenylation signal, AATAAA, exists 19 bp upstream from the poly(A) tail. The cDNA sequenced in this study has only the about 30 bp upstream from the N-terminal codon and seems to lack most of the sequence for the pre and pro region. We have already obtained some larger cDNA fragments of the alkaline protease, and at present we are determining its sequence.

It is known that the alkaline protease from *Aspergillus oryzae* is a serine protease.⁴⁾ We have compared the primary structures of various serine proteases from microorganisms using the Micro Genie program (Beckman). The alkaline protease of *Aspergillus oryzae* shows 43.9%, 41.2%, 43.3% and 29.0% homology with the protease B from *Saccharomyces cerevisiae*,⁵⁾ the alkaline extracellular protease from *Yarrowia lipolytica*,⁶⁾ the proteinase K from *Tritirachium album* Limber⁷⁾ and the subtilisin from *Bacillus amyloliquefaciens*,⁸⁾ respectively. The three amino acid residues (Asp 32, His 64 and Ser 221 of the subtilisin) which compose the active center

are conserved and amino acid sequences encompassing these residues are highly homologous among these five alkaline proteases. The details will be given elsewhere.

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

REFERENCES

- 1) H. Tatsumi, T. Masuda and E. Nakano, *Agric. Biol. Chem.*, **52**, 1123 (1988).
- 2) T. Maniatis, E. F. Fritsch and J. Sambrook, "Molecular Cloning," Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y., 1982, pp. 329~352.
- 3) T. Nakadai, S. Nasuno and N. Iguchi, *Agric. Biol. Chem.*, **37**, 2685 (1973).
- 4) Y. Nakagawa, *Meth. Enzymol.*, **19**, 581 (1970).
- 5) C. M. Mochile, R. Tizard, S. K. Lemmon, J. Smart and E. W. Jones, *Mol. Cell. Biol.*, **7**, 4390 (1987).
- 6) L. S. Davidow, M. M. O'Donnell, F. S. Kaczmarek, D. A. Pereira, J. R. DeZeeuw and A. E. Franke, *J. Bacteriol.*, **169**, 4621 (1987).
- 7) K. -D. Jany, G. Lederer and B. Mayer, *FEBS Lett.*, **199**, 139 (1986).
- 8) J. A. Wells, E. Ferrari, D. J. Henner, D. A. Estell and E. Y. Chen, *Nucleic Acids Res.*, **11**, 7911 (1983).