Processing of the Initiation Methionine from Proteins: Properties of the *Escherichia coli* Methionine Aminopeptidase and Its Gene Structure

ARIE BEN-BASSAT,^{1*} KEITH BAUER,¹ SHENG-YUNG CHANG,² KEN MYAMBO,² ALBERT BOOSMAN,³
AND SHING CHANG²

Departments of Fermentation Research and Development,¹ Microbial Genetics,² and Analytical Development,³ Cetus Corporation, Emeryville, California 94608

Received 14 July 1986/Accepted 21 November 1986

Methionine aminopeptidase (MAP) catalyzes the removal of amino-terminal methionine from proteins. The *Escherichia coli* map gene encoding this enzyme was cloned; it consists of 264 codons and encodes a monomeric enzyme of 29,333 daltons. In vitro analyses with purified enzyme indicated that MAP is a metallo-oligopeptidase with absolute specificity for the amino-terminal methionine. The methionine residues from the amino-terminal end of the recombinant proteins interleukin-2 (Met-Ala-Pro-IL-2) and ricin A (Met-Ile-Phe-ricin A) could be removed either in vitro with purified MAP enzyme or in vivo in MAP-hyperproducing strains of *E. coli*. In vitro analyses of the substrate preference of the *E. coli* MAP indicated that the residues adjacent to the initiation methionine could significantly influence the methionine cleavage process. This conclusion is consistent, in general, with the deduced specificity of the enzyme based on the analysis of known amino-terminal sequences of intracellular proteins (S. Tsunasawa, J. W. Stewart, and F. Sherman, J. Biol. Chem. 260:5382-5391, 1985).

Protein synthesis is always initiated with either methionine or *N*-formylmethionine. The *N*-formyl moiety that is present on proteins from organelles and bacteria is removed by deformylases, leaving methionine at the amino terminus. For a significant fraction of the intracellular proteins, the amino-terminal methionine is removed enzymatically after the initiation of translation (1, 2, 5, 10, 24, 33, 34, 38, 40; for a review, see reference 28). Although other mechanisms may also be involved in the post- or cotranslational modifications of the amino-terminal residues of some proteins (e.g., signal peptidases that remove the secretory signal peptide sequences from certain transported proteins, transacetylases that acetylate the amino-terminal residues of some eucaryotic proteins), methionine is removed from the majority of proteins. However, very little is known about the structure and properties of the methionine aminopeptidase (MAP) that is responsible for removing the amino-terminal methionine residue from the cellular proteins. From a systematic analysis of the amino-terminal sequences data for various mutant forms of yeast iso-1-cytochrome *c*, as well as from the data for 82 mature intracellular proteins, Sherman et al. (28) concluded that the MAPs from various procaryotes and eucaryotes share similar substrate specificity. Surveys of the protein sequence data bank by Boissel et al. (4) and Sherman et al. (28) provided similar deduced general sequence features in amino-terminal methionine processing.

For *Escherichia coli* and *Salmonella typhimurium*, many peptidase genes have been mapped, and their cognate peptidases have been characterized (6, 9, 22, 23, 29, 30, 32, 35, 36). However, none reported so far corresponds to the activity postulated for MAP. We took a direct approach to study the MAP of *E. coli* by molecular cloning of its gene (*map*) and identified the gene based on the postulated properties of the gene product. An *E. coli* strain deficient in five peptidases was transformed with a plasmid preparation

containing a library of cloned genomic DNA fragments from the same strain. The lysates from individual transformants were then prepared and analyzed for the presence of elevated MAP activity. This allowed us to identify the *E. coli* *map* gene from the gene library. We determined the coding sequences of the *map* gene and purified and characterized the *E. coli* MAP from a hyperproducer strain.

MATERIALS AND METHODS

Bacterial strains and media. *E. coli* CM89 [*leu-9* Δ (*pro-lac*) *met thyA pepN102 pepA11 pepB1 pepQ10*], deficient in peptidases A, B, D, N, and Q, was provided by C. G. Miller (23). Strain MM294 (19) was used as the host for routine transformation experiments. The *lacI*^q strains DG99 and DG98 (provided by D. Gelfand [16]) were used as the hosts for pUC18 and pUC19 plasmids and for M13 phage, respectively. The vectors pUC18 and pUC19 (39) and coliphages M13mp10 and M13mp11 (20) were used for molecular cloning and sequence analysis. Plasmid pSYC795 is a derivative of plasmid pSYC423 (15) that contains the promoterless penicillinase gene (*penP*), except that pSYC795 contains the Cys-to-Ser substitution at position 27 in *penP* (8) and that the *Bam*HI site distal to *penP* on pSYC423 was removed. Plasmids pSYC1143 and pRAP229 contain the cDNA sequences encoding human interleukin-2 (IL-2) and ricin A chain, respectively. They are both derived from pACYC184 (7), which is compatible with pUC18. Plasmid pSYC1143 carries a 0.95-kilobase-pair (kb) *Eco*RI fragment inserted at the unique *Eco*RI site in pACYC184; this insert, which originated from plasmid pFC51.T (provided by F. Laywer and D. Gelfand), contains the IL-2 sequence (26) under the control of the *E. coli* *trp* promoter. Plasmid pRAP229 (provided by M. Piatak) contains the coding sequence for ricin A chain under the control of the *phoA* promoter (M. Piatak, personal communication). *E. coli* cells were routinely grown in Penassay broth (Difco Laboratories) or on Penassay base agar (Difco) unless otherwise specified.

* Corresponding author.

Ampicillin was added to a final concentration of 50 $\mu\text{g/ml}$ for selection of resistant cells. Supplemented E medium is the E medium of Vogel and Bonner (37) supplemented with the following (each at 5% [vol/vol]): 2 $\times$ tryptose broth (Difco) plus 1% NaCl, 10% Casamino Acids (Difco), and 10 $\times$ yeast nitrogen base medium (Difco).

Molecular cloning and sequencing techniques. All restriction endonucleases, alkaline phosphatase, DNA polymerase, and nucleases were purchased from either Bethesda Research Laboratories, Inc., or New England BioLabs, Inc. Standard techniques for molecular cloning were used (13). The dideoxynucleotide chain termination method of DNA sequencing (27) was used to determine the sequence of both strands of the *map* gene fragment.

Construction and screening of the genomic library. Plasmid pUC18 was digested with *Bam*HI endonuclease to completion and then treated with alkaline phosphatase. Total DNA from strain CM89 was isolated, partially digested with endonuclease *Sau*3A, and fractionated by centrifugation in a sucrose gradient. Fractions containing 4- to 8-kb DNA fragments were pooled and used for the construction of the gene library. Size-selected chromosomal fragments were ligated to dephosphorylated pUC18 vector DNA (4:1, wt/wt) and transformed into *E. coli* DG99. Analysis of a small sample of the transformants showed that more than 95% of the plasmids contained inserts that ranged between 4 and 8 kb. Plasmid DNA prepared from these transformants was used to transform competent cells of strain CM89 for further screening.

Individual colonies of CM89 cells transformed with the pUC18 plasmids from the gene library were picked from plates and then transferred into the wells of microtiter plates containing 200 μl of supplemented E medium. After overnight growth at 37°C, cells were washed twice in 0.1 M Tris hydrochloride (pH 7.4). The cells were lysed by the addition of 20 μl of lysozyme solution (1 mg/ml in the same buffer) and subjected to three cycles of freezing and thawing; then 180 μl of 0.1 M potassium phosphate buffer (pH 7.4) containing 0.2 mM CoCl_2 , 72 μg of Met-Gly-Met-Met, 36 μg of L-amino acid oxidase, 5 μg of horseradish peroxidase, and 18 μg of *o*-dianisidine dihydrochloride was added to each well. Rapid color development indicates the presence of high peptidase activity in the lysates.

MAP activity assay. The enzymatic assay was a modification of the method described by Carter and Miller (6). The protein solution (10 μl) to be assayed was added to 90 μl of the substrate solution containing 4 mM Met-Gly-Met-Met, 0.1 M potassium phosphate buffer (pH 7.5), and 0.2 mM CoCl_2 . The tubes (13 by 100 mm) containing the reaction mixtures were incubated at 30°C for 10 min. The reaction was stopped by placing the tubes in a boiling water bath for 2 min. After the addition of 0.9 ml of the color development mixture (0.1 M Tris hydrochloride [pH 7.4] containing 0.2 mg of L-amino acid oxidase, 0.03 mg of horseradish peroxidase, and 0.2 mg of *o*-dianisidine dihydrochloride), the tubes were incubated for 30 to 60 min at 30°C, and the optical density at 440 nm was recorded. The readings, corrected against an enzyme blank, were expressed as μmol s of product by using the empirically determined conversion factor of 1 μmol of methionine per ml = 8.6 A_{440} . One unit of activity is defined as 1 μmol of amino acids produced per min under the assay conditions used.

Alternatively, peptidase activities were assayed by thin-layer chromatography using silica gel plates with a solvent mixture consisted of *n*-butanol-acetic acid-water (60:25:15, vol/vol). Amino acids released were detected and identified

after the plates were sprayed with ninhydrin reagent and oven dried. Thin-layer chromatography was used as a qualitative assay for MAP activity to detect the release of amino acids that do not react with L-amino acid oxidase (3) and to verify the enzymatic assay results.

The oligopeptides either were purchased from Sigma Chemical Co., Research Plus, or Chemical Dynamics Corporation or were synthesized by the solid-phase synthesis method (17, 18).

Purification of MAP from the hyperproducer strain. About 1.3 g of wet cells was suspended in 25 ml of 0.2 M potassium phosphate buffer (pH 7.5) containing 0.2 mM CoCl_2 and 0.1 mM phenylmethylsulfonyl fluoride. The cell suspension was sonicated, and the extracts were centrifuged at 15,000 $\times g$ for 15 min. The proteins in the supernatant were fractionated by DEAE-Sepharose column chromatography, and the MAP was eluted with a 0-to-0.25 M NaCl linear gradient. Fractions containing peak MAP activity (0.1 M NaCl) were pooled. The sample was concentrated in an Amicon pressure cell with PM10 membrane and further fractionated with an S-200 Sephacryl column equilibrated with sonication buffer containing 1 mM methionine.

In vivo and in vitro processing of recombinant proteins by MAP. Plasmid pSYC1174 directs the constitutive synthesis of high levels of MAP in *E. coli*. *E. coli* MM294(pSYC1143) and MM294(pRAP229) express the cloned IL-2 under the *trp* promoter and the cloned ricin A chain under the *phoA* promoter, respectively. Plasmid pSYC1174 was introduced into these strains by transformation. To induce the synthesis of the recombinant proteins, strains carrying the IL-2 cDNA were induced by tryptophan starvation (26), and strains carrying the ricin A chain cDNA were induced by phosphate limitation as described by Michaelis et al. (21). The recombinant proteins were purified from the strains with and without plasmid pSYC1174, and their N-terminal sequences were determined by quantitative Edman degradation.

To determine the in vitro processing of the recombinant proteins by MAP, each purified recombinant protein (100 μg) prepared from strain MM294(pSYC1143) or from strain MM294(pRAP229) was incubated with 6 μg of MAP in a reaction mixture containing 0.2 mM CoCl_2 in 100 mM potassium phosphate buffer (pH 7.5). After overnight incubation at 30°C, the reactions were stopped by heating. No kinetic studies were conducted to investigate the optimal conditions and the time courses of these reactions. The proteins were desalted, and the N-terminal sequence was determined.

RESULTS

Isolation of the *E. coli map* gene. *E. coli* K-12 CM89 is deficient in peptidases A, B, D, N, and Q. It was chosen for this work because of its low peptidase activity against Met-Gly-Met-Met and Leu-Gly-Gly, the screening substrates. Peptidases A, B, and N have broad specificity and can degrade these substrates as well (22, 23). We prepared a gene library of the CM89 genome by using plasmid pUC18 as the cloning vector. This library of plasmids that contained inserts ranging mostly between 4 and 8 kb was introduced into CM89 cells by transformation. About 1,000 transformants were picked and cultured individually in the wells of microtiter plates. We prepared lysates of these cultures and determined the ability of crude extracts to release methionine from the tetrapeptide Met-Gly-Met-Met in a semi-quantitative assay. Ten clones that showed significantly higher levels of activity were identified from these cultures.

FIG. 1. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis profiles. (A) Total cellular proteins from strains CM89 (lane 1), CM89(pSYC1174) (= strain 18E7) (lane 2), CM89(pSYC1187) (lane 3), and CM89(pSYC1188) (lane 4). (B) Purification of MAP from strain 18E7. Lane 1, total cell extract; lanes 2 and 3, duplicated samples from pooled active fractions after DEAE-Sepharose column purification, lanes 4 and 5, duplicate from pooled active fractions after S-200 Sephacryl column purification (see Table 1). Samples were fractionated on 12.5% (A) and 15% (B) acrylamide gels according to the method of Laemmli (11), and protein bands were visualized by Coomassie blue staining. The sizes (kDa) of the molecular weight markers are shown.

They were further assayed for their activity to release leucine from the tripeptide Leu-Gly-Gly (a substrate for peptidase T [32]). One clone (isolate 18E7) among the 10 tested showed the parental (i.e., CM89 strain) level of activity against this tripeptide; it was saved for further studies. The other nine clones exhibited elevated peptidase activity but showed no specificity for the amino-terminal methionine; they probably carry a cloned gene or genes encoding other peptidases.

Preliminary analysis of strain 18E7 showed that it expressed a high level of peptidase activity against the amino-terminal methionine residues from a variety of oligopeptide substrates. Using the synthetic tetrapeptide Met-Gly-Met-Met as the substrate in an *in vitro* aminopeptidase assay, we found that the MAP level in strain 18E7 was about 100-fold higher than that found in the parental CM89 strain. The total

cellular protein from strain 18E7 was prepared and fractionated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and compared with that of the parent strain (Fig. 1). In addition to the 30-kDa band corresponding to the beta-lactamase encoded by the *bla* gene on the pUC18 vector, there was another prominent protein band present in the extracts of strain 18E7. This protein, having an apparent size of 32 kDa, was identified as MAP by activity assay and deletion analysis (described below). The MAP represents about 15 to 20% of the total cellular protein as estimated by densitometry measurement of the stained gel. The hyperproduction of MAP in strain 18E7 probably results from the amplification of the *map* gene copy number.

Structure of the *map* gene. The recombinant plasmid in strain 18E7, designated pSYC1174, was isolated for detailed analysis. It carries a chromosome insert of 3.2 kb in the pUC18 vector at the *Bam*HI site. The restriction map of the insert is shown in Fig. 2. Three derivatives of pSYC1174 were constructed for deletion mapping of the cloned *map* gene; they each contained a deletion spanning a different region within the insert.

The strains harboring these plasmids were analyzed for MAP overproduction, and the deletion mapping results localized the *map* gene between the 5' end of the insert (adjacent to the *Eco*RI site on the vector) and the first *Nco*I site (Fig. 2). We subsequently excised the 1.2-kb fragment within this region by digestion with *Eco*RI and *Cla*I endonucleases and inserted it into the pUC18 and pUC19 plasmids between the *Sma*I and the *Eco*RI sites, generating plasmids pSYC1187 and pSYC1188, respectively. CM89 strains harboring these plasmids, as well as their parental plasmid, produced comparably high levels of MAP (Fig. 1). The results of these analyses suggest that the *lac* promoter on the vector is not required for high-level *map* expression and that both the coding sequence of the *map* gene and the *map* promoter are located within this 1.2-kb *Eco*RI-*Cla*I fragment.

The entire nucleotide sequence of the 1.2-kb fragment containing the *E. coli map* gene was determined (Fig. 3). A long open reading frame starting with ATG is present between positions 219 and 1010; this open reading frame encodes a protein of 264 amino acids. The calculated molecular weight is 29,333, which is slightly smaller than the estimated size of 32 kDa for the *E. coli* MAP. The amino acid sequence of the amino-terminal 64 residues of the purified MAP was determined; it corresponds to the deduced amino acid sequence of the longest open reading frame in the region

FIG. 2. The cloned *map* fragment in plasmid pSYC1174. (A) Restriction map. The restriction sites within the fragment (those in parentheses have not been precisely mapped), the location, and the direction of transcription of the *map* gene (open box and the arrow) are shown. The polylinker sites from the pUC18 vector are also indicated at the ends of the fragment. (B) Deletion map. Three deletions were created from pSYC1174, each covering a unique region (dashed lines). The MAP hyperproduction phenotypes of the resulting plasmids are indicated by + (observed) or - (not observed).

FIG. 3. Structure of the *map* gene. The 1,197-base-pair fragment containing the *map* gene and the predicted protein sequence of the MAP enzyme are shown. The location of the *Bss*HII, *Sma*I, and *Cla*I sites and the two regions containing the inverted repeats (<<<< . . . >>>>) are indicated. The positions of the nucleotides in the fragment and the predicted amino acids in MAP are indicated at the left and the right margins, respectively.

(Fig. 3), except that the purified protein does not have the amino-terminal methionine residue (data not shown). These results establish the precise location of the *map* gene within the cloned chromosomal DNA fragment.

The 1.2-kb *Eco*RI-to-*Cla*I fragment from plasmid pSYC1174 was used to probe the genomic DNA of *E. coli* by the method of Southern (31). It hybridized to unique bands of 3.6 and 1.35 kb that were generated by digestions with *Pst*I and *Bss*HII, respectively. The *map* sequence did not hybridize to other regions of the chromosome, indicating that the *map* gene is present as a single copy in *E. coli* and that it does not share a high degree of homology with any of the other peptidase genes at the DNA level.

At the end of the *map* coding sequence, there is an inverted repeat, AAGCCGGC----GCCGGCTT (positions 1018 to 1037), with four more T's at the end of the repeat; this feature is characteristic of rho-independent transcription terminators (25). In addition, nonsense codons are present in all three reading frames after the putative terminator sequence. Thus, we consider it likely that there are no cotranscribed genes downstream from *map*. Upstream from the coding region, there are only 218 nucleotides that are derived from chromosomal DNA in pSYC1174. Within this region lie the promoters for the *map* gene (see below).

***map* is transcribed from tandem promoters.** The location of the transcription start site of *map* was investigated by deletion analyses. We first deleted the sequence between positions 1 and 150 on pSYC1174 by digestion with *Sma*I endonucleases (the upstream *Sma*I site is derived from the pUC18 polylinker sequence [Fig. 2]). The resulting plasmid, pSYC1224, directed high-level MAP synthesis in both the CM89 host and the *lacI*^q strain DG99 as well as did its parent plasmid, suggesting that the promoter for *map* is probably located within the sequence between positions 151 and 218. However, when we inserted the nominally 150-base-pair *Sma*I-*Sma*I fragment in front of the promoterless *penP* (penicillinase) gene on the promoter-testing vector, pSYC795, with the *Eco*RI site distal to *penP*, we obtained *penP* expression in *E. coli* transformants. This result sug-

gests that there were at least two tandem *map* gene promoters separated by the *Sma*I site at position 150.

Characterization of MAP. We prepared a cell extract from strain CM89(pSYC1174) and purified the MAP protein by using a DEAE-Sepharose column. The MAP activity eluted at 0.10 M NaCl. The active fractions were pooled and concentrated and then were fractionated on an S-200 Sephacryl column; the results are summarized in Table 1. The increase in the total activity of MAP after the DEAE-Sepharose purification step might be due to the removal of inhibitory compounds from the cell extract. The *E. coli* MAP prepared by the above procedure was judged to be more than 95% pure (Fig. 1).

Purified MAP was further analyzed (Table 2). MAP is inactive in the presence of EDTA; therefore, it is a metalloproteinase. Cobalt (Co²⁺) is essential for enzyme activity, and Mn²⁺, Cu²⁺, Zn²⁺, or Mg²⁺ ions could not substitute for Co²⁺. The activity of MAP was severely reduced when we substituted Tris buffer for potassium phosphate buffer, but the addition of sodium or potassium salts restored the MAP activity in Tris buffer. Based on gel filtration data (not shown), this native enzyme behaves as a monomer.

Substrate specificity of MAP. The substrate specificity of the purified MAP was investigated (Table 3). MAP utilized only tripeptide or larger substrates and had an absolute specificity for a methionine residue at the amino terminus; Met(sulfoxide)-Ala-Ser was not a substrate for MAP. The ability to remove the amino-terminal methionine by the *E.*

TABLE 1. Purification of MAP from strain CM89(pSYC1174)

Purification stage	Vol (ml)	Total protein (mg)	Total activity (U) ^a	Sp act (U/mg)	Purification (fold)	Yield (%)
Crude extract	21.0	103.0	144.9	1.41	1.0	100
DEAE-Sepharose	2.5	18.8	198.0	10.56	7.5	136
S-200 Sephacryl	1.8	10.8	122.0	11.30	8.01	84

^a One unit of activity equals 1 μmol of amino acids produced per min.

TABLE 2. Reaction conditions for MAP^a

Buffer system (concn, mM)	Relative activity (% of control)
A	
Complete; in KPB.....	100
-CoCl ₂	16
+EDTA (2.5).....	3
-CoCl ₂ + ZnCl ₂ (1).....	5
-CoCl ₂ + MnCl ₂ (1).....	16
-CoCl ₂ + MgCl ₂ (1).....	15
B	
Complete; in Tris hydrochloride.....	0.6
+ KCl (25).....	76
+ K ₂ SO ₄ (25).....	94
+ Na ₂ SO ₄ (50).....	76
+ KPB (25).....	113
- Tris hydrochloride + KPB (50).....	100
Tris hydrochloride (6 instead of 25).....	4
Tris hydrochloride (6 instead of 25) + CoCl ₂ (10).....	2.3

^a Buffer system A consisted of 50 mM potassium phosphate buffer (KPB) (pH 7.5), 0.2 mM CoCl₂, 4 mM Met-Gly-Met-Met, and 1 µg of purified MAP. All activities are expressed relative to this set of assay conditions. In system B, 25 mM Tris hydrochloride buffer (pH 7.5) was substituted for potassium phosphate of system A. The enzyme preparation (1 mg of protein per ml) used for the assays contained 0.2 mM CoCl₂. Assay conditions are described in Materials and Methods.

coli enzyme was significantly dependent on the properties of the adjacent amino acids. For example, MAP cleaved the methionine when the second residue was glycine, alanine, or proline, but not when the second residue was phenylalanine, leucine, methionine, glutamic acid, arginine, or lysine. The turnover rate of MAP with Met-Gly-Met-Met as a substrate was estimated to be 6 mol of product per mol of MAP per s. With other commercially available oligopeptides as the substrates, the rates were similar or lower (Table 3).

Cleavage of methionine by MAP from the amino-terminal end of proteins was tested in vitro by incubating MAP with purified recombinant proteins and in vivo by expression of the genes encoding the recombinant proteins in MAP-hyperproducing *E. coli* strains. *E. coli* strains expressing

human IL-2 (Met-Ala-Pro-IL-2) and *Ricinus communis* var. ricin A chain (Met-Ile-Phe-ricin A) sequences produced recombinant proteins that partially contained amino-terminal methionine. The purified recombinant proteins were used as substrates for the in vitro assay. In addition, for the in vivo MAP activity assay, the cloned *map* gene was transferred into the *E. coli* MM294 strains that contained the IL-2 or the ricin gene on a compatible plasmid. The results (Table 4) showed that MAP can catalyze methionine removal from these proteins in vitro and in vivo. In the case of Met-Ala-Pro-IL-2, a significant fraction (60%) of the in vivo-processed protein also lost the alanine residue, while no alanine removal was detected from the in vitro MAP reaction. Similar in vivo processing products were observed in the MM294 host and in the CM89 strain that is deficient in five peptidases. This result suggests that another aminopeptidase(s) might be responsible for the removal of the alanine residue from the recombinant IL-2 protein in vivo. It is interesting to note that a small fraction of the recombinant proteins, after being exposed to MAP either in vitro or in vivo, still retained their terminal methionines (Table 4). One possible explanation is that a small fraction of these proteins are folded such that their amino termini are not accessible to MAP.

DISCUSSION

Our data on the substrate specificity of MAP conform, in general, to the hypothesis of Sherman et al. (28). They proposed that the ability to remove the amino-terminal methionine by MAP is dependent on the structure of its adjacent residue. In most instances, an adjacent residue with a side chain size (radius of gyration) smaller than 0.129 nm is preferred; for side chain sizes larger than 0.143 nm, MAP is unable to cleave the methionine. Most of our experimental results support these conclusions. Glycine, alanine, and proline have 0-, 0.077-, and 0.125-nm radii of gyration, respectively, while arginine, lysine, phenylalanine, methionine, glutamic acid, and leucine have 0.238-, 0.208-, 0.190-, 0.180-, 0.177-, and 0.154-nm radii of gyration, respectively

TABLE 3. Substrate specificity of *E. coli* MAP on oligopeptides^a

Substrate sizes	Substrates cleaved	Sp act of MAP (U/mg)	Substrates not cleaved
Dipeptides			Met-Ala, Met-Ser, Met-Phe, Leu-Gly
Tripeptides	Met-Ala-Met	175	Met-[0]-Ala-Ser
	Met-Ala-Ser	108	Met-Phe-Gly
	Met-Gly-Gly	12	Met-Leu-Phe
	Met-Gly-Met	36	Met-Met-Met
	(+)	(+)	Met-Arg-Phe-(acetate), Leu-Gly-Gly, Thr-Gly-Gly, Val-Gly-Gly, Trp-Gly-Gly, Phe-Gly-Gly, Tyr-Gly-Gly, Ile-Gly-Gly, Arg-Gly-Gly, Leu-Leu-Leu, Ser-Ser-Ser, Ala-Ala-Ala, Thr-Val-Leu, Glu-Gly-Phe, N-formyl-Met-Met-Met
Tetrapeptides	Met-Gly-Met-Met	175	Met-Phe-Ala-Gly
Larger oligopeptides	Met-Pro-Thr-Ser-Ser-Ser-Thr-Lys-Lys-Thr-Gln-Leu-Cys,	(+)	Gly-Gly-Gly-Gly-Gly, Met-Glu-His-Phe-Arg-Trp-Gly,
	Met-Ala-Pro-Thr-Ser-Ser-Ser-Thr-Lys-Lys-Thr-Gln-Leu	(+)	Met-Lys-Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg, Pro-Thr-Ser-Ser-Ser-Thr-Lys-Lys-Thr-Gln-Leu-Cys, Ala-Pro-Thr-Ser-Ser-Ser-Thr-Lys-Lys-Thr-Gln-Leu, Leu-Ala-Pro-Thr-Ser-Ser-Ser-Thr-Lys-Lys-Thr-Gln-Leu

^a Activity of MAP was determined as described in Materials and Methods. One unit of activity equals 1 µmol of amino acids produced per min. Those substrates against which the MAP activity was below detection (<0.05 U/mg) are classified as "not cleaved." (+) indicates that the specific activities of MAP against these substrates were significantly higher than the detection limit, but they were not quantitatively determined due to the presence of some impurities in the substrates. Met-[0] is methionine sulfoxide.

TABLE 4. Processing of terminal methionine by MAP in vivo and in vitro^a

Processing conditions	Amino-terminal residues at methionine (%)	
	Met-Ala-Pro-IL-2	Met-Ile-Phe-Ricin A
In vitro		
Without MAP treatment	70	40
After MAP treatment	6	8
In vivo		
MM294(pSYC1143)	70	
MM294(pSYC1143)(pSYC1174)	5	
MM294(pRAP229)		40
MM294(pRAP229)(pSYC1174)		9

^a In vitro processing of proteins was tested by incubating purified MAP with purified recombinant protein. In vivo processing was tested in the MAP-hyperproducer hosts. The recombinant proteins were purified from these strains and subjected to Edman degradation for quantitative sequence analysis as described in Materials and Methods.

(12). An exception is the cleavage of methionine from Met-Ile-Phe-ricin A; the radius of gyration for isoleucine is 0.156 nm. It is possible that intermediate sizes of side chains have effects on cleavage that are influenced significantly by other neighboring amino acids and the tertiary structure of the substrate. In addition, the model of Sherman et al. (28) is based in part on qualitative results that would classify partial methionine cleavage for intermediate cases as noncleavage (see Table 4 for Met-Ile-Phe-ricin A).

The specificity of MAP is probably influenced by factors other than a geometric property of the adjacent residue. In addition to the exception noted above, we also noticed that despite the fact that glycine has no side chain, Met-Gly-Met is less preferred as a substrate for MAP in vitro than is Met-Ala-Met (Table 3). Furthermore, significant variations in MAP activity were observed among substrates that have identical residues at the first two positions. For example, among oligopeptides that start with the sequence Met-Gly-, *E. coli* MAP shows the following relative specificity: Met-Gly-Met-Met > Met-Gly-Met > Met-Gly-Gly, with more than threefold differences between each pair (Table 3).

The specificity of MAP is unique among other known aminopeptidases in *E. coli* and *S. typhimurium*. Peptidases A, B, D, N, and T are broad-specificity peptidases with overlapping activities against different peptides (22, 23, 32). Peptides P and Q are specific to X-Pro peptides (14), and peptidase E is specific to Asp-X peptides (6). Based on substrate size, peptidases D, Q, and probably E are dipeptidases (6, 22, 23), peptidase T is a tripeptidase (32), and the rest are oligopeptidases (22, 23).

The methionine residues at the amino termini of many proteins can be removed cotranslationally (2, 10). The physiological function of this process is unknown. The *map* gene we have cloned from *E. coli* could encode the MAP that is responsible for this process in vivo. Consistent with this hypothesis are the findings that (i) MAP is specific for the amino-terminal methionine and (ii) hyperproduction of the *E. coli* MAP enhances the removal of the amino-terminal methionine from the recombinant IL-2 and ricin A chain proteins in vivo. It would be necessary to investigate the effects of mutations in the *map* gene to definitively assign the physiological roles of the MAP enzyme.

Many of the hyperproduced recombinant proteins synthesized in *E. coli* still contain amino-terminal methionine. For those proteins that partially retained the methionine residue,

the purified MAP enzyme could be useful to "polish" the frayed amino terminal sequences and to generate more homogeneous protein products. Alternatively, as we have demonstrated here, this process can be achieved in vivo in a MAP hyperproducer host.

ACKNOWLEDGMENTS

We are very grateful to C. Miller for furnishing the peptidase-deficient CM89 strain, for communicating unpublished results on the phosphate requirement of the *Salmonella* peptidase M, and for consultation on the screening strategy. We thank our colleagues D. Nitecki and L. Aldwin for peptide synthesis, C. Levenson and the Cetus DNA Synthesis Group for oligonucleotide synthesis, M. Innis for input in DNA sequencing, J. Thompson and B. Ferris for help in protein purification, D. Gelfand for critical reading of the manuscript and for providing some of the plasmids and strains used in this study, and E. McCallan and J. Davis for preparation of the manuscript.

LITERATURE CITED

- Adams, J. M. 1968. On the release of the formyl group from nascent proteins. *J. Mol. Biol.* **33**:571-589.
- Ball, L. A., and P. Kaesberg. 1973. Cleavage of the N-terminal formylmethionine residue from a bacteriophage coat protein *in vitro*. *J. Mol. Biol.* **79**:531-537.
- Barman, T. E. 1969. Enzyme handbook, vol. 1, p. 176-177. Springer-Verlag, New York.
- Biossel, J.-P., T. J. Kasper, S. Shah, J. I. Malone, and H. F. Bunn. 1985. Amino-terminal processing of proteins: hemoglobin South Florida, a variant with retention of initiator methionine and N-acetylation. *Proc. Natl. Acad. Sci. USA* **82**:8448-8452.
- Capecchi, M. R. 1966. Initiation of *E. coli* proteins. *Biochemistry* **55**:1517-1524.
- Carter, T. H., and C. G. Miller. 1984. Aspartate-specific peptidase in *Salmonella typhimurium*. *J. Bacteriol.* **159**:453-459.
- Chang, A. C. Y., and S. N. Cohen. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* **134**:1141-1156.
- Hayashi, S., S.-Y. Chang, S. Chang, and H. C. Wu. 1984. Modification and processing of *Bacillus licheniformis* prepenicillinase in *Escherichia coli*. Fate of mutant penicillinase lacking lipoprotein modification site. *J. Biol. Chem.* **259**:10448-10454.
- Hermisdorf, C. L. 1978. Tripeptide-specific aminopeptidase from *Escherichia coli* AJ005. *Biochemistry* **17**:3370-3376.
- Housman, D., D. Gillespie, and H. Lodish. 1972. Removal of formyl-methionine from nascent bacteriophage f2 protein. *J. Mol. Biol.* **65**:163-166.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**:680-685.
- Levitt, M. 1976. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **104**:59-107.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning—a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- McHugh, G. L., and C. G. Miller. 1974. Isolation and characterization of proline peptidase mutants of *Salmonella typhimurium*. *J. Bacteriol.* **120**:364-371.
- McLaughlin, J. R., S.-Y. Chang, and S. Chang. 1982. Transcriptional analyses of the *Bacillus licheniformis* *penP* gene. *Nucleic Acids Res.* **10**:3905-3919.
- McLaughlin, J. R., H. C. Wong, Y. E. Ting, J. van Arsdell, and S. Chang. 1986. Control of lysogeny and immunity by *Bacillus subtilis* temperate phage SP8 α gene. *J. Bacteriol.* **167**:952-959.
- Merrifield, R. B. 1969. Solid-phase peptide synthesis. *Adv. Enzymol.* **32**:221-296.
- Merrifield, R. B. 1985. Solid phase synthesis. *Biosci. Rep.* **5**:353-376.
- Meselson, M., and R. Yuan. 1968. DNA restriction enzyme from

- E. coli*. Nature (London) 217:1110-1114.
20. Messing, J., and J. Vieira. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* 19:269-276.
 21. Michaelis, S., L. Guarente, and J. Beckwith. 1983. In vitro construction and characterization of *phoA-lacZ* gene fusions in *Escherichia coli*. *J. Bacteriol.* 154:356-365.
 22. Miller, C. G., and K. MacKinnon. 1974. Peptidase mutants of *Salmonella typhimurium*. *J. Bacteriol.* 120:355-363.
 23. Miller, C. G., and G. Schwartz. 1978. Peptidase-deficient mutants of *Escherichia coli*. *J. Bacteriol.* 135:603-611.
 24. Pine, M. J. 1969. Kinetics of maturation of the amino termini of the cell proteins of *Escherichia coli*. *Biochim. Biophys. Acta* 174:359-372.
 25. Rosenberg, M., and D. Court. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* 13:319-353.
 26. Rosenberg, S., E. A. Grimm, M. McGrogan, M. Doyle, E. Kawasaki, K. Kothe, and D. Mark. 1984. Biological activity of recombinant human interleukin-2 produced in *Escherichia coli*. *Science* 223:1412-1415.
 27. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
 28. Sherman, F., J. W. Stewart, and S. Tsunasawa. 1985. Methionine or not methionine at the beginning of a protein. *BioEssays* 3:27-31.
 29. Simmonds, S. 1970. Garvan awards of the American Chemical Society: peptidase activity and peptide metabolism in *Escherichia coli* K-12. *Biochemistry* 9:1-8.
 30. Simmonds, S., K. S. Szeto, and C. G. Fletteric. 1976. Soluble tri- and dipeptidases in *Escherichia coli* K-12. *Biochemistry* 15:261-271.
 31. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
 32. Strauch, K. L., and C. G. Miller. 1983. Isolation and characterization of *Salmonella typhimurium* mutants lacking a tripeptidase (peptidase T). *J. Bacteriol.* 154:763-771.
 33. Takeda M., and R. E. Webster. 1968. Protein chain initiation and deformylation in *B. subtilis* homogenates. *Proc. Natl. Acad. Sci. USA* 60:1487-1494.
 34. Tsunasawa, S., J. W. Stewart, and F. Sherman. 1985. Amino-terminal processing of mutant forms of yeast iso-1-cytochrome c. *J. Biol. Chem.* 260:5382-5391.
 35. Vimr, E. R., L. Green, and C. G. Miller. 1983. Oligopeptidase-deficient mutants of *Salmonella typhimurium*. *J. Bacteriol.* 153:1259-1265.
 36. Vimr, E. R., and C. G. Miller. 1983. Dipeptidyl carboxypeptidase-deficient mutants of *Salmonella typhimurium*. *J. Bacteriol.* 153:1252-1258.
 37. Vogel, H. J., and D. M. Bonner. 1956. Acetylornithine of *Escherichia coli*: partial purification and some properties. *J. Biol. Chem.* 218:97-106.
 38. Walker, J. P. 1963. The NH₂-terminal residue of the proteins from cell-free extract of *E. coli*. *J. Mol. Biol.* 7:483-496.
 39. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33:103-119.
 40. Yoshida, A., and M. Lin. 1972. NH₂-terminal formylmethionine and NH₂-terminal methionine cleaving enzymes in rabbits. *J. Biol. Chem.* 247:952-957.