

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/36605 A2

(51) International Patent Classification⁷: **C12N 9/00**

(21) International Application Number: PCT/US00/31873

(22) International Filing Date:
20 November 2000 (20.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/443,795 19 November 1999 (19.11.1999) US

(71) Applicant (for all designated States except US): **MILLENNIUM PHARMACEUTICALS, INC.** [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **KAPELLER-LIBERMANN, Rosana** [BR/US]; 86 Beacon Street, Chestnut Hill, MA 02467 (US). **WHITE, David** [US/US]; 35 Hollingsworth Avenue, Braintree, MA 02148 (US). **MACBETH, Kyle, J.** [US/US]; 688 Tremont Street, #4, Boston, MA 02118 (US).

(74) Agents: **COULTER, Kathryn, L.** et al.; Alston & Bird LLP, P.O. Drawer 34009, Charlotte, NC 28234-4009 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: 2786, A NOVEL HUMAN AMINOPEPTIDASE

(57) Abstract: The present invention relates to a newly identified human aminopeptidase. The invention also relates to polynucleotides encoding the aminopeptidase. The invention further relates to methods using the aminopeptidase polypeptides and polynucleotides as a target for diagnosis and treatment in aminopeptidase-related disorders. The invention further relates to drug-screening methods using the aminopeptidase polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the aminopeptidase polypeptides and polynucleotides. The invention further relates to procedures for producing the aminopeptidase polypeptides and polynucleotides.



WO 01/36605 A2

2786, A NOVEL HUMAN AMINOPEPTIDASE

FIELD OF THE INVENTION

The present invention relates to a newly identified human aminopeptidase. The invention also relates to polynucleotides encoding the aminopeptidase. The invention further relates to methods using the aminopeptidase polypeptides and polynucleotides as a target for diagnosis and treatment in aminopeptidase-related disorders. The invention
5 further relates to drug-screening methods using the aminopeptidase polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the aminopeptidase polypeptides and polynucleotides. The invention further relates to procedures for producing the aminopeptidase polypeptides and polynucleotides.

10

BACKGROUND OF THE INVENTION

Proteases may function in carcinogenesis by inactivating or activating regulators of the cell cycle, differentiation, programmed cell death, or other processes affecting
15 cancer development and/or progression. Consistent with the model involving protease activity and tumor progression, certain protease inhibitors have been shown to be effective inhibitors of carcinogenesis both *in vitro* and *in vivo*.

Aminopeptidases

Aminopeptidases (APs) are a group of widely distributed exopeptidases that
20 catalyse the hydrolysis of amino acid residues from the amino-terminus of polypeptides and proteins. The enzymes are found in plant and animal tissue, in eukaryotes and prokaryotes, and in secreted and soluble forms. Biological functions of aminopeptidases include protein maturation, terminal degradation of proteins, hormone level regulation, and cell-cycle control.

25 The enzymes are implicated in a host of conditions and disorders including aging, cancers, cataracts, cystic fibrosis and leukemias. In eukaryotes, APs are associated with removal of the initiator methionine. In prokaryotes the methionine is removed by methionine aminopeptidase subsequent to removal of the *N*-formyl group from the initiator *N*-formyl methionine, facilitating subsequent modifications such as

N-acetylation and *N*-myristoylation. In *E. coli* AP-A (*pepA*), the *xerB* gene product is required for stabilization of unstable plasmid multimers.

APs are also involved in the metabolism of secreted regulatory molecules, such as hormones and neurotransmitters, and modulation of cell-cell interactions. In mammalian cells and tissues, the enzymes are apparently required for terminal stages of protein degradation, and EGF-induced cell-cycle control; and may have a role in protein turnover and selective elimination of obsolete or defective proteins. Furthermore, the enzymes are implicated in the supply of amino acids and energy during starvation and/or differentiation, and degradation of transported exogenous peptides to amino acids for nutrition. As leukotriene A4 hydrolase may be an aminopeptidase, APs may further have a role in inflammation. Industrial uses of the enzymes include modification of amino termini in recombinantly expressed proteins. See A. Taylor (1993) *TIBS* 18: 1993:167-172.

A variety of aminopeptidases have been identified from a wide variety of tissues and organisms, including zinc aminopeptidase and aminopeptidase M from kidney; arginine aminopeptidase from liver; aminopeptidase N^b from muscle; leucine aminopeptidase (LAP) from lens and kidney; aminopeptidase A (*xerB* gene product) from *E. coli*; *yscI* APE1/LAP4 and aminopeptidase A (*pep4* gene product) from *S. cerevisiae*; LAP from *Aeromonas*; dipeptidase from mouse ascites; methionine aminopeptidase from *Salmonella*, *E. coli*, *S. cerevisiae* and hog liver; and D-amino acid aminopeptidase from *Ochrobactrum anthropi* SCRC C1-38.

Of these aminopeptidases, the structure of bovine lens leucine aminopeptidase (bLAP) is well-characterized and consists of a homohexamer synthesized as a large precursor, each monomer containing two thirds of the protein in a major lobe and one third in a minor lobe. The minor lobe contains the N-terminal 150 residues. All putative active site residues, presumably also the inhibitor bestatin-binding site, are found in the C-terminal lobe and include Ala-333, Asn-330, Leu360, Asp332, Asp255, Glu-334, Lys250, Asp273, Met-454, Ala-451, Gly362, Thr-359, Met270, Lys262, Gly362 and Ile-421.

Many aminopeptidases are metalloenzymes, requiring divalent cations, with specificities for Zn²⁺ or Co²⁺. However, particular sites of certain aminopeptidases can readily utilize Mn²⁺ and Mg²⁺. Residues used to ligand Zn²⁺ include the His His Glu and Asp Glu Lys configurations. In addition to bestatin, boronic and phosphonic

acids, α -methylleucine and isoamylthioamide are identified as competitive inhibitors for most aminopeptidases. See A. Taylor (1993) *TIBS* 18: 1993:167-172; Burley *et al.* (1992) *J. Mol. Biol.* 224:113-140; Taylor *et al.* (1993) *Biochemistry* 32:784-790.

Aminopeptidases from various organisms and various tissues within an
5 organism have high degrees of primary sequence homology, as indicated by immunological assays. Some enzymes also exhibit very similar kinetic profiles. Direct amino acid sequence comparison of bLLAP and aminopeptidase-A from *E. coli* shows 18, 44 and 35% identity for the amino- and carboxy- terminals, and the entire protein, respectively. The comparison shows 46, 66, and 60% identity for the
10 respective regions. See Burley *et al.* (1992) *J. Mol. Biol.* 224:113-140.

Bovine lens leucine aminopeptidase (bLLAP), bovine kidney LAP, human lens and liver LAPs, hog lens, kidney and intestine LAPs, proline-AP, *E. coli* AP-A, AP-I and the *S. typhimurium pepA* gene product have been categorized as belonging to the family of zinc aminopeptidases which utilize the residues Asp Glu Lys for zinc
15 binding and the active site amino acid configuration described above for bovine LAP for substrate binding. This family, possibly also including *Aeromonas* LAP, is suggested to be distinguished from zinc proteases which utilize His His Glu in zinc binding and Arg in substrate binding. The *Saccharomyces* methionine-AP is characterized to contain two zinc finger like motifs in the amino-terminus and shares
20 little homology with bLLAP. See A. Taylor (1993) *TIBS* 18:167-171; Watt *et al.* (1989) *J. Biol. Chem.* 264:5480-5487.

Leucine aminopeptidase expression is regulated at the transcriptional level, evidenced by enhancement of both activity and mRNA upon removal of serum in *in vitro* aged and/or transforming lens epithelial cells. Furthermore, LAP mRNA and
25 protein are induced by interferon γ in human ACHN renal carcinoma, A549 lung carcinoma, HS153 fibroblasts and A375 melanoma. Regulation by development and growth is also implicated. The *E. coli pepN* gene is transcriptionally regulated upon anaerobiosis and phosphate starvation. Membrane bound AP-N (CD13) is expressed in a lineage-restricted manner by subsets of normal and malignant cells, apparently
30 through regulation by physically distinct promoters. Expression of the yeast *yscI* product APE1 is dependent upon the levels of *yscA* and *PEP4* gene products. Synthesis of APE1 is sensitive to media glucose levels, and the activity of yeast aminopeptidase is sensitive to substitution of ammonia rather than peptone as the

source of nitrogen. See Harris *et al.* (1992) *J. Biol. Chem.* 267:6865-6869; Jones *et al.* (1982) *Genetics* 102:665-677.

Aminopeptidase B

Aminopeptidase B is an exopeptidase that removes arginine and/or lysine from various amino terminal peptide substrates. This enzyme is structurally related to leukotriene A₄ hydrolase. The activity of aminopeptidase B is dependent upon Zn²⁺. With respect to primary structure, the enzyme isolated from rat testis exhibits an amino terminal potential signal peptide and a Zn²⁺ binding consensus sequence (HEXXHX₁₈E). In view of the fact that the enzyme contains this consensus sequence, the enzyme can be classified as a M1 family metallopeptidase.

In the M1 family, the aminopeptidase is most closely related to the leukotriene A₄ hydrolase. Accordingly, in addition to the aminopeptidase activity, the enzyme contains leukotriene A₄ epoxide hydrolase activity. The specific enzyme from rat testis also contains a fully conserved ribonucleoprotein binding site ([RK]-G-[AFILMNQSTVWY]-[AGSCI]-[FY]-[LIVA]-X-[FYM]). The enzyme also contains a potential nuclear localisation (KKK) and a consensus (KGYCFVSY) ribonucleoprotein binding motif. Based on these features it has been proposed that the enzyme could be transported to the nucleus where it could be interacting with specific nuclear proteins. See Foulon *et al.* (1999) *Int. J. Biochem. Cell Biol.* 31:747-750.

Aminopeptidase B exopeptidase activity was originally identified using L-aminoacyl β-naphthylamide and L-amino acid 7-amido-4-methylcoumarins. It was also shown that the enzyme could cleave the lysine residue from the amino terminus of kallidin but could not hydrolyse the arginine-proline of bradykinin. The enzyme purified from rat testis acts on substrates including Arg⁰-Leu-enkephalin, Arg⁰-Met-enkephalin and Arg¹-Lys⁶-somatostatin-14.

Foulon *et al.* characterize aminopeptidase B as specifically removing basic amino acid residues from aminoacyl β-naphthylamides and from the amino terminus of various peptides such as kallidin 10, Met⁵-enkephalin, [Arg⁰]-Leu⁵-enkephalin, [Arg¹-Lys⁰]-somatostatin-14, neurokinin, natriuretic factor and thymopentin. Further, the enzyme is characterized as being sensitive to classical inhibitors of aminopeptidases such as bestatin and arphamenine. Further, it is able to hydrolyse

leukotriene A₄, the natural substrate of leukotriene A₄ hydrolase into leukotriene B₄, a lipid mediator of inflammation.

Based on the structural and biochemical information, Foulon *et al.* have proposed that the aminopeptidase B is associated with post-translational maturation in the trans-Golgi network and regulatory processes in the plasma membrane, including
5 extracellular processing of various peptide substrates. It was further proposed that in the acrosome of spermatids, the aminopeptidase could participate in maturation of proenkephalin and procholecystokinin. In the inflammatory process, it is suggested that potential substrates of the aminopeptidase B include kallidins, enkephalins and
10 somatostatin.

Accordingly, aminopeptidases are a major target for drug action and development. Therefore, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown aminopeptidases. The present invention advances the state of the art by providing a previously unidentified human
15 aminopeptidase.

SUMMARY OF THE INVENTION

It is an object of the invention to identify novel aminopeptidases.

20 It is a further object of the invention to provide novel aminopeptidase polypeptides that are useful as reagents or targets in aminopeptidase assays applicable to treatment and diagnosis of aminopeptidase-related disorders.

It is a further object of the invention to provide polynucleotides corresponding to the novel aminopeptidase polypeptides that are useful as targets and reagents in
25 aminopeptidase assays applicable to treatment and diagnosis of aminopeptidase-related disorders and useful for producing novel aminopeptidase polypeptides by recombinant methods.

A specific object of the invention is to identify compounds that act as agonists and antagonists and modulate the expression of the novel aminopeptidase.

30 A further specific object of the invention is to provide compounds that modulate expression of the aminopeptidase for treatment and diagnosis of aminopeptidase-related disorders.

The invention is thus based on the identification of a novel human aminopeptidase, a human ortholog of rat aminopeptidase B. The amino acid sequence is shown in SEQ ID NO: 1. The nucleotide sequence is shown as SEQ ID NO: 2.

5 The invention provides isolated aminopeptidase polypeptides, including a polypeptide having the amino acid sequence shown in SEQ ID NO: 1 or the amino acid sequence encoded by the cDNA deposited as ATCC No. _____ on _____ (“the deposited cDNA”).

The invention also provides isolated aminopeptidase nucleic acid molecules having the sequence shown in SEQ ID NO: 2 or in the deposited cDNA.

10 The invention also provides variant polypeptides having an amino acid sequence that is substantially homologous to the amino acid sequence shown in SEQ ID NO: 1 or encoded by the deposited cDNA.

The invention also provides variant nucleic acid sequences that are substantially homologous to the nucleotide sequence shown in SEQ ID NO: 2 or in the deposited
15 cDNA.

The invention also provides fragments of the polypeptide shown in SEQ ID NO: 1 and nucleotide sequence shown in SEQ ID NO: 2, as well as substantially homologous fragments of the polypeptide or nucleic acid.

20 The invention further provides nucleic acid constructs comprising the nucleic acid molecules described herein. In a preferred embodiment, the nucleic acid molecules of the invention are operatively linked to a regulatory sequence.

The invention also provides vectors and host cells for expressing the aminopeptidase nucleic acid molecules and polypeptides, and particularly recombinant vectors and host cells.

25 The invention also provides methods of making the vectors and host cells and methods for using them to produce the aminopeptidase nucleic acid molecules and polypeptides.

The invention also provides antibodies or antigen-binding fragments thereof that selectively bind the aminopeptidase polypeptides and fragments.

30 The invention also provides methods of screening for compounds that modulate expression or activity of the aminopeptidase polypeptides or nucleic acid (RNA or DNA).

The invention also provides a process for modulating aminopeptidase polypeptide or nucleic acid expression or activity, especially using the screened compounds. Modulation may be used to treat conditions related to aberrant activity or expression of the aminopeptidase polypeptides or nucleic acids.

5 The invention also provides assays for determining the activity of or the presence or absence of the aminopeptidase polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

The invention also provides assays for determining the presence of a mutation in the polypeptides or nucleic acid molecules, including for disease diagnosis.

10 In still a further embodiment, the invention provides a computer readable means containing the nucleotide and/or amino acid sequences of the nucleic acids and polypeptides of the invention, respectively.

DESCRIPTION OF THE DRAWINGS

15

Figure 1 shows an analysis of the aminopeptidase amino acid sequence: α turn and coil regions; hydrophilicity; amphipathic regions; flexible regions; antigenic index; and surface probability plot.

Figure 2 shows a hydrophobicity plot of the aminopeptidase.

20

Figure 3 shows an analysis of the 2786 aminopeptidase open reading frame for amino acids corresponding to specific functional sites. A cAMP and cGMP-dependent protein kinase phosphorylation site is found from about amino acids 356-359 with the actual modified residue being the last amino acid. A protein kinase C phosphorylation site is found from about amino acids 141-143 and 374-376 with the actual modified residue being the first amino acid. Casein kinase II phosphorylation sites are found from about amino acids 208-211, 318-321, 368-371, 386-389, 408-411, 412-415 and 496-499 with the actual modified residue being the first amino acid. N-myristoylation sites are found from about amino acids 9-14, 58-63, 119-124, 333-338, 364-369, and 614-619 with the actual modified residue being the first amino acid. An amidation site is found from about amino acids 159-162. A eukaryotic putative RNA-binding region RNP-1 signature is found from about amino acids 416-423. The protein also contains a zinc binding region signature found in neutral zinc metalloproteinases at about amino acids 322-331. This last site also corresponds to the

25

30

consensus sequence HEXXHX₁₈E reported in Foulon *et al.* In addition, a putative nuclear localization site (KKK) occurs at amino acids 161-163.

Figure 4 shows 2786 expression (relative to esophagus) in various normal human tissues.

5 Figure 5 shows 2786 expression (relative to vein) in various normal human tissues. Note the high levels of expression in osteoclasts.

Figure 6 shows 2786 expression in 7 normal breast tissue samples and 15 breast carcinoma samples as determined by microarray hybridization analysis. The breast carcinoma samples included 5 invasive lobular carcinoma (ILC) samples, 5 invasive
10 ductal carcinoma (IDC) samples, and 5 ductal carcinoma in situ (DCIS) samples. 2786 expression levels were elevated in all but one of the breast carcinoma samples.

Figure 7 shows 2786 expression in a panel of oncology samples as determined by quantitative PCR (Taqman® brand quantitative PCR kit, Applied Biosystems). The quantitative PCR reactions were performed according to the kit manufacturer's
15 instructions, using the untranslated region of the 2786 cDNA as the target. Eight out of nine breast carcinoma samples showed elevated 2786 expression in comparison with normal breast tissue.

Figure 8 shows 2786 expression in a panel of oncology tissues as determined by quantitative PCR (Taqman® brand quantitative PCR kit, Applied Biosystems). The
20 quantitative PCR reactions were performed according to the kit manufacturer's instructions, using the coding sequence of the 2786 cDNA as the target. Four out of seven breast carcinoma samples showed elevated 2786 expression in comparison with normal breast tissue.

Figure 9 shows the identification of 2786 substrates. The 2786 polypeptide was
25 expressed from three independent Baculovirus expression vectors in insect cells, and crude lysates from these cells were assayed for their ability to hydrolyze various AMC-labeled amino acid substrates as described in the experimental. Lysates from cells expressing a negative control (non-peptidase) protein, from cells infected with the Baculovirus vector alone, or from mock-infected control cells were also analyzed. The
30 2786 aminopeptidase showed hydrolytic specificity for the AMC-labeled tyrosine, proline, arginine, and lysine substrates.

DETAILED DESCRIPTION OF THE INVENTION

Polypeptides

5 The invention is based on the discovery of a novel human aminopeptidase. Specifically, an expressed sequence tag (EST) was selected based on homology to aminopeptidase sequences. This EST was used to design primers based on sequences that it contains and used to identify a cDNA from a human cDNA library. Positive clones were sequenced and the overlapping fragments were assembled. Analysis of the
10 assembled sequence revealed that the cloned cDNA molecule encodes an aminopeptidase, a human ortholog of rat aminopeptidase B.

 The invention thus relates to a novel aminopeptidase having the deduced amino acid sequence shown in SEQ ID NO:1 or having the amino acid sequence encoded by the deposited cDNA, ATCC No. _____ or ATCC No. _____.

15 The deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms. The deposit is provided as a convenience to those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. § 112. The deposited sequence, as well as the polypeptides encoded by the sequence, is incorporated herein by reference and controls in the event of any
20 conflict, such as a sequencing error, with description in this application.

 “Aminopeptidase polypeptide” or “aminopeptidase protein” refers to the polypeptide in SEQ ID NO: 1 or encoded by the deposited cDNA. The term “aminopeptidase protein” or “aminopeptidase polypeptide”, however, further includes the numerous variants described herein, as well as fragments derived from the full-length
25 aminopeptidase and variants.

 Tissues and/or cells in which the aminopeptidase is found include, but are not limited to, the tissues shown in Figures 4-8. In normal human tissues, the highest expression was shown in fetal kidney, fetal liver, and osteoclasts. Moderate expression was shown in prostate, breast, brain, testis and undifferentiated osteoblasts.
30 Expression of 2786 in breast tissue was confirmed by *in situ* hybridization with a 2786 antisense probe.

 Expression of 2786 in breast carcinoma was high relative to normal breast tissue. 2786 expression was elevated in all three types of breast carcinoma examined

(invasive lobular carcinomas, invasive ductal carcinomas, and ductal carcinomas in situ). The increase in 2786 expression in breast carcinomas was observed regardless of whether the 2786 coding sequence (Figure 8) or untranslated region (Figure 7) was used as the quantitative PCR target.

5 The present invention thus provides an isolated or purified aminopeptidase polypeptide and variants and fragments thereof.

 Based on a BLAST search, 2786 showed highest was most similar to aminopeptidase B. Amino acids 32-417 of the 2786 polypeptide sequence (shown in SEQ ID NO:1) constitute a peptidase family M1 domain that shares 88% identity and
10 94% similarity to the rat aminopeptidase B (GenBank Accession No. D87515) M1 domain.

 As used herein, a polypeptide is said to be "isolated" or "purified" when it is substantially free of cellular material when it is isolated from recombinant and non-recombinant cells, or free of chemical precursors or other chemicals when it is
15 chemically synthesized. A polypeptide, however, can be joined to another polypeptide with which it is not normally associated in a cell and still be considered "isolated" or "purified."

 The aminopeptidase polypeptides can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to
20 homogeneity are useful and considered to contain an isolated form of the polypeptide. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity.

 In one embodiment, the language "substantially free of cellular material"
25 includes preparations of the aminopeptidase having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the polypeptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20%, less than about 10%, or less than about 5% of
30 the volume of the protein preparation.

 An aminopeptidase polypeptide is also considered to be isolated when it is part of a membrane preparation or is purified and then reconstituted with membrane vesicles or liposomes.

The language “substantially free of chemical precursors or other chemicals” includes preparations of the aminopeptidase polypeptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language “substantially free of chemical precursors or other chemicals” includes preparations of the polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

In one embodiment, the aminopeptidase polypeptide comprises the amino acid sequence shown in SEQ ID NO:1. However, the invention also encompasses sequence variants. Variants include a substantially homologous protein encoded by the same genetic locus in an organism, i.e., an allelic variant. Variants also encompass proteins derived from other genetic loci in an organism, but having substantial homology to the aminopeptidase of SEQ ID NO: 1. Variants also include proteins substantially homologous to the aminopeptidase but derived from another organism, i.e., an ortholog. Variants also include proteins that are substantially homologous to the aminopeptidase that are produced by chemical synthesis. Variants also include proteins that are substantially homologous to the aminopeptidase that are produced by recombinant methods. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

As used herein, two proteins (or a region of the proteins) are substantially homologous when the amino acid sequences are at least about 60-65%, 65-70%, 70-75%, 75-80%, typically at least about 80-85% or 85-90%, and most typically at least about 90-95% or more homologous. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence hybridizing to the nucleic acid sequence, or portion thereof, of the sequence shown in SEQ ID NO: 2 under stringent conditions as more fully described below.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at

least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence.

The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is
5 occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid “identity” is equivalent to amino acid or nucleic acid “homology”). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the
10 number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by the aminopeptidase. Similarity is determined by conserved amino acid
15 substitution. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp
20 and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

TABLE 1. Conservative Amino Acid Substitutions.

Aromatic	Phenylalanine
	Tryptophan
	Tyrosine
Hydrophobic	Leucine
	Isoleucine
	Valine
Polar	Glutamine
	Asparagine
Basic	Arginine
	Lysine
	Histidine
Acidic	Aspartic Acid
	Glutamic Acid
Small	Alanine
	Serine
	Threonine
	Methionine
	Glycine

The comparison of sequences and determination of percent identity and
 5 similarity between two sequences can be accomplished using a mathematical
 algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University
 Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W.,
 ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*,
 Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence*
 10 *Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence*
Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York,
 1991).

A preferred, non-limiting example of such a mathematical algorithm is described in Karlin *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402. When utilizing BLAST and
5 Gapped BLAST programs, the default parameters of the respective programs (e.g., NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>. In one embodiment, parameters for sequence comparison can be set at score = 100, wordlength = 12, or can be varied (e.g., W = 5 or W = 20).

In a preferred embodiment, the percent identity between two amino acid
10 sequences is determined using the Needleman *et al.* (1970) (*J. Mol. Biol.* 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a BLOSUM 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two
15 nucleotide sequences is determined using the GAP program in the GCG software package (Devereux *et al.* (1984) *Nucleic Acids Res.* 12(1):387) (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6.

Another preferred, non-limiting example of a mathematical algorithm utilized for
20 the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the CGC sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis
25 are known in the art and include ADVANCE and ADAM as described in Torellis *et al.* (1994) *Comput. Appl. Biosci.* 10:3-5; and FASTA described in Pearson *et al.* (1988) *PNAS* 85:2444-8.

A variant polypeptide can differ in amino acid sequence by one or more substitutions, deletions, insertions, inversions, fusions, and truncations or a combination
30 of any of these.

Variant polypeptides can be fully functional or can lack function in one or more activities. Thus, in the present case, variations can affect the function, for example, of one or more of the regions corresponding to the catalytic regions, regulatory regions,

substrate binding regions, zinc binding regions, regions involved in membrane association, regions involved in ribonucleoprotein binding, regions involved in nuclear localisation, and regions involved in enzyme modification, for example, by phosphorylation, amidation, or N-myristoylation. Catalytic regions include regions
5 involved in broad specificity exopeptidase activity and regions involved in leukotriene A₄ epoxide hydrolase activity.

Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids, which results in no change or an insignificant
10 change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

15 As indicated, variants can be naturally-occurring or can be made by recombinant means or chemical synthesis to provide useful and novel characteristics for the aminopeptidase polypeptides. This includes preventing immunogenicity from pharmaceutical formulations by preventing protein aggregation.

Useful variations further include alteration of catalytic activity. For example,
20 one embodiment involves a variation at the peptide binding site that results in binding but not hydrolysis of the peptide substrate. A further useful variation at the same site can result in altered affinity for the peptide substrate. Useful variations also include changes that provide specificity for another peptide substrate or the ability to cleave peptide substrates at amino acids other than lysine or arginine. Another useful variation
25 provides a fusion protein in which one or more domains or subregions are operationally fused to one or more domains or subregions from another aminopeptidase, such as regions that determine hydrolysis sites, for example the ability to cleave peptide bonds at other than arginine and lysine.

Amino acids that are essential for function can be identified by methods known
30 in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.* (1985) *Science* 244:1081-1085). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity, such as peptide bond hydrolysis *in vitro*

or related biological activity, such as proliferative or inflammatory activity. Sites that are critical for binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.* (1992) *J. Mol. Biol.* 224:899-904; de Vos *et al.* (1992) *Science* 255:306-312).

5 Substantial homology can be to the entire nucleic acid or amino acid sequence or to fragments of these sequences.

The invention thus also includes polypeptide fragments of the aminopeptidase. Fragments can be derived from the amino acid sequence shown in SEQ ID NO:1. However, the invention also encompasses fragments of the variants of the

10 aminopeptidase as described herein.

The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed prior to the present invention.

Accordingly, a fragment can comprise at least about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more contiguous amino acids. Fragments can retain one or more of the
15 biological activities of the protein, for example the ability to bind to or hydrolyze target peptides, as well as fragments that can be used as an immunogen to generate aminopeptidase antibodies.

Biologically active fragments (peptides which are, for example, 5, 7, 10, 12, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) can
20 comprise a functional site. Such sites include but are not limited to the catalytic sites, regulatory sites, sites important for substrate recognition or binding, zinc binding region, the site(s) contributing to exopeptidase specificity, nuclear localization site, ribonucleoprotein binding site, membrane binding site, phosphorylation sites, glycosylation sites, and other functional sites disclosed herein. Catalytic sites include
25 broad specificity exopeptidase sites and leukotriene A₄ hydrolase sites. In the present case, these sites include, but are not limited to, KKK, HEISH WLNE, and KGFCFVSY.

Such sites or motifs can be identified by means of routine computerized homology searching procedures.

30 Fragments, for example, can extend in one or both directions from the functional site to encompass 5, 10, 15, 20, 30, 40, 50, or up to 100 amino acids. Further, fragments can include sub-fragments of the specific sites or regions disclosed herein, which sub-fragments retain the function of the site or region from which they are derived.

These regions can be identified by well-known methods involving computerized homology analysis.

The invention also provides fragments with immunogenic properties. These contain an epitope-bearing portion of the aminopeptidase and variants. These epitope-bearing peptides are useful to raise antibodies that bind specifically to an aminopeptidase polypeptide or region or fragment. These peptides can contain at least 10, 12, at least 14, or between at least about 15 to about 30 amino acids.

Non-limiting examples of antigenic polypeptides that can be used to generate antibodies include but are not limited to peptides derived from extracellular regions.

Regions having a high antigenicity index are shown in Figure 1. However, intracellularly-made antibodies ("intrabodies") are also encompassed, which would recognize intracellular peptide regions.

The epitope-bearing aminopeptidase polypeptides may be produced by any conventional means (Houghten, R.A. (1985) *Proc. Natl. Acad. Sci. USA* 82:5131-5135).

Simultaneous multiple peptide synthesis is described in U.S. Patent No. 4,631,211.

Fragments can be discrete (not fused to other amino acids or polypeptides) or can be within a larger polypeptide. Further, several fragments can be comprised within a single larger polypeptide. In one embodiment a fragment designed for expression in a host can have heterologous pre- and pro-polypeptide regions fused to the amino terminus of the aminopeptidase fragment and an additional region fused to the carboxyl terminus of the fragment.

The invention thus provides chimeric or fusion proteins. These comprise an aminopeptidase peptide sequence operatively linked to a heterologous peptide having an amino acid sequence not substantially homologous to the aminopeptidase. "Operatively linked" indicates that the aminopeptidase peptide and the heterologous peptide are fused in-frame. The heterologous peptide can be fused to the N-terminus or C-terminus of the aminopeptidase or can be internally located.

In one embodiment the fusion protein does not affect aminopeptidase function *per se*. For example, the fusion protein can be a GST-fusion protein in which the aminopeptidase sequences are fused to the N- or C-terminus of the GST sequences. Other types of fusion proteins include, but are not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL4 fusions, poly-His fusions and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate

the purification of recombinant aminopeptidase. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence. Therefore, in another embodiment, the fusion protein contains a heterologous signal sequence at its C- or N-terminus.

5 EP-A-O 464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists (Bennett *et al.* (1995) *J. Mol.*
10 *Recog.* 8:52-58 (1995) and Johanson *et al.* *J. Biol. Chem.* 270:9459-9471). Thus, this invention also encompasses soluble fusion proteins containing an aminopeptidase polypeptide and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclass (IgG, IgM, IgA, IgE). Preferred as immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1,
15 where fusion takes place at the hinge region. For some uses it is desirable to remove the Fc after the fusion protein has been used for its intended purpose, for example when the fusion protein is to be used as antigen for immunizations. In a particular embodiment, the Fc part can be removed in a simple way by a cleavage sequence, which is also incorporated and can be cleaved with factor Xa.

20 A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can
25 be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.* (1992) *Current Protocols in Molecular Biology*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). An aminopeptidase-
30 encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the aminopeptidase.

Another form of fusion protein is one that directly affects aminopeptidase functions. Accordingly, an aminopeptidase polypeptide is encompassed by the present

invention in which one or more of the aminopeptidase regions (or parts thereof) has been replaced by homologous regions (or parts thereof) from another aminopeptidase.

Accordingly, various permutations are possible. Thus, chimeric aminopeptidases can be formed in which one or more of the native domains or subregions has been replaced by another.

Additionally, chimeric aminopeptidase proteins can be produced in which one or more functional sites is derived from a different aminopeptidase, including but not limited to those aminopeptidases disclosed above in the background discussion. Sites include but are not limited to exopeptidase specificity, exopeptidase binding, sites required for hydrolysis, Zn^{2+} binding sites, leukotriene A_4 hydrolase activity sites, membrane binding sites, ribonucleoprotein binding sites, and nuclear localization sites. It is understood however that sites could be derived from aminopeptidases that occur in the mammalian genome but which have not yet been discovered or characterized.

The isolated aminopeptidase protein can be purified from cells that naturally express it, such as from osteoclasts or from any of the tissues shown in Figures 4-8, including but not limited to brain, testis, fetal kidney, fetal liver, and breast tumor tissue, especially purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods.

In one embodiment, the protein is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the aminopeptidase polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally-occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in polypeptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art.

Accordingly, the polypeptides also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a

substituent group is included, in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature polypeptide or a pro-protein sequence.

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (1990) *Meth. Enzymol.* 182: 626-646) and Rattan *et al.* (1992) *Ann. N.Y. Acad. Sci.* 663:48-62).

As is also well known, polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of post-translation events, including natural processing events and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translational natural processes and by synthetic methods.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. Blockage of

the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally-occurring and synthetic polypeptides. For instance, the aminoterminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

- 5 The modifications can be a function of how the protein is made. For recombinant polypeptides, for example, the modifications will be determined by the host cell posttranslational modification capacity and the modification signals in the polypeptide amino acid sequence. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell.
- 10 Insect cells often carry out the same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to efficiently express mammalian proteins having native patterns of glycosylation. Similar considerations apply to other modifications.

- The same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain more than one type of modification.
- 15

Polypeptide Uses

- The protein sequences of the present invention can be used as a “query sequence” to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.
- 20
- 25
- 30

 The aminopeptidase polypeptides are useful for producing antibodies specific for the aminopeptidase, regions, or fragments. Regions having a high antigenicity index

score are shown in Figure 1.

The aminopeptidase polypeptides are useful for biological assays related to aminopeptidases in general and to the 2786 protein specifically. Such assays involve any of the known functions or activities or properties useful for diagnosis and treatment of 2786 aminopeptidase-related conditions and screening for agents that modulate activity or levels. Functions include but are not limited to broadly specific exopeptidase activity for basic amino acids (lysine and/or arginine), dependence upon Zn^{2+} for exopeptidase activity, leukotriene A_4 epoxide hydrolase activity, and the ability to be inhibited by classical aminopeptidase inhibitors such as bestatin and arphamenine A and B. Assays that are useful for determining exopeptidase activity include but are not limited to those known in the art, such as cleavage of Arg^0 -Leu-enkephalin, Arg^0 -Met-enkephalin, Arg^1 -Lys⁶-somatostatin-14, leukotriene $A_4 \rightarrow$ leukotriene B_4 , the removal of basic amino acids from L-amino acyl β -naphthylamides, the removal of basic amino L-amino acid-7-amido-4-methylcoumarins, cleavage of lysine from the aminotermminus of kallidins, Arg^0 -neurokinin A, Arg^0 - α -atrial natriuretic factor, and thymopentin. The substrates also include, but are not limited to, those described in the background section above, and in the references cited therein, which are incorporated by reference for the assays disclosed therein.

The aminopeptidase polypeptides are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the aminopeptidase, as a biopsy or expanded in cell culture. In one embodiment, however, cell-based assays involve recombinant host cells expressing the aminopeptidase.

Determining the ability of the test compound to interact with the aminopeptidase can also comprise determining the ability of the test compound to preferentially bind to the polypeptide as compared to the ability of a known binding molecule to bind to the polypeptide.

The polypeptides can be used to identify compounds that modulate aminopeptidase activity. Such compounds, for example, can increase or decrease affinity or rate of binding to peptide substrate, compete with peptide substrate for binding to the aminopeptidase, or displace peptide substrate bound to the aminopeptidase. Both aminopeptidase and appropriate variants and fragments can be

used in high-throughput screens to assay candidate compounds for the ability to bind to the aminopeptidase. These compounds can be further screened against a functional aminopeptidase to determine the effect of the compound on the aminopeptidase activity. Compounds can be identified that activate (agonist) or inactivate (antagonist) the aminopeptidase to a desired degree. Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject).

The aminopeptidase polypeptides can be used to screen a compound for the ability to stimulate or inhibit interaction between the aminopeptidase protein and a target molecule that normally interacts with the aminopeptidase protein, for example, substrate-peptide or zinc component. The assay includes the steps of combining the aminopeptidase protein with a candidate compound under conditions that allow the aminopeptidase protein or fragment to interact with the target molecule, and to detect the formation of a complex between the aminopeptidase protein and the target or to detect the biochemical consequence of the interaction with the aminopeptidase and the target.

Determining the ability of the aminopeptidase to bind to a target molecule can also be accomplished using a technology such as real-time Bimolecular Interaction Analysis (BIA). Sjolander *et al.* (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore™). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in

the art, for example in DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233. Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 97:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra*).

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.* (1991) *Nature* 354:82-84; Houghten *et al.* (1991) *Nature* 354:84-86) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.* (1993) *Cell* 72:767-778); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')₂, Fab expression library fragments, and epitope-binding fragments of antibodies); 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries; 5) zinc analogs; 6) leukotriene A₄ and derivatives; 7) classical aminopeptidase inhibitors and derivatives of such inhibitors, such as bestatin and arphamenine A and B and derivatives; 8) artificial peptide substrates and other substrates, such as those disclosed hereinabove and derivatives thereof; and 9) other inhibitors and derivatives thereof, such as those disclosed in the background above.

One candidate compound is a soluble full-length aminopeptidase or fragment that competes for peptide binding. Other candidate compounds include mutant aminopeptidases or appropriate fragments containing mutations that affect aminopeptidase function and compete for peptide substrate. Accordingly, a fragment that competes for substrate, for example with a higher affinity, or a fragment that binds substrate but does not hydrolyze it, is encompassed by the invention.

The invention provides other end points to identify compounds that modulate (stimulate or inhibit) aminopeptidase activity. The assays typically involve an assay of cellular events that indicate aminopeptidase activity. Thus, the expression of genes that are up- or down-regulated in response to the aminopeptidase activity can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase. Alternatively, modification of the aminopeptidase could also be measured.

Any of the biological or biochemical functions mediated by the aminopeptidase can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art. These include but are not limited to amino acid hydrolysis products from hydrolysis of the amino terminus of polypeptides and proteins, and more global biological effects such as hormone level and function, neurotransmitter level and function, cell cycle control, modulation of cell-cell interaction, general protein turnover, and effects on inflammatory processes. In one embodiment, bone resorption (i.e., osteoclast function) can be measured.

Binding and/or activating compounds can also be screened by using chimeric aminopeptidase proteins in which one or more regions, segments, sites, and the like, as disclosed herein, or parts thereof, can be replaced by their heterologous counterparts derived from other aminopeptidases. For example, a catalytic region can be used that interacts with a different peptide sequence specificity and/or affinity than the native aminopeptidase. Accordingly, a different substrate is available as an end-point assay for activation. As a further alternative, the site of modification by an effector protein, for example phosphorylation, can be replaced with the site for a different effector protein. Activation can also be detected by a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native pathway in which the aminopeptidase is involved.

The aminopeptidase polypeptides are also useful in competition binding assays in methods designed to discover compounds that interact with the aminopeptidase. Thus, a compound is exposed to an aminopeptidase polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble aminopeptidase polypeptide is also added to the mixture. If the test compound interacts

with the soluble aminopeptidase polypeptide, it decreases the amount of complex formed or activity from the aminopeptidase target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the aminopeptidase. Thus, the soluble polypeptide that competes with the target
5 aminopeptidase region is designed to contain peptide sequences corresponding to the region of interest.

Another type of competition-binding assay can be used to discover compounds that interact with specific functional sites. As an example, bindable zinc and a candidate compound can be added to a sample of the aminopeptidase. Compounds that interact
10 with the aminopeptidase at the same site as the zinc will reduce the amount of complex formed between the aminopeptidase and the zinc. Accordingly, it is possible to discover a compound that specifically prevents interaction between the aminopeptidase and the zinc component. Another example involves adding a candidate compound to a sample of aminopeptidase and substrate. These substrates include peptides with which the
15 aminopeptidase normally interacts as well as the various substrates for which exopeptidase activity has been shown, discussed hereinabove. A compound that competes with the substrate will reduce the amount of hydrolysis or binding of the substrate to the aminopeptidase. Accordingly, compounds can be discovered that directly interact with the aminopeptidase and compete with the substrate. Such assays
20 can involve any other component that interacts with the aminopeptidase, including but not limited to membrane components responsible for membrane binding, nuclear components responsible for ribonucleoprotein binding, and nuclear components responsible for nuclear localization.

To perform cell free drug screening assays, it is desirable to immobilize either
25 the aminopeptidase, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a
30 domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/aminopeptidase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ³⁵S-labeled) and the candidate

compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes is dissociated.

- 5 Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of aminopeptidase-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies
- 10 reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of an aminopeptidase-binding target component, such as a peptide or zinc component, and a candidate compound are incubated in the aminopeptidase-presenting wells and the amount of complex trapped in
- 15 the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the aminopeptidase target molecule, or which are reactive with aminopeptidase and compete with the target molecule; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the
- 20 target molecule.

- Modulators of aminopeptidase activity identified according to these drug screening assays can be used to treat a subject with a disorder related to the aminopeptidase, by treating cells that express the aminopeptidase, such as any of those shown in Figures 4-8. In one embodiment, the cells that are modulated are found in
- 25 the prostate, breast, liver, brain, testis, fetal kidney, and fetal liver. In another embodiment, the cells in which the activity is modulated are cells in a breast carcinoma, lung carcinoma, colon carcinoma, or colonic metastases to the liver. In another embodiment, modulation occurs in osteoclasts. These methods of treatment include the steps of administering the modulators of aminopeptidase activity in a
- 30 pharmaceutical composition as described herein, to a subject in need of such treatment.

Disorders involving the colon include, but are not limited to, congenital anomalies, such as atresia and stenosis, Meckel diverticulum, congenital aganglionic megacolon-Hirschsprung disease; enterocolitis, such as diarrhea and dysentery,

infectious enterocolitis, including viral gastroenteritis, bacterial enterocolitis, necrotizing enterocolitis, antibiotic-associated colitis (pseudomembranous colitis), and collagenous and lymphocytic colitis, miscellaneous intestinal inflammatory disorders, including parasites and protozoa, acquired immunodeficiency syndrome, 5 transplantation, drug-induced intestinal injury, radiation enterocolitis, neutropenic colitis (typhlitis), and diversion colitis; idiopathic inflammatory bowel disease, such as Crohn disease and ulcerative colitis; tumors of the colon, such as non-neoplastic polyps, adenomas, familial syndromes, colorectal carcinogenesis, colorectal carcinoma, and carcinoid tumors.

10 Disorders involving the prostate include, but are not limited to, inflammations, benign enlargement, for example, nodular hyperplasia (benign prostatic hypertrophy or hyperplasia), and tumors such as carcinoma.

Disorders involving the liver include, but are not limited to, hepatic injury; jaundice and cholestasis, such as bilirubin and bile formation; hepatic failure and 15 cirrhosis, such as cirrhosis, portal hypertension, including ascites, portosystemic shunts, and splenomegaly; infectious disorders, such as viral hepatitis, including hepatitis A-E infection and infection by other hepatitis viruses, clinicopathologic syndromes, such as the carrier state, asymptomatic infection, acute viral hepatitis, chronic viral hepatitis, and fulminant hepatitis; autoimmune hepatitis; drug- and 20 toxin-induced liver disease, such as alcoholic liver disease; inborn errors of metabolism and pediatric liver disease, such as hemochromatosis, Wilson disease, α_1 -antitrypsin deficiency, and neonatal hepatitis; intrahepatic biliary tract disease, such as secondary biliary cirrhosis, primary biliary cirrhosis, primary sclerosing cholangitis, and anomalies of the biliary tree; circulatory disorders, such as impaired blood flow 25 into the liver, including hepatic artery compromise and portal vein obstruction and thrombosis, impaired blood flow through the liver, including passive congestion and centrilobular necrosis and peliosis hepatis, hepatic vein outflow obstruction, including hepatic vein thrombosis (Budd-Chiari syndrome) and veno-occlusive disease; hepatic disease associated with pregnancy, such as preeclampsia and eclampsia, acute fatty 30 liver of pregnancy, and intrahepatic cholestasis of pregnancy; hepatic complications of organ or bone marrow transplantation, such as drug toxicity after bone marrow transplantation, graft-versus-host disease and liver rejection, and nonimmunologic damage to liver allografts; tumors and tumorous conditions, such as nodular

hyperplasias, adenomas, and malignant tumors, including primary carcinoma of the liver and metastatic tumors.

Disorders of the breast include, but are not limited to, disorders of development; inflammations, including but not limited to, acute mastitis, periductal mastitis, periductal mastitis (recurrent subareolar abscess, squamous metaplasia of lactiferous ducts), mammary duct ectasia, fat necrosis, granulomatous mastitis, and pathologies associated with silicone breast implants; fibrocystic changes; proliferative breast disease including, but not limited to, epithelial hyperplasia, sclerosing adenosis, and small duct papillomas; tumors including, but not limited to, stromal tumors such as fibroadenoma, phyllodes tumor, and sarcomas, and epithelial tumors such as large duct papilloma; carcinoma of the breast including in situ (noninvasive) carcinoma that includes ductal carcinoma in situ (including Paget's disease) and lobular carcinoma in situ, and invasive (infiltrating) carcinoma including, but not limited to, invasive ductal carcinoma, no special type, invasive lobular carcinoma, medullary carcinoma, colloid (mucinous) carcinoma, tubular carcinoma, and invasive papillary carcinoma, and miscellaneous malignant neoplasms.

Disorders in the male breast include, but are not limited to, gynecomastia and carcinoma.

Disorders involving the skeletal muscle include tumors such as rhabdomyosarcoma.

Disorders involving the brain include, but are not limited to, disorders involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states--global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive cerebrovascular disease, including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain

abscess, subdural empyema, and extradural abscess, chronic bacterial meningoencephalitis, including tuberculosis and mycobacterioses, neurosyphilis, and neuroborreliosis (Lyme disease), viral meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2, *Varicella-zoster* virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral neuropathy, and AIDS in children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute necrotizing hemorrhagic encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer disease and Pick disease, degenerative diseases of basal ganglia and brain stem, including Parkinsonism, idiopathic Parkinson disease (paralysis agitans), progressive supranuclear palsy, corticobasal degeneration, multiple system atrophy, including striatonigral degeneration, Shy-Drager syndrome, and olivopontocerebellar atrophy, and Huntington disease; spinocerebellar degenerations, including spinocerebellar ataxias, including Friedreich ataxia, and ataxia-telangiectasia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; inborn errors of metabolism, such as leukodystrophies, including Krabbe disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, and Canavan disease, mitochondrial encephalomyopathies, including Leigh disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic diseases, including vitamin deficiencies such as thiamine (vitamin B₁) deficiency and vitamin B₁₂ deficiency, neurologic sequelae of metabolic disturbances, including hypoglycemia, hyperglycemia, and hepatic encephatopathy, toxic disorders, including carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate and radiation-induced injury; tumors, such as gliomas, including astrocytoma, including fibrillary (diffuse) astrocytoma and glioblastoma multiforme, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma,

oligodendroglioma, and ependymoma and related paraventricular mass lesions, neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors, including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including neurofibromatosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease.

10 Disorders involving the testis and epididymis include, but are not limited to, congenital anomalies such as cryptorchidism, regressive changes such as atrophy, inflammations such as nonspecific epididymitis and orchitis, granulomatous (autoimmune) orchitis, and specific inflammations including, but not limited to, gonorrhea, mumps, tuberculosis, and syphilis, vascular disturbances including torsion, testicular tumors including germ cell tumors that include, but are not limited to, seminoma, spermatocytic seminoma, embryonal carcinoma, yolk sac tumor choriocarcinoma, teratoma, and mixed tumors, tumors of sex cord-gonadal stroma including, but not limited to, leydig (interstitial) cell tumors and sertoli cell tumors (androblastoma), and testicular lymphoma, and miscellaneous lesions of tunica vaginalis.

Bone-forming cells include the osteoprogenitor cells, osteoblasts, and osteocytes. The disorders of the bone are complex because they may have an impact on the skeleton during any of its stages of development. Hence, the disorders may have variable manifestations and may involve one, multiple or all bones of the body.

25 Such disorders include, congenital malformations, achondroplasia and thanatophoric dwarfism, diseases associated with abnormal matrix such as type 1 collagen disease, osteoporosis, Paget's disease, rickets, osteomalacia, high-turnover osteodystrophy, low-turnover of aplastic disease, osteonecrosis, pyogenic osteomyelitis, tuberculous osteomyelitis, osteoma, osteoid osteoma, osteoblastoma, osteosarcoma, osteochondroma, chondromas, chondroblastoma, chondromyxoid fibroma, chondrosarcoma, fibrous cortical defects, fibrous dysplasia, fibrosarcoma, malignant fibrous histiocytoma, Ewing's sarcoma, primitive neuroectodermal tumor, giant cell tumor, and metastatic tumors.

Disorders in which aminopeptidase expression is especially relevant include, but are not limited to, breast carcinoma. The aminopeptidase is also expressed differentially in lung and colon carcinoma. As such, the gene is relevant for the treatment of these disorders, where increasing or inhibiting expression of the gene could affect tumor
5 development and/or progression.

The 2786 aminopeptidase is highly expressed in osteoclasts. Further, it displays homology to leukotriene A₄ hydrolase. Leukotriene A₄ hydrolase catalyzes the conversion of leukotriene A₄ to B₄. B₄ is a known stimulator of osteoclasts activity *in vitro* and *in vivo*. Accordingly, modulation of this gene is particularly important in
10 conditions and disorders involving bone resorption. Inhibition of the gene for example could function to prevent or decrease bone resorption *in vivo*. Conversely, increasing the level or activity of the gene could be used to increase resorption *in vivo*. The gene is also differentially expressed in differentiated and undifferentiated osteoblasts. Therefore, the gene may function in bone formation. Modulation of the
15 gene, therefore, may be relevant to this process.

The aminopeptidase polypeptides are thus useful for treating an aminopeptidase-associated disorder characterized by aberrant expression or activity of an aminopeptidase. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of
20 agents that modulates (e.g., upregulates or downregulates) expression or activity of the protein. In another embodiment, the method involves administering the aminopeptidase as therapy to compensate for reduced or aberrant expression or activity of the protein.

Methods for treatment include but are not limited to the use of soluble
25 aminopeptidase or fragments of the aminopeptidase protein that compete for substrate or any other component that directly interacts with the aminopeptidase, such as zinc or any of the enzymes that modify the aminopeptidase. These aminopeptidases or fragments can have a higher affinity for the target so as to provide effective competition.

Stimulation of activity is desirable in situations in which the protein is
30 abnormally downregulated and/or in which increased activity is likely to have a beneficial effect. Likewise, inhibition of activity is desirable in situations in which the protein is abnormally upregulated and/or in which decreased activity is likely to have a beneficial effect. In one example of such a situation, a subject has a disorder

characterized by aberrant development or cellular differentiation, for example bone development or differentiation. In another example, the subject has a proliferative disease (e.g., cancer) or a disorder characterized by an aberrant hematopoietic response. In another example, it is desirable to achieve tissue regeneration in a subject (e.g., where a subject has undergone brain or spinal cord injury and it is desirable to regenerate neuronal tissue in a regulated manner).

In yet another aspect of the invention, the proteins of the invention can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO 94/10300), to identify other proteins (captured proteins) which bind to or interact with the proteins of the invention and modulate their activity.

The aminopeptidase polypeptides also are useful to provide a target for diagnosing a disease or predisposition to disease mediated by the aminopeptidase, including, but not limited to, those diseases discussed herein, and particularly breast carcinoma and excessive or deficient bone mass. Targets are useful for diagnosing a disease or predisposition to disease mediated by the aminopeptidase, in the tissues shown in Figures 4-8. Accordingly, methods are provided for detecting the presence, or levels of, the aminopeptidase in a cell, tissue, or organism. The method involves contacting a biological sample with a compound capable of interacting with the aminopeptidase such that the interaction can be detected.

One agent for detecting aminopeptidase is an antibody capable of selectively binding to aminopeptidase. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The aminopeptidase also provides a target for diagnosing active disease, or predisposition to disease, in a patient having a variant aminopeptidase. Thus, aminopeptidase can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in an aberrant protein. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered aminopeptidase activity in cell-based or cell-free assay, alteration in peptide binding or degradation, zinc binding or

antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein in general or in an aminopeptidase specifically.

In vitro techniques for detection of aminopeptidase include enzyme linked
5 immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. Alternatively, the protein can be detected *in vivo* in a subject by introducing into the subject a labeled anti-aminopeptidase antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods,
10 which detect the allelic variant of the aminopeptidase expressed in a subject, and methods, which detect fragments of the aminopeptidase in a sample.

The aminopeptidase polypeptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See,
15 e.g., Eichelbaum, M. (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985, and Linder, M.W. (1997) *Clin. Chem.* 43(2):254-266. The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic
20 compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes affects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms
25 in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the aminopeptidase in which one or
30 more of the aminopeptidase functions in one population is different from those in another population. The polypeptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a peptide-based treatment, polymorphism may give rise to catalytic regions that are more or less active.

Accordingly, dosage would necessarily be modified to maximize the therapeutic effect within a given population containing the polymorphism. As an alternative to genotyping, specific polymorphic polypeptides could be identified.

The aminopeptidase polypeptides are also useful for monitoring therapeutic effects during clinical trials and other treatment. Thus, the therapeutic effectiveness of an agent that is designed to increase or decrease gene expression, protein levels or aminopeptidase activity can be monitored over the course of treatment using the aminopeptidase polypeptides as an end-point target. The monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression or activity of the protein in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the protein in the post-administration samples; (v) comparing the level of expression or activity of the protein in the pre-administration sample with the protein in the post-administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

Antibodies

The invention also provides antibodies that selectively bind to the aminopeptidase and its variants and fragments. An antibody is considered to selectively bind, even if it also binds to other proteins that are not substantially homologous with the aminopeptidase. These other proteins share homology with a fragment or domain of the aminopeptidase. This conservation in specific regions gives rise to antibodies that bind to both proteins by virtue of the homologous sequence. In this case, it would be understood that antibody binding to the aminopeptidase is still selective.

To generate antibodies, an isolated aminopeptidase polypeptide is used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. Either the full-length protein or antigenic peptide fragment can be used. Regions having a high antigenicity index are shown in Figure 1.

Antibodies are preferably prepared from these regions or from discrete fragments in these regions. However, antibodies can be prepared from any region of the peptide as described herein. A preferred fragment produces an antibody that diminishes or completely prevents peptide hydrolysis or binding. Antibodies can be

developed against the entire aminopeptidase or domains of the aminopeptidase as described herein, for example, the zinc binding region, sites contributing to exopeptidase specificity, and the peptidase domain or subregions thereof. Antibodies can also be developed against specific functional sites as disclosed herein.

5 The antigenic peptide can comprise a contiguous sequence of at least 12, 14, 15, or 30 amino acid residues. In one embodiment, fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions. These fragments are not to be construed, however, as encompassing any fragments, which may be disclosed prior to the invention.

10 Antibodies can be polyclonal or monoclonal. An intact antibody, or a fragment thereof (e.g. Fab or F(ab')₂) can be used.

Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent
15 materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl
20 chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ¹²⁵I, ¹³¹I, ³⁵S or ³H.

An appropriate immunogenic preparation can be derived from native, recombinantly expressed, or chemically synthesized peptides.

25

Antibody Uses

The antibodies can be used to isolate an aminopeptidase by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural aminopeptidase from cells and recombinantly produced
30 aminopeptidase expressed in host cells.

The antibodies are useful to detect the presence of aminopeptidase in cells or tissues to determine the pattern of expression of the aminopeptidase among various tissues in an organism and over the course of normal development.

The antibodies can be used to detect aminopeptidase *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression.

The antibodies can be used to assess abnormal tissue distribution or abnormal expression during development.

5 Antibody detection of circulating fragments of the full length aminopeptidase can be used to identify aminopeptidase turnover.

Further, the antibodies can be used to assess aminopeptidase expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to aminopeptidase function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, or level of expression of the aminopeptidase protein, the antibody can be prepared against the normal aminopeptidase protein. If a disorder is characterized by a specific mutation in the aminopeptidase, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant aminopeptidase. However, intracellularly-made antibodies (“intrabodies”) are also encompassed, which would recognize intracellular aminopeptidase peptide regions.

15 The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Antibodies can be developed against the whole aminopeptidase or portions of the aminopeptidase, for example, the functional sites disclosed herein.

20 The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting aminopeptidase expression level or the presence of aberrant aminopeptidases and aberrant tissue distribution or developmental expression, antibodies directed against the aminopeptidase or relevant fragments can be used to monitor therapeutic efficacy.

25 Antibodies accordingly can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen.

30 Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic aminopeptidase can be used to identify individuals that require modified treatment modalities.

The antibodies are also useful as diagnostic tools as an immunological marker for aberrant aminopeptidase analyzed by electrophoretic mobility, isoelectric point,

tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Thus, where a specific aminopeptidase has been correlated with expression in a specific tissue, antibodies that are specific for this aminopeptidase can be used to identify a tissue type.

5 The antibodies are also useful in forensic identification. Accordingly, where an individual has been correlated with a specific genetic polymorphism resulting in a specific polymorphic protein, an antibody specific for the polymorphic protein can be used as an aid in identification.

10 The antibodies are also useful for inhibiting aminopeptidase function, for example, zinc binding, and peptide binding and/or hydrolysis.

These uses can also be applied in a therapeutic context in which treatment involves inhibiting aminopeptidase function. An antibody can be used, for example, to block peptide binding. Antibodies can be prepared against specific fragments containing sites required for function or against intact aminopeptidase associated with a cell.

15 Completely human antibodies are particularly desirable for therapeutic treatment of human patients. For an overview of this technology for producing human antibodies, see Lonberg *et al.* (1995) *Int. Rev. Immunol.* 13:65-93. For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, e.g., U.S. Patent 5,625,126;
20 U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806.

25 The invention also encompasses kits for using antibodies to detect the presence of an aminopeptidase protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting aminopeptidase in a biological sample; means for determining the amount of aminopeptidase in the sample; and means for comparing the amount of aminopeptidase in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect aminopeptidase.

30

Polynucleotides

The nucleotide sequence in SEQ ID NO:2 was obtained by sequencing the deposited human cDNA. Accordingly, the sequence of the deposited clone is controlling

as to any discrepancies between the two and any reference to the sequence of SEQ ID NO:2 includes reference to the sequence of the deposited cDNA.

The specifically disclosed cDNA comprises the coding region and 5' and 3' untranslated sequences in SEQ ID NO:2.

5 The invention provides isolated polynucleotides encoding the novel aminopeptidase. The term "aminopeptidase polynucleotide" or "aminopeptidase nucleic acid" refers to the sequence shown in SEQ ID NO:2 or in the deposited cDNA. The term "aminopeptidase polynucleotide" or "aminopeptidase nucleic acid" further includes variants and fragments of the aminopeptidase polynucleotides.

10 An "isolated" aminopeptidase nucleic acid is one that is separated from other nucleic acid present in the natural source of the aminopeptidase nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the aminopeptidase nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there
15 can be some flanking nucleotide sequences, for example up to about 5KB. The important point is that the aminopeptidase nucleic acid is isolated from flanking sequences such that it can be subjected to the specific manipulations described herein, such as recombinant expression, preparation of probes and primers, and other uses specific to the aminopeptidase nucleic acid sequences.

20 Moreover, an "isolated" nucleic acid molecule, such as a cDNA or RNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

25 In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 % (on a
30 molar basis) of all macromolecular species present.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially)

DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

5 In some instances, the isolated material will form part of a composition (or example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90% (on a molar basis) of all
10 macromolecular species present.

 The aminopeptidase polynucleotides can encode the mature protein plus additional amino or carboxy terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature
15 form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

 The aminopeptidase polynucleotides include, but are not limited to, the sequence
20 encoding the mature polypeptide alone, the sequence encoding the mature polypeptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature polypeptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated
25 sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the polynucleotide may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

 Aminopeptidase polynucleotides can be in the form of RNA, such as mRNA, or
30 in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

Aminopeptidase nucleic acid can comprise the nucleotide sequences shown in SEQ ID NO:2.

In one embodiment, the aminopeptidase nucleic acid comprises only the coding region.

5 The invention further provides variant aminopeptidase polynucleotides, and fragments thereof, that differ from the nucleotide sequence shown in SEQ ID NO:2 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence shown in SEQ ID NO: 2.

10 The invention also provides aminopeptidase nucleic acid molecules encoding the variant polypeptides described herein. Such polynucleotides may be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms.

15 Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions.

Typically, variants have a substantial identity with a nucleic acid molecules of SEQ ID NO: 2 and the complements thereof. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

20

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. These variants comprise a nucleotide sequence encoding an aminopeptidase that is typically at least about 50-55%, 55-60%, 60-65%, 65-70%, 70-75%, more typically at least about 75-80% or 80-85%, and most typically at least about

25 85-90% or 90-95% or more homologous to the nucleotide sequence shown in SEQ ID NO: 2 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO: 2 or a fragment of the sequence. It is understood that stringent hybridization does not indicate substantial homology where it is due to general

30 homology, such as poly A sequences, or sequences common to all or most proteins, metalloproteases, zinc binding proteins, proteins in the M1 family, aminopeptidases, aminopeptidase B family, or leukotriene A₄ epoxide hydrolases. Moreover, it is

understood that variants do not include any of the nucleic acid sequences that may have been disclosed prior to the invention.

As used herein, the term “hybridizes under stringent conditions” is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a polypeptide at least 50-55% or 55-60% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95% or more identical to each other remain hybridized to one another. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6, incorporated by reference. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. In another non-limiting example, nucleic acid molecules are allowed to hybridize in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more low stringency washes in 0.2X SSC/0.1% SDS at room temperature, or by one or more moderate stringency washes in 0.2X SSC/0.1% SDS at 42°C, or washed in 0.2X SSC/0.1% SDS at 65°C for high stringency. In one embodiment, an isolated nucleic acid molecule that hybridizes under stringent conditions to the sequence of SEQ ID NO: 2 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a “naturally-occurring” nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

As understood by those of ordinary skill, the exact conditions can be determined empirically and depend on ionic strength, temperature and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS. Other factors considered in determining the desired hybridization conditions include the length of the nucleic acid sequences, base composition, percent mismatch between the hybridizing sequences and the frequency of occurrence of subsets of the sequences within other non-identical sequences. Thus, equivalent conditions can be determined by varying one or more of these parameters while maintaining a similar degree of identity or similarity between the two nucleic acid molecules.

The present invention also provides isolated nucleic acids that contain a single or double stranded fragment or portion that hybridizes under stringent conditions to

the nucleotide sequence of SEQ ID NO: 2 or the complement of SEQ ID NO: 2. In one embodiment, the nucleic acid consists of a portion of the nucleotide sequence of SEQ ID NO: 2 and the complement of SEQ ID NO: 2. The nucleic acid fragments of the invention are at least about 10 or 15, preferably at least about 18, 20, 23 or 25
5 nucleotides, and can be 30, 40, 50, 100, 200, 500 or more nucleotides in length. Longer fragments, for example, 30 or more nucleotides in length, which encode antigenic proteins or polypeptides described herein are useful.

Furthermore, the invention provides polynucleotides that comprise a fragment of the full-length aminopeptidase polynucleotides. The fragment can be single or
10 double-stranded and can comprise DNA or RNA. The fragment can be derived from either the coding or the non-coding sequence.

In another embodiment an isolated aminopeptidase nucleic acid encodes the entire coding region. In another embodiment the isolated aminopeptidase nucleic acid encodes a sequence corresponding to the mature protein that may be from about amino
15 acid 6 to the last amino acid. Other fragments include nucleotide sequences encoding the amino acid fragments described herein.

Thus, aminopeptidase nucleic acid fragments further include sequences corresponding to the regions described herein, subregions also described, and specific functional sites. Aminopeptidase nucleic acid fragments also include combinations of
20 the regions, segments, motifs, and other functional sites described above. A person of ordinary skill in the art would be aware of the many permutations that are possible.

Where the location of the regions or sites have been predicted by computer analysis, one of ordinary skill would appreciate that the amino acid residues constituting these regions can vary depending on the criteria used to define the regions.

25 However, it is understood that an aminopeptidase fragment includes any nucleic acid sequence that does not include the entire gene.

The invention also provides aminopeptidase nucleic acid fragments that encode epitope bearing regions of the aminopeptidase proteins described herein.

Nucleic acid fragments, according to the present invention, are not to be
30 construed as encompassing those fragments that may have been disclosed prior to the invention.

Polynucleotide Uses

The nucleotide sequences of the present invention can be used as a “query sequence” to perform a search against public databases, for example, to identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

The nucleic acid fragments of the invention provide probes or primers in assays such as those described below. “Probes” are oligonucleotides that hybridize in a base-specific manner to a complementary strand of nucleic acid. Such probes include polypeptide nucleic acids, as described in Nielsen *et al.* (1991) *Science* 254:1497-1500. Typically, a probe comprises a region of nucleotide sequence that hybridizes under highly stringent conditions to at least about 15, typically about 20-25, and more typically about 40, 50 or 75 consecutive nucleotides of the nucleic acid sequence shown in SEQ ID NO: 2 and the complements thereof. More typically, the probe further comprises a label, e.g., radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

As used herein, the term “primer” refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis using well-known methods (e.g., PCR, LCR) including, but not limited to those described herein. The appropriate length of the primer depends on the particular use, but typically ranges from about 15 to 30 nucleotides. The term “primer site” refers to the area of the target DNA to which a primer hybridizes. The term “primer pair” refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5' end of the nucleic acid sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the sequence to be amplified.

The aminopeptidase polynucleotides are thus useful for probes, primers, and in biological assays.

Where the polynucleotides are used to assess aminopeptidase properties or functions, such as in the assays described herein, all or less than all of the entire cDNA can be useful. Assays specifically directed to aminopeptidase functions, such as assessing agonist or antagonist activity, encompass the use of known fragments. Further, diagnostic methods for assessing aminopeptidase function can also be practiced with any fragment, including those fragments that may have been known prior to the invention. Similarly, in methods involving treatment of aminopeptidase dysfunction, all fragments are encompassed including those, which may have been known in the art.

The aminopeptidase polynucleotides are useful as a hybridization probe for cDNA and genomic DNA to isolate a full-length cDNA and genomic clones encoding the polypeptides described in SEQ ID NO: 1 and to isolate cDNA and genomic clones that correspond to variants producing the same polypeptides shown in SEQ ID NO: 1 or the other variants described herein. Variants can be isolated from the same tissue and organism from which the polypeptides shown in SEQ ID NO: 1 were isolated, different tissues from the same organism, or from different organisms. This method is useful for isolating genes and cDNA that are developmentally-controlled and therefore may be expressed in the same tissue or different tissues at different points in the development of an organism.

The probe can correspond to any sequence along the entire length of the gene encoding the aminopeptidase. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions.

The nucleic acid probe can be, for example, the full-length cDNA of SEQ ID NO: 2, or a fragment thereof, such as an oligonucleotide of at least 12, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to mRNA or DNA.

Fragments of the polynucleotides described herein are also useful to synthesize larger fragments or full-length polynucleotides described herein. For example, a fragment can be hybridized to any portion of an mRNA and a larger or full-length cDNA can be produced.

The fragments are also useful to synthesize antisense molecules of desired length and sequence.

Antisense nucleic acids of the invention can be designed using the nucleotide sequences of SEQ ID NO: 2, and constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest).

Additionally, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4:5). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of

PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996), *supra*; Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:14670.

- 5 PNAs can be further modified, e.g., to enhance their stability, specificity or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), *supra*, Finn *et al.* (1996) *Nucleic Acids Res.* 24(17):3357-10 63, Mag *et al.* (1989) *Nucleic Acids Res.* 17:5973, and Peterser *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5:1119.

- The nucleic acid molecules and fragments of the invention can also include other appended groups such as peptides (e.g., for targeting host cell aminopeptidases *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger 15 *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO 88/0918) or the blood brain barrier (see, e.g., PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents (see, e.g., 20 Zon (1988) *Pharm. Res.* 5:539-549).

The aminopeptidase polynucleotides are also useful as primers for PCR to amplify any given region of an aminopeptidase polynucleotide.

- The aminopeptidase polynucleotides are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the 25 aminopeptidase polypeptides. Vectors also include insertion vectors, used to integrate into another polynucleotide sequence, such as into the cellular genome, to alter *in situ* expression of aminopeptidase genes and gene products. For example, an endogenous aminopeptidase coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

- 30 The aminopeptidase polynucleotides are also useful for expressing antigenic portions of the aminopeptidase proteins.

The aminopeptidase polynucleotides are also useful as probes for determining the chromosomal positions of the aminopeptidase polynucleotides by means of *in situ*

hybridization methods, such as FISH. (For a review of this technique, see Verma *et al.* (1988) *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York), and PCR mapping of somatic cell hybrids. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes
5 associated with disease.

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding
10 sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in *Mendelian Inheritance in Man*, V.
15 McKusick, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland *et al.* ((1987) *Nature* 325:783-787).

Moreover, differences in the DNA sequences between individuals affected and
20 unaffected with a disease associated with a specified gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations, that are
25 visible from chromosome spreads, or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

The aminopeptidase polynucleotide probes are also useful to determine patterns
30 of the presence of the gene encoding the aminopeptidases and their variants with respect to tissue distribution, for example, whether gene duplication has occurred and whether the duplication occurs in all or only a subset of tissues. The genes can be naturally occurring or can have been introduced into a cell, tissue, or organism exogenously.

The aminopeptidase polynucleotides are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from genes encoding the polynucleotides described herein.

5 The aminopeptidase polynucleotides are also useful for constructing host cells expressing a part, or all, of the aminopeptidase polynucleotides and polypeptides.

The aminopeptidase polynucleotides are also useful for constructing transgenic animals expressing all, or a part, of the aminopeptidase polynucleotides and polypeptides.

10 The aminopeptidase polynucleotides are also useful for making vectors that express part, or all, of the aminopeptidase polypeptides.

The aminopeptidase polynucleotides are also useful as hybridization probes for determining the level of aminopeptidase nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, aminopeptidase nucleic acid in cells, tissues, and in organisms. The nucleic acid whose level is
15 determined can be DNA or RNA. Accordingly, probes corresponding to the polypeptides described herein can be used to assess gene copy number in a given cell, tissue, or organism. This is particularly relevant in cases in which there has been an amplification of the aminopeptidase genes.

Alternatively, the probe can be used in an *in situ* hybridization context to assess
20 the position of extra copies of the aminopeptidase genes, as on extrachromosomal elements or as integrated into chromosomes in which the aminopeptidase gene is not normally found, for example as a homogeneously staining region.

These uses are relevant for diagnosis of disorders involving an increase or decrease in aminopeptidase expression relative to normal, such as a proliferative
25 disorder, a differentiative or developmental disorder, or a hematopoietic disorder.

Disorders in which the aminopeptidase expression is relevant include, but are not limited to, breast, lung and colon carcinomas and disorders involving abnormal or undesirable bone mass.

The aminopeptidase is expressed in the tissues shown in Figures 4-8. As such,
30 the gene is particularly relevant for the treatment of disorders involving these tissues, especially prostate, brain, testis, osteoblasts, fetal kidney, and fetal liver. Expression is also highly relevant in osteoclasts.

Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant expression or activity of aminopeptidase nucleic acid, in which a test sample is obtained from a subject and nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of the nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the nucleic acid.

One aspect of the invention relates to diagnostic assays for determining nucleic acid expression as well as activity in the context of a biological sample (e.g., blood, serum, cells, tissue) to determine whether an individual has a disease or disorder, or is at risk of developing a disease or disorder, associated with aberrant nucleic acid expression or activity. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with expression or activity of the nucleic acid molecules.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express the aminopeptidase, such as by measuring the level of an aminopeptidase-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if the aminopeptidase gene has been mutated.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate aminopeptidase nucleic acid expression (e.g., antisense, polypeptides, peptidomimetics, small molecules or other drugs). A cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of the mRNA in the presence of the candidate compound is compared to the level of expression of the mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. The modulator can bind to the nucleic acid or indirectly modulate expression, such as by interacting with other cellular components that affect nucleic acid expression

Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject) in patients or in transgenic animals.

5 The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the aminopeptidase gene. The method typically includes assaying the ability of the compound to modulate the expression of the aminopeptidase nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired aminopeptidase nucleic acid expression.

10 The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the aminopeptidase nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

Alternatively, candidate compounds can be assayed *in vivo* in patients or in transgenic animals.

15 The assay for aminopeptidase nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds (such as peptide hydrolysis). Further, the expression of genes that are up- or down-regulated in response to the aminopeptidase activity can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

20 Thus, modulators of aminopeptidase gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of aminopeptidase mRNA in the presence of the candidate compound is compared to the level of expression of aminopeptidase mRNA in the absence of the candidate compound. The candidate compound can then be
25 identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically
30 significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

Accordingly, the invention provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to

modulate aminopeptidase nucleic acid expression. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or effects on nucleic acid activity (e.g. when nucleic acid is mutated or improperly modified). Treatment is of disorders characterized by aberrant expression or
5 activity of the nucleic acid.

Disorders in which the aminopeptidase expression is relevant include, but are not limited to, those discussed herein and particularly in breast carcinoma and disorders involving undesirable or aberrant bone mass.

Alternatively, a modulator for aminopeptidase nucleic acid expression can be a
10 small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the aminopeptidase nucleic acid expression.

The aminopeptidase polynucleotides are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the aminopeptidase gene in clinical trials or in a treatment regimen. Thus, the gene
15 expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the
20 administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

Monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of
25 expression of a specified mRNA or genomic DNA of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the mRNA or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the mRNA or genomic DNA in the pre-administration sample with the
30 mRNA or genomic DNA in the post-administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

The aminopeptidase polynucleotides are also useful in diagnostic assays for qualitative changes in aminopeptidase nucleic acid, and particularly in qualitative

changes that lead to pathology. The polynucleotides can be used to detect mutations in aminopeptidase genes and gene expression products such as mRNA. The polynucleotides can be used as hybridization probes to detect naturally-occurring genetic mutations in the aminopeptidase gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the aminopeptidase gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of an aminopeptidase.

Mutations in the aminopeptidase gene can be detected at the nucleic acid level by a variety of techniques. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way.

In certain embodiments, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *PNAS* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.* (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.* (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well-known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

Alternatively, mutations in an aminopeptidase gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method.

Furthermore, sequence differences between a mutant aminopeptidase gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.* (1985) *Science* 230:1242); Cotton *et al.* (1988) *PNAS* 85:4397; Saleeba *et al.* (1992) *Meth. Enzymol.* 217:286-295), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.* (1989) *PNAS* 86:2766; Cotton *et al.* (1993) *Mutat. Res.* 285:125-144; and Hayashi *et al.* (1992) *Genet. Anal. Tech. Appl.* 9:73-79), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.* (1985) *Nature* 313:495). The sensitivity of

the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991)

- 5 *Trends Genet.* 7:5). Examples of other techniques for detecting point mutations include, selective oligonucleotide hybridization, selective amplification, and selective primer extension.

- In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays
10 containing hundreds or thousands of oligonucleotide probes (Cronin *et al.* (1996) *Human Mutation* 7:244-255; Kozal *et al.* (1996) *Nature Medicine* 2:753-759). For example, genetic mutations can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a
15 sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is
20 composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

- The aminopeptidase polynucleotides are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the polynucleotides can be used to study the relationship
25 between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). In the present case, for example, a mutation in the aminopeptidase gene that results in altered affinity for a zinc compound or peptide substrate could result in an excessive or decreased drug effect with standard concentrations of the zinc compound or peptide substrate. Accordingly, the
30 aminopeptidase polynucleotides described herein can be used to assess the mutation content of the gene in an individual in order to select an appropriate compound or dosage regimen for treatment.

Thus polynucleotides displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

5 The methods can involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting mRNA, or genomic DNA, such that the presence of mRNA or genomic DNA is detected in the biological sample, and comparing the presence of mRNA or genomic DNA in the control sample with the presence of mRNA or genomic DNA in the test
10 sample.

 The aminopeptidase polynucleotides are also useful for chromosome identification when the sequence is identified with an individual chromosome and to a particular location on the chromosome. First, the DNA sequence is matched to the chromosome by *in situ* or other chromosome-specific hybridization. Sequences can
15 also be correlated to specific chromosomes by preparing PCR primers that can be used for PCR screening of somatic cell hybrids containing individual chromosomes from the desired species. Only hybrids containing the chromosome containing the gene homologous to the primer will yield an amplified fragment. Sublocalization can be achieved using chromosomal fragments. Other strategies include prescreening with
20 labeled flow-sorted chromosomes and preselection by hybridization to chromosome-specific libraries. Further mapping strategies include fluorescence *in situ* hybridization, which allows hybridization with probes shorter than those traditionally used. Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on the chromosome, or panels of reagents can be used for marking multiple
25 sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

 The aminopeptidase polynucleotides can also be used to identify individuals
30 from small biological samples. This can be done for example using restriction fragment-length polymorphism (RFLP) to identify an individual. Thus, the polynucleotides described herein are useful as DNA markers for RFLP (See U.S. Patent No. 5,272,057).

Furthermore, the aminopeptidase sequence can be used to provide an alternative technique, which determines the actual DNA sequence of selected fragments in the genome of an individual. Thus, the aminopeptidase sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These
5 primers can then be used to amplify DNA from an individual for subsequent sequencing.

Panels of corresponding DNA sequences from individuals prepared in this manner can provide unique individual identifications, as each individual will have a unique set of such DNA sequences. It is estimated that allelic variation in humans occurs with a frequency of about once per each 500 bases. Allelic variation occurs to
10 some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. The aminopeptidase sequences can be used to obtain such identification sequences from individuals and from tissue. The sequences represent unique fragments of the human genome. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be
15 compared for identification purposes.

If a panel of reagents from the sequences is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue
20 samples.

The aminopeptidase polynucleotides can also be used in forensic identification procedures. PCR technology can be used to amplify DNA sequences taken from very small biological samples, such as a single hair follicle, body fluids (e.g. blood, saliva, or semen). The amplified sequence can then be compared to a standard allowing
25 identification of the origin of the sample.

The aminopeptidase polynucleotides can thus be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular
30 individual). As described above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to the noncoding region are particularly useful

since greater polymorphism occurs in the noncoding regions, making it easier to differentiate individuals using this technique.

The aminopeptidase polynucleotides can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This is useful in cases in which a forensic pathologist is presented with a tissue of unknown origin. Panels of aminopeptidase probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these primers and probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Alternatively, the aminopeptidase polynucleotides can be used directly to block transcription or translation of aminopeptidase gene sequences by means of antisense or ribozyme constructs. Thus, in a disorder characterized by abnormally high or undesirable aminopeptidase gene expression, nucleic acids can be directly used for treatment.

The aminopeptidase polynucleotides are thus useful as antisense constructs to control aminopeptidase gene expression in cells, tissues, and organisms. A DNA antisense polynucleotide is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of aminopeptidase protein. An antisense RNA or DNA polynucleotide would hybridize to the mRNA and thus block translation of mRNA into aminopeptidase protein.

Examples of antisense molecules useful to inhibit nucleic acid expression include antisense molecules complementary to a fragment of the 5' untranslated region of SEQ ID NO: 2 which also includes the start codon and antisense molecules which are complementary to a fragment of the 3' untranslated region of SEQ ID NO: 2.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of aminopeptidase nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired aminopeptidase nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding

regions and particularly coding regions corresponding to the catalytic and other functional activities of the aminopeptidase protein.

The aminopeptidase polynucleotides also provide vectors for gene therapy in patients containing cells that are aberrant in aminopeptidase gene expression. Thus, 5 recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired aminopeptidase protein to treat the individual.

The invention also encompasses kits for detecting the presence of an aminopeptidase nucleic acid in a biological sample. For example, the kit can comprise 10 reagents such as a labeled or labelable nucleic acid or agent capable of detecting aminopeptidase nucleic acid in a biological sample; means for determining the amount of aminopeptidase nucleic acid in the sample; and means for comparing the amount of aminopeptidase nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using 15 the kit to detect aminopeptidase mRNA or DNA.

Computer Readable Means

The nucleotide or amino acid sequences of the invention are also provided in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a 20 manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a nucleotide or amino acid sequence of the present invention. Such a manufacture provides the nucleotide or amino acid sequences, or a subset thereof (e.g., a subset of open reading frames (ORFs)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the 25 nucleotide or amino acid sequences, or a subset thereof, as they exists in nature or in purified form.

In one application of this embodiment, a nucleotide or amino acid sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed 30 directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. The skilled

artisan will readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on
5 computer readable medium. The skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide or amino acid sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for
10 creating a computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The
15 sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of data processor structuring formats (e.g., text file or database) in order to obtain
20 computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the
25 nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

30 As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length

of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

- 5 As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences.
- 10 Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed

15 publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA).

For example, software which implements the BLAST (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410) and BLAZE (Brutlag *et al.* (1993) *Comp. Chem.* 17:203-207)

20 search algorithms on a Sybase system can be used to identify open reading frames (ORFs) of the sequences of the invention which contain homology to ORFs or proteins from other libraries. Such ORFs are protein encoding fragments and are useful in producing commercially important proteins such as enzymes used in various

25 reactions and in the production of commercially useful metabolites.

Vectors/Host Cells

The invention also provides vectors containing the aminopeptidase polynucleotides. The term "vector" refers to a vehicle, preferably a nucleic acid

30 molecule that can transport the aminopeptidase polynucleotides. When the vector is a nucleic acid molecule, the aminopeptidase polynucleotides are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid,

single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the aminopeptidase polynucleotides. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the aminopeptidase polynucleotides when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the aminopeptidase polynucleotides. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the aminopeptidase polynucleotides such that transcription of the polynucleotides is allowed in a host cell. The polynucleotides can be introduced into the host cell with a separate polynucleotide capable of affecting transcription. Thus, the second polynucleotide may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the aminopeptidase polynucleotides from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself.

It is understood, however, that in some embodiments, transcription and/or translation of the aminopeptidase polynucleotides can occur in a cell-free system.

The regulatory sequence to which the polynucleotides described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control

elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

A variety of expression vectors can be used to express an aminopeptidase polynucleotide. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The aminopeptidase polynucleotides can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate polynucleotide can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the polypeptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the aminopeptidase polypeptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired polypeptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.* (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.* (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.* (1990) *Gene Expression Technology: Methods in Enzymology* 185:60-89).

Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S. (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California 119-128). Alternatively, the sequence of the polynucleotide of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.* (1992) *Nucleic Acids Res.* 20:2111-2118).

The aminopeptidase polynucleotides can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari *et al.* (1987) *EMBO J.* 6:229-234), pMFa (Kurjan *et al.* (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.* (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA).

The aminopeptidase polynucleotides can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow *et al.* (1989) *Virology* 170:31-39).

In certain embodiments of the invention, the polynucleotides described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195).

5 The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the aminopeptidase polynucleotides. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the polynucleotides described herein. These are found for example in Sambrook *et al.*
10 (1989) *Molecular Cloning: A Laboratory Manual 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

 The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense
15 transcript can be produced to all, or to a portion, of the polynucleotide sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

20 The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

 The recombinant host cells are prepared by introducing the vector constructs
25 described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, 2d ed., Cold
30 Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

 Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the

aminopeptidase polynucleotides can be introduced either alone or with other polynucleotides that are not related to the aminopeptidase polynucleotides such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or
5 joined to the aminopeptidase polynucleotide vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that
10 complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the polynucleotides described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes
15 for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free
20 transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the polypeptide is desired, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the aminopeptidase polypeptides or heterologous to these polypeptides.

25 Where the polypeptide is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The polypeptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange
30 chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the polypeptides described herein, the polypeptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the polypeptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of Vectors and Host Cells

It is understood that “host cells” and “recombinant host cells” refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

The host cells expressing the polypeptides described herein, and particularly recombinant host cells, have a variety of uses. First, the cells are useful for producing aminopeptidase proteins or polypeptides that can be further purified to produce desired amounts of aminopeptidase protein or fragments. Thus, host cells containing expression vectors are useful for polypeptide production.

Host cells are also useful for conducting cell-based assays involving the aminopeptidase or aminopeptidase fragments. Thus, a recombinant host cell expressing a native aminopeptidase is useful to assay for compounds that stimulate or inhibit aminopeptidase function. This includes zinc or peptide binding, gene expression at the level of transcription or translation, and interaction with other cellular components.

Host cells are also useful for identifying aminopeptidase mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant aminopeptidase (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native aminopeptidase.

Recombinant host cells are also useful for expressing the chimeric polypeptides described herein to assess compounds that activate or suppress activation by means of a heterologous domain, segment, site, and the like, as disclosed herein.

Further, mutant aminopeptidases can be designed in which one or more of the various functions is engineered to be increased or decreased and used to augment or replace aminopeptidase proteins in an individual. Thus, host cells can provide a

therapeutic benefit by replacing an aberrant aminopeptidase or providing an aberrant aminopeptidase that provides a therapeutic result. In one embodiment, the cells provide aminopeptidases that are abnormally active.

In another embodiment, the cells provide aminopeptidases that are abnormally
5 inactive. These aminopeptidases can compete with endogenous aminopeptidases in the individual.

In another embodiment, cells expressing aminopeptidases that cannot be activated, are introduced into an individual in order to compete with endogenous aminopeptidases for zinc or peptide. For example, in the case in which excessive zinc
10 compound is part of a treatment modality, it may be necessary to effectively inactivate the compound at a specific point in treatment. Providing cells that compete for the molecule, but which cannot be affected by aminopeptidase activation would be beneficial.

Homologously recombinant host cells can also be produced that allow the *in situ*
15 alteration of endogenous aminopeptidase polynucleotide sequences in a host cell genome. The host cell includes, but is not limited to, a stable cell line, cell *in vivo*, or cloned microorganism. This technology is more fully described in WO 93/09222, WO 91/12650, WO 91/06667, U.S. 5,272,071, and U.S. 5,641,670. Briefly, specific polynucleotide sequences corresponding to the aminopeptidase polynucleotides or
20 sequences proximal or distal to an aminopeptidase gene are allowed to integrate into a host cell genome by homologous recombination where expression of the gene can be affected. In one embodiment, regulatory sequences are introduced that either increase or decrease expression of an endogenous sequence. Accordingly, an aminopeptidase protein can be produced in a cell not normally producing it. Alternatively, increased
25 expression of aminopeptidase protein can be effected in a cell normally producing the protein at a specific level. Further, expression can be decreased or eliminated by introducing a specific regulatory sequence. The regulatory sequence can be heterologous to the aminopeptidase protein sequence or can be a homologous sequence with a desired mutation that affects expression. Alternatively, the entire gene can be
30 deleted. The regulatory sequence can be specific to the host cell or capable of functioning in more than one cell type. Still further, specific mutations can be introduced into any desired region of the gene to produce mutant aminopeptidase proteins. Such mutations could be introduced, for example, into the specific functional

regions such as the ligand-binding site.

In one embodiment, the host cell can be a fertilized oocyte or embryonic stem cell that can be used to produce a transgenic animal containing the altered aminopeptidase gene. Alternatively, the host cell can be a stem cell or other early tissue precursor that gives rise to a specific subset of cells and can be used to produce transgenic tissues in an animal. See also Thomas *et al.*, *Cell* 51:503 (1987) for a description of homologous recombination vectors. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous aminopeptidase gene is selected (see e.g., Li, E. *et al.* (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinions in Biotechnology* 2:823-829 and in PCT International Publication Nos. WO 90/11354; WO 91/01140; and WO 93/04169.

The genetically engineered host cells can be used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of an aminopeptidase protein and identifying and evaluating modulators of aminopeptidase protein activity.

Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

In one embodiment, a host cell is a fertilized oocyte or an embryonic stem cell into which aminopeptidase polynucleotide sequences have been introduced.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the aminopeptidase nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the aminopeptidase protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems, which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al.* (1992) *PNAS* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g.,

by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.* (1997) *Nature* 385:810-813 and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to a pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the polypeptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could affect binding or activation, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* aminopeptidase function, including peptide interaction, the effect of specific mutant aminopeptidases on aminopeptidase function and peptide interaction, and the effect of chimeric aminopeptidases. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more aminopeptidase functions.

In general, methods for producing transgenic animals include introducing a nucleic acid sequence according to the present invention, the nucleic acid sequence capable of expressing the aminopeptidase protein in a transgenic animal, into a cell in culture or *in vivo*. When introduced *in vivo*, the nucleic acid is introduced into an intact organism such that one or more cell types and, accordingly, one or more tissue types, express the nucleic acid encoding the aminopeptidase protein. Alternatively, the nucleic acid can be introduced into virtually all cells in an organism by transfecting a cell in culture, such as an embryonic stem cell, as described herein for the production of transgenic animals, and this cell can be used to produce an entire transgenic organism. As described, in a further embodiment, the host cell can be a fertilized oocyte. Such cells are then allowed to develop in a female foster animal to

produce the transgenic organism.

Pharmaceutical Compositions

5 The aminopeptidase nucleic acid molecules, protein, modulators of the protein, and antibodies (also referred to herein as "active compounds") can be incorporated into pharmaceutical compositions suitable for administration to a subject, e.g., a human. Such compositions typically comprise the nucleic acid molecule, protein, modulator, or antibody and a pharmaceutically acceptable carrier.

10 The term "administer" is used in its broadest sense and includes any method of introducing the compositions of the present invention into a subject. This includes producing polypeptides or polynucleotides *in vivo* by *in vivo* transcription or translation of polynucleotides that have been exogenously introduced into a subject. Thus, polypeptides or nucleic acids produced in the subject from the exogenous compositions are encompassed in the term "administer."

15 As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, such media can be used in the
20 compositions of the invention. Supplementary active compounds can also be incorporated into the compositions. A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration.
25 Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as
30 ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The

parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the

5 extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage

10 and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance

15 of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the

20 composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., an aminopeptidase protein or anti- aminopeptidase antibody) in the

25 required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions,

30 the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For oral administration, the agent can be contained in enteric forms to survive the stomach or further coated or mixed to be released in a particular region of the GI tract by known methods. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser, which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems.

Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova
5 Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in
10 dosage unit form for ease of administration and uniformity of dosage. "Dosage unit form" as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are
15 dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for
20 example, intravenous injection, local administration (U.S. 5,328,470) or by stereotactic injection (see e.g., Chen *et al.* (1994) *PNAS* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be
25 produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

As defined herein, a therapeutically effective amount of protein or polypeptide
30 (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight.

The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments. In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or

sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL

Generation of a 2786-Specific Antibody

In order to generate an antibody to the 2786 polypeptide, an expression vector was constructed that fused the coding sequence for amino acids 450-650 of 2786 (SEQ ID NO:1) in frame with the coding sequence for GST (glutathione S-transferase). This portion of the 2786 polypeptide was chosen because it does not comprise the peptidase family M1 domain and thus the antibody it elicits is less likely to cross-react with other peptidases. The GST fusion protein was produced in *E. coli* BL₂₁, and soluble proteins were purified using glutathione sepharose. The production of the 2786-GST fusion protein was verified by anti-GST western blotting. The purified fusion protein was dialyzed, and 3 mg were used for polyclonal antibody production.

Two lots of antisera produced as described above were tested for their ability to bind the 2786 polypeptide by western blotting. The target was amino terminal histidine-tagged 2786 polypeptide that had been produced by Baculovirus expression vector infection of insect cells. Both lots of antisera (but not the pre-immune sera) showed high affinity for the 2786 polypeptide.

2786 Substrate Identification

The amino acid substrate hydrolytic specificity of the 2786 aminopeptidase was determined using crude cell lysates from insect cell lines infected with three independent bacmid expression vectors (clones 1, 2, and 4) encoding amino terminal histidine-tagged 2786. The substrates tested in the assay were 7-amino-4-methylcoumarin (AMC) labeled amino acids. The assay was performed essentially as described in Yaron *et al.* (1979), *Anal. Biochem.* 95: 228-235. Briefly, 2.5 µl of crude cell extract was incubated with 2.5 µl (25 µM) of AMC labeled amino acid substrate and 45 µl reaction buffer (50 mM Tris-HCl, pH 7.5 and 5mM ZnCl). The real time progressive kinetic curves for each substrate were monitored using a BMG-Galaxy fluorescent plate reader. Significant levels of 2786 hydrolytic activity were observed for AMC-labeled tyrosine and proline substrates, and high levels of hydrolytic activity were observed for AMC-labeled arginine and lysine substrates. See figure 9.

20

This invention may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will fully convey the invention to those skilled in the art. Many modifications and other embodiments of the invention will come to mind in one skilled in the art to which this invention pertains having the benefit of the teachings presented in the foregoing description. Although specific terms are employed, they are used as in the art unless otherwise indicated.

25

Applicant's or agent's file reference	35800/206247	International application No.	PCT/US00/31873
--	--------------	-------------------------------	----------------

**INDICATIONS RELATING TO DEPOSITED MICROORGANISM
OR OTHER BIOLOGICAL MATERIAL**

(PCT Rule 13bis)

A. The indications made below relate to the deposited microorganism or other biological material referred to in the description on page 6, line 6, 9, 12 and 14	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 10801 University Blvd. Manassas, VA 20110-2209 US	
Date of deposit 15 December 2000 (15.12.00)	Accession Number PTA-2811
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet ☐	
Page 9, lines 14-16 and 22; Page 38, line 33; Page 39, lines 2 and 7; Page 80, lines 7, 11, 17, 19, 23 and 28; Page 81, lines 5, 10, 20, 22 and 32	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indicators are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

For receiving Office use only
☐ This sheet was received with the international application
Authorized officer

For International Bureau use only
☐ This sheet was received with the International Bureau on:
Authorized officer

THAT WHICH IS CLAIMED:

1. An isolated polypeptide having an amino acid sequence selected from the group consisting of:
 - 5 (a) The amino acid sequence shown in SEQ ID NO: 1;
 - (b) The amino acid sequence encoded by the cDNA contained in ATCC Deposit No. _____;
 - (c) The amino acid sequence of an allelic variant of the amino acid sequence shown in SEQ ID NO: 1;
 - 10 (d) The amino acid sequence of an allelic variant of the amino acid sequence encoded by the cDNA contained in ATCC Deposit No. _____;
 - (e) The amino acid sequence of a sequence variant of the amino acid sequence shown in SEQ ID NO: 1, wherein the sequence variant is encoded by a nucleic acid molecule hybridizing to the nucleic acid molecule shown in SEQ ID NO: 2 under
15 stringent conditions;
 - (f) The amino acid sequence of a sequence variant of the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. ____, wherein the sequence variant is encoded by a nucleic acid molecule hybridizing under stringent conditions to the cDNA contained in ATCC Deposit No. _____;
 - 20 (g) A fragment of the amino acid sequence shown in SEQ ID NO: 1, wherein the fragment comprises at least 12 contiguous amino acids;
 - (h) A fragment of the amino acid sequence encoded by the cDNA contained in ATCC Deposit No. _____, wherein the fragment comprises at least 12 contiguous amino acids;
 - 25 (i) The amino acid sequence of the mature polypeptide from about amino acid 6 to the last amino acid shown in SEQ ID NO: 1;
 - (j) The amino acid sequence of the mature polypeptide from about amino acid 6 to the last amino acid encoded by the cDNA clone contained in ATCC Deposit No. _____;
 - 30 (k) The amino acid sequence of an epitope bearing region of any one of the polypeptides of (a)-(j).

2. An isolated antibody that binds to a polypeptide of claim 1.

3. An isolated nucleic acid molecule having a nucleotide sequence selected from the group consisting of:
- (a) The nucleotide sequence shown in SEQ ID NO: 2;
 - 5 (b) The nucleotide sequence in the cDNA contained in ATCC Deposit No. ____;
 - (c) A nucleotide sequence encoding the amino acid sequence shown in SEQ ID NO: 1;
 - (d) A nucleotide sequence encoding the amino acid sequence encoded by the
10 cDNA contained in ATCC Deposit No. ____; and
 - (e) A nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c), or (d).
4. An isolated nucleic acid molecule having a nucleotide sequence selected from the group consisting of:
- (a) A nucleotide sequence encoding an amino acid sequence of a sequence variant of the amino acid sequence shown in SEQ ID NO: 1 that hybridizes to the nucleotide sequence shown in SEQ ID NO: 2 under stringent conditions;
 - (b) A nucleotide sequence encoding the amino acid sequence of a sequence
20 variant of the amino acid sequence encoded by the cDNA contained in ATCC Deposit No. ____, the nucleic acid sequence of the sequence variant hybridizing to the cDNA contained in ATCC Deposit No. ____ under stringent conditions; and
 - (c) A nucleotide sequence complementary to either of the nucleotide sequences in (a) or (b).
- 25
5. An isolated nucleic acid molecule having a polynucleotide having a nucleotide sequence selected from the group consisting of:
- (a) A nucleotide sequence encoding a fragment of the amino acid sequence shown in SEQ ID NO: 1, wherein the fragment comprises at least 12 contiguous amino
30 acids;
 - (b) A nucleotide sequence encoding a fragment of the amino acid sequence encoded by the cDNA contained in ATCC Deposit No. ____, wherein the fragment comprises at least 12 contiguous amino acids;

(c) A nucleotide sequence complementary to either of the nucleotide sequences in (a) or (b).

6. A nucleic acid vector having the nucleic acid sequences of any one of
5 claims 3-5.

7. A host cell containing the vector of claim 6.

8. A method for producing any of the polypeptides in claim 1 comprising
10 introducing a nucleotide sequence encoding any of the polypeptide sequences in (a)-(k) into a host cell, and culturing the host cell under conditions in which the proteins are expressed from the nucleic acid.

9. A method for detecting the presence of any of the polypeptides of
15 claim 1 in a sample, said method comprising contacting said sample with an agent that specifically allows detection of the presence of the polypeptide in the sample and then detecting the presence of the polypeptide.

10. The method of claim 9, wherein said agent is capable of selective
20 physical association with said polypeptide.

11. The method of claim 10, wherein said agent binds to said polypeptide.

12. The method of claim 11, wherein said agent is an antibody.
25

13. The method of claim 11, wherein said agent is a ligand.

14. A kit having reagents for use in the method of claim 9, wherein the reagents comprise an agent that specifically binds to said polypeptide.
30

15. A method for detecting the presence of any of the nucleic acid molecules of any one of claims 3-5 in a sample, the method comprising the steps of contacting said sample with an agent that specifically allows detection of the presence

of the nucleic acid molecule in the sample and then detecting the presence of the nucleic acid molecule.

16. The method of claim 15, wherein said agent is an polynucleotide that
5 hybridizes to the nucleic acid molecules of any one of claims 3-5 under stringent conditions.

17. The method of claim 15, wherein the nucleic acid molecule is mRNA.

10 18. A kit having reagents used for the method of claim 15, wherein the reagents comprise a compound that hybridizes to the nucleic acid molecules of any one of claims 3-5 under stringent conditions.

15 19. A method for identifying an agent that interacts with any of the polypeptides of claim 1 in a cell, said method comprising the steps of
(a) contacting said agent with a cell capable of allowing an interaction between said polypeptide and said agent such that said polypeptide can interact with said agent; and

(b) measuring the interaction between said polypeptide and said agent.
20

20. A method of screening a cell to identify an agent that interacts with any of the polypeptides of claim 1 in a cell, said method comprising contacting said agent with a cell capable of allowing an interaction between said polypeptide and said agent such that said polypeptide can interact with said agent, and measuring the
25 interaction.

21. A method for identifying an agent that binds to any of the polypeptides in claim 1, said method comprising contacting the polypeptide with an agent that binds to the polypeptide and assaying the complex formed with the agent bound to the
30 polypeptide.

22. The method of claim 21, wherein a fragment of the polypeptide is contacted.

23. A method of screening a cell to identify an agent that modulates the level or activity of any of the polypeptides of claim 1 in a cell, said method comprising contacting said agent with a cell capable of expressing said polypeptide
5 such that said polypeptide level or activity can be modulated in said cell by said agent and measuring said polypeptide level or activity.

24. The method of claim 23, wherein said cell is an osteoclast.

10 25. The method of claim 23, wherein said cell is a breast carcinoma cell.

26. The method of claim 23, wherein said agent increases the level or activity of said polypeptide.

15 27. The method of claim 23, wherein said agent decreases the level or activity of said polypeptide.

28. The method of claim 19, wherein said method comprises the additional steps of

20 (a) adding competing polypeptide that can interact with said agent; and
(b) comparing the amount of interaction between said agent and said polypeptide to the amount of interaction in the absence of said competing polypeptide.

25 29. The method of claim 19, wherein said interaction is binding.

30. A method of screening a cell to identify an agent that modulates the level or activity of any of the polypeptides of claim 1 in a cell by increasing the interaction between said polypeptide and a target molecule for said polypeptide, said
30 method comprising combining said polypeptide with said agent under conditions that allow said polypeptide to interact with said target molecule; and detecting the formation of a complex between said polypeptide and said target molecule or activity of said polypeptide as a result of interaction of said polypeptide with said target

molecule.

31. A method of screening a cell to identify an agent that modulates the level or activity of any of the polypeptides of claim 1 in a cell by decreasing the interaction between said polypeptide and a target molecule for said polypeptide, said method comprising combining said polypeptide with said agent under conditions that allow said polypeptide to interact with said target molecule; and detecting the formation of a complex between said polypeptide and said target molecule or activity of said polypeptide as a result of interaction of said polypeptide with said target molecule.

32. The method of claim 23, wherein said cell is *in vivo*.

33. The method of claim 32, wherein said cell is in a transgenic animal.

34. The method of claim 33, wherein said cell is in a non-transgenic subject.

35. The method of claim 23, wherein said cell is *in vitro*.

36. The method of claim 35, wherein said cell has been disrupted.

37. The method of claim 35, wherein said cell is in a biopsy.

38. The method of claim 35, wherein said cell is in cell culture.

39. The method of claim 38, wherein said cell is naturally-occurring or recombinant.

40. The method of claim 23, wherein said agent is selected from the group consisting of a peptide; phosphopeptide; antibody; organic molecule; and inorganic molecule.

41. A method for modulating the level or activity of any of the polypeptides of claim 1, said method comprising contacting said polypeptide with an agent under conditions that allow the agent to modulate the level or activity of the polypeptide.

5

42. A method for identifying an agent that modulates the level or activity of any of the polypeptides of claim 1 in a cell, said method comprising contacting said agent with a cell capable of expressing said polypeptide such that said polypeptide level or activity can be modulated in said cell by said agent and measuring said polypeptide level or activity.

10

43. A method for identifying an agent that modulates the level or activity of any of the nucleic acid molecules of any one of claims 3-5 in a cell, said method comprising contacting said agent with the cell capable of expressing said nucleic acid molecule such that said nucleic acid molecule level or activity can be modulated in said cell by said agent and measuring said nucleic acid molecule level or activity.

15

44. A method of screening a cell to identify an agent that modulates the level or activity of any of the nucleic acid molecules in any one of claims 3-5 in said cell, said method comprising contacting said agent with the cell capable of expressing said nucleic acid molecule such that said nucleic acid molecule level or activity can be modulated in said cell by said agent and measuring nucleic acid molecule level or activity.

20

45. A method for identifying an agent that interacts with any of the nucleic acid molecules of any one of claims 3-5 in a cell, said method comprising contacting said agent with a cell capable of allowing an interaction between said nucleic acid molecule and said agent such that said nucleic acid molecule can interact with said agent in measuring the interaction.

25

30

46. A method of screening a cell to identify an agent that interacts with any of the nucleic acid molecules of any one of claims 3-5 in a cell, said method comprising contacting said agent with a cell capable of allowing an interaction

between said nucleic acid molecule and said agent such that said nucleic acid molecule can interact with said agent and measuring the interaction.

47. A method for modulating the level or activity of any one of the nucleic acid molecules any one of claims 3-5, said method comprising contacting said nucleic acid molecule with an agent under conditions that allow the agent to modulate the level or activity of the nucleic acid molecule.

48. The method of claim 47, wherein said modulation is in cells derived from tissue selected from the group consisting of brain, breast, fetal kidney, fetal liver, or osteoclasts.

49. The method of claim 47, wherein said modulation is *in vivo*.

50. The method of claim 49, wherein said modulation is in a patient having a disorder involving brain, breast, fetal kidney, fetal liver, or bone mass.

51. The method of claim 49, wherein said modulation is in a patient having abnormal or undesirable bone mass development.

52. The method of claim 49, wherein said modulation is in a patient having breast cancer.

53. A method of treating a disorder involving brain, brain, fetal kidney, fetal liver, or bone mass in a subject in need of such treatment, said method comprising administering any of the polypeptides of claim 1 to said subject in a therapeutically effective amount.

54. A method of treating osteoporosis or osteopetrosis in a subject in need of such treatment, said method comprising administering any of the polypeptides of claim 1 to said subject in therapeutically effective amounts.

55. A method of treating breast cancer in a subject in need of such

treatment, said method comprising administering any of the polypeptides of claim 1 to said subject in therapeutically effective amounts.

56. A pharmaceutical composition containing at least one of the
5 polypeptides in claim 1 in a pharmaceutically acceptable carrier.

57. A pharmaceutically acceptable composition containing at least one
nucleic acid molecules of any one of claims 3-5 in a pharmaceutically acceptable
carrier.
10

58. A nonhuman transgenic animal, wherein one or more cells of said
animal contains at least one nucleic acid molecule of any one of claims 3-5.

59. A nonhuman transgenic animal, wherein one or more cells of said
15 animal contains at least one nucleic acid molecule of any one of claims 3-5, wherein
said cell expresses at least one of the polypeptides of claim 1.

60. A method for producing a transgenic animal according to claim 59,
said method comprising introducing at least one of the nucleic acid sequences of any
20 one of claims 3-5 into a cell, wherein said cell is present in said animal or gives rise to
said animal.

61. An agent identified by any one of the methods of claims 19-40.

25 62. An agent identified by any one of the methods of claims 42-46.

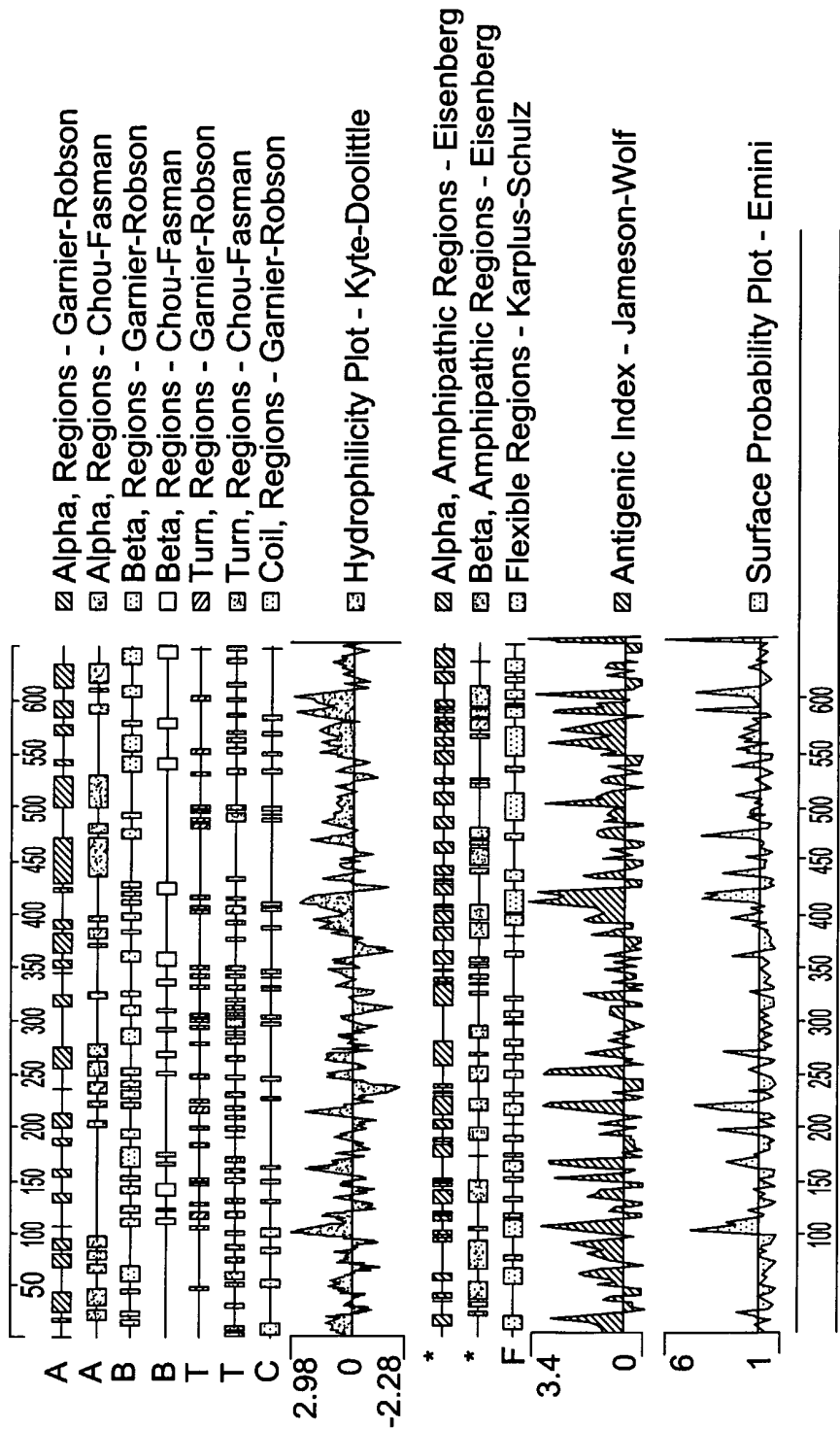


FIG. 1.

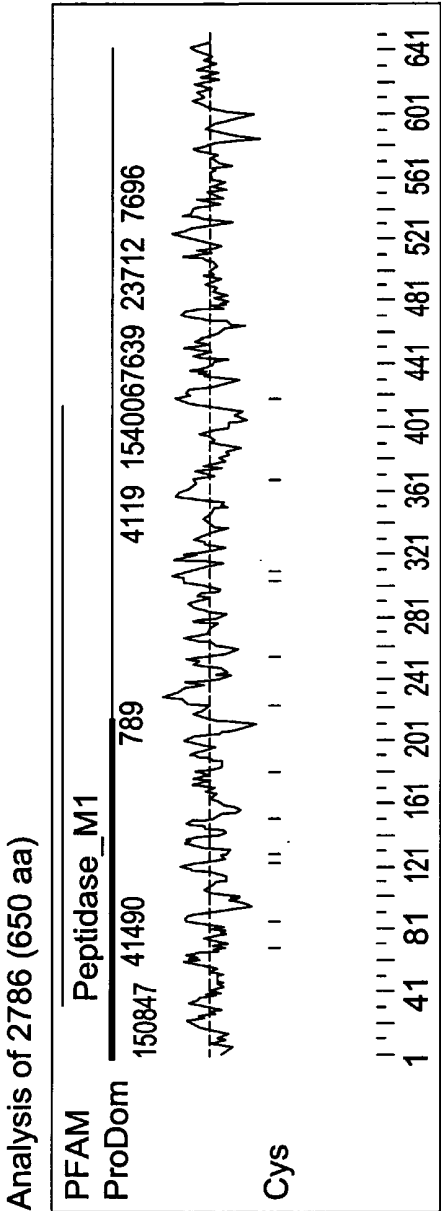


FIG. 2.

3/10

Prosite Pattern Matches for 2786

Prosite version: Release 12.2 of February 1995

>PS00004/PDOC00004/CAMP_PHOSPHO_SITE CAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 356 RRIS 359

>PS00005/PDOC00005/PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 141 TYR 143

Query: 374 TGR 376

>PS00006/PDOC00006/CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query: 208 STWE 211

Query: 318 SLAD 321

Query: 368 TCLE 371

Query: 386 TGEE 389

Query: 408 TYNE 411

Query: 412 TPYE 415

Query: 496 SPGD 499

>PS00008/PDOC00008/MYRISTYL N-myristoylation site.

Query: 9 GSGAAR 14

Query: 58 GLSGTA 63

Query: 119 GQALCV 124

Query: 333 GNLVTN 338

Query: 364 GAAYTC 369

Query: 614 GSEVAQ 619

>PS00009/PDOC00009/AMIDATION Amidation site.

Query: 159 AGKK 162

>PS00030/PDOC00030/RNP_1 Eukaryotic putative RNA-binding region RNP-1 signature.

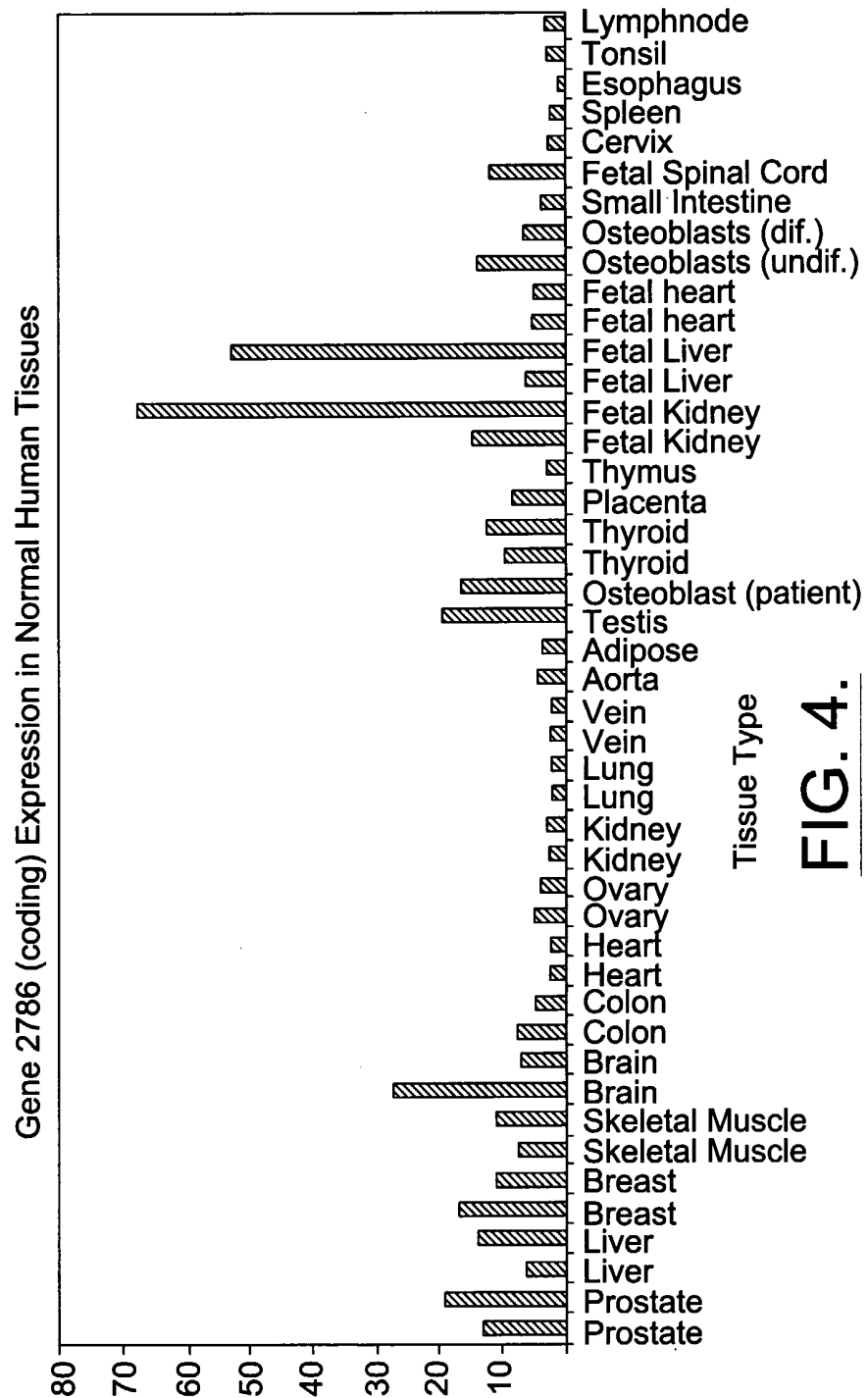
Query: 416 KGFCVSY 423

>PS00142/PDOC00129/ZINC_PROTEASE Neutral zinc metalloproteases, zinc-binding region signature.

Query: 322 VIIHEISHSW 331

FIG. 3.

4/10



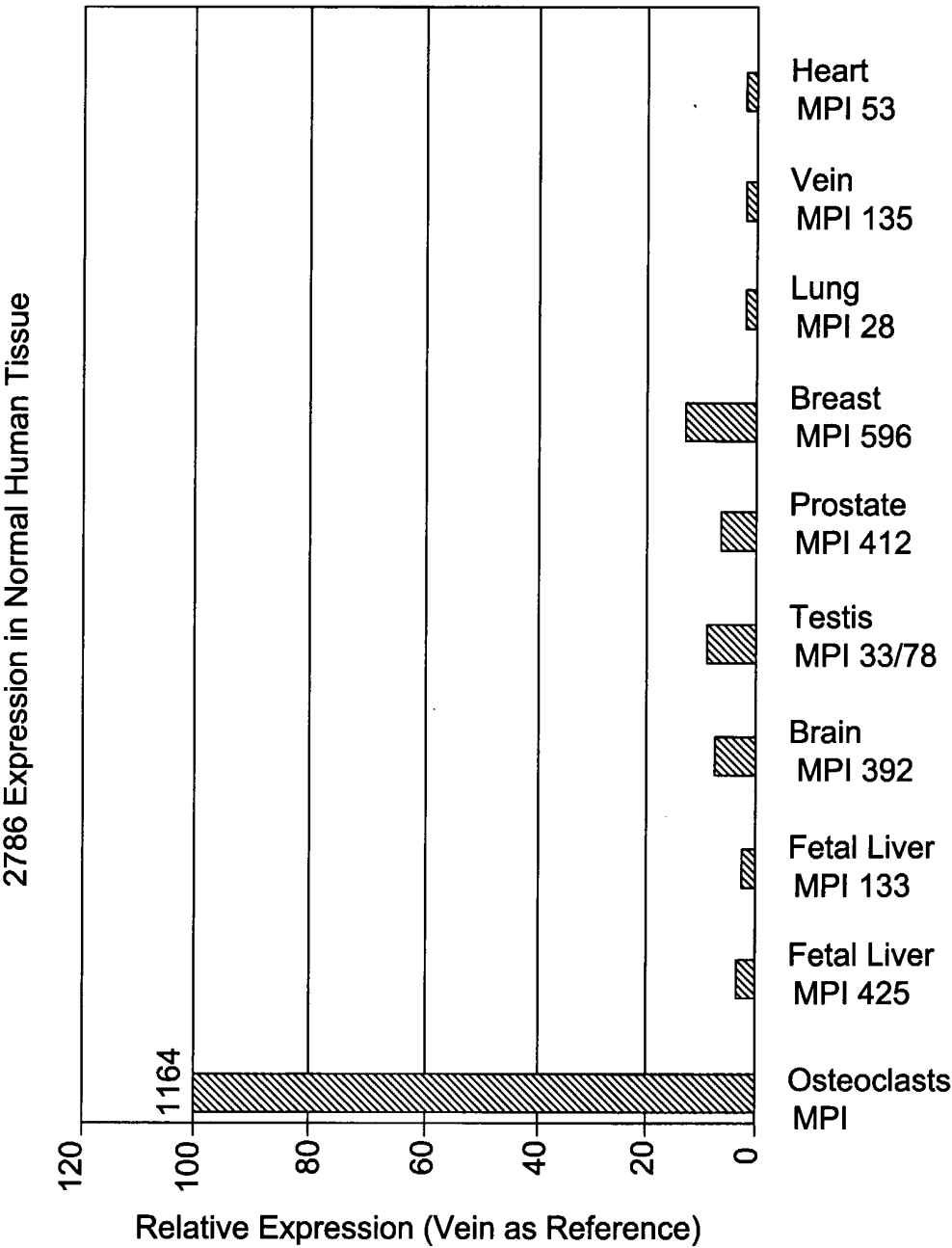


FIG. 5.

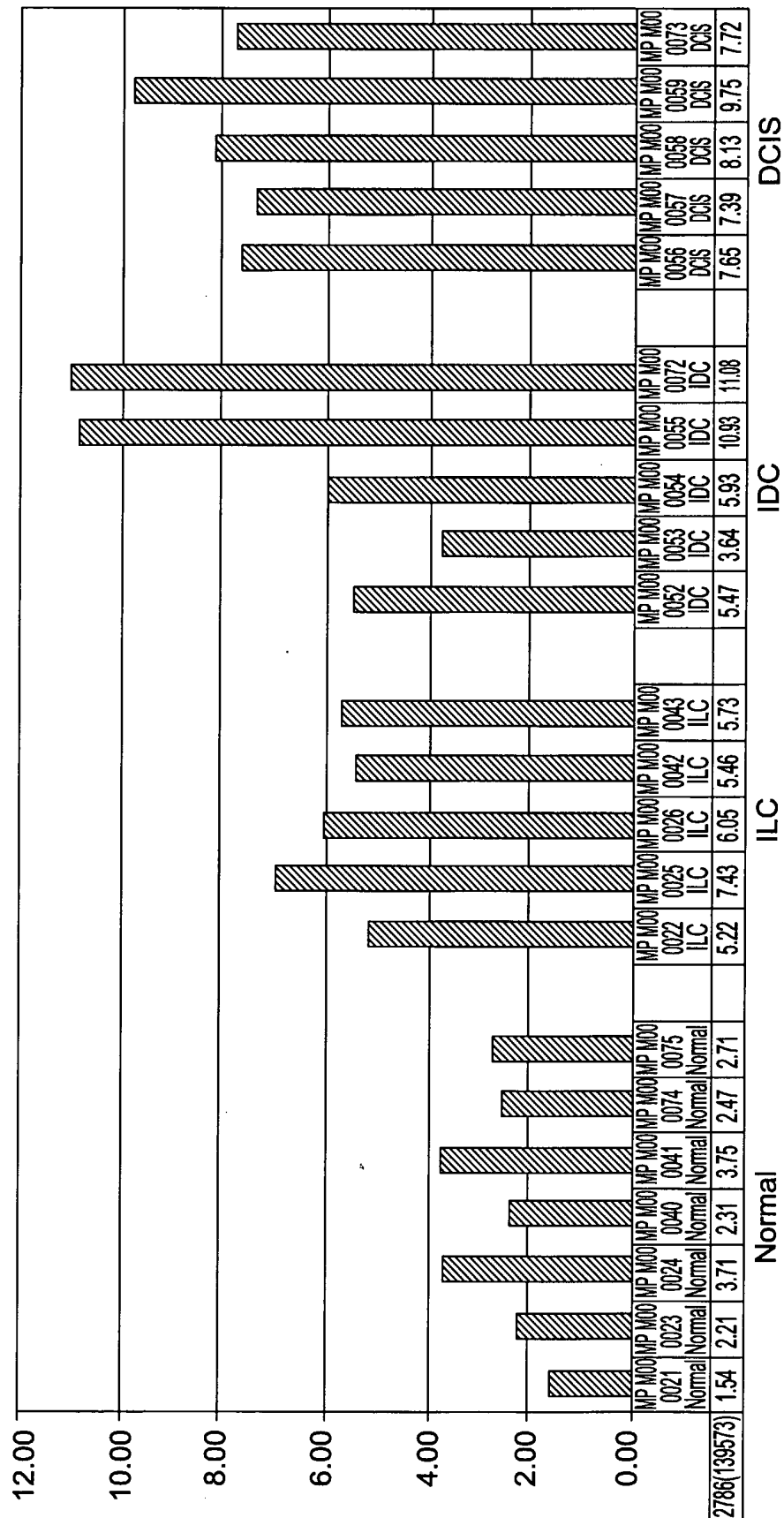


FIG. 6.

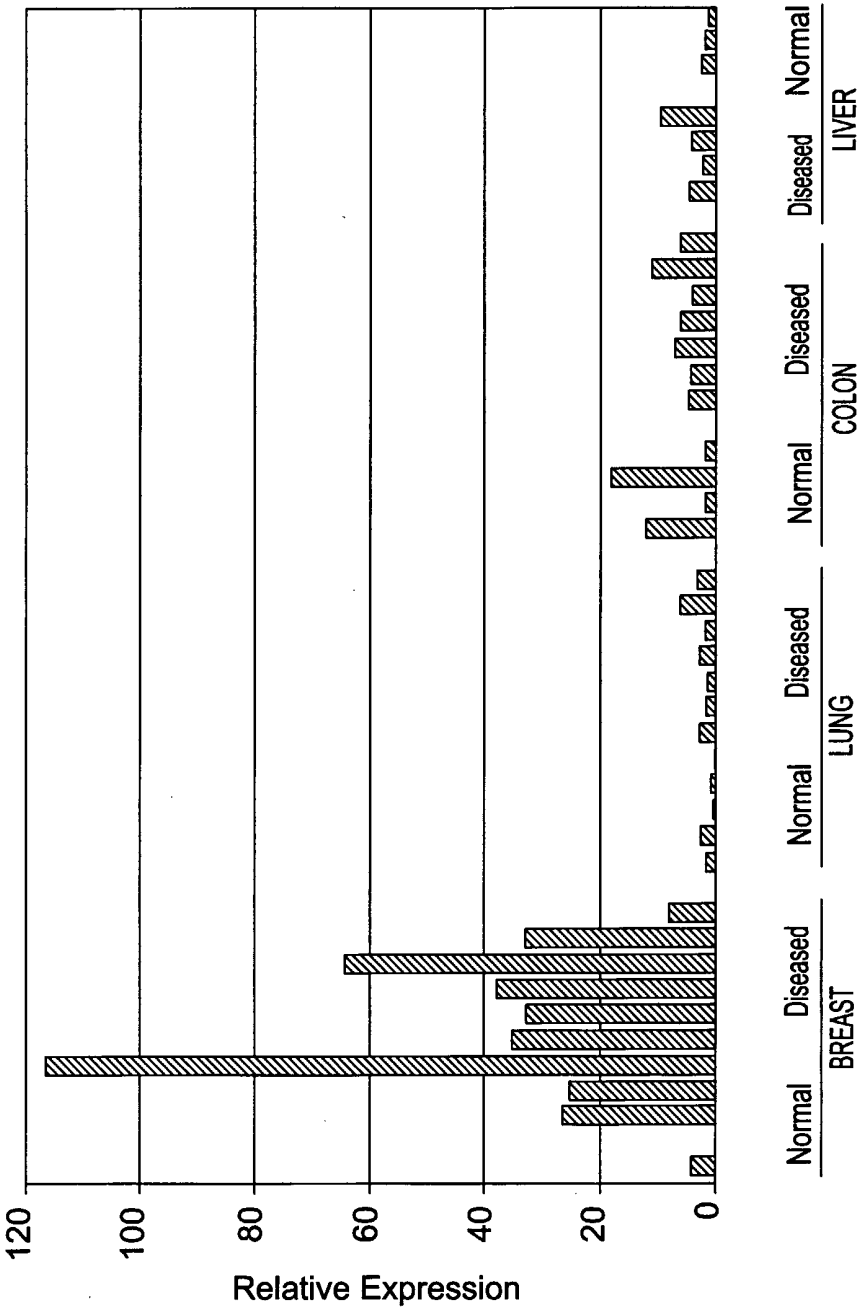
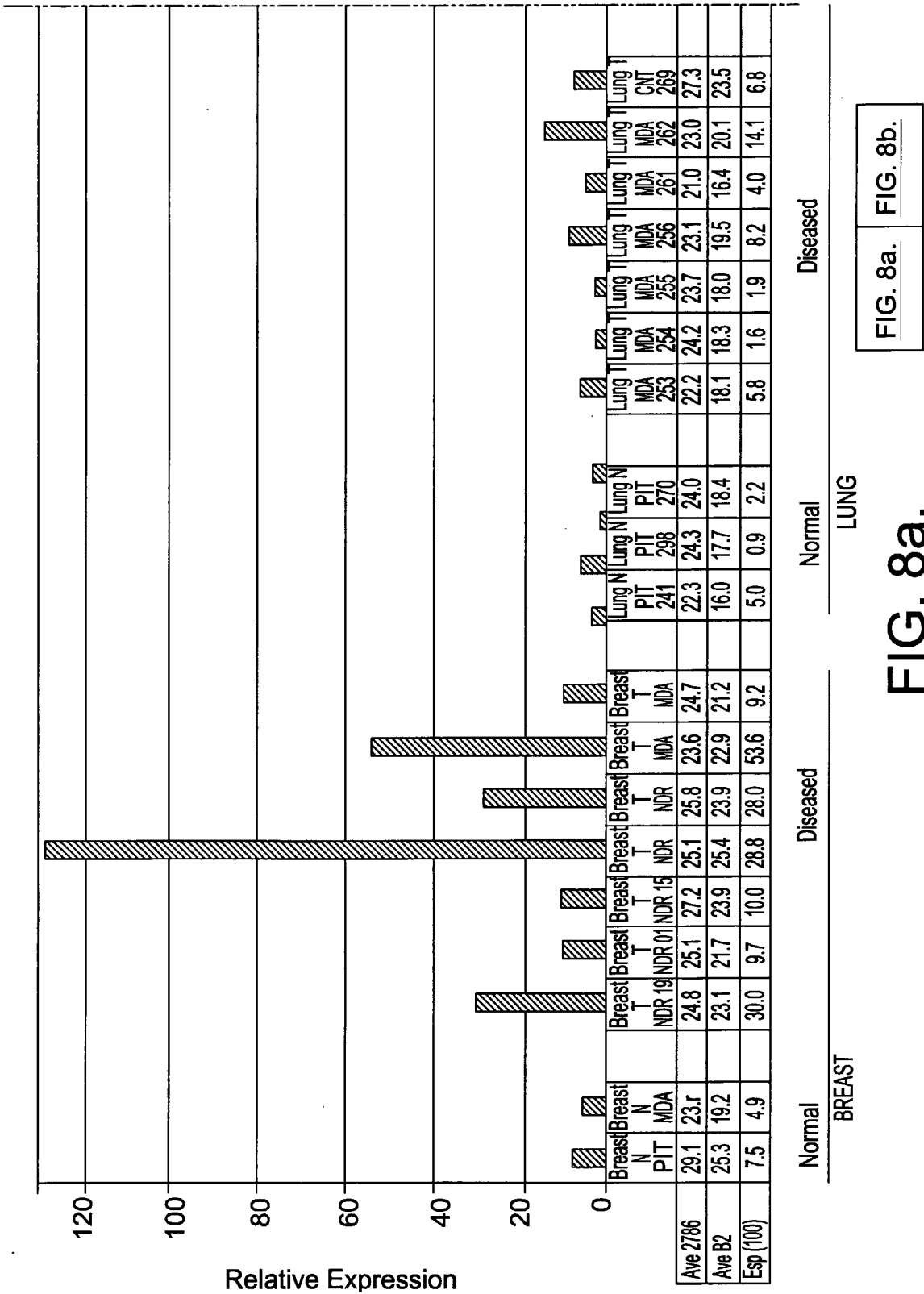


FIG. 7.

TO FIG. 8b.



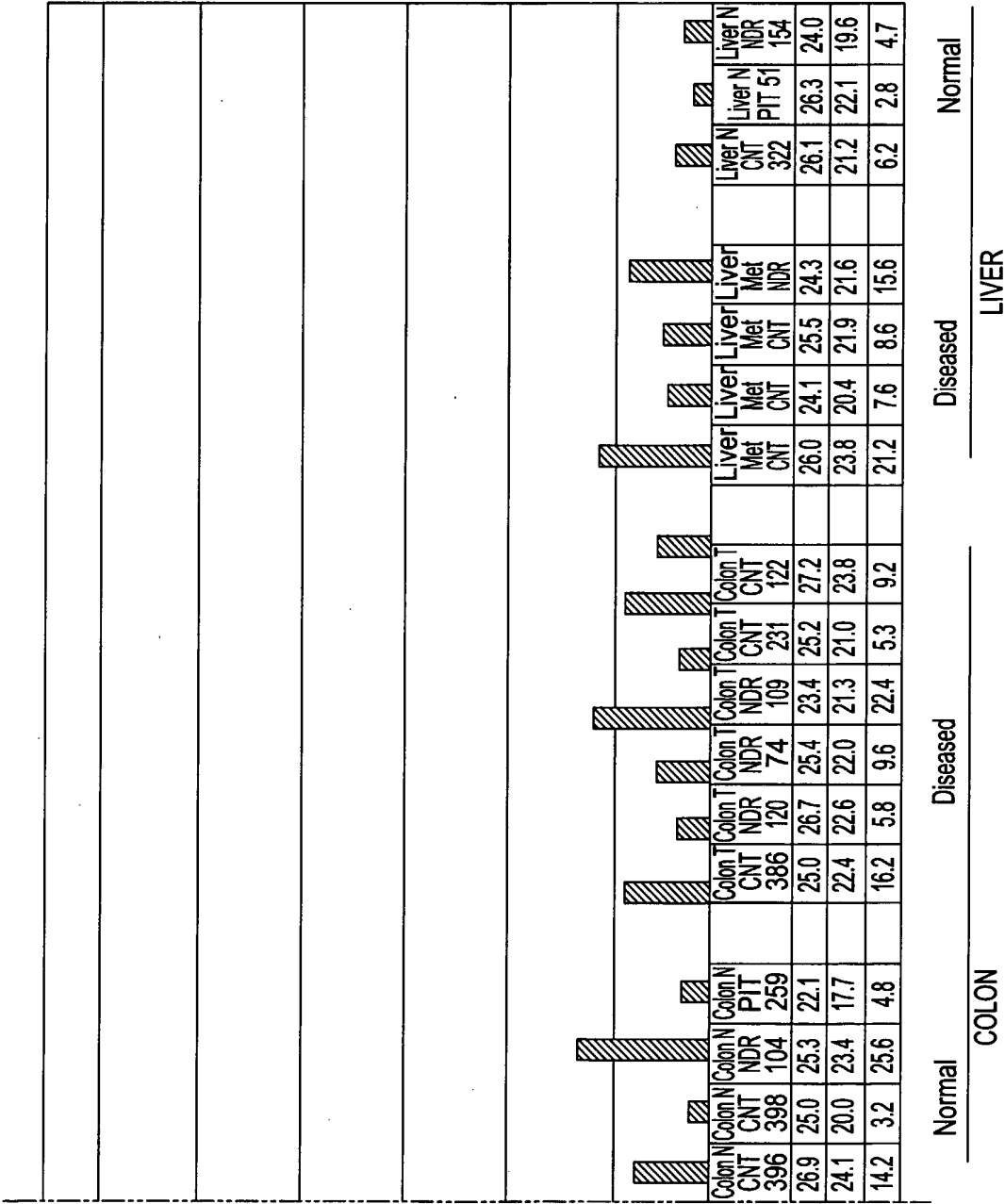
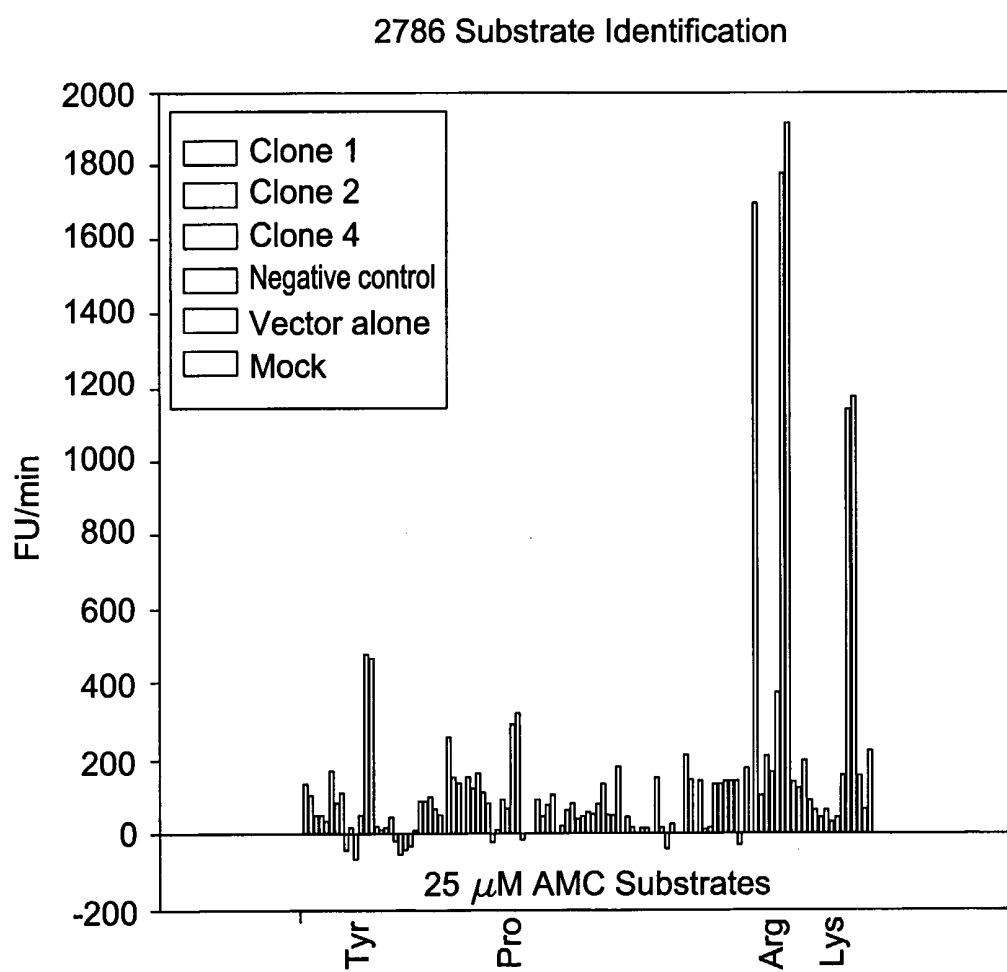


FIG. 8b.

10/10

**FIG. 9.**

SEQUENCE LISTING

<110> Kapeller Libermann, Rosana
White, David
MacBeth, Kyle J.

<120> 2786, A NOVEL HUMAN AMINOPEPTIDASE

<130> 5800-62-1

<160> 2

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 650

<212> PRT

<213> Homo sapiens

<400> 1

```

Met Ala Ser Gly Glu His Ser Pro Gly Ser Gly Ala Ala Arg Arg Pro
 1          5          10          15
Leu His Ser Ala Gln Ala Val Asp Val Ala Ser Ala Ser Asn Phe Arg
 20          25          30
Ala Phe Glu Leu Leu His Leu His Leu Asp Leu Arg Ala Glu Phe Gly
 35          40          45
Pro Pro Gly Pro Gly Ala Gly Ser Arg Gly Leu Ser Gly Thr Ala Val
 50          55          60
Leu Asp Leu Arg Cys Leu Glu Pro Glu Gly Ala Ala Glu Leu Arg Leu
 65          70          75          80
Asp Ser His Pro Cys Leu Glu Val Thr Ala Ala Ala Leu Arg Arg Glu
 85          90          95
Arg Pro Gly Ser Glu Glu Pro Pro Ala Glu Pro Val Ser Phe Tyr Thr
100          105          110
Gln Pro Phe Ser His Tyr Gly Gln Ala Leu Cys Val Ser Phe Pro Gln
115          120          125
Pro Cys Arg Ala Ala Glu Arg Leu Gln Val Leu Leu Thr Tyr Arg Val
130          135          140
Gly Glu Gly Pro Gly Val Cys Trp Leu Ala Pro Glu Gln Thr Ala Gly
145          150          155          160
Lys Lys Lys Pro Phe Val Tyr Thr Gln Gly Gln Ala Val Leu Asn Arg
165          170          175
Ala Phe Phe Pro Cys Phe Asp Thr Pro Ala Val Lys Tyr Lys Tyr Ser
180          185          190
Ala Leu Ile Glu Val Pro Asp Gly Phe Thr Ala Val Met Ser Ala Ser
195          200          205
Thr Trp Glu Lys Arg Gly Pro Asn Lys Phe Phe Phe Gln Met Cys Gln
210          215          220
Pro Ile Pro Ser Tyr Leu Ile Ala Leu Ala Ile Gly Asp Leu Val Ser
225          230          235          240
Ala Glu Val Gly Pro Arg Ser Arg Val Trp Ala Glu Pro Cys Leu Ile
245          250          255
Asp Ala Ala Asn Glu Glu Tyr Asn Gly Val Ile Glu Glu Phe Leu Ala
260          265          270
Thr Gly Glu Lys Leu Phe Gly Pro Tyr Val Trp Gly Arg Tyr Asp Leu
275          280          285
Leu Phe Met Pro Pro Ser Phe Pro Phe Gly Gly Met Glu Asn Pro Cys
290          295          300
Leu Thr Phe Val Thr Pro Cys Leu Leu Ala Gly Asp Arg Ser Leu Ala
305          310          315          320
Asp Val Ile Ile His Glu Ile Ser His Ser Trp Phe Gly Asn Leu Val
325          330          335
Thr Asn Ala Asn Trp Gly Glu Phe Trp Leu Asn Glu Gly Phe Thr Met
340          345          350
Tyr Ala Gln Arg Arg Ile Ser Thr Ile Leu Phe Gly Ala Ala Tyr Thr

```

355 360 365
 Cys Leu Glu Ala Ala Thr Gly Arg Ala Leu Leu Arg Gln His Met Asp
 370 375 380
 Ile Thr Gly Glu Glu Asn Pro Leu Asn Lys Leu Arg Val Lys Ile Glu
 385 390 395 400
 Pro Gly Val Asp Pro Asp Asp Thr Tyr Asn Glu Thr Pro Tyr Glu Lys
 405 410 415
 Gly Phe Cys Phe Val Ser Tyr Leu Ala His Leu Val Gly Asp Gln Asp
 420 425 430
 Gln Phe Asp Ser Phe Leu Lys Ala Tyr Val His Glu Phe Lys Phe Arg
 435 440 445
 Ser Ile Leu Ala Asp Asp Phe Leu Asp Phe Tyr Leu Glu Tyr Phe Pro
 450 455 460
 Glu Leu Lys Lys Lys Arg Val Asp Ile Ile Pro Gly Phe Glu Phe Asp
 465 470 475 480
 Arg Trp Leu Asn Thr Pro Gly Trp Pro Pro Tyr Leu Pro Asp Leu Ser
 485 490 495
 Pro Gly Asp Ser Leu Met Lys Pro Ala Glu Glu Leu Ala Gln Leu Trp
 500 505 510
 Ala Ala Glu Glu Leu Asp Met Lys Ala Ile Glu Ala Val Ala Ile Ser
 515 520 525
 Pro Trp Lys Thr Tyr Gln Leu Val Tyr Phe Leu Asp Lys Ile Leu Gln
 530 535 540
 Lys Ser Pro Leu Pro Pro Gly Asn Val Lys Lys Leu Gly Asp Thr Tyr
 545 550 555 560
 Pro Ser Ile Ser Asn Ala Arg Asn Ala Glu Leu Arg Leu Arg Trp Gly
 565 570 575
 Gln Ile Val Leu Lys Asn Asp His Gln Glu Asp Phe Trp Lys Val Lys
 580 585 590
 Glu Phe Leu His Asn Gln Gly Lys Gln Lys Tyr Thr Leu Pro Leu Tyr
 595 600 605
 His Ala Met Met Gly Gly Ser Glu Val Ala Gln Thr Leu Ala Lys Glu
 610 615 620
 Thr Phe Ala Ser Thr Ala Ser Gln Leu His Ser Asn Val Val Asn Tyr
 625 630 635 640
 Val Gln Gln Ile Val Ala Pro Lys Gly Ser
 645 650

<210> 2

<211> 2459

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (62)...(2011)

<221> misc_feature

<222> (1)...(3107)

<223> n = A,T,C or G

<400> 2

gcggccgcgt cgacctcccc tcgggttcgc ggcccggccg gtgagcaacg gctctgcggc 60
 c atg gcg agc ggc gag cat tcc ccc ggc agc ggc gcg gcc cgg cgg ccg 109
 Met Ala Ser Gly Glu His Ser Pro Gly Ser Gly Ala Ala Arg Arg Pro
 1 5 10 15
 ctg cac tcc gcg cag gct gtg gac gtg gcc tcg gcc tcc aac ttc cgg 157
 Leu His Ser Ala Gln Ala Val Asp Val Ala Ser Ala Ser Asn Phe Arg
 20 25 30
 gcc ttt gag ctg ctg cac ttg cac ctg gac ctg cgg gct gag ttc ggg 205
 Ala Phe Glu Leu Leu His Leu His Leu Asp Leu Arg Ala Glu Phe Gly
 35 40 45
 cct cca ggg ccc ggc gca ggg agc cgg ggg ctg agc ggc acc gcg gtc 253
 Pro Pro Gly Pro Gly Ala Gly Ser Arg Gly Leu Ser Gly Thr Ala Val

50	55	60	
ctg gac ctg cgc tgc ctg gag ccc gag ggc gcc gcc gag ctg cgg ctg Leu Asp Leu Arg Cys Leu Glu Pro Glu Gly Ala Ala Glu Leu Arg Leu 65 70 75 80			301
gac tcg cac ccg tgc ctg gag gtg acg gcg gcg gcg ctg cgg cgg gag Asp Ser His Pro Cys Leu Glu Val Thr Ala Ala Leu Arg Arg Glu 85 90 95			349
cgg ccc ggc tcg gag gag ccg cct gcg gag ccc gtg agc ttc tac acg Arg Pro Gly Ser Glu Glu Pro Pro Ala Glu Pro Val Ser Phe Tyr Thr 100 105 110			397
cag ccc ttc tcg cac tat ggc cag gcc ctg tgc gtg tcc ttc ccg cag Gln Pro Phe Ser His Tyr Gly Gln Ala Leu Cys Val Ser Phe Pro Gln 115 120 125			445
ccc tgc cgc gcc gcc gag cgc ctc cag gtg ctg ctc acc tac cgc gtc Pro Cys Arg Ala Ala Glu Arg Leu Gln Val Leu Leu Thr Tyr Arg Val 130 135 140			493
ggg gag gga ccc ggg gtt tgc tgg ttg gct ccc gag cag aca gca gga Gly Glu Gly Pro Gly Val Cys Trp Leu Ala Pro Glu Gln Thr Ala Gly 145 150 155 160			541
aag aag aag ccc ttc gtg tac acc cag ggc cag gct gtc cta aac cgg Lys Lys Lys Pro Phe Val Tyr Thr Gln Gly Gln Ala Val Leu Asn Arg 165 170 175			589
gcc ttc ttc cct tgc ttc gac acg cct gct gtt aaa tac aag tat tca Ala Phe Phe Pro Cys Phe Asp Thr Pro Ala Val Lys Tyr Lys Tyr Ser 180 185 190			637
gct ctt att gag gtc cca gat ggc ttc aca gct gtg atg agt gct agc Ala Leu Ile Glu Val Pro Asp Gly Phe Thr Ala Val Met Ser Ala Ser 195 200 205			685
acc tgg gag aag aga ggt cca aat aag ttc ttc ttc cag atg tgt cag Thr Trp Glu Lys Arg Gly Pro Asn Lys Phe Phe Phe Gln Met Cys Gln 210 215 220			733
ccc atc ccc tcc tat ctg ata gct ttg gcc atc gga gat ctg gtt tcg Pro Ile Pro Ser Tyr Leu Ile Ala Leu Ala Ile Gly Asp Leu Val Ser 225 230 235 240			781
gct gaa gtt gga ccc agg agc cgg gtg tgg gct gag ccc tgc ctg att Ala Glu Val Gly Pro Arg Ser Arg Val Trp Ala Glu Pro Cys Leu Ile 245 250 255			829
gat gct gcc aat gag gag tac aac ggg gtg ata gaa gaa ttt ttg gca Asp Ala Ala Asn Glu Glu Tyr Asn Gly Val Ile Glu Glu Phe Leu Ala 260 265 270			877
aca gga gag aag ctt ttt gga cct tat gtt tgg gga agg tat gac ttg Thr Gly Glu Lys Leu Phe Gly Pro Tyr Val Trp Gly Arg Tyr Asp Leu 275 280 285			925
ctc ttc atg cca ccg tcc ttt cca ttt gga gga atg gag aac cct tgt Leu Phe Met Pro Pro Ser Phe Pro Phe Gly Gly Met Glu Asn Pro Cys 290 295 300			973
ctg acc ttt gtc acc ccc tgc ctg cta gct ggg gac cgc tcc ttg gca Leu Thr Phe Val Thr Pro Cys Leu Leu Ala Gly Asp Arg Ser Leu Ala 305 310 315 320			1021
gat gtc atc atc cat gag atc tcc cac agt tgg ttt ggg aac ctg gtc			1069

Asp Val Ile Ile His Glu Ile Ser His Ser Trp Phe Gly Asn Leu Val	
325 330 335	
acc aac gcc aac tgg ggt gaa ttc tgg ctc aat gaa ggt ttc acc atg	1117
Thr Asn Ala Asn Trp Gly Glu Phe Trp Leu Asn Glu Gly Phe Thr Met	
340 345 350	
tac gcc cag agg agg atc tcc acc atc ctc ttt ggc gct gcg tac acc	1165
Tyr Ala Gln Arg Arg Ile Ser Thr Ile Leu Phe Gly Ala Ala Tyr Thr	
355 360 365	
tgc ttg gag gct gca acg ggg cgg gct ctg ctg cgt caa cac atg gac	1213
Cys Leu Glu Ala Ala Thr Gly Arg Ala Leu Leu Arg Gln His Met Asp	
370 375 380	
atc act gga gag gaa aac cca ctc aac aag ctc cgc gtg aag att gaa	1261
Ile Thr Gly Glu Glu Asn Pro Leu Asn Lys Leu Arg Val Lys Ile Glu	
385 390 395 400	
cca ggc gtt gac ccg gac gac acc tat aat gag acc ccc tac gag aaa	1309
Pro Gly Val Asp Pro Asp Asp Thr Tyr Asn Glu Thr Pro Tyr Glu Lys	
405 410 415	
ggg ttc tgc ttt gtc tca tac ctg gcc cac ttg gtg ggt gat cag gat	1357
Gly Phe Cys Phe Val Ser Tyr Leu Ala His Leu Val Gly Asp Gln Asp	
420 425 430	
cag ttt gac agt ttt ctc aag gcc tat gtg cat gaa ttc aaa ttc cga	1405
Gln Phe Asp Ser Phe Leu Lys Ala Tyr Val His Glu Phe Lys Phe Arg	
435 440 445	
agc atc tta gcc gat gac ttt ctg gac ttc tac ttg gaa tat ttc cct	1453
Ser Ile Leu Ala Asp Asp Phe Leu Asp Phe Tyr Leu Glu Tyr Phe Pro	
450 455 460	
gag ctt aag aaa aag aga gtg gat atc att cca ggt ttt gag ttt gat	1501
Glu Leu Lys Lys Lys Arg Val Asp Ile Ile Pro Gly Phe Glu Phe Asp	
465 470 475 480	
cga tgg ctg aat acc ccc ggc tgg ccc ccg tac ctc cct gat ctc tcc	1549
Arg Trp Leu Asn Thr Pro Gly Trp Pro Pro Tyr Leu Pro Asp Leu Ser	
485 490 495	
cct ggg gac tca ctc atg aag cct gct gaa gag cta gcc caa ctg tgg	1597
Pro Gly Asp Ser Leu Met Lys Pro Ala Glu Glu Leu Ala Gln Leu Trp	
500 505 510	
gca gcc gag gag ctg gac atg aag gcc att gaa gcc gtg gcc atc tct	1645
Ala Ala Glu Glu Leu Asp Met Lys Ala Ile Glu Ala Val Ala Ile Ser	
515 520 525	
ccc tgg aag acc tac cag ctg gtc tac ttc ctg gat aag atc ctc cag	1693
Pro Trp Lys Thr Tyr Gln Leu Val Tyr Phe Leu Asp Lys Ile Leu Gln	
530 535 540	
aaa tcc cct ctc cct cct ggg aat gtg aaa aaa ctt gga gac aca tac	1741
Lys Ser Pro Leu Pro Pro Gly Asn Val Lys Lys Leu Gly Asp Thr Tyr	
545 550 555 560	
cca agt atc tca aat gcc cgg aat gca gag ctc cgg ctg cga tgg ggc	1789
Pro Ser Ile Ser Asn Ala Arg Asn Ala Glu Leu Arg Leu Arg Trp Gly	
565 570 575	
caa atc gtc ctt aag aac gac cac cag gaa gat ttc tgg aaa gtg aag	1837
Gln Ile Val Leu Lys Asn Asp His Gln Glu Asp Phe Trp Lys Val Lys	
580 585 590	

gag ttc ctg cat aac cag ggg aag cag aag tat aca ctt ccg ctg tac	1885
Glu Phe Leu His Asn Gln Gly Lys Gln Lys Tyr Thr Leu Pro Leu Tyr	
595 600 605	
cac gca atg atg ggt ggc agt gag gtg gcc cag acc ctc gcc aag gag	1933
His Ala Met Met Gly Gly Ser Glu Val Ala Gln Thr Leu Ala Lys Glu	
610 615 620	
act ttt gca tcc acc gcc tcc cag ctc cac agc aat gtt gtc aac tat	1981
Thr Phe Ala Ser Thr Ala Ser Gln Leu His Ser Asn Val Val Asn Tyr	
625 630 635 640	
gtc cag cag atc gtg gca ccc aag ggc agt tagaggctcg tgtgcatggc	2031
Val Gln Gln Ile Val Ala Pro Lys Gly Ser	
645 650	
ccctgcctct tcaggctctc caggctttca gaataattgt ttgttcccaa attcctgttc	2091
cctgatcaac ttctggaggt ttatatcccc tcaggataat ctattctcta gcttaggtat	2151
ctgtgactct tgggcctctg ctctgggtggg aacttacttc tctatagccc actgagcccc	2211
gagacagaga acctgcccac agctctcccc gctacaggct gcaggcactg cagggcagcg	2271
ggtattctcc tccccaccta agtctctggg aagaagtgga gaggactgat gctcttcttt	2331
tttctctttc tgtccttttt cttgctgatt ttatgcaaag ggctggcatt ctgattgttc	2391
ttttttcagg tttaatcctt attttaataa agttttcaag caaaaattaa aaaaaaaaaa	2451
aaaaaaaa	2459