Note

Molecular Cloning of a cDNA for Proctase B from Aspergillus niger var. macrosporus and Sequence Comparison with Other Aspergillopepsin I

Jing-Fang Lu, Hideshi Inoue, Takao Kimura, Osamu Makabe, and Kenji Takahashi

Department of Biophysics and Biochemistry, Faculty of Science, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Pharmaceutical Research Center, Meiji Seika Kaisha, Ltd., Kohoku-ku, Yokohama 222, Japan

Received December 19, 1994

A cDNA for proctase B from Aspergillus niger var. macrosporus was isolated and sequenced. The deduced amino acid sequence (394 residues) of the preproform of the enzyme was highly homologous (98% identity) with those of aspergillopepsins I from A. awamori and A. saitoi, and moderately homologous (68% identity) with that of A. oryzae. Most of the sequence differences were found in the carboxyl-terminal domain.

Aspergillus niger var. macrosporus excretes two types of acid endopeptidases. One is proctase B (aspergillopepsin I, EC 3.4.23.18) and the minor one is proctase A (aspergillopepsin II, EC 3.4.23.19). They have been used as a digest, and has been also used to solubilize insoluble collagen. We have determined the complete amino acid sequence of proctase A and isolated its genomic DNA and cDNA. Proctase A is not homologous with any of the ordinary aspartic proteinases in primary structure and is insensitive to their specific inhibitors. Therefore, it is called a non-Asp proteinase. On the other hand, proctase B has properties of peptidase-type aspartic proteinases; it has a molecular mass of about 35 kDa and is susceptible to specific inhibitors of proteinases such as pepstatin A, diazoacetyl-b-norleucine methyl ester in the presence of cupric ions, and 1,2-epoxy-3-(p-nitrophenoxyl)propene. The primary structure of proctase B, however, remained unknown. In this study, we isolated the cDNA coding for proctase B and deduced its primary structure as a basis for further studies on the structure/function relationships of aspergillopepsins I.

The N-terminus of proctase B (Seikagaku Kogyo Co.) and three of the tryptic fragments purified by HPLC were sequenced. One of the fragments contained a consensus sequence of the active site of aspartic proteinases, Asp-Thr-Gly-Ser. The locations of the tryptic fragments in proctase B were presumed based on the homology with penicillopepsin (EC 3.4.23.20), since the complete amino acid sequence of A. awamori pepsin, deduced from the nucleotide sequences of the genomic DNAs, had not been published before this study. The primers for polymerase chain reaction were designed based upon the partial amino acid sequences obtained above (Fig. 1, double-underlines), and the DNA probe used in cDNA cloning of proctase B was specifically amplified. The amplified DNA was approximately 440 bp long, which was consistent with the predicted length from the homology with penicillopepsin. A cDNA library was constructed using a cDNA cloning system (Agt10 Amersham Corp.) and RNA extracted from A. niger var. macrosporus. Using the amplified DNA as a probe, five positive clones out of 14,000 were obtained by screening of the cDNA library and the longest cDNA clone was sequenced (Fig. 1).

The cDNA encoded 394 amino acid residues containing all of the amino acid sequences identified at the protein level. The deduced amino acid sequence of preproctase B is very similar to that of the preproform of A. awamori pepsin, with only six residue differences in the enzyme part (Fig. 2). Preproctase B was also highly homologous with A. saitoi pepsin with only one-, two-, and seven residue differences in the signal sequences, prosequences and enzyme parts between them (Fig. 2). Further, the deduced amino acid sequence of mature proctase B was 72% and 69% identical with A. oryzae pepsin and penicillopepsin, respectively (Fig. 2), and the active site sequences including the

* The nucleotide sequence reported in this paper has been submitted to the GenBank/EMBL Data Bank with the accession number D45177.

* To whom correspondence should be addressed.

Fig. 1. The Nucleotide Sequence of the cDNA Coding for Proctase B and the Deduced Amino Acid Sequence.

Both the sense and antisense strands were completely sequenced. The amino acid residues identified by Edman degradation of proctase B and its three triptic peptides are underlined. The two active site aspartyl residues are represented by D*.

The amino acid sequences of residues 1-20 and 21-69 are the putative signal sequences and prosequences, respectively. The sequences used for design of the primers for polymerase chain reaction are shown by double-underlines.
two-catalytic aspartic acid residues and the locations of two cytistine residues, which will form a disulfide bond, are completely conserved among these enzymes. On the other hand, procat B has less similarity with rhizopuspepsin (34%) \(^{12}\) and human pepsin (27%). \(^{13}\)

The nucleotide sequence of the coding region of the procat B cDNA also shared high identity with that of the genes for aspergillopепsin I from _A. awamori_ (98%) and _A. saitoi_ (95%). Moreover, the upstream and the downstream untranslated sequences of the cDNA for procat B were completely and 93%, respectively, identical with the corresponding sequences in the genomic DNA of _A. awamori_ pепsin. \(^{31}\) The upstream non-coding sequence in the cDNA for procat B was also identical with that in the genomic DNA coding for _A. saitoi_ pепsin, \(^{10}\) although their downstream non-coding sequences were not so homologous with each other. On the other hand, the pre-pro-form of _A. oryzae_ pепsin was fairly different in amino acid sequence from those of the above three aspergillopепsins; there were 9, 28, and 91 residue differences in the signal sequences, prosegments and enzyme parts, respectively, as compared with the sequence of preprocat B pепsin. The nucleotide sequence of the coding region of the _A. oryzae_ pепsin gene was 69% identical with that of the procat B cDNA. Moreover, neither the upstream nor downstream non-coding sequence in the cDNA for procat B was homologous with those in the genomic DNA for _A. oryzae_ pепsin. \(^{13}\) These results are consistent with the fact that _A. niger_ var. _macroasperus_, _A. awamori_, and _A. saitoi_ are genetically very close to one another, while _A. oryzae_ is more distant from the other three. The codon usage, however, was similar among these aspergillopепsins I.

Since the three-dimensional structures of aspergillopепsins I are considered to be similar to that of penicillopепsin, five out of the six and six out of the seven differences in amino acids between procat B and _A. awamori_ pепsin and between procat B and _A. saitoi_ pепsin, respectively, are presumed to be in the β-sheet V or in the loops between β-sheets III and V of the carboxyl-terminal domain \(^{14}\) (Fig. 2). Further, there are ten residue differences between _A. awamori_ pепsin and _A. saitoi_ pепсин. Eight out of the ten differences are also presumed to be in the β-sheet V or in a loop between β-sheets III and V. Therefore, some differences in their properties \(^{30}\) seem to be partly attributed to the structural differences in or around the β-sheet V.

**References**


