Targeting of Endopeptidase 24.16 to Different Subcellular Compartments by Alternative Promoter Usage*

(Received for publication, February 28, 1997)

Akira Kato‡, Naoaki Sugiiura§, Yohko Saruta, Takehiko Hosoiri, Hiroshi Yasue¶, and Shigehisa Hirose

From the Department of Biological Sciences, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226, Japan, the Department of Biochemistry I, National Defense Medical College, Tokorosawa, Saitama 359, Japan, and the Animal Genome Research Group, National Institute of Animal Industry, Inashiki-gun, Ibaraki 305, Japan

Endopeptidase 24.16 or mitochondrial oligopeptidase, abbreviated here as EP 24.16 (MOP), is a thiol- and metal-dependent oligopeptidase that is found in multiple intracellular compartments in mammalian cells. From an analysis of the corresponding gene, we found that the distribution of the enzyme to appropriate subcellular locations is achieved by the use of alternative sites for the initiation of transcription. The pig EP 24.16 (MOP) gene spans over 100 kilobases and is organized into 16 exons. The core protein sequence is encoded by exons 5–16 which match perfectly with exons 2–13 of the gene for endopeptidase 24.15, another member of the thimet oligopeptidase family. These two sets of 11 exons share their proximal sites. Their expression in vitro in COS-1 cells indicated that they encoded two isoforms (long and short) which differed only at their amino termini: the long form contained a cleavable mitochondrial targeting sequence and was directed to mitochondria; the short form, lacking such a signal sequence, remained in the cytosol. The complex structure of the EP 24.16 (MOP) gene thus allows, by alternative promoter usage, a fine transcriptional regulation of coordinate expression, in the different subcellular compartments, of the two isoforms arising from a single gene.

Metalloendopeptidases form a large family of peptidases that have a His-Glu-X-His (HEXXH) zinc-binding motif and preferentially cleave short substrates. For example, endopeptidase 24.15 (EP1 24.15), a member of this family, acts on peptides of 6–18 amino acid residues and exhibits no or only very weak proteolytic activity against proteins (1–3). Among the members of this family, thimet oligopeptidase (TOP) or EP 24.151 and oligopeptidase M (MOP or EP 24.16) are unique in their sensitivities to thiol reagents and they constitute a subfamily, the thimet (thiol- and metal-dependent) oligopeptidase subfamily. Recent molecular cloning revealed the presence of a cysteine residue unique to members of this subfamily near position 483. This residue is absent from the other members that exhibit no thiol dependence (4, 5). In addition to the members of this family of mammalian origin, certain oligopeptidases of microbial origin that belong to this family have also been identified, including oligopeptidase A (OpdA) and dipeptidyl carboxypeptidase (Dcp) of Escherichia coli and Salmonella typhimurium (6), peptide F of Lactococcus lactis (7), mitochondrial intermediate peptidase of rat and yeast (8, 9), and saccharolysin (YCL57w or proteinase yscD) of yeast (10). This report deals with the two best characterized mammalian enzymes, namely, EP 24.15 (TOP) and EP 24.16 (MOP), which are members of the thimet oligopeptidase family. This family has also been called the M3 family of metalloendopeptidases in the classification of Rawlings and Barrett (11, 12).

EP 24.15 (TOP) was first identified as a collagenase-like peptidase or Pz-peptidase in experiments with the Pz-peptide that was originally designed by Wünsch and Heidrich (13) as a substrate for collagenase. Although the Pz-peptide was a good substrate for clostridial collagenase, it turned out not to be a substrate for avian and mammalian collagenases (14). The Pz-peptide hydrolyzing activities found in avian and mammalian tissues have, therefore, been designated collagenase-like peptidases or simply Pz-peptidases. Independent studies on the metabolism of brain peptides led to the discovery of two enzymes: one was described by Camargo et al. (15) in 1972 and was named neutral endopeptidase and, later, endo-oligopeptidase A; and the other, first described by Orłowski et al. (16) in 1983, was initially named soluble metalloendopeptidase and subsequently endopeptidase 24.15. All these enzymes turned out to be the same and are now known as thimet oligopeptidase (17). In this report we use the abbreviated designation EP 24.15 (TOP). cDNA sequences for the mammalian enzyme are now available for the rat (4, 18, 19), pig (20), and human (21).

EP 24.16 (MOP) was also discovered independently in several different laboratories. 1) Heidrich et al. (22) demonstrated a Pz-peptide hydrolyzing activity in a mitochondrial fraction of rat liver, which was later shown to be distinct from EP 24.15 (TOP) by both biochemical characterization (23) and partial

*This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas (Intracellular Proteolysis) from the Ministry of Education, Science, Sports and Culture of Japan and by research grants from SRF, the Cosmetology Research Foundation, and the Life Insurance Association of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1754 solely to indicate this fact. The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AB000170–AB000175, AB000411–AB000425, and AB000426–AB000438.

‡ Supported by a Research Fellowship for Young Scientists from the Japan Society for the Promotion of Science.

§To whom correspondence should be addressed: Dept. of Biological Sciences, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226, Japan. Tel.: 81-45-924-5726; Fax: 81-45-924-5824.

1The abbreviations used are: TOP, thimet oligopeptidase; EP, endopeptidase; MOP, oligopeptidase M; PAGE, polyacrylamide gel electrophoresis; bp, base pair(s); SINE(s), short interspersed repetitive element(s); RACE, rapid amplification of cDNA ends; PIPES, 1,4-piperazinediethanesulfonic acid; nt, nucleotide(s).
amino acid sequencing of the purified enzyme; it was named oligopeptidase M (24). 2) We (25) and Kiron and Soffer (26) identified a soluble angiotensin-binding protein in pig and rabbit liver during the course of studies aimed at identifying hepatic receptors for angiotensin II. After our publication of the cDNA sequence of the binding protein from pig (27), McKie et al. (18) pointed out the strong similarity between our sequence and that of rat EP 24.15 (TOP) which had been determined by Pierotti et al. (4, 19). We then obtained a second cDNA clone which was very similar to but clearly different from that of the cDNA for the binding protein, and we showed that the second clone represented the pig homolog of rat EP 24.15 (TOP) (20). The angiotensin-binding protein, although originally identified as a binding protein, did indeed have thiol- and metal-dependent oligopeptidase activity (20). At that time, therefore, the binding protein appeared to represent a new member of the thimet oligopeptidase family since the amino acid sequence of oligopeptidase M or EP 24.16 (MOP) from no mammalian species had yet been determined. 3) Kawabata et al. (28, 29) isolated an endopeptidase and the corresponding cDNA clone as a candidate for an enzyme responsible for the post-transcriptional processing of γ-carboxyglutamic acid-containing blood coagulation factors. They failed to notice the strong similarity to our binding protein, which was later pointed out by McKie et al. (18). 4) Chedler et al. (20, 31) demonstrated the presence of a novel proteolytic activity capable of inactivating neurotensin. They purified the peptidase from rat brain synaptic membranes and characterized it (32). The enzyme, termed neurelysin or endopeptidase 24.16, was shown to be distinct from EP 24.15 (TOP) and nephrilysin (also known as enkephalilnase or endopeptidase 3.4.24.11) and to have a relatively broad substrate-specificity and tissue distribution. Recent determination of its amino acid sequence by cDNA cloning clearly indicated that neurelysin is identical to the three enzymes mentioned above (33). Thus, four separate lines of research have converged in the discovery of a single new member of the thimet oligopeptidase family. In this report we use the abbreviation EP 24.16 (MOP) for this protein, whose identity has been recently established.

EP 24.15 (TOP) and EP 24.16 (MOP) are very similar in terms of size and enzymatic properties: both are intracellular proteins of 78–80 kDa, consisting of about 680–700 amino acids, and their sequences are 65% homologous (20). They are, however, clearly distinguishable in several respects. For example, they have different specificities for inhibitors, different immunoreactivity, and different cleavage-site specificities. EP 24.15 (TOP) hydrolyzes neurelysin exclusively at the Arg-Arg bond whereas EP 24.16 (MOP) cleaves it at the Pro-Tyr bond (16, 24, 32). Another difference is found in the subcellular localization of these enzymes. EP 24.15 (TOP) is found in the cytosol while EP 24.16 (MOP) is found in both the cytosolic and mitochondrial compartments. How can the product of a single gene be localized to more than one intracellular compartment? To answer this question and to characterize evolutionary relationships among the members of the thimet oligopeptidase family, we investigated the structural organization of the pig genes for EP 24.15 (TOP) and EP 24.16 (MOP) and of their 5′-proximal flanking regions. We discovered six species of mRNA for EP 24.16 (MOP) that are generated from a single gene as a result of the utilization of alternative sites for the initiation of transcription. The six species of mRNA can be classified into two categories: those containing an additional sequence that encodes a mitochondrial targeting sequence and those that lack such a sequence. The use of different promoters for the eventual targeting of proteins to appropriate subcellular compartments appears to be a useful mechanism for adjusting local concentrations of proteins that function at different intracellular sites in response to the physiological requirements of the cell.

**EXPERIMENTAL PROCEDURES**

Identification of 5′-Terminal Sequences of cDNAs for EP 24.16 (MOP)—The 5′-ends of cDNAs for EP 24.16 (MOP) were cloned with the 5′-RACE (rapid amplification of cDNA ends) system (CLONTECH, Palo Alto, CA). Two μg of poly(A)⁺ RNA, isolated from pig liver (27), were reverse-transcribed with a specific primer for the cDNA for pig EP 24.16 (MOP), 5RA-1 (5′-CTCTAGATGCTCTGTCCTC-3′), and avian myeloblastosis virus reverse transcriptase. The first-strand cDNA was ligated at the 3′-end with an anchor (5′-CAGGATCTCATCGAGTCTGGAACATTTCAGAGG−3′) by T4 RNA ligase. A nested specific primer for the cDNA for EP 24.16 (MOP), 5RA-2 (5′-CTGCCTGACCCTACTCACCCTTACCTC−3′), was used with an anchor primer (5′-CTGTTGCGCCACCTCCTAGGAGGTTCCAGAATGCTCCTG) for amplification of the 5′-ends of the cDNAs by polymerase chain reaction. The products of polymerase chain reaction were fractionated on a 3% agarose gel, and fragments of 300–650 bp were isolated and cloned into pBluescript II (Strategene, La Jolla, CA). Positive clones were identified by colony hybridization, with the 32P-labeled EcoRI-EcoRV 592-bp fragment of pAB-L1 (27) as probe, and sequenced.

Sequencing of DNA—DNA was sequenced by the dideoxy chain termination method of Sanger et al. (34) with double-stranded plasmids as templates. Termination reactions were performed with Sequitherm DNA polymerase (Epicient Technologies, Madison, WI) and HIRD1-labeled M13 universal or reverse primer (LI-COR, Lincoln, NE). The products were analyzed with a DNA sequencer (model 4000; LI-COR). Sequences were organized and analyzed with GENETYX-Mac programs (Software Development, Tokyo, Japan).

Isolation of Genomic Clones for Pig EP 24.16 (MOP) and EP 24.15 (TOP)—A pig liver genomic library constructed in AEBL3 SP6/T7 (CLONTECH) was screened with the 2.7-kilobase EcoRI-EcoRI fragment of a cDNA for EP 24.16 (MOP) clone (PAB-L1), (27), or with the 2.5-kilobase pair EcoRI-EcoRI fragment of a cDNA clone for EP 24.15 (TOP) (clones JH-1; (20), both of which had been labeled with [γ-32P]dCTP (Amersham, Little Chalfont, UK) with a random priming kit (Takara, Kyoto, Japan). Phage clones (2 x 10⁶) were plated at a density of 30,000 plaque-forming units per 135 x 55-mm plate on E. coli NM538, from which duplicate replications were made on cellulose-nitrate filters (Schleicher & Schuell, Dassel, Germany) and allowed to hybridize with the 32P-labeled probe in a solution of 6 x SSPE (1 x SSPE = 15 mM NaCl, 15 mM sodium citrate, pH 7.0), 0.1% SDS, 0.1% formamide, 0.1% SDS, and 5 x Denhardt's solution at 42 °C for 16 h. The filters were rinsed twice at room temperature in 2 x SSC (1 x SSC is 0.15 M NaCl, 15 mM sodium citrate, pH 7.0) that contained 0.1% SDS and washed twice at 60 °C in 1 x SSC that contained 0.1% SDS for 1 h. Positive plaques were identified by autoradiography and purified with the additional rounds of screening.

Restriction Mapping of Phages—Positions of the EcoRI, SalI, and XbaI restriction sites in genomic clones were determined by complete or partial digestion with restriction enzymes and subsequent Southern blot analysis. UV irradiation and formation of pyrimidine dimers were used for preparation of incompletely digested genomic clones. AEBL3 SP6/T7 T7 contains two unique SfiI or Sall sites and bacteriophage promoters (SP6 and T7) that flank the insert. Arms were separated with SfiI or SalI from the inserts, which still contain promoter sequences at the both ends. DNA samples were UV-irradiated for 0 or 20 min with UV Stratalinker 2400 (Stratagene) in 10 mm Tris, pH 7.5, 10 mm MgCl₂, and 1 mm dithiothreitol. UV-irradiated samples (500 ng) were digested incompletely with EcoRI, SalI, or XbaI (10 units) for 1 h at 37 °C, fractionated on a 0.7% agarose gel, and transferred to nylon membranes (MagnaGraph, MSI, Westboro, MA). A set of filters was prepared and allowed to hybridize with end-labeled oligoprobes for T7 or SP6 promoter sequence, for 14 h at 37 °C in 6 x SSC, 5 x Denhardt's solution, 0.5% SDS, and 100 μg/ml herring sperm DNA. The filters were washed twice in 1 x SSC, 0.1% SDS at 42 °C for 30 min, exposed to imaging plates, and analyzed with a Bioimage Analyzer (model BAS 2000; Fuji Film, Tokyo, Japan).

**3′ Nucleotide Protection Assay—**Three primers, namely, 108L (CAAGCTTTGCGGCGGCCTAGCAAAAGGAGGACACAG) for exon 1; 107L (GGGCGCCCTCGGGTGATACGGTGTCGTTGA) for exon 2; and 106B (GTCCTTCATGAGTAATGCTCCT) for exon 3, were designed for the synthesis of single-stranded antisense DNA probes that would protect pig 5′-ends of mRNAs for EP 24.16 (MOP). Ten pmol of each primer were labeled with [γ-32P]ATP (Amersham) by polynucleotide kinase and added to a transcription reaction mixture containing 200 μg/ml herring sperm DNA. After incubation at 37 °C for 1 h in 6 x SSC, 5 x Denhardt's solution, 0.5% SDS, and 100 μg/ml herring sperm DNA. The filters were washed twice in 1 x SSC, 0.1% SDS at 42 °C for 30 min, exposed to imaging plates, and analyzed with a Bioimage Analyzer (model BAS 2000; Fuji Film, Tokyo, Japan).
Alternative Promoters and Protein Targeting

Fig. 1. Nucleotide sequences of the six types of 5'-end of cDNAs for pig EP 24.16 (MOP). The six different sequences revealed by 5'-RACE are aligned. Candidates for codons for initiation of translation are boxed and labeled M1, M2, and M3, respectively. These initiation sites are aligned with the consensus sequence for sites of initiation of translation (GCCA/GCCATGG) (45) which is shown in italics. Three nucleotide replacements, due possibly to allelic polymorphism, were found in exon 2 (underlined). Exon boundaries, determined by comparison of genomic and cDNA sequences (Fig. 4D), are indicated by vertical arrows.

...
The extreme 3' are mutually exclusive (Fig. 6B), and by the alternative splicing of exon 4; exons 1, 2, and 3 are used to assemble the functional domain of the enzyme.

A zinc-binding motif HEFGH is encoded by exon 12 (Fig. 4B), which is rich in basic amino acids and can be expected to form an amphipathic helix (36). The common exons 5 through 16 are used to assemble the functional domain of the enzyme.

There appears to be a "pseudo-exon" that encodes a protein that resembles a ribosomal protein (11.5 kDa, L44 (37)) in reverse orientation (3' to 5') within the untranslated region of the 3'-most exon (Fig. 4B). The sequence encoding the homolog of ribosomal L44 is flanked by the direct repeat TGTTTTATTTT and has a poly(A) tract, suggesting that the pseudogene might have arisen as a result of retroposition.

Structural Comparison of the Genes for EP 24.15 (TOP) and EP 24.16 (MOP)—We wondered whether the complexity of organization of the gene for EP 24.16 (MOP) might be reflected in the genes for other members of the thimet oligopeptidase family and, to this end, we also characterized the gene for EP 24.15 (TOP). The gene for EP 24.15 (TOP) was isolated from the same pig genomic DNA library as that used for isolation of genes for EP 24.16 (MOP), and it was found to have a much simpler structure in its 5'-region (Figs. 3 and 4B). The gene exists as a single copy, as revealed by Southern blot analysis (data not shown); it spans approximately 45 kilobase pairs (Fig. 4A); and it is organized into 13 exons. The overall organization of the two genes is very similar with the exception of the length of introns and the 5'-leader and untranslated exons (Fig. 4B). For example, exons 2–12 of the gene for EP 24.15 (TOP) correspond precisely to exons 5–15 of the gene for EP 24.16 (MOP) and there is strong conservation of the respective exon-intron boundaries (Fig. 4D), suggesting evolution from a common ancestor. The zinc-binding motif HEGFH is encoded within

FIG. 2. Structure of the cDNA for pig EP 24.16 (MOP) in relation to the structure of the gene (A), exon-intron organization of the gene (B), restriction maps for EcoRI, SacI, and XbaI (C), and the relative locations of genomic clones used for the analysis (D). The structure of cDNA for pig EP 24.16 (MOP) is described elsewhere (27).

FIG. 3. Structure of the pig gene for EP 24.15 (TOP). A, structural relationship to the cDNA. The structure of cDNA for pig EP 24.15 (TOP) is described elsewhere (20). B, exon-intron organization. C, restriction maps. The following abbreviations are used for the restriction enzymes: E, EcoRI; S, SacI; X, XbaI. D, relative locations of genomic clones.
exon 9. The 3' terminal exon 13 is composed of a short coding sequence, the termination codon, and the entire 3' untranslated sequence. The promoter region of the gene for EP 24.15 (TOP) lacks the TATA box but contains several putative binding sites for ubiquitous factors including one CCAAT box, three Sp1 sites, one NF-1 site, one AP-1 site, and two AP-2 sites (data not shown).

Identification of Three Major Sites of Transcription Initiation—Characterization of the 5' ends of mRNAs for EP 24.16 (MOP) and EP 24.15 (TOP). A comparison of the lengths of pig genes for EP 24.16 (MOP) and EP 24.15 (TOP). The exons corresponding to the zinc-binding motifs (exon 12 of EP 24.16 (MOP) and exon 9 of EP 24.15 (TOP)) are aligned (dotted vertical line). B, comparison of the exon-intron organization of the gene for EP 24.16 (MOP) with that of EP 24.15 (TOP). Exons are indicated by boxes and are numbered in bold type. Coding regions for the peptides are shown by filled boxes and amino-terminal extension sequences generated from the alternative sites of initiation of translation of EP 24.16 (MOP) are hatched. Numbers indicate the sizes of exons and introns in base pairs and kilobase pairs, respectively. C, the extent of sequence homology between the product of the gene for EP 24.16 (MOP) and related peptides. Coding exons of the gene for EP 24.16 (MOP) (exon 5-17) were converted into deduced amino acid sequences and compared with those of other peptides, such as pig EP 24.15 (TOP), saccharolysin from S. cerevisiae (10), rat mitochondrial intermediate peptidase (MIP) (8), oligopeptidase A (OpdA) from E. coli (52), and dipeptidyl carboxypeptidase (Dcp) from E. coli (53). The extent of homology is indicated, as a percentage, below the corresponding exons. Bold numbers indicate relatively strongly conserved regions, with homology scores more than five points above the average. Italic numbers indicate poorly conserved regions, with homology scores more than five points below the average. D, nucleotide sequences of exon-intron junctions. The exons are shown in capital letters and the introns in lowercase letters. The 5'-flanking regions of three sites of initiation of transcription are shown in italics.
Identification of site of initiation of transcription of the pig gene for EP 24.16 (MOP) by S1 nuclease mapping (A) and sequences of promoter regions (B and C). A, 5 µg of poly(A)⁺ RNA from pig liver (lanes 2, 4, and 6) or yeast tRNA (lanes 1, 3, and 5) were
strated to serve as a Myb-suppressible promoter (44).

**Differential Subcellular Localization of EP 24.16 (MOP) Directed by Alternatively Generated Species of mRNA—**The results described above suggest that the organization of the 5′-region of the genes for EP 24.16 (MOP) is unusually complex and that six mRNA species with different 5′ termini are generated as a consequence of the use of separate promoters (Fig. 5) and the splicing of the 5′-leader exons 1 through 3 (in a mutually exclusive manner) and of exon 4. The cDNA sequences corresponding to the six species of mRNA are shown schematically in Fig. 6A, and they were used for the expression experiments described below. It should be noted that exon 1 has an in-frame ATG codon (designated M1), when connected directly to exon 5, and the open reading frame in exon 1 encodes a putative signal peptide for import into mitochondria; exon 2 also has an in-frame ATG codon (M2) in an appropriate context for the initiation of translation (Fig. 1) (45) and the open reading frame predicts an enzyme with 64 more amino acids at its amino terminus than the product generated by the open reading frame that starts with an ATG codon (M3) in the common exon 5 (Figs. 1 and 6A). The fact that exon 1 could encode an amino-terminal leader sequence for targeting to mitochondria strongly suggests that, upon selection or elimination of the sequence of exon 1 via differential utilization of the multiple promoters, the subcellular localization of the products of the gene for EP 24.16 (MOP) is strictly and efficiently controlled. To confirm this possibility, we carried out the following experiments.

The six cDNA constructs depicted on the right side of Fig. 6A (labeled types 1 through 3 and 1′ through 3′) were inserted separately into the mammalian expression vector pcDNA3 and used to transfect COS-1 cells. Then subcellular organelles were isolated from the transfectants and the levels of EP 24.16 (MOP) in these organelles were examined by Western blotting (Fig. 6, B and C). The type 1 (1-4-[5–16]) construct directed the synthesis of EP 24.16 (MOP) that was targeted to mitochondria (Fig. 6, B, lane 3, and C, MOP(M1b)); the mitochondrial enzyme was slightly smaller than the unprocessed precursor that remained, as a consequence of overexpression of the protein, in the cytosol (Fig. 6, B, lane 2, and C, MOP(M1a)). This difference in size indicates that the amino-terminal mitochondrial targeting sequence is cleaved after translocation of the protein into mitochondria. Type 1′ (1-4-[5–16]), in which the connection between exons 1 and 5 is interrupted by insertion of exon 4 which includes a stop codon (Fig. 6D), yielded only the cytosolic form of EP 24.16 (MOP) generated from the ATG initiation codon (M3) in exon 5 (Fig. 6B, lanes 5–8). Type 2 (2-5-[16]) allowed the synthesis of an amino-terminally extended cytosolic form (Fig. 6B, lane 10, upper band). Again, as seen with type 2′ (2-4-[5–16]), insertion of exon 4 generated a stop codon and only the short cytosolic form was expressed (Fig. 6B, lanes 9–16). With type 1 and type 2, products of translation from the ATG codon in exon 5 (M3) were also detected (Fig. 6, B, lane 10, lower band, and C, MOP(M3)), suggesting that these mRNAs generate two isoforms of the protein by alternative usage of codons for the initiation of translation (M1 and M3 for type 1 and M2 and M3 for type 2). The constructs having exon 3 as the 5′-leader exon (types 3 and 3′) produced only the cytosolic form of the enzyme (Fig. 6B, lanes 17–24), as expected from the fact that exon 3 contains no in-frame ATG codon.

**DISCUSSION**

In this study, we demonstrated the heterogeneity at the 5′-end of the mRNA for EP 24.16 (MOP). Moreover, we showed that the heterogeneity is generated by alternative usage of promoters and splicing of multiple 5′-leader and untranslated exons and that it is responsible for the differential subcellular localization of the products of translation.

**Targeting of Proteins to Different Subcellular Locations by Alternative Usage of Promoters: EP 24.16 (MOP) Represents the First Example of Such a Mechanism for Intracellular Peptidases/Proteinases—**Proteins, after their synthesis, must be delivered to their sites of action. Delivery is usually accomplished with the help of terminal or internal targeting sequences. Sequences for the targeting proteins to the following sites have been identified: mitochondria, endoplasmic reticulum, lysosomes, nuclei, and peroxisomes.

The presence of a putative mitochondrial targeting sequence at the amino terminus of the precursor to EP 24.16 (MOP) was first deduced by Serizawa et al. (24) from the potential ability of this sequence to form an amphipathic α-helix with a hydrophobic and a positively charged face of the type expected for a mitochondrial leader sequence (36, 46). This scenario explains the presence of EP 24.16 (MOP) in mitochondria. The enzyme is, however, known also to be present in the cytosol and, prior to the present study, the mechanism responsible for this distribution of EP 24.16 (MOP) has remained unclear. Discovery of 5′-end variants of the mRNA for EP 24.16 (MOP) by the 5′-RACE technique led us to investigate the genetic basis for such diversity. Through an analysis of the structure of the gene, which led to the identification of the three 5′-leader exons that are selected, in a mutually exclusive manner, by use of alternative promoters and splicing, we provided the following resolution of this problem (Fig. 7). If promoter 1 is used, the mitochondrial isoform of EP 24.16 (MOP) is generated by splicing of exon 1, which has a sequence that encodes a signal for transport to mitochondria, to exon 5, which is the beginning of the common translated region that encodes the mature portion of the protein (type 1 in Fig. 6, A and D). The precursor form (704 amino acid residues) with the mitochondrial targeting sequence is processed to the mature mitochondrial form of 667 residues (Fig. 7). The type 1 transcript can also yield the cytosolic form of 681 amino acids when the M3 site of initiation of translation is used instead of the M1 site. If promoter 3 is used and exon 3, which lacks an in-frame ATG codon, is joined to exon 5 (type 5 in Fig. 6, A and D), the cytosolic isoform is produced from the ATG initiation codon in exon 5. The use of promoter 2, which directs the synthesis of a cytosolic variant, is discussed below.

Similar scenarios have been reported for several other enzymes that are known to occur and function in more than one subcellular compartment (for a recent review, see Ref. 47). Typical examples are the histidine and valine tRNA synthetases of *Saccharomyces cerevisiae* that are involved in protein synthesis in the cytosol and the mitochondria (48, 49). In these cases, two types of transcript (long and short) are produced by
FIG. 6. Generation of cytosolic and mitochondrial forms of EP 24.16 (MOP) from a single gene by alternative usage of three promoters (P1–P3) and three codons for initiation of translation (M1–M3). A, schematic representation of the six isoforms of the mRNA for EP 24.16 (MOP) and the organization of the 5′-region of the gene for EP 24.16 (MOP), showing how the various isoforms are generated. Exons
Subcellular fractions were obtained by differential centrifugation, as follows: cytosol identified by subcellular localization of the products of translation of isoforms of the cDNA for EP 24.16 (MOP) expressed in COS-1 cells. Six cDNA species, 

alternative usage of promoters, and the long transcript yields the mitochondrial isoform exclusively, while the short transcript yields the cytosolic enzyme. In this way, adjustment of the levels of the proteins to the needs of each compartment is possible. Although the biological significance of this mechanism in the present case is not immediately apparent since the true substrates of the enzyme have not yet been identified, the general regulation of expression of the gene for an oligopeptidase by transcription regulatory factors and the unique regulation, as reported herein, of targeting of the product by use of alternative promoters seems to provide a powerful method by which cells can modulate the concentration of specific peptides in certain intracellular compartments to reflect the metabolic state of the cell. 

Complex Organization of the Pig Gene for EP 24.16 (MOP)—As compared with the gene for EP 24.15 (TOP), another member of the thimet oligopeptidase family, the gene for EP 24.16 (MOP) have quite a complicated structure in its 5'- and 3'-regions. The two genes do, however, exhibit extensive similarity in the regions that encode the mature proteins, which consist of 11 exons, namely, exons 2 through 12 in the case of the gene for EP 24.15 (TOP) and exons 5 through 15 in the case of the gene for EP 24.16 (MOP). The similarity suggests that these two genes and, probably, the genes for other members of this family were generated from an ancestral gene as distinct sequences as a consequence of gene duplication. The presence of a SINE in the 3'-untranslated region of the gene for EP 24.16 (MOP) suggests that insertion of a SINE after the gene duplication might have destabilized the gene for EP 24.16 (MOP) and stimulated extensive diversification of 5'- and 3'-regions by recruiting the entire gene for ribosomal protein L44 (in the reverse orientation) into the 3'-most exon (exon 16) and the 5'-leader exons into the 5'-flanking region by, perhaps, retroposition and gene conversion.

Long and Short Forms of EP 24.16 (MOP)—The use of promoter 2 of the gene for EP 24.16 (MOP) yields the type 2 transcript, which is predicted to have an amino-terminally elongated product (Fig. 7). Consistent with this prediction, we are indicated by boxes and numbered (box patterns: black, coding regions for the peptidase; hatched, reading frames encoding the amino-terminal extensions; and white, non-coding regions). Three sites for initiation of transcription (Fig. 5) are indicated by arrows. The deduced amino acid sequences corresponding to the reading frames of exons 1, 2, and 5 are shown in single letter code. are indicated by bold and large letters. Basic amino acid residues which are necessary for mitochondrial targeting sequences are indicated by bold letters. Alternative initiation of the transcription of exons 1–3 (P1–P3) and alternative splicing of exon 4 generate six isoforms of the mRNA. B, subcellular localization of the products of translation of isoforms of the cDNA for EP 24.16 (MOP) expressed in COS-1 cells. Six cDNA species, identified by 5'-RACE (Figs. 1 and panel A, of this figure), were expressed in COS-1 cells and the products were detected by Western blotting. Subcellular fractions were obtained by differential centrifugation, as follows: cytosol (100,000 × g supernatant); mit, mitochondria (7,000 × g pellet); mic, microsomes (100,000 × g pellet). Construction of cDNA used for expression is indicated by boxes on the right of panel (A). C, resolution of mitochondrial and cytosolic forms of EP 24.16 (MOP) by SDS-PAGE in the high-resolution buffer system described in the text. M1a, precursor of the mitochondrial form generated by use of the first site (M1) for initiation of translation; M1b, the processed mature form imported into mitochondria; M3, the cytosolic form generated by translation from the M3 site of initiation of translation (for details see Fig. 7). D, nucleotide and amino acid sequences of the type 1, 1', 2, and 3 isoforms. Three codons for initiation of translation, M1, M2, and M3, are present in exon 1, exon 2, and exon 5, respectively. The mitochondrial targeting sequence of EP 24.16 (MOP), containing six arginine residues (24), is underlined.
detected a long form of the protein in the cytosol of COS-1 cells transfected with the type 2 construct after SDS-PAGE and Western blotting (Fig. 6D, lane 10). Although the relative abundance of the corresponding mRNA, as estimated from the data after 5'-RACE, in the pig liver is low (<10%), the physiological significance of this form clearly merits further study. The roles of the extended amino-terminal region of 64 amino acid residues (Fig. 6D) could include stabilization of the enzyme, modulation of the substrate specificity, and/or mediation of interactions with other cytosolic proteins.

In our analysis, we also noticed the presence of a splice variant that lacked the sequence of exon 15. This variant should encode a protein with a short and slightly different carboxyl-terminal tail. The functional significance of this variant and the tissue- and development-specific regulation of the splicing will be the subject of further research. The presence of at least two forms of EP 24.16 (MOP) has also been demonstrated in purified preparations of the enzyme from rabbit and pig liver (50, 51).

The type 1, type 2, and type 3 species of mRNA all have splice variants, designated type 1', type 2', and type 3', respectively, with an extra exon sequence (5'-untranslated exon 4, Fig. 7), but none of them results in a long open reading frame from the first initiation codon. It is unknown whether, whether these variants have any biological significance. However, the possibility exists that the variant species of mRNA might contribute to the regulation of rates of translation. Alternatively, they might produce short peptides with as yet unidentified functions. In the case of the type 1' transcript, the insertion of the insertion of the extended amino-terminal region of 64 amino acid residues (Fig. 7), the translation of the corresponding mRNA, as estimated from the data after 5'-RACE, in the pig liver is low (<10%), the physiological significance of this form clearly merits further study.

Conclusions—Analysis of the gene for EP 24.16 (MOP) revealed the very complex organization of the gene and the presence of a variety of transcripts generated by differential use of multiple sites of initiation of transcription and by alternative splicing of exons 2, 3, 4, and 15. In contrast to these complexities, a simple and definitive answer was obtained to the question of how the product of a single gene for EP 24.16 (MOP) is delivered to two different cellular compartments, namely, the cytosol and the mitochondria.

Acknowledgments—We thank Makoto Itakura, Takeshi Ihara, and Hiromi Hagiwara for helpful discussions, and Setuko Satoh and Kazuko Tanaka for expert secretarial and technical assistance.

REFERENCES

1. Bond, J. S., and Beynon, R. J. (1985) Int. J. Biochem. 17, 565–574
2. Tisijar, U. (1993) Bioch. Biophys. Hoppe Seyler 374, 91–100
3. Barrett, A. J., Brown, M. A., Dando, P. M., Knight, C. G., McKie, N., Rawlings, N. D., and Serizawa, A. (1995) Methods Enzymol. 246, 529–556
4. Pierotti, A., Dong, K. W., Glucksman, M. J., Orlowski, M., and Roberts, J. L. (1990) Biochemistry 29, 10323–10329
5. Orlowski, M., Reznik, S., Ayala, J., and Pierotti, A. R. (1989) Biochem. J. 261, 567–579
6. Conlin, C. A., and Miller, C. G. (1995) Methods Enzymol. 246, 567–579
7. Monnet, V. (1995) Methods Enzymol. 246, 579–592
8. Isaya, G., Kaloupek, P., and Rosenberg, L. E. (1992) Proc. Natl. Acad. Sci.

* A. Kato, N. Sugiuira, and S. Hirose, unpublished observation.