Cloning of Genes of the Aminopeptidase T Family from *Thermus thermophilus* HB8 and *Bacillus stearothermophilus* NCIB8924: Apparent Similarity to the Leucyl Aminopeptidase Family

Hidemasa MOTOSHIMA,† Etsuo MINAGAWA, Fuji TSUKASAKI, and Shuichi KAMINOGAWA*
Research Center, Yotsuba Milk Products Co., Ltd., 465–1, Watsus, Kitahiroshima, Hokkaido 061–12, Japan
*Department of Applied Biochemistry, The University of Tokyo, 1–1–1 Yayoi, Bunkyo-ku, Tokyo 113, Japan
Received April 21, 1997

To obtain genes with sequence similarity to aminopeptidase T (AP-T) of *Thermus aquaticus* YT-1, we cloned the genes encoding aminopeptidase Th (AP-Th) from *Thermus thermophilus* HB8 and aminopeptidase II (APII) from *Bacillus stearothermophilus* NCIB8924. The AP-Th gene encoded a polypeptide of 408 amino acid residues and the deduced molecular weight of this subunit was 45,015. The APII gene encoded a polypeptide of 413 amino acid residues with a deduced molecular weight of 46,207. The extent of amino acid sequence similarity between AP-Th and AP-T was 86%, and that between APII and AP-T was 43%. The substrate specificities of these expressed enzymes were similar, and each efficiently hydrolyzed leucyl- or phenyl-peptide substrates. Since the deduced amino acid sequence of these enzymes show no similarity to other known aminopeptidases, they appear to comprise an independent family of peptidases, designated the AP-T family. However, a conserved region within the enzymes of the AP-T family shows similarity to the active site signature of the leucyl aminopeptidase family, suggesting that these enzymes may belong to the leucyl aminopeptidase superfamily.

**Key words:** aminopeptidase; *Bacillus stearothermophilus*; *Thermus*; aminopeptidase T family

An aminopeptidase is an exopeptidase which releases amino acids from the N-terminus of a peptide, and such enzymes are ubiquitous in nature. In cooperation with endopeptidases, aminopeptidases function to hydrolyze proteins or peptides completely to amino acids. This type of enzyme is important for turnover of amino acids and for regulation of peptide hormones in biological systems.1) We have previously purified and characterized a heat-stable aminopeptidase from an extremely thermophilic bacterium, *Thermus aquaticus* YT-1,2) and we named this enzyme aminopeptidase T (AP-T).3) We also cloned the AP-T gene and discovered its entire sequence.4) It is a homodimeric enzyme with a subunit molecular weight of 44,820, it is metallo-dependent (activated by cobalt or magnesium ions), and its optimal temperature for activity is above 70°C. It releases amino acids from the N-terminus of peptides with broad substrate specificity, and especially from leucyl and phenyl peptides.

We have found that AP-T contains a region with similarity (43%) to the N-terminal amino acid sequence (15 a.a.) of aminopeptidase II (APII) from *Bacillus stearothermophilus* NCIB8924.4,5) This strain is known to produce at least three different types of aminopeptidases: API, APII, and APIII, and has been studied extensively by Zubler et al.6) API is a heat-stable membrane-associated enzyme with a high molecular weight (400,000), and it consists of 12 subunits of two different types (α and β, with an identical molecular weight of 36,500 ± 4000).7) APII and APIII are more heat-labile, low molecular weight enzymes located in the cytoplasm and each is composed of two identical subunits of molecular weight 46,000 and 47,500, respectively.8) Many characteristics other than the heat stability of APII are similar to AP-T. The gene for APII has not been cloned yet, so the entire sequences of these enzymes could not be compared and the similarity to other known aminopeptidases is unclear.

The purpose of this experiment was to clone genes similar to the AP-T gene in an effort to understand the relationships between these aminopeptidases and other known aminopeptidases.

First, we cloned a gene similar to the AP-T gene from *Thermus thermophilus* HB89), which, as in the case of the YT-1 strain, has been extensively studied in molecular biology. We named the protein product of this gene in strain HB8 aminopeptidase Th (AP-Th) to distinguish it from AP-T. Secondly, we cloned the APII gene from *B. stearothermophilus* strain NCIB8924. We compared the sequences of AP-Th and APII with AP-T, consequently these aminopeptidases were found to comprise a family (AP-T family). Furthermore, we compared the conserved regions within these enzymes of the AP-T family with reference to the PROSITE data base, and a region with similarity to the active site signature of the leucyl (cytosol) aminopeptidase family was identified.

Here, we also compare the flanking regions upstream and downstream of the AP-T family genes.

**Materials and Methods**

**Chemicals and enzymes.** Restriction enzymes were obtained from Boehringer Mannheim Biochemicals. T4 DNA polymerase and T4 DNA ligase were obtained from Takara Syuzo Co., Ltd. (Japan). Enzymes used in the PCR reaction were obtained from Perkin-Elmer Cetus. An ECL direct DNA hybridization kit used in Southern hybridization was...
obtained from Amersham Corp. (Japan). Oligonucleotides used as primers or probes were synthesized using an Applied Biosystems DNA Synthesizer. Fluorescent peptide substrates for aminopeptidase, i.e., t-leucine 4-methylcoumarin-7-amide (Leu-MCA), Phe-MCA, Arg-MCA, Met-MCA, Ala-MCA, Lys-MCA, L-prolylglutamic acid 4-methylcoumaryl-7-amide (PyMCA), Lys-Ala-MCA, Gly-Pro-MCA, Pro-Phe-Arg-MCA, were obtained from the Peptide Institute, Inc. (Osaka, Japan).

Bacterial strains and plasmids. *T. thermophilus* HB8 was obtained as *T. aquatius* NCIB11244 (Oshima HB8, ATCC27634 strain) from The National Collections of Industrial and Marine Bacteria Limited (NCIMB) (Scotland, United Kingdom). *R. stearothermophilus* NCIB8924 was also obtained from NCIMB, *Escherichia coli* MV1184 (Pharmacia, Upsalla, Sweden) was used as a host for the vectors pUC118 and pUC119 (Pharmacia) and the expression vector pEXP7.120 *E. coli* XLI-Blue and lambda ZAPII vectors were from Stratagene (La Jolla, Calif., U.S.A.). Plasmids pAT15A and pAT15G containing the AP-T gene, and the AP-T expression plasmid pKL4 were previously reported.49

Media and cultivation. *T. thermophilus* HB8 was cultivated aerobically (shaking in a water bath) at 75°C for 2 days in TY salts medium (NCIB medium no. 161). *B. stearothermophilus* NCIB8924 was cultured in broth (BBL Polypeptone 5.0 g, BBL Bact extract 3.0 g, NaCl 0.8 g/liter, pH 7.3) aerobiocally at 55°C overnight. *E. coli* strains were usually grown aerobiocically in Luria broth containing ampicillin (50 μg/ml) at 37°C.

DNA manipulation. The techniques used to manipulate plasmid DNA were essentially those of Sambrook et al.11 Genomic DNA of *T. thermophilus* HB8 and *B. stearothermophilus* NCIB8924 were prepared by the modified method of Saito and Miura.12 The restriction endonuclease-digested fragments were separated in a 0.8% agarose gel and transferred to a nylon membrane (Hybond N+, Amerham, U.S.) using a VacuGene vacuum blotting system (Miles, 2016, Pharmacia, LKB) in 20 × SSC (1 × SSC is 0.15 M NaCl plus 0.015 M sodium citrate) according to the manufacturer's protocol. Genomic hybridization was done by the protocol of an ECL direct nucleic acid labeling and detection system (Amerham Corp., In vitro packaging was done using GigapackIII gold (Clontec Laboratories, Inc., U.S.A.).

Conditions of the polymerase chain reaction (PCR). PCR was done as described by Saki13 using a Hoei thermal cycler Model TC-100 (Hoei Science, Tokyo, Japan). The general conditions were as follows: the total 100 μl of reaction solution contained 10 μl of PCR buffer,13 1 μl of target DNA (1.6 μg/μl) as a template, 1 μl of each primer (100 pmol/μl), 4 μl of dNTPs solution (2.5 mM), 0.5 μl of Taq polymerase (Ampli-Taq, Perkin Elmer Cetus, 2.5 U/μl). Denaturation step, 96°C, 15 s; annealing step, 52°C, 30 s; extension step, 72°C, 1 min. 30 s. After the 30th cycle, an additional incubation at 72°C for 5 min was done.

DNA sequencing. The nucleotides of pTh1 and pBAP1 were sequenced using an Applied Biosystems 373A DNA sequencer (Applied Biosystems, U.S.A.) with a Taq Dye Primer Cycle Sequencing Kit (Applied Biosystems). Deletion plasmids for sequencing were also prepared using a Kilo-sequencing (Takara Suyou Co., Ltd., Japan). All regions were sequenced in both directions. The detailed strategy is not presented here.

DNA analysis and homology search. DNA sequence analysis was done using Geneworks version 2.2 (IntelliGenetics, Inc., U.S.A.). The similarity search was done using DNAIS software (Hitachi Software Engineering Co., Ltd., Japan) with GeneBank release 86.0, December, 1994 and Swiss Protein data base release 30.0, October, 1994. Analysis of the amino acid sequence motif was done with Prosite Data Base release 12.1, October 1994.14

Expression of the aminopeptidase gene in *E. coli*. Expression of the AP-T gene was described previously.45 For the expression of the AP-T gene, we used the tac-promoter expression plasmid pEXP7, in the same way as described in a previous paper.46 We synthesized 5'GGGAAATTCTGGGA-CGCCTTCAAGGCGAACCTTG3' (5'-coding region of AP-T gene attached to an EcoRI site) as a forward primer and 5'GGGAAAGCTTCT-ACACCCCCACCCCCCAGGC3' (3'-coding region of the AP-T gene attached to a HindIII site) as a reverse primer. Using these primers under the standard PCR conditions we amplified the AP-T coding region from the total DNA of strain HB8 as a template, and then digested the PCR products with EcoRI and HindIII. The resulting sticky fragments were ligated into the EcoRI/HindIII site of the tac promoter plasmid pEXP7, then consequently we obtained the AP-T gene expression plasmid pEXTH1. E. coli transformed with pEXTH1 expressed heat-stable aminopeptidase activity at 70°C in response to induction with isopropyl-β-D-thiogalactopyranoside (IPTG). Expression methods used were the same as those previously described.44 The cells expressing AP-T were sonicated and heated at 70°C for 20 min and then centrifuged. The supernatant was further purified on a MonoQ (Pharmacia) column. The active fraction was used as the crude enzyme for characterization. (Detailed data are not shown here.)

For the expression of the APII gene, we synthesized the forward primer 5'GGGAAATTCTGACCGTGTGAGAACTG3' (5'-coding region of AP-II gene attached to an EcoRI site) and a reverse primer; 5'GGAAGCTTCTACATATGTTAGCTG3' (3'-coding region of the AP-II gene attached to a HindIII site). We amplified the APII coding region by PCR using the genomic DNA of NCIB8924, in the same manner as mentioned above. The PCR product was digested and ligated into an EcoRI/HindIII site of pEXP7 to construct pAP2EX1. The expression method was the same as described above except for heat treatment at 60°C for 20 min. The expressed enzyme was purified on a MonoQ column. The detailed method is not presented here.

Aminopeptidase activity with fluorescent substrate. Ten μl of enzyme solution (about 20 U/ml) was added to 2960 μl of 50 mM Tris–HCl buffer (pH 7.5, containing 0.1 mM CoCl2), at 70°C for AP-T or AP-II, and at 55°C for API, and then 30 μl of 10 mM pepstatin-MCA substrate in DMSO solution was added to the mixture in a cuvette with continuous mixing by a stirrer. The change in fluorescence per minute after 5 seconds was measured using a Fluorophotometer (Hitachi model F2000, excitation at 370 nm; emission at 440 nm; bandwidth, excitation, 10; emission, 10; lamp energy, 400). The activity with Leu-MCA as substrate was defined as 100%.

Nucleotide sequence accession number. The nucleotide sequence data of AP-T, API and AP-II shown in this paper have been submitted to the DDBJ, EMBL, and GenBank Nucleotide Sequence Databases and were assigned the accession numbers D13386 (for AP-T), D13385 (for APII), D00814 and D57664 (for AP-T).

Results

Molecular cloning of the AP-T gene from *T. thermophilus* HB8

A XhoI fragment (1185 bp) of the AP-T gene was obtained from the plasmid pAT15G, which contains the entire AP-T gene.43 This fragment was labeled by the ECL labeling system and used as a hybridization probe. Southern hybridization showed that a 2.1-kb HindIII fragment of genomic DNA of *T. thermophilus* HB8 hybridized with the probe (XhoI fragment of the AP-T gene). The fragment was cloned into the HindIII site of pUC119, and the resulting plasmid named pTH1.

The restriction map of pTH1 is shown in Fig. 1A. The nucleotide sequence (total 2058 bp) of the HindIII insert in pTH1 containing a gene with a sequence similar to the AP-T gene sequence, is shown in Fig. 2. An open reading frame from 811 bp in Fig. 2 encodes the AP-T gene. The GC content of the fragment was 66%, and that of the coding region was 69.9%. The AP-T gene starts with a GTG codon and encodes 408 amino acids. The deduced amino acid sequence showed 86% homology to that of AP-T. The guanine-plus-cytosine (GC) content of the 1st, 2nd, and 3rd positions of the codons was 71.8%, 46.1%, and 91.9%, respectively, which is consistent with the known characteristic tendency in *Thermus* species. The recombinant *E. coli* harboring the pTH1 plasmid expressed a heat-stable aminopeptidase activity, indicating that the gene actually encodes an active heat-stable enzyme similar to that of the AP-T gene.
Molecular cloning of the APII gene from Bacillus stea- 
thermophilus NCIB8924

To clone the APII gene of *B. stea- 
thermophilus* NCIB- 
8924, a degenerated primer (5’GGAAATCATATTCG- 
CTGGGA(AG)AAAAGA(AG)(TC)TGGG(TC)AA(AG)T- 
A(TC)GCT3’) was synthesized as a forward primer corre- 
sponding to the N-terminal sequence (1–12) of the APII gene as found by Stoll et al.,5 with a HindIII site added at 
the 5’ end. Since we had no data relating to the amino acid 
sequence of the internal portion of APII, the primer (5’- 
CCGAATTCCCGGCC(AG)TT(TC)TCA(TC)AA(AG)AA3’) 
was synthesized and used as a reverse primer, corre- 
sponding to the amino acid sequence of the internal portion of AP-T (residues 338–343) with an EcoRI site added at the 5’ end. To reduce the mixed numbers of the primer, we referred to the codon usage of the lactate dehydrogenase gene from 
the NCIB8924 strain. The PCR reaction was done using 
genomic DNA from *B. stea- 
thermophilus* as a template. After the 30-cycle reaction, 
a combination of the forward primer and the reverse primer, a fragment of about 1.0 kb was amplified as a result and became visible upon analysis by agarose gel electrophoresis. The amplified frag- 
ment was digested with EcoRI and HindIII, and then ligated into pUC18 for sequencing. The nucleotide sequences of the resulting PCR products indicated that each of these fragments correctly incorporated a part of the APII gene. Therefore the probe AP74 (5’TAGCTGACGACCCG- 
CCTTGTGAAACCGAGCCGATC- 
ACGTTGCCTTGTTAATTGAGCGCCGCTT3’), prepared 
according to the sequence of the PCR product, was syn- 
thesized for use as a hybridization probe for cloning for 
the complete APII gene. Genomic hybridization demon- 
strated that the probe AP74 hybridized to an EcoRI frag- 
ment of genomic DNA about 4.5 kb in size. This fragment 
was extracted from an 0.8% agarose gel and ligated into 
the EcoRI site of the lambda ZAPII vector. The cloned 
4.5-kb fragment was subcloned into the EcoRI site of 
pUC118, and the resulting plasmid was named pBAP1.

The restriction map and location of the APII gene in 
pBAP are shown in Fig. 1C. The nucleotide sequence of 
the complete 4549-bp EcoRI fragment was analyzed (acces- 
sion number; D13385), but only the nucleotide sequence of 
the SacI–EcoRI region (corresponding to the region from 
positions 2328 to 4549 bp within the entire sequence) of 
the APII gene is shown in Fig. 3. The open reading frame 
encoding the APII gene consists of 1242 bp including a stop 
codon, and it encodes 413 amino acid residues. The GC 
content of the 4.5 kb EcoRI fragment is 53.9%, and that 
of the coding region is 56.4%. The deduced amino acid 
sequence of the APII gene is shown under the nucleotide 
sequence in Fig. 3. The N-terminal amino acid sequence 
found by Stoll et al.5 is underlined and is completely iden- 
tical to that deduced here. A typical ribosome binding 
sequence (AAAGGAG), present at a position 9 bp upstream 
of the start codon ATG, was complementary to a sequence 
near the 3’ end of the 16S rRNA of *B. stea- 
thermophilus*.6

Fig. 1. Restriction Map of a 2058 bp HindIII Fragment Containing the Aminopeptidase Th Gene of *T. thermophillus* HB8 (A), a 5556 bp SacI–HindIII Region Containing the Aminopeptidase T Gene of *Thermus aquaticus* YT-1 (B), and a 4549 bp EcoRI Fragment Containing the Aminopeptidase II Gene of *B. stea- 
thermophilus* NCIB8924 (C).
Fig. 2. Nucleotide Sequence of a 2058 bp-HindIII Fragment Containing the Aminopeptidase Th Gene of T. thermophila and Its Duced Amino Acid Sequence.

The transcription start point and the promoter sequence were not located in this study. An inverted repeat (1511–1538) that might function as a rho-dependent transcription terminator was evident.

Nucleotide sequencing of the flanking region of the AP-T gene from Thermus aquaticus YT-1

In a previous paper,11 we reported the nucleotide sequence of the SpH1–XhoI region (1520–3096 in Fig. 1B, accession number: D00814) of a 5.2 kb HindIII fragment containing the AP-T gene. Because we expected that conserved ORFs were present in the upstream regions of both the AP-Th gene and the AP-T gene, we sequenced the fragments SacI(1)–SpH1(1520) and XhoI(3096)–HindIII(5556) as shown in Fig. 1B. The sequence analyzed represents a total of 5556 bp, including the previously reported sequence (accession number: D87664). The details of the sequencing strategy will not be presented here.

Only the first 1800 bp of this sequence containing part of S' region of the AP-Th gene is shown in Fig. 4.

Comparison of amino acid composition and molecular weight of AP-Th, and APII, and AP-T

The amino acid compositions of AP-Th, APII, and AP-T, as deduced from the sequence data are shown in Table I in comparison with those found chemically by Stoll et al.5 Lys was abundant in APII and the contents of Ala, Arg, and Gly were less than that in AP-Th and AP-T. These differences might be a reflection mainly of the genomic GC content. In Thermus species, there is a higher content of amino acids encoded by codons that tend to increase the GC content of the genomic DNA.

Amino acid sequence similarities among AP-Th, APII, and AP-T

Multiple sequence alignment of the deduced amino acid sequences is shown in Fig. 5. AP-Th showed 86% similarity to AP-T and APII showed 43% similarity to AP-T. These aminopeptidases display sequence similarity throughout the molecule and conserved sequences are present particularly in the C-terminals (the region of Asn242-Thr253, Leu304-Thr321, Thr336-Ala343, Val378-Gly385). Thus, these aminopeptidases comprise a single aminopeptidase family (the aminopeptidase T family).

Substrate specificity of AP-Th, APIII, and AP-T

Table II shows the relative activities of the aminopeptidases expressed in E. coli observed with several MCA substrates. All of the enzymes efficiently hydrolyzed Leu-MCA, Phe-MCA, Arg-MCA, Met-MCA, and Lys-MCA and a similar profile of specificity was displayed in each instance. These findings indicate that these aminopeptidases were essentially members of the same enzyme group.

Discussion

In this study, we have demonstrated that AP-Th,
Fig. 3. Nucleotide Sequence of the Soc1-EcoRI Region of an 5.2 kb-EcoRI Fragment Containing the Aminopeptidase II Gene of B. stearothermophilus and Its Deduced Amino Acid Sequence.

RBS, possible ribosome binding site. The mRNA sequence analyzed chemically by Stoll et al. [7] is underlined. X in a sequence indicates the inverted repeats of a putative terminator sequence. Numbering starts from the first base of the SocI site (corresponding to position 2532 in the whole sequence). No hypothetical open reading frame is indicated.

H. Motosu et al.

Fig. 4. Nucleotide Sequence of the Upstream Flanking Region of the Aminopeptidase T Gene of T. aquaticus YT-1 and the hypothetical open Reading Frames. The aminopeptidase T gene and the hypothetical open reading frames (ORF1, ORF2, and ORF3) are indicated. The deduced amino acid sequence of ORF1, indicated in italics, is similar (29.7%) to the DNA polymerase beta of rats and humans (see the text).
Table I. Comparison of Amino Acid Compositions of Aminopeptidase Th, Aminopeptidase T and Aminopeptidase II

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>No. of residues deduced from the DNA sequence found in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td>38 37 27 (26)</td>
</tr>
<tr>
<td>Ala</td>
<td>58 53 46 (45)</td>
</tr>
<tr>
<td>Val</td>
<td>30 35 28 (29)</td>
</tr>
<tr>
<td>Leu</td>
<td>41 43 36 (36)</td>
</tr>
<tr>
<td>Ile</td>
<td>12 11 21 (20)</td>
</tr>
<tr>
<td>Ser</td>
<td>8 10 16 (16)</td>
</tr>
<tr>
<td>Thr</td>
<td>17 18 21 (22)</td>
</tr>
<tr>
<td>Phe</td>
<td>19 18 18 (15)</td>
</tr>
<tr>
<td>Tyr</td>
<td>7 5 13 (13)</td>
</tr>
<tr>
<td>Trp</td>
<td>8 9 8 (7)</td>
</tr>
<tr>
<td>Cys</td>
<td>1 1 2 (3)</td>
</tr>
<tr>
<td>Met</td>
<td>5 5 6 (6)</td>
</tr>
<tr>
<td>Pro</td>
<td>21 22 21 (20)</td>
</tr>
<tr>
<td>Asn</td>
<td>12 15 26 (-)</td>
</tr>
<tr>
<td>Glu</td>
<td>8 12 12 (-)</td>
</tr>
<tr>
<td>Asp</td>
<td>17 17 17 (33)</td>
</tr>
<tr>
<td>Glu (Glx)</td>
<td>44 42 30 (44)</td>
</tr>
<tr>
<td>Lys</td>
<td>16 14 22 (21)</td>
</tr>
<tr>
<td>Arg</td>
<td>37 30 21 (30)</td>
</tr>
<tr>
<td>His</td>
<td>9 11 11 (11)</td>
</tr>
</tbody>
</table>

Total no. of residues: 408 408 413

Subunit molecular weight: 45,015 44,818 46,207

* No. in parentheses are those found chemically by Stoll et al.5)

We compared the most conserved regions of aminopeptidases of the AP-T family with the sequences of other aminopeptidases, and found a region of residues Asp306-Arg312 [Asp-Thr-Asp(Glu or Asp)-Gly-Ala-Arg] having low but apparent similarity to the amino acid signature of cytosol (leucyl) aminopeptidase (Asn-Thr-Asp-Ala-Glu-Arg-Leu) in the PROSITE database.14) Cytosol aminopeptidase (EC 3.4.11.1) is commonly called leucyl aminopeptidase, and these enzymes are ubiquitous in nature. One of the most extensively studied of the leucyl aminopeptidases is bovine lens leucine aminopeptidase (AMPL... BOVIN).19–22) The primary amino acid sequence of AMPL... BOVIN has been reported19) and its three-dimensional structure has been analyzed both in native from and in complex with bestatin21) or amastatin.22) A mechanism for hydrolysis of the peptide bond has been proposed.23,24) In AMPL... BOVIN, the mechanism includes roles for Asp253 as a general base, for Arg336 as additional electrophilic substrate activator and transition state stabilizer, and for Lys262 as a proton shuttle.23) The leucyl aminopeptidase signature is a region containing the active site Arg336. The primary structure of some other members of the leucyl aminopeptidase group, as deduced from the DNA sequence, has been reported including aminopeptidase from mouse-ear cress (AMPL...).

Table II. Relative Activities of the Aminopeptidases Expressed in E. coli Transformants as Determined Using Various Peptidyl-MCA Substrates

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Relative activity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP-Th</td>
<td>AP-T</td>
</tr>
<tr>
<td>Leu-MCA</td>
<td>100.0</td>
</tr>
<tr>
<td>Phe-MCA</td>
<td>102.6</td>
</tr>
<tr>
<td>Arg-MCA</td>
<td>8.4</td>
</tr>
<tr>
<td>Met-MCA</td>
<td>7.7</td>
</tr>
<tr>
<td>Ala-MCA</td>
<td>0</td>
</tr>
<tr>
<td>Lys-MCA</td>
<td>2.7</td>
</tr>
<tr>
<td>Phe-MCA</td>
<td>0</td>
</tr>
<tr>
<td>Lys-Ala-MCA</td>
<td>0</td>
</tr>
<tr>
<td>Gly-Pro-MCA</td>
<td>0</td>
</tr>
<tr>
<td>Pro-Phe-MCA</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 5. Multiple Alignment of Aminopeptidase II from B. stearothermophilus, Aminopeptidase T from Thermus aquaticus YT1, and Aminopeptidase Th from Thermus thermophilus HB8.

Multiple alignment was done using the GeneWorks program (IntelliGenetics, Inc., CA, U.S.A.) based on a PAM-250 matrix. Asterisks along the bottom lane indicate completely conserved amino acid sequences in these aminopeptidases.
Leucyl aminopeptidase family

**Leucyl aminopeptidase signature**

<table>
<thead>
<tr>
<th>Leucyl aminopeptidase family</th>
<th>B Z Z +</th>
<th>BB B</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMPL BOVIN (P00727) 323</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMPL ARATH (P30184) 364</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMPL SOLTU (P31427) 398</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMPL ECOLI (P11648) 343</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMPL RICPR (P27888) 338</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Aminopeptidase T family**

<table>
<thead>
<tr>
<th>Aminopeptidase T family</th>
<th>AP-T (P23341) 299</th>
<th>AP-Th (P42778) 299</th>
<th>APII (D13385) 299</th>
<th>APS (U51911) 299</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EVLKE ALE GQGGE</td>
<td>EVLKE ALE GQGGE</td>
<td>ETIKK IEL GQGGE</td>
<td>DILKE VEGSGL</td>
</tr>
<tr>
<td></td>
<td>VPF DNP IA KEG</td>
<td>VPF DNP IA KEG</td>
<td>VPF SPVS LSN</td>
<td>VPF DNP IA KEG</td>
</tr>
<tr>
<td></td>
<td>FFD FDEN AASH</td>
<td>FFD FDEN AASH</td>
<td>FYN FDEN AASH</td>
<td>FDP SIS QSNLFY</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FDEN AS HH</td>
</tr>
</tbody>
</table>

Fig. 6. Comparison of the Conserved Regions in the Leucyl Aminopeptidase Family and the Aminopeptidase T Family.

The frame displays the leucyl (cytosol) aminopeptidase signature which represents a consensus pattern in all members of the cytosol aminopeptidase group (Prosite Data Base accession no. PS06331). Amino acid numbering starts from the N-terminus of the mature enzyme. AMPL BOVIN: bovine lens leucine aminopeptidase, AMPL ARATH: aminopeptidase of Arabidopsis thaliana, AMPL SOLTU: aminopeptidase from Potato, AMPL ECOLI: aminopeptidase A from E. coli, AMPL RICPR: aminopeptidase A from Ricettsia prowazekii, AMPL T: aminopeptidase Th (This paper), APS: aminopeptidase S from Bacillus subtilis (U51911).

ARATH, potato (AMPL..SOLTU), E. coli (AMPA..ECOLI), Ricettsia prowazekii (AMPL..RICPR), and others. All of these aminopeptidases display a completely conserved leucyl aminopeptidase signature. In Fig. 6, we compare the conserved regions in aminopeptidases of the AP-T family as determined in this experiment and those of several known leucyl aminopeptidases. There was no sequence similarity observed comparing the total sequences, but a zinc binding residue Asp332 in AMPL BOVIN was conserved in the AP-T family as residue Asp308. Positively charged Arg336 in AMPL BOVIN was conserved as Arg312 in all enzymes of the AP-T family, and this might be a catalytic residue of the active site. Thr359 and Leu360 in AMPL BOVIN are residues involved in stabilizing the enzyme-bestatin complex, according to data concerning the three-dimensional structure of this complex. Also these are conserved as Thr336 and Leu337 in the AP-T family. Lys262 in AMPL BOVIN probably corresponds to Lys237 in AP-T, though it is not conserved in AP-II.

AP-T is sensitive to inhibition by amastatin, which is a tetrapeptide analog inhibitor specific to the leucyl aminopeptidase family, therefore the 3D-structure of the substrate-enzyme binding site of the AP-T family should be similar to that of the leucyl aminopeptidase family. Many characteristics of AP-T are similar to those of the leucyl aminopeptidases, including the following: broad substrate specificity with preference displayed especially to leucyl- or phenyl-peptide substrates; inability to hydrolyze peptides containing proline in the penultimate position; metallo-dependent with two metal ions (probably zinc ions) per subunit; conservation of the amino acid sequence in the C-terminal domain. In spite of almost no sequence similarity among the AP-T and leucyl aminopeptidases, the members of the AP-T family display similarity to leucyl aminopeptidases at the active site, so it is thought that these are related families (i.e., a “clan”). In other words, the AP-T family seems to be a member of the leucyl (cytosol) aminopeptidase superfamily.

Recently, the product of the aminopeptidase S (APS) gene (ampS) of Bacillus subtilis was reported as a putative protein based on the similarity of ampS to the APII gene (EMBL: D37799, U51911). This gene encodes 410 amino acids. The similarity to AP-T is 44%, that to AP-II is 53%. The sequence of APS is shown Fig. 6. It is apparent that APS is a member of AP-T family. Characteristically in APS, Arg312 is not conserved and is changed to His312. His is a positively charged residue, like the Arg residue, and it might function as a catalytic residue. The AP-T family is probably ubiquitous in bacteria other than thermophiles.

Dot matrix analysis with Geneworks software showed that there are conserved regions upstream of the AP-Th and AP-T genes. These conserved regions are thought to encode hypothetical open reading frames (ORFs); ORF1, ORF2, and ORF3 as shown in Fig. 1. The complete nucleotide sequence of the ORF1 upstream of the AP-Th gene and a partial sequence of the ORF1 upstream of the AP-T gene were analyzed and the deduced amino acid sequence was similar to that of the DNA polymerase beta of rats and humans as indicated by the similarity search. The polymerase beta of these vertebrates is thought to function in DNA repair, and these enzymes are considered to belong to the DNA polymerase family X on the basis of sequence similarity to DNA polymerase (EC 2.7.7.7). Further work is needed to find whether the product of ORF1 actually functions in DNA repair or has other DNA polymerase activity.

On the other hand, ORF2 and ORF3 are only hypothetical and were identified based on their similarity to conserved sequences known to be present in strains HB8 and YT-1. These ORFs might encode regulatory proteins but this remains to be seen in a further study.

The ORF4 is similar to the Met17 gene encoding O-acetylserine sulphydrylase (OAS SHLase, EC 4.2.9.8) in Saccharomyces cerevisiae. This enzyme converts O-acetylhomoserine to homocysteine in the biosynthetic pathways of sulfur-containing amino acids such as methionine and cysteine.

In this study, we sequenced the nucleotides of an aminopeptidase from T. thermophilus and B. steatornerophilus, and it showed that the AP-T family is related to the leucyl aminopeptidase family.

References

3. E. Minagawa, S. Kaminogawa, H. Matsuoka, T. Ohda, and K.


