J. Biochem. 122, 843-850 (1997)

Methionine Aminopeptidase from the Hyperthermophilic Archaeon *Pyrococcus furiosus*: Molecular Cloning and Overexperssion in *Escherichia coli* of the Gene, and Characteristics of the Enzyme

Susumu Tsunasawa,[1] Yukiko Izu, Masaru Miyagi, and Ikunoshin Kato

Biotechnology Research Laboratories, Takara Shuzo Co., Ltd., Sunaike 2257, Nojicho, Kusatsu, Shiga 525

Received for publication, June 12, 1997

A gene for a methionine aminopeptidase (MAP; EC 3.4.11.18), which catalyzes the removal of amino-terminal methionine from the growing peptide chain on the ribosome, has been cloned from the hyperthermophilic Archaeon, *Pyrococcus furiosus*, by a novel method effectively using its cosmid protein library, sequenced and expressed in *Escherichia coli*. The DNA sequence encodes a protein containing 295 amino acid residues with methionine at the N-terminus. From protein analyses of the recombinant protein expressed in *E. coli*, by using both amino acid sequence analysis from the N-terminus by automated Edman degradation and analyses of molecular masses of the peptides generated by two enzymatic cleavages performed independently, digestions with lysylendopeptidase and Endoproteinase Asp-N, with ionspray mass spectrometry, the primary structure of the protein has been elucidated to be completely identical with that deduced from its DNA sequence. Comparison of the amino acid sequence of *P. furiosus* MAP (*P.f.* MAP) with those of other MAPs from Eukarya and Bacteria showed that the protein has a high degree of sequence homology in the stretches surrounding the five cobalt-binding residues fully preserved in all of MAPs determined so far, but *P.f.* MAP belongs to Type II because it has an extra long insertion of about 60 amino acid residues between the fourth and fifth cobalt-binding ligands, similar to MAPs from human and rat, and to Met-AP2 from *Saccharomyces cerevisiae*, in comparison to Type I MAPs from Bacteria. Therefore, *P.f.* MAP seems to be rather close to those from Eukarya, although it is distinct in lacking the N-terminal extension of about 90-150 residues universally found in MAPs from Eukarya. These findings suggest that *P.f.* MAP is evolutionally located at the Eukarya-Bacteria boundary. The enzyme expressed in *E. coli* exhibits a considerable thermostability, with a half-life of approximately 4.5 h at 90°C and an optimum temperature of around 90°C.

Key words: amino acid sequence, hyperthermostable enzyme, ionspray mass spectrometry, methionine aminopeptidase, *Pyrococcus furiosus*.

Amino-terminal processing of nascent peptides is a very common event (*1*). In Bacteria, mitochondria and chloroplasts, protein synthsis is initiated with formyl methionine, and the formyl group is then removed from the N-terminus by the action of deformylase, but in Eukarya, translation of proteins is initiated with methionine. The amino-terminal methionine residue on more than half of both Eukarya and Bacteria proteins is then cleaved by methionine aminopeptidase (MAP) depending on the penultimate amino acid residue (*2*). It is significant in the evolutional analysis of these microorganisms, to ascertain whether the common processing by a putative MAP occurs at the amino-termini of proteins in Archaea,the third kingdom of life, presuming the most ancient known forms of life on earth (*3*), and to elucidate the presence of the MAP in Archaea and its protein structure. Furthermore, we focused on the enzyme from hyperthermophilic Archaea *Pyrococcus furiosus* (*4*) in order to understand the molecular basis for hyperthermostability of enzymes in this microorganism. In this paper, we show the presence of a MAP having similar substrate specificities in *P. furiosus*, its cloning and DNA sequencing. In addition, we report several characteristics of the recombinant protein expressed in *Escherichia coli*, and comparison of its amino acid sequence with those of MAPs previously identified in other living cells (Eukarya, Bacteria and Archaea).

[1] To whom correspondence should be addressed. Fax: +81-775-65-6965, E-mail: s-tsunas@mx.biwa.or.jp

Abbreviations: MAP, methionine aminopeptidase; TFA, trifluoroacetic acid; PIPES, piperazine-*N*,*N*′-bis(2-ethanesulfonic acid); *P.f.* MAP, *Pyrococcus furiosus* methionine aminopeptidase; PTH-, phenylthiohydantoin-.

MATERIALS AND METHODS

Materials—Culture of *P. furiosus* (DM 3638) were grown as described previously (*5*). *E. coli* JM109, enzymes for *in vitro* manipulations of DNA and their kits, and vector DNAs were the products of Takara Shuzo. *E. coli* DH5αMCR was from Bethesda Research Laboratories. Gigapack II Gold kit and Triple Helix Cosmid Vector were from Stratagene. Lysylendopeptidase from *Achromobacter lyticus* [EC 3.4.21.50] was obtained from Wako Pure Chemical Industries. Endoproteinase Asp-N was from

Boehringer Mannheim. Peptide substrates were synthesized by *t*-butoxycarbonic methods on a peptide synthesizer (Perkin-Elmer 430A). Recombinant human fibronectin fragment (C281) expressed in *E. coli*, corresponding to residues 1270-1548 with additional amino acid residues consisting of an initiator methionine and an alanine, at the N-terminus, was kindly provided by Dr. K. Hashino of our laboratory. The sources of other specific chemicals and reagents are shown in the text.

Preparation of Pyrococcus furiosus Cosmid Protein Library—The *P. furiosus* genome DNA (400 μg) was partially digested with *Sau*3AI in a buffer [50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM dithiothreitol, 100 mM NaCl] and fractionated according to the size by sucrose density gradient centrifugation. Triple Helix Cosmid Vector (1 μg) was cleaved with *Bam*HI and mixed with the genome DNA fragments (140 μg) of 35-50 kb fractionated as described above. After ligation by use of a Ligation kit, the *P. furiosus* genome DNA fragments in the cosmid vector were packaged into λ-phage particles by the *in vitro* packaging method using Gigapack II Gold. By using a part of the phage solution thus obtained, *E. coli* DH5αMCR was transformed to give a cosmid library. Then 500 colonies were selected and cultured independently in an L-broth containing 100 μg/ml ampicillin at 37°C with shaking for 16 h. The culture was centrifuged and cells were harvested. These cells were ultrasonicated and further treated at 100°C for 10 min. After centrifugation, the supernatants were used as a protein library to screen for the target activity.

Measurement of MAP Activity—To an appropriate amount of each sample (usually 5 μl) the substrate solution [Na-PIPES buffer (pH 7.2) containing 1 mM Met-Pro-Ala-Ala-Gly and 0.5 mM $CoCl_2$, usually 45 μl] was added, and the reaction mixture was incubated at 75°C for 5 min. The reaction was stopped by adding 100 mM EDTA solution (pH 7.5, usually 10 μl), then the release of the N- terminal methionine was quantified by measuring the absorbance at 450 nm developed by incubation with the color reagent [100 mM Na-phosphate buffer (pH 7.5) containing 18 μg of L-amino acid oxidase (Sigma Chemical), 9 μg of *o*-dianisidine (Sigma Chemical) and 2.5 μg of horseradish peroxidase (Sigma Chemical), usually 50 μl] at 37°C for 10 min. One unit of the activity was defined as the amount of enzyme that released 1 μmol of methionine per min under the conditions describd above. For screening the cosmids and clones having the MAP activity, amino acid-releasing activity for a peptide, Leu-Pro-Ala-Ala-Gly-, was analyzed at the same time, and the cosmids and clones showing activities against both substrates were rejected because their activities seemed to be attributable to other aminopeptidases and proteases.

Screening of the MAP Gene of Pyrococcus furiosus—The MAP activity in the cosmid protein library prepared as described above was measured. Of six transformants showing the MAP activity, a cosmid DNA was prepared and digested with *Xho*I, inserted into the *Sal*I site of the plasmid vector pUC18, then introduced into *E. coli* JM109. From the transformants, a colony showing the MAP activity in the supernatant of cells prepared in a similar manner to as described in the cosmid protein library was further selected, and the resultant plasmid was designated pMAP1. The pMAP1 was digested with *Bln*I, and the resulting DNA fragments were inserted into the *Xba*I site of the pUC118, and the plasmid pMAP2 was obtained by measuring the MAP activity in *E. coli* transformants in the same way as described above. The plasmid pMAP2P was prepared by digestion of the plasmid pMAP2 with *Not*I and *Bln*I, followed by insertion of the fragment into the *Not*I-*Xba*I site of pUC19 and its introduction of the plasmid into *E. coli* JM109. The plasmid pMAP2p was digested with *Eco*RI and *Xho*I, and electrophoresed on an agarose gel. A DNA fragment was extracted from the agarose gel and inserted into the *Eco*RI-*Sal*I site of the plasmid vector pUC18, and *E. coli* JM109 was transformed with this plasmid. *E. coli* JM109/MAP8 thus obtained showed strong and hyperthermostable activity of MAP, and therefore a DNA fragment of about 1.3-kb prepared from the plasmid pMAP8 was subjected to DNA sequencing to deduce primary structure of MAP from *P. furiosus*.

DNA Sequencing—The nucleotide sequences of the DNA fragments formed by digestion with several restriction enzymes were analyzed by the dideoxy chain termination method basically as described by Sanger *et al.* (*6*). BcaBEST DNA polymerase (*7*) was used for the chain elongation reaction with fluorescent dideoxy terminators and the products were analyzed with an autosequencer (Perkin-Elmer 370A).

Purification of the Recombinant Pyrococcus furiosus MAP Expressed in E. coli—*E. coli* JM109 carrying pMAP8 was grown at 37°C for 16 h with shaking in 500 ml of an L-broth medium containing 0.01% of ampicillin. After centrifuging the culture (6,000 rpm$\times$10 min), the cells thus collected (9.1 g wet weight) were suspended in 45 ml of 20 mM Tris-HCl buffer (pH 7.5), ultrasonicated, and treated at 100°C for 10 min. After centrifuging (12,000 rpm$\times$20 min), the supernatant was used as a crude enzyme source. The supernatant was dialyzed against 20 mM K-phosphate buffer (pH 8.0) containing 0.1 mM $CoCl_2$, and the dialysate (55 ml) was loaded on a DEAE-Sepharose CL-6B (Pharmacia) column (2.5$\times$16 cm), equilibrated with the same buffer. The column was developed with a linear gradient from 0 to 0.5 M of NaCl in the same buffer, and active fractions were collected (98 ml) and dialyzed against 10 mM K-phosphate buffer (pH 7.0) containing 0.2 mM $CoCl_2$.The dialyzate (111 ml) was loaded on a CM-Sepharose CL-6B (Pharmacia) column (1.5$\times$12 cm), equilibrated with the same buffer. The active fractions were recovered by eluting with a linear gradient from 0 to 0.5 M of NaCl. The enzyme thus obtained showed a single band on SDS-PAGE. Protein was measured by the method of Bradford (*8*).

Protein Analysis—The N-terminal sequences of the recombinant *P. furiosus* MAP and the substrate protein were analyzed by the direct sequencing of proteins with a pulse-liquid phase protein sequencer (Perkin-Elmer 477A). Amino acid analysis of the protein was performed with an amino acid analyzer (Hitachi L-8500S) for the hydrolyzate with 5.7 M HCl containing 0.2% phenol or with 4 M methanesulfonic acid containing 0.2% 3-(2-aminoethyl)indole at 100°C for 24 h (*9*). A triple quadrupole equipped with an ionspray ion source mass spectrometer (Perkin-Elmer-Sciex, API-III) was used to determine the molecular mass of the protein by introducing the protein sample (200 pmol), dialyzed against 0.01% TFA, lyophilized, and redissolved in 100 μl of 0.1% formic acid-50%

acetonitrile, into the mass spectrometer through a fused silica tube (100 μm i.d.) at a flow rate of 2 μl/min. The quadrupole was scanned from 350 to 2,000 Da, using a step size of 0.1 Da and a 0.5 ms dwell time per step. Ionspray voltage was set at 5 kV, and the orifice voltage was 80 V. In advance of LC-MS analysis, the recombinant *P.f.* MAP was denatured by incubation in 70% TFA at 4°C for 8 h and lyophilized. Digestion with each protease was performed at 30°C for 24 h under the conditions indicated in parentheses (buffer; molar substrate-to-enzyme ratio), lysylendopeptidase [0.1 M Tris-HCl buffer (pH 9.0) containing 2 M urea; 200:1], Endoproteinase Asp-N [0.1 M Tris-HCl buffer (pH 7.5) containing 2 M urea; 200:1]. Aliquots of each digest (25 pmol each) were loaded onto a packed capillary reversed phase column (Poros II R/H, 0.3×100 mm, LC-Packing) connected to the ionspray mass spectrometer. A solvent flow of 10 μl/min was generated by a Waters 625 LC system. Solvent A was 0.05% TFA in H_2O, and a solvent B was 0.05% TFA in acetonitrile. Separation of the peptides obtained was effected with a gradient of 0-60% B over 45 min. The column effluent, except for the unadsorbed fraction, was passed into an ionspray probe of the mass spectrometer. The total ion chromatograms were recorded in the single quadrupole mode.

Enzymatic Properties of the MAP—The substrate specificity was examined by using synthetic peptides, Met-X-Ala-Ala-Ala (X = 19 natural amino acids other than cystine and cysteine), and the recombinant human fibronectin fragment (C281). The pH optimum and stability were determined at 75°C in a buffer containing 0.5 mM $CoCl_2$ over the pH range 4.0-10.6. The buffers used were 20 mM Na-acetate for pH 4.0-5.0, 20 mM Na-PIPES for pH 5.8-7.2, 20 mM Na-borate for pH 8.0-9.5, and disodium hydrogenphosphate-NaOH for pH 9.9-10.6. The pH stability was measured with an aliquot of each solution withdrawn after incubation in a buffer described above at 75°C for 1 h by using the standard assay system. The effect of temperature on the activity was determined by incubating an appropriate amount of the enzyme in a 20 mM Na-PIPES buffer (pH 7.2) containing 0.5 mM $CoCl_2$ at temperatures varying from 20 to 100°C for 5 min. Thermostability of the enzyme was determined by measuring the remaining activity in the standard assay system after incubation in the same buffer described above at temperatures varying from 20 to 100°C for 1 h. The effects of metal ions and denaturants on the enzyme activity were also examined.

Computer Analysis—A homology search for the amino acid sequence of *P.f.* MAP was done by GenBank using BLAST search program.

RESULTS AND DISCUSSION

Cloning of the MAP Gene from Pyrococcus furiosus—To screen the map gene in *P. furiosus*, the cosmid protein library described in "MATERIALS AND METHODS" was effectively used. By using cosmid vectors, the probability that the target gene coding is divided by a restriction enzyme digestion within its coding region decreases, and the numbers of clones necessary to be tested could be reduced. Furthermore, the background of the assay could be greatly diminished by denaturation of either proteins or inhibitors from host cells with heat treatment at 100°C for 10 min. From the cosmid protein library made from 500 different clones, the supernatants of six transformants showed high MAP activities. Cosmid DNAs were prepared from these transformants, and the restriction patterns of each insert DNA fragment were compared. All of the inserts had a *Bln*I-*Xho*I fragment of about 1.3 kb long. A cosmid DNA was thus selected and subjected to screening and sequencing of the *map* gene. From subcloning and assay of its heat-stable MAP activity as described in "MATERIALS AND METHODS," the *map* gene was located within the 1.3 kb *Bln*I-*Xho*I fragment. The plasmid containing the *Bln*I-*Xho*I restriction fragment was designated pMAP8. The restriction map of the 1.3-kb insert and its complete nucleotide sequence are shown in Figs. 1 and 2, respectively. Two ATG codons were found at nucleotides 308-310 and 326-328, downstream from the in-frame terminator TGA (nucleotides 296-298) in the determined sequence, and the first one (nucleotides 308-310) was concluded to be an initiation codon from the N-terminal sequence analysis of the MAP protein described as follows. The *map* gene thus encompasses 888 nucleotides, with the initiation codon ATG and the termination codon TGA. From analysis of the nucleotide sequence flanking the *map* gene, four regions having conserved box A Archaeal promoter sequences (*10*), TTATA at nucleotides 112 and 231 and TTAAA at 57 and 269 to the initiation codon, were found.

Experssion of the MAP Gene in E. coli—The recombinant *E. coli* JM109/pMAP8 constructed as described in "MATERIALS AND METHODS" produced 6.7 units/mg protein of *P. furiosus* MAP in the crude extract heated at 100°C for 10 min with induction by isopropyl-1 thio-β-D-galactopyranoside. The enzyme was purified to be homogeneity by heat treatment, followed by DEAE-Sepharose, and CM-Sepharose column chromatographies (Table I).

Characterization of the Recombinant Pyrococcus furiosus MAP—The N-terminal 35 amino acid residues of the recombinant *P.f.* MAP were determined by direct sequencing of the intact protein. The result indicates the initiation codon of *P.f. map* gene corresponds to the nucleotides 308-310 (ATG) in Fig. 2. The molecular weight (32,848 Da) determined with an ionspray triple quadrupole mass spectorometer was in close agreement with the value (32,843 Da) calculated from the translated nucleotide sequence by the ATG described above as the initiation codon. A more detailed structural analysis of the protein was performed by LC-MS analysis of each peptide produced by either lysylendopeptidase or Endoprotinease Asp-N digestion. Both total ion chromatograms are shown in Fig. 3, and the

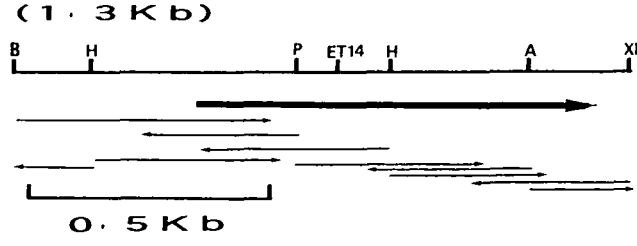

Fig. 1. **Restriction map of the 1.3 kb DNA fragment containing the *map* gene from *P. furiosus* and the sequencing strategy.** The open reading frame is indicated by the thick arrow. Arrows below the DNA fragment show the direction of sequencing and the region sequenced. Abbreviations: B, *Bln*I; H, *Hin*dIII; P, *Pst*I; ET14, *Eco*T14I; A, *Afl*II; Xh, *Xho*I.

observed masses of the resulting peptides are listed in Table II with the expected average masses of amino acid residues deduced from the nucleotide sequence of the *map* gene. The observed mass in each case is within 0.5 Da of the expected mass. Several small peptides of less than 6 residues from both digests, which might be eluted in the unadsorbed fraction on the LC column used, could not be analyzed, but by mass analyses of the peptides corresponding to the overlaps of each digest, the entire amino acid sequence of the recombinant *P.f.* MAP was elucidated to be identical with that deduced from the nucleotide sequence of its gene. Figure 4 shows the amino acid sequence of the recombinant *P.f.* MAP with each peptide assigned on both total ion chromatograms.

Enzymatic Properties of the Recombinant Pyrococcus furiosus MAP—The effects of temperature and pH on the

TABLE I. **Purification of the recombinant *P.f.* MAP from *E. coli*.** Values are from 9.1 g (wet weight) of *E. coli* JM109/pMAP8.

Step	Protein (mg)	Activity[a] (U)	Purification (fold)	Recovery (%)
Crude extract (*with heat treatment*)	113	761	1	100
DEAE-Sepharose	22.1	536	3.73	70.4
CM-Sepharose	16.6	502	4.48	65.9

[a]One unit of the activity is defined as the amount of the enzyme that releases 1 μmol of methionine from the synthetic peptide Met-Pro-Ala-Ala-Gly per min at pH 7.2 at 75°C.

```
   1 CTAGGTGCATTCAGTTTGGCTATTGTCCCGAGAGAGATCTAATGCCTCCTGGGTGCTTAA   60
  61 AGAAAACTAGAAAAAAGTGGGAAAAAGTTGCGGAAAGTAAGAGCTAAATTGTTATATTGA  120
 121 GTAAAAGCTTTCTTTCTTTATTTGTCTTTATGGCAAAATCCCAGAAGTTCAGCTATTGAA  180
 181 TTAGAGAACTGTTCGTCACTGAAAGTAAACTTCTATGGGATTCTTCTGAATTATATGGTA  240
 241 AGGTTTGGAAAATTTGGACATAAAAGTCTTAAAGTTTCCTTTTTCAACTCTAAACTAGGG  300
 301 TGAGCTAATGGATACTGAAAAACTTATGAAAGCCGGAGAAATAGCAAAAAAAGTAAGAGA  360
   1         M  D  T  E  K  L  M  K  A  G  E  I  A  K  K  V  R  E   18
 361 GAAAGCTATTAAACTTGCTAGACCTGGGATGTTGTTGTTAGAACTTGCAGAGTCTATAGA  420
  19   K  A  I  K  L  A  R  P  G  M  L  L  L  E  L  A  E  S  I  E   38
 421 AAAGATGATAATGGAACTTGGGGGTAAACCTGCTTTCCCAGTAAATTTATCAATTAATGA  480
  39   K  M  I  M  E  L  G  G  K  P  A  F  P  V  N  L  S  I  N  E   58
 481 AATTGCAGCTCACTATACTCCTTACAAGGGAGATACTACTGTTCTGAAAGAGGGGGATTA  540
  59   I  A  A  H  Y  T  P  Y  K  G  D  T  T  V  L  K  E  G  D  Y   78
 541 TCTAAAGATCGACGTGGGGGTTCACATAGATGGATTTATAGCAGATACTGCAGTTACAGT  600
  79   L  K  I  D  V  G  V  H  I  D  G  F  I  A  D  T  A  V  T  V   98
 601 TAGAGTAGGGATGGAAGAAGATGAGCTTATGGAGGCTGCCAAGGAAGCGTTAAACGCCGC  660
  99   R  V  G  M  E  E  D  E  L  M  E  A  A  K  E  A  L  N  A  A  118
 661 AATTTCTGTAGCTAGGGCGGGAGTGGAGATAAAGGAACTAGGAAAGGCAATAGAAAATGA  720
 119   I  S  V  A  R  A  G  V  E  I  K  E  L  G  K  A  I  E  N  E  138
 721 AATTAGGAAGAGAGGATTCAAACCAATAGTTAATCTAAGTGGGCACAAGATAGAAAGATA  780
 139   I  R  K  R  G  F  K  P  I  V  N  L  S  G  H  K  I  E  R  Y  158
 781 CAAGCTTCATGCAGGGATTAGCATTCCGAACATTTATAGACCGCATGATAACTATGTTTT  840
 159   K  L  H  A  G  I  S  I  P  N  I  Y  R  P  H  D  N  Y  V  L  178
 841 AAAGGAAGGAGATGTTTTCGCAATTGAGCCTTTCGCTACTATAGGTGCTGGTCAAGTAAT  900
 179   K  E  G  D  V  F  A  I  E  P  F  A  T  I  G  A  G  Q  V  I  198
 901 TGAGGTTCCCCCAACCTTAATCTACATGTACGTTAGAGATGTTCCAGTTAGAGTGGCCCA  960
 199   E  V  P  P  T  L  I  Y  M  Y  V  R  D  V  P  V  R  V  A  Q  218
 961 AGCTAGGTTCCTTTTGGCTAAGATAAAAAGGGAATATGGAACCCTACCCTTTGCCTATAG 1020
 219   A  R  F  L  L  A  K  I  K  R  E  Y  G  T  L  P  F  A  Y  R  238
1021 GTGGCTTCAGAATGACATGCCAGAAGGACAGCTTAAGTTGGCCCTAAAAACCCTCGAAAA 1080
 239   W  L  Q  N  D  M  P  E  G  Q  L  K  L  A  L  K  T  L  E  K  258
1081 GGCTGGAGCTATATATGGCTATCCAGTGCTTAAAGAAATTAGAAATGGCATTGTGGCACA 1140
 259   A  G  A  I  Y  G  Y  P  V  L  K  E  I  R  N  G  I  V  A  Q  278
1141 ATTTGAGCACACAATCATTGTTGAAAAGGATTCTGTGATAGTGACGACAGAATGAGTTAA 1200
 279   F  E  H  T  I  I  V  E  K  D  S  V  I  V  T  T  E  *        295
1201 ACTTTATAAGTTCTCATGTATCAAGAAATTGGGAGCGCCGGGGTAGCCTAGTCAGGGAAG 1260
1261 GCGCGGGACTCGA                                                 1273
```

Fig. 2. **Nucleotide sequence of the *map* gene regions and the deduced amino acid sequence of the *P.f.* MAP.** The nucleotide numbering starts with the 5′-untranslated region of the DNA fragment. Amino acid numbering is shown beneath the nucleotide numbers. The amino acid residues directly identified by sequence analysis of the N-terminus of the intact *P.f.* MAP are underlined.

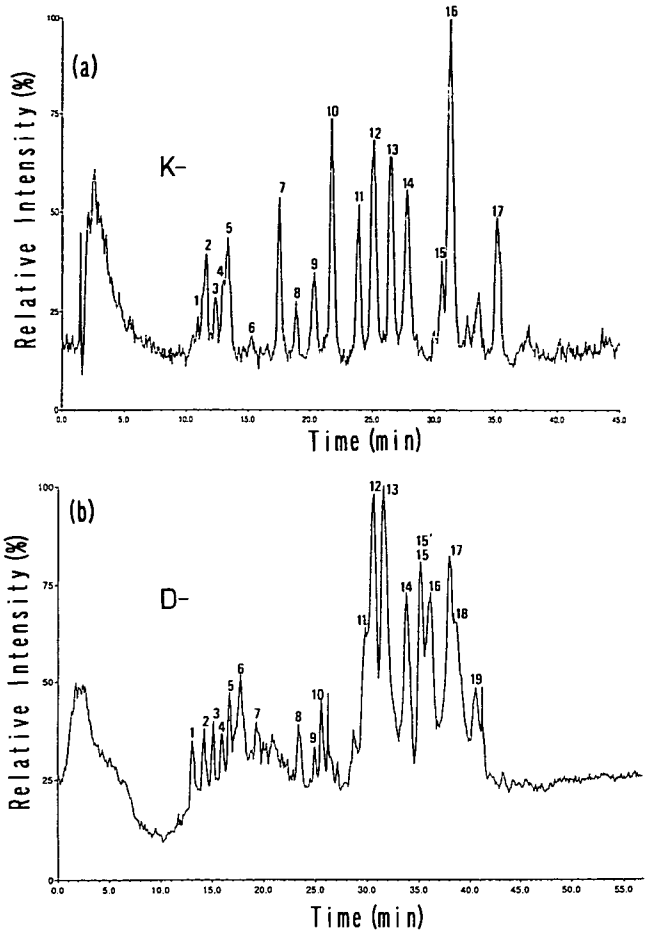

Fig. 3. **Total ion current chromatograms for the digests from the recombinant *P.f.* MAP with lysylendopeptidase (a), and with Endoproteinase Asp-N (b).** Peaks are numbered according to their order in elution on HPLC.

TABLE II. **Isolated peptides by digestion of the recombinant *P.f.* MAP with lysylendopeptidase (K1-K17), and with Endoproteinase Asp-N (D1-D19) and their molecular weights.**

Peptide No.	Position	Theoretical mass	Observed mass
D11[a]	M1-A34	3,796.7	3,797.0
D17[a]	M1-G76	8,372.0	8,372.3
D18	M1-G68	7,528,1	7,528.8
K16	L23-K39	1,883.3	1,882.7
K7	M40-K48	878.1	877.8
K13	P49-K67	2,245.6	2,245.6
D5[a]	E58-G68	1,249.7	1,248.9
K2	G68-K74	732.8	732.5
K3	E75-K80	723.8	723.5
D7	D77-I81	650.8	650.5
D8[a]	D77-I87	1,271.5	1,270.9
D5	D82-I87	638.7	638.2
D1	D88-A92	521.6	521.2
K15	I81-K112	3,431.9	3,432.0
D6	D93-E104	1,306.5	1,305.9
D15	D105-H173	7,676.0	7,675.6
D9[a]	E109-K129	2,111.4	2,111.7
D13[a]	E109-H173	7,187.4	7,187.8
K10	E113-K129	1,712.0	1,712.0
K4	A134-K141	972.1	971.8
K8[a]	R142-K154	1,451.8	1,452.4
K6	P146-K154	964.1	964.0
K1	I155-K159	707.8	708.0
K11	L160-K179	2,320.7	2,320.5
D4	D174-G181	937.0	936.8
K17	E180-K225	5,062.0	5,062.7
D10[a]	E180-N242	7,217.5	7,217.0
D14	D182-R210	3,210.8	3,210.4
D19[a]	D182-E295	12,866.1	12,865.7
D12	D211-N242	3,820.5	3,820.6
K14	R228-K250	2,813.2	2,813.7
D16[a]	D243-E295	5,870.9	5,870.0
D15′	D243-K287	5,026.0	5,025.4
K9	A259-K269	1,151.4	1,151.0
K12	E270-K278	2,096.4	2,097.4
K5	D288-E295	862.9	862.8
D3	D288-E295	862.9	862.5

[a]Peptides produced by either partial or non-specific cleavages.

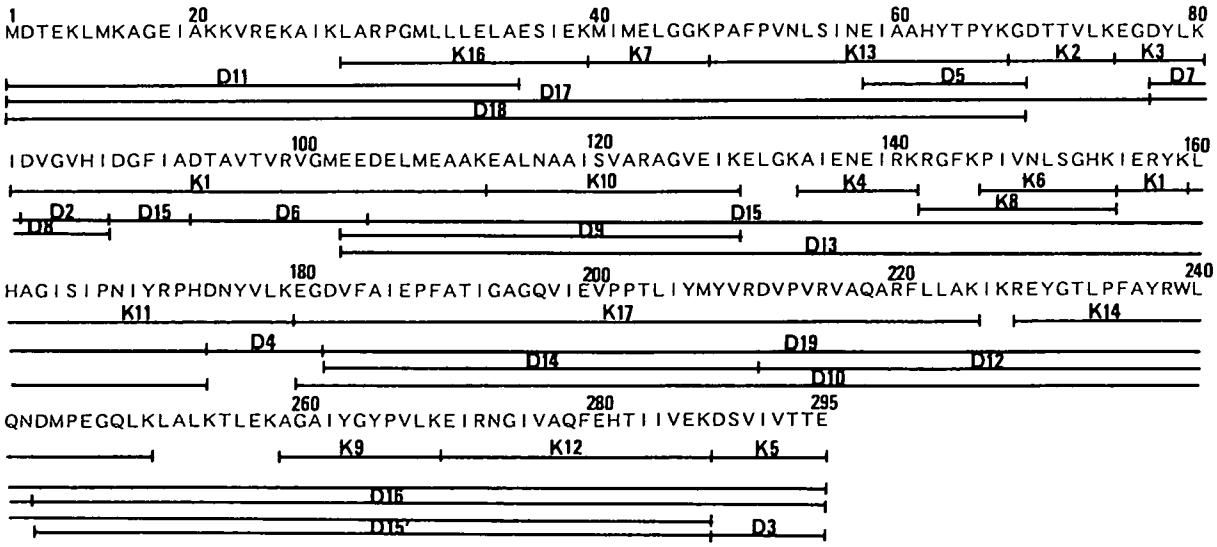

Fig. 4. **Confirmed amino acid sequence of the *P.f.* MAP.** Peptides produced by digestion with lysylendopeptidase, and with Endoproteinase Asp-N are numbered K1 to K17, and D1 to D19, respectively, as described in Fig. 3.

recombinant *P.f.* MAP were examined. The optimum pH was 7-8 and the optimum temprature was around 90°C (Fig. 5). The MAP activity was completely retained after heating at 75°C for 60 min in the range of pH 4.5 to 10.5, but its initial value was reduced to one-half after 4.5 h at 90°C at pH 7.5. Since the MAP partially purified from crude extract of *P. furiosus* by the same procedures as for the recombinant protein exhibited the same properties against temprature and pH as the recombinant enzyme, the recombinant MAP was used in the following experiments. By dialysis against 50 mM Na-PIPES buffer (pH 7.5) containing 0.1 mM EDTA, the enzyme activity was completely lost, but addition of 1 mM $CoCl_2$, followed by incubation at 95°C for 10 min, restored 94.6% of the activity. Of other metal ions, Mg^{2+} restored 33% of the original enzyme activity, whereas Zn^{2+} and Mn^{2+} showed no effect. This behavior suggests that *P.f.* MAP from Archaea is a metalloenzyme utilizing Co^{2+} ion for expression of the activity similarly to other MAPs from both Eukarya and Bacteria. The resistance of the enzyme activity against several denaturants was examined. No loss of the activity was observed on preincubation with 0.5 mM $CoCl_2$, 20 mM Na-PIPES buffer (pH 7.5) containing either 0.01% SDS or 8 M urea at 37°C for 1 h, whereas about 75% of the original activity was lost when 1 M guanidine-HCl was used as the denaturant. The substrate specificity of the enzyme was tested using a series of peptides, Met-X-Ala-Ala- Ala, and it was found that the enzyme cleaved N-terminal methionine from these substrates whose penultimate amino acid residues were Gly, Ala, Ser, Thr, Pro, and Val. This completely agrees with the specificities of other MAPs reported so far (*11-14*). The specificity of the enzyme was also examined for a protein of which the N-terminal methionine was partially retained because of its hyperexpression in the *E. coli* host. As such a protein, the recombinant human fibronectin fragment (C281) expressed in *E. coli* was used. Table III shows the N-terminal amino acid sequences of the recombinant proteins before and after treatment with the recombinant *P.f.* MAP in 10 mM Na-PIPES buffer (pH 7.5) containing 0.2 mM $CoCl_2$ at 75°C for 5 h at a molar substrate-to-enzyme ratio of 300 to 1. These results indicate the MAP releases only the N-terminal methionine residue of protein substrates and has no other endo-/exoproteolytic activities.

Sequence Comparison with MAPs from Other Species— Several features of the structural, functional, and evolutional characteristics of *P.f.* MAP are apparent from its amino acid sequence. From amino acid sequences of MAPs determined so far, as shown in Fig. 6 (*11, 14, 15-20*), five cobalt-binding residues (R1-R5) in the catalytic domain are conserved in all of them, but they have been classified to two types (Type II or I) depending on whether or not an extra long insertion of approximately 60 residues is present between R4 and R5 residues (*14, 18*). Type I consists of about 260 amino acid residues and includes MAPs from *E. coli* (*11*), *Salmonella typhimurium* (*15*), *Bacillus subtilis* (*16*), and *Haemophilus influenzae* (*19*), and Type II includes those from human (*18*) and rat (*13*). In *Saccharomyces cerevisiae*, a member of the Eukarya, two MAPs, one each of Types I and II, have been found (*14*). One more striking characteristic of eukaryotic MAPs is the N-terminal extension of about 90-150 residues containing basic and/or acidic clusters or a zinc-finger domain, which has little effect on the catalytic properties of enzymes, whether of Type I or II. Based on these facts, *P.f.* MAP seems to be close to the eukaryotic type, although it lacks the N-terminal extension observed in all eukaryotic MAPs found so far. The amino acid sequence of the MAP from *Methanococcus jannaschii*, a member of the Archaea, recently deduced from its gene (*20*) has been found to be highly homologous to *P.f.* MAP. These facts suggest that MAPs from Archaea are located at the Eukarya-Bacteria boundary. At this stage, we cannot speculate why MAPs diverged into two types with completely the same substrate specificities or what the functional importance is of the N-terminal extension observed in MAPs of Eukarya, but the results obtained here coincide with the report that the proteins involved in transcription, translation, and replication in Archaea are more similar to those found in Eukarya (*20*). From a model-building study performed for human MAP on the basis of the 3-D structure of the MAP from *E. coli* (*18-22*) and the result of our preliminary X-ray analysis for *P.f.* MAP, it appears that the insertion of about 60 residues between R4 and R5 in Type II itself has an ordered

TABLE III. **N-terminal sequence analyses of the recombinant human fibronectin fragment (C281) with or without treatment by *P.f.* MAP.**

Sequence	Met	Ala	Pro	Thr	Asp	Leu
Without treatment	137.6[a]	156.6	154.7	109.3	126.5	151.8
With treatment	103.5	83.4	65.6	58.9	83.5	

[a]Values represent the raw amounts (pmol) of PTH-amino amino acids obtained by sequence analyses of the fibronectin fragments (C281, 680 pmol each) with or without the treatment by *P.f.* MAP.

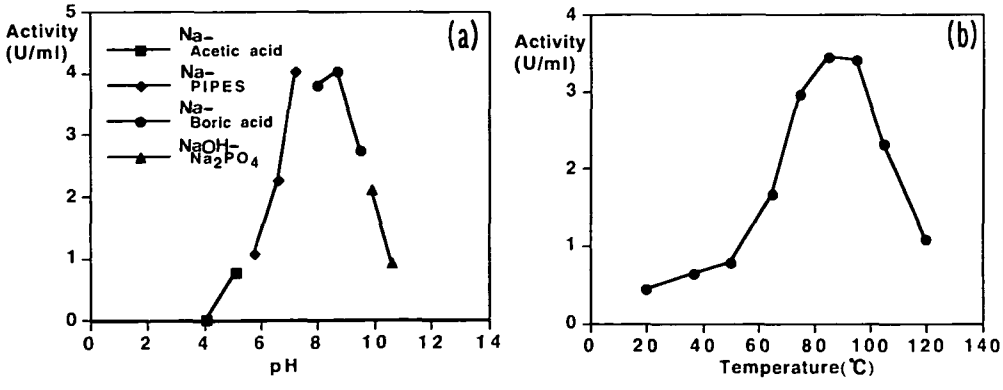

Fig. 5. **Effects of pH (a) and temperature (b) on the activity of *P.f.* MAP.** Details are described in the text.

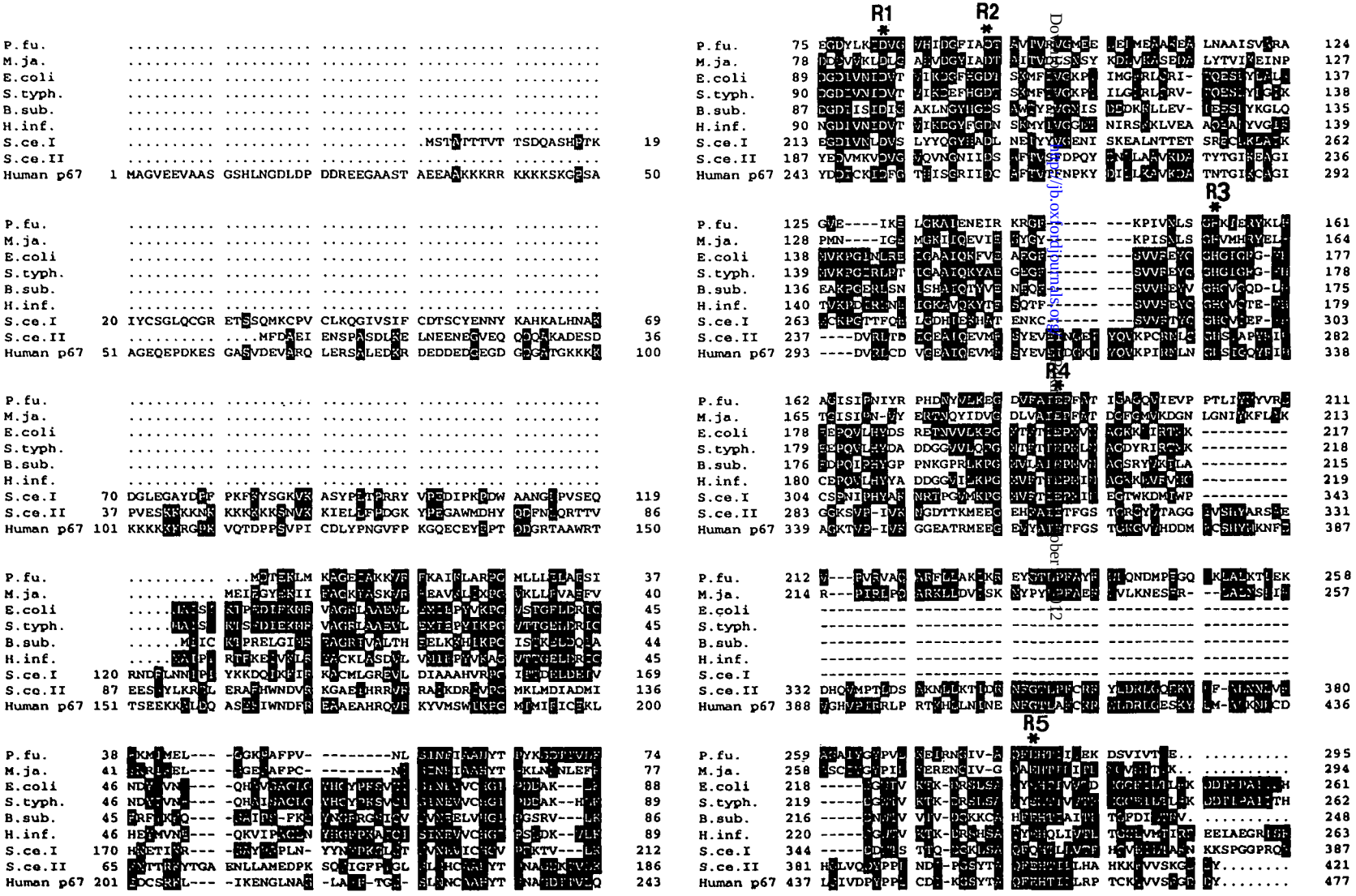

Fig. 6. **Amino acid sequence alignment of *P.f.* MAP with those of MAPs from other species.** Residues that are cobalt ligands estimated from the structure of *E. coli* MAP are indicated by asterisks (R1–R5). Shaded residues are identical in two or more MAPs. Abbreviations are as follows: P.fu, *Pyrococcus furiosus*; M.ja, *Methanococcus jannaschii*; S.typh, *Salmonella typhimurium*; B.sub, *Bacillus subtilis*; H.inf; *Haemophilus infulenzae*; S.ce.I, *Saccharomyces cerevisiae* AP1 (Type I); S.ce.II, *Saccharomyces cerevisiae* AP2 (Type II).

structure, but the backbone of Type II is retained exactly as in the Type I structure reported for *E. coli*. Therefore, it is of importance for understanding the evolution of cotranslational processing including MAP activity *in vivo* to clarify the functions of both the N-terminal extension and the insertion between R4 and R5 observed in MAPs of Eukarya in comparison with MAPs from Bacteria and Archaea. On the other hand, according to the structural basis of hyperthermostability of *P.f.* MAP, it is notable that the contents of charged amino acids, especially Glu and Lys residues, are higher in the part outside the extra insertion of 62 residues than in the MAP from *E. coli*. This suggests the contribution of electrostatic interaction in stabilizing *P.f.* MAP. However, to elucidate the structural characteristics of this hyperthermophilic enzyme, biophysical analyses for thermal stability of various designed mutants of *P.f.* MAP under different conditions of pH, ionic strength, temperature *etc.* by differential scanning calorimetry and CD spectrometry are required, in addition to information of the 3-D structure on the enzyme. Such studies are now under progress and will be published elsewhere. Recombinant proteins produced at high levels in *E. coli* often retain an N-terminal methionine residue (*23*). Such phenomena may be the result of a compromise, either by saturation of the MAP from host cells or by inclusion-body formation of the recombinant proteins. This is illustrated by the facts that some purified N-terminally methionylated recombinant proteins can be readily demethionylated *in vitro* with MAPs (*11, 24*). The MAP therefore can be applied to the *in vitro* removal of unprocessed N-terminal methionine from proteins produced by recombinant techniques. In particular, the *P.f.* MAP produced in *E. coli* here will be the most suitable for such industrial application in terms of its easier purification with high yield, enhanced thermostability, and resistance against such denaturants as 8 M urea and 0.01% SDS, and higher specific activity even at moderate temperature in comparison with other MAPs obtained so far.

REFERENCES

1. Tsunasawa, S. and Sakiyama, F. (1992) Amino-terminal acetylation of proteins in *The Posttranslational Modification of Proteins* (Tsuboi, S., Taniguchi, N., and Katsunuma, N., eds.) pp. 113-121, CRC Press, Tokyo
2. Moerschell, R.P., Hosokawa, Y., Tsunasawa, S., and Sherman, F. (1990) The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine *in vivo*. *J. Biol. Chem.* **265**, 19638-19643
3. Woesse, C.R., Kandler, O., and Wheelis, M.L. (1990) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eukarya. *Proc. Natl. Acad. Sci. USA* **87**, 4576-4579
4. Fiala, G. and Stetter, K.O. (1986) *Pyrococcus furiosus*, sp. nov., represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Arch. Microbiol.* **145**, 56-61
5. Laderman, K.A., Davis, B.R., Krutzsch, H.C., Lewis, M.C., and Anfinsen, C.B. (1993) The purification and characterization of an extremely thermostable α-amylase from the hyperthermophilic Archaebacterium *Pyroccous furiosus*. *J. Biol. Chem.* **268**, 24394-24401
6. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467
7. Ishino, Y. (1992) Rapid and reliable DNA sequencing with a dideoxy sequencing kit. *Am. Biotechnol. Lab.* **10**, 47
8. Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254
9. Simpson, R.J., Neuberger, M.R., and Liu, T.-Y. (1976) Complete amino acid analysis of proteins from a single hydrolysate. *J. Biol. Chem.* **251**, 1936-1940
10. Reiter, W.-D., Huneepohl, U., and Zilling, W. (1990) Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection *in vitro*. *Proc. Natl. Acad. Sci. USA* **87**, 9509-9513
11. Ben-Bassat, A., Bauer, K., Chang, S.-Y., Myambo, K., Boosman, A., and Chang, S. (1987) Processing of the initiation methionine from proteins: Properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. *J. Bactriol.* **169**, 751-757
12. Chang, Y.-H., Teichert, U., and Smith, J.A. (1990) Purification and characterization of a methionine aminopeptidase from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**, 19892-19897
13. Kendall, R.L. and Bradshaw, R.A. (1992) Isolation and characterization of the methionine aminopeptidase from porcine liver responsible for the cotranslational processing of proteins. *J. Biol. Chem.* **267**, 20667-20673
14. Li, X. and Chang, Y.-H. (1995) Amino-terminal protein processing in *Saccharomyces cerevisiae* is an essential function that requires two distinct methionine aminopeptidases. *Proc. Natl. Acad. Sci. USA* **92**, 12357-12361
15. Wingfield, P., Graber, P., Turacatti, G., Movva, N.R., Pelletier, M., Craig, S., Rose, K., and Miller, C.G. (1989) Purification and characterization of a methionine-specific aminopeptidase from *Salmonella typhimurium*. *Eur. J. Biochem.* **180**, 23-32
16. Nakamura, K., Nakamura, A., Takamatsu, H., Yosikawa, H., and Yamane, K. (1990) Cloning and characterization of a *Bacillus subtilis* gene homologous to *E. coli* SecY. *J. Biochem.* **107**, 603-607
17. Chang, Y.-H., Teichert, U., and Smith, J.A. (1992) Molecular cloning, sequencing, deletion, and overexpression of a methionine aminopeptidase gene from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **267**, 8007-8011
18. Arfin, S.M., Kendall, R.L., Hall, L., Weaver, L.H., Stewart, A.E., Matthews, B.W., and Bradshaw, R.A. (1995) Eukaryotic methionyl aminopeptidases: Two classes of cobalt-dependent enzymes. *Proc. Natl. Acad. Sci. USA* **92**, 7714-7718
19. Fleischmann, R.D., Adams, M.D. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512
20. Bult, C.J., White, O. *et al.* (1996) Complete genome sequence of the methanogenic Archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058-1073
21. Roderick, S.L. and Matthews, B.W. (1993) Structure of the cobalt-dependent methionine aminopeptidase from *Escherichia coli*: a new type of proteolytic enzyme. *Biochemistry* **32**, 3907-3912
22. Bazan, J.F., Weaver, L.H., Roderick, S.L., Huber, R., and Matthews, B.W. (1994) Sequence and structure comparison suggest that methionine aminopeptidase, prolidase, aminopeptidase P, and creatinase share a common fold. *Proc. Natl. Acad. Sci. USA* **91**, 2473-2477
23. Rose, K., Savoy, L.-A., Simona, M.G., Offord, R.E., Wingfield, P.T., Mattaliano, R.J., and Thatcher, D.R. (1987) The state of the N-terminus of recombinant proteins: determination of N-terminal methionine (formylated, acetylated, or free). *Anal. Biochem.* **165**, 59-69
24. Miller, C.G., Strauch, K.L., Kukral, A.M., Miller, J.L., Wingfield, P.T., Mazzei, G.J., Werlen, R.C., Graber, P., and Movva, N.R. (1987). N-terminal methionine-specific peptidase in *Salmonella typhimurium*. *Proc. Natl. Acad. Sci. USA* **84**, 2718-2722