PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/57, 9/64, 15/62 A61K 37/54, C12N 1/21 // (C12N 1/21, C12R 1:00)	A2	(11) International Publication Number: WO 93/18159 (43) International Publication Date: 16 September 1993 (16.09.93)

(21) International Application Number: PCT/US93/02147

(22) International Filing Date: 9 March 1993 (09.03.93)

(30) Priority data:
07/848,629 9 March 1992 (09.03.92) US

(71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

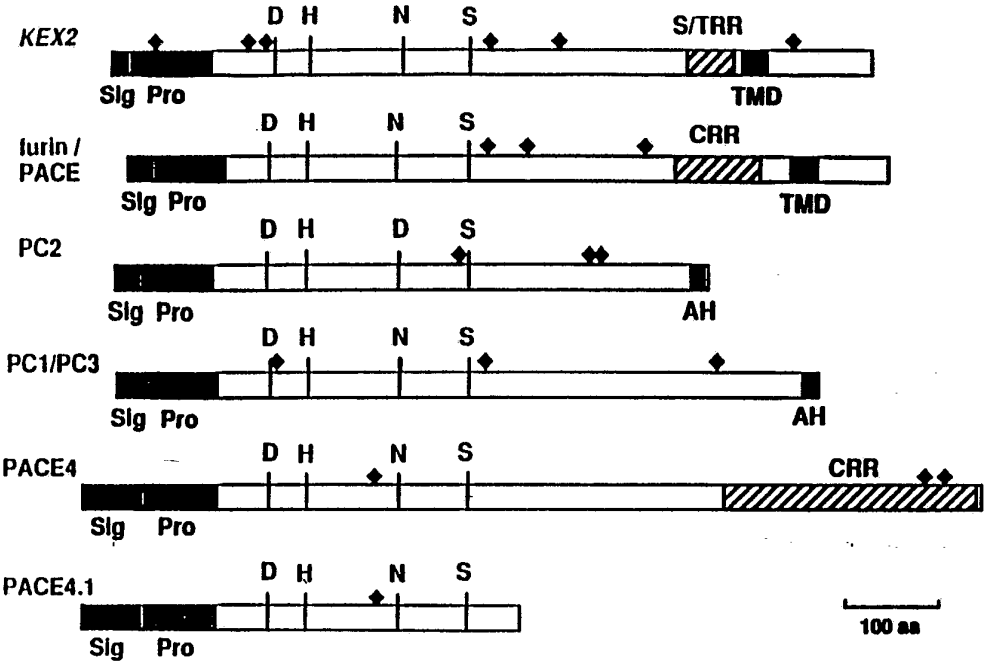
(72) Inventors: BARR, Philip, J. ; 152 Hillcrest Road, Berkeley, CA 94705 (US). KIEFER, Michael, C. ; 401 Wright Court, Clayton, CA 94517 (US).

(74) Agents: CHUNG, Ling-Fong et al.; Chiron Corporation, Intellectual Property Department, 4560 Horton Street, Emeryville, CA 94608 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

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

(54) Title: COMPOSITIONS AND METHODS FOR PACE 4 AND 4.1 GENE AND POLYPEPTIDES IN CELLS



(57) Abstract

Compositions and methods are provided for endopeptidases and their production, and for enhanced efficiencies of processing heterologous precursor polypeptides to mature polypeptides. These compositions and methods utilize recombinant PACE 4 and 4.1, mammalian endopeptidases that are specific for dibasic amino acid sites. Therapeutic compositions and methods employing PACE 4 or 4.1 or their inhibitors are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SK	Slovak Republic
CI	Côte d'Ivoire	KZ	Kazakhstan	SN	Senegal
CM	Cameroon	LI	Liechtenstein	SU	Soviet Union
CS	Czechoslovakia	LK	Sri Lanka	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	UA	Ukraine
DK	Denmark	MG	Madagascar	US	United States of America
ES	Spain	ML	Mali	VN	Viet Nam
FI	Finland	MN	Mongolia		

COMPOSITIONS AND METHODS FOR PACE 4 AND 4.1 GENE AND POLYPEPTIDES IN CELLS

BACKGROUND OF THE INVENTION

1. Field of the Disclosure

This invention relates generally to proteins and their production in recombinant host cells. More particularly, it relates to polynucleotide sequences that encode novel paired basic amino acid converting enzymes (generically referred to as "PACE," with the two new enzymes being PACE 4 and 4.1), to the production of PACE 4 and 4.1 in transformed cells, and to compositions comprising the PACE 4 and 4.1 enzymes. Materials and methods are also provided for the production of mature forms of proteins from heterologous precursor polypeptides using PACE 4 or 4.1, which are expressed in recombinant host cells.

2. Description of Related Art

Many eukaryotic proteins synthesized in bacteria fold incorrectly and, consequently, exhibit low specific activities. Post-translational proteolysis is a common mechanism required for the synthesis of biologically active proteins and peptides in all eukaryotes examined, including yeast (R.S. Fuller et al., (1988), Ann. Rev. Physiol. 50:345), invertebrates (R.H, Scheller et al. (1983), Cell 32:7), and mammalian cells (J. Douglass et al. (1984), Ann. Rev. Biochem. 53:665; W.S. Sossin et al. (1989), Neuron 2, 1407).

One of the early events in precursor protein maturation is endoproteolysis at the carboxyl side of pairs of basic amino acid sequences (e.g. -LysArg- and -ArgArg-). This kind of endoproteolytic cleavage was initially inferred from the sequences of several endocrine and neuroendocrine precursor proteins and was first proposed from studies of proinsulin (D.F. Steiner et al. (1968), Science 157:697; R.E. Chance et al. (1968), Science 161:165) and the ACTH/ß-endorphin precursor, proopiomelanocortin (POMC)(M. Chretien and C.H. Li (1967), Can. J. Biochem. 45:1163). Subsequent studies have revealed a broad spectrum of precursor proteins that require

endoproteolysis at pairs of basic amino acids to yield mature peptides including serum factors (A.K. Bentley et al. (1986), Cell 45:343), viral proteins (C.M. Rice et al. (1986), Virology 151:1; C.M. Rice et al. (1985), Science 229:726; J.M. McCune et al. (1988), Cell 53:55), growth factors (L.E. Gentry et al. (1988), Mol. Cell Biol. 8:4162; K. Sharples et al. (1987), DNA 6:239; M. Yanagisawa et al. (1988), Nature 332:411; and Gray et al. (1983), Nature 303:722) and receptors (Y. Yosimasa (1988), Science 240:784).

Several candidate enzymes that are capable of cleaving at single or paired basic residues in vitro have been proposed as authentic mammalian precursor endoproteases. See, for example, Y.P. Loh and H. Gainer, in Brain Peptides, D.T. Krieger, M.J. Brownstein, J.B. Martin, Eds. (Wiley-Interscience, New York, 1983), pp.76-116; M. Chretien, et al. in Cell Biology of the Secretory Process (Karger, Basel, Switzerland, 1983), pp.214-246; A.J. Mason, et al. (1983), Nature 303:300; P.J. Isackson et al. (1987), J. Cell. Biochem. 33:65; I. Lindberg et al., (1984), J. Neurochem 42:1411; J.A. Cromlish et al. (1986), J. Biol. Chem. 261:10850; K. Docherty et al. (1984), J. Biol. Chem. 259:6041; T.C. Chang and Y.P. Loh (1984), Endocrinology 114, 2092; B.P. Noe et al. (1984), J. Cell. Biol. 99:578; U.P. Loh (1986), J. Biol. Chem. 261:11949; H.W. Davidson et al. (1987), Biochem. J. 246:279; P. Gluschankof et al. (1987), J. Biol. Chem. 262:9615; C. Clamigrand et al. (1987), Biochem 26:6018; S.O. Brennan and R.J. Peach (1988), FEBS Letters 229:167; R.S. Fuller et al. (1989), Proc. Natl. Acad. Sci. USA 86:1434; K. Mizuno et al. (1989), Biochem. Biophys. Res. Comm. 159:305; I.C. Bathurst et al. (1987), Science 235:348; and G. Thomas et al. (1988), Science 241:226. However, none of these candidate activities have been shown to be a bona fide precursor cleaving endoprotease in vivo.

The yeast enzyme Kex2, a membrane-bound, Ca^{2+}-dependent serine protease (K. Mizuno et al. (1988), Biochem. Biophys. Res. Commun. 156:246; R.S. Fuller et al. (1989), Proc. Natl. Acad. Sci. USA 86:1434), has been considered to be a prototypic proprotein convertase. The Kex2 endoprotease, which is encoded by the KEX2 gene, functions late in the secretory pathway of Saccharomyces cerevisiae, and cleaves the polypeptide chains of prepro-killer toxin and prepro-α-factor at the paired basic amino acid sequences Lys-Arg and Arg-Arg (D. Julius et al. (1984), Cell 36:309). Furthermore, co-expression of the KEX2 gene with POMC in BSC-40 cells (a cell line

which is incapable of processing this peptide precursor) resulted in the generation, by proteolysis at pairs of basic amino acids, of authentic product peptides, including b-LPH and ß-endorphin (Thomas et al. (1988), id.).

Two human DNA sequences, fur and PC2, share some structural homology with each other and with the KEX2 gene sequence. A.M.W. van den Ouweland et al. (1990), Nucleic Acids Res. 18:664 (presenting a cDNA coding sequence for fur; the cDNA sequence of Figure 1 differs from this sequence in the region encoded by nucleotides 1-238); R.S. Fuller et al., Science 246:482 (1989); S.P. Smeekens (1990), J. Biol. Chem. 265:2997. The fur locus was initially identified by its proximity to the fes/fps proto-oncogene (A.J.M. Roebroek et al. (1986), EMBO J. 5:2197). PC2 was identified by amplification of a human insulinoma library by the polymerase chain reaction using KEX2-derived primers; it shares a partial homology with Kex2, especially in the putative active site domains (S.P. Smeekens and D.F. Steiner, (1990) J. Biol. Chem. 265:2997).

Another related cDNA encoding a protein variously called PC1 (Seidah et al. (1991) Mol. Endocrinol. 5:111-122) or PC3 (Smeekens et al. (1991) Proc. Natl. Acad. Sci. USA 88:340-344) has been isolated from murine pituitary cells. PC1/PC3 also shares amino acid similarity with KEX2, particularly in the putative catalytic domains.

Recently a functional activity has been demonstrated for PC2 and PC1/PC3, using as a substrate the prohormone proopiomelanocortin (POMC). In the anterior lobe of the pituitary, POMC is processed to yield the hormones ACTH, ß-lipotropin, and ß-endorphin, while in the intermediate lobe it is processed to yield α-melanocyte stimulating hormone (α-MSH) and variant forms of ACTH and ß-lipotropin. PC1/PC3 expressed in processing deficient cells cleaved POMC to give ACTH, but was less efficient in cleaving ß-lipotropin to give ß-endorphin. This is similar to the activity found in pituitary corticotrophic cells. Coexpression of PC2 and PC1/PC3 efficiently converted ß-lipotropin, as is found in melanotrophic cells. Coexpression of POMC and PC2 in adrenomedullary chromaffin cells resulted in secretion of ß-endorphin and α-MSH. Therefore PC1/PC3 and PC2 have differing substrate selectivities which yield physiologically relevant processed polypeptides.

RELEVANT LITERATURE

AVIV, H., and LEDER, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**, 1408-1412.

BARR, P.J. (1991). Mammalian subtilisins: The long-sought dibasic processing endo-proteases. Cell **66**, 1-3.

BARR, P.J., MASON, O.B., LANDSBERG, K.E., WONG, P.A., KIEFER, M.C., and BRAKE, A.J. (1991). cDNA and gene structure for a human subtilisin-like protease with cleavage specificity for paired basic amino acid residues. DNA Cell Biol. **10**, 319-328.

BARR, P.J., THAYER, R.M., LAYBOURN, P., NAJARIAN, R.C., SEELA, F., and TOLAN, D.R. (1986). 7-Deaza-2'-Deoxyguanosine-5'-Triphosphate: Enhanced resolution in M13 dideoxy sequencing. Biotechniques **4**, 428-432.

BENJANNET, S., RONDEAU, N., DAY, R., CHRETIEN, M., and SEIDAH, N.G. (1991). PC1 and PC2 are proprotein convertases capable of cleaving proopiomelanocortin at distinct pairs of basic residues. Proc. Natl. Acad. Sci. USA **88**, 3564-3568.

BRESNAHAN, P.A., LEDUC, R., THOMAS, L., THORNER, J., GIBSON, H.L., BRAKE, A.J., BARR, P.J., and THOMAS, G. (1990). Human fur gene encodes a yeast KEX-2 like endoprotease that cleaves pro-b-NGF in vivo. J. Cell. Biol. **111**, 2851-2859.

CHIRGWIN, J.M., PRZBYLA, A.E., MACDONALD, R.J., AND RUTTER, W.J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**, 5294-5299.

CHRISTIE, D.L., BATCHELOR, D.C., and PALMER, D.J. (1991). Identification of kex2 related proteases in chromaffin granules by partial amino acid sequence analysis. J. Biol. Chem. **266**, 15679-15683.

FEINBERG, A.P., and VOGELSTEIN, B. (1984). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**, 266-267.

FREEMAN, G.J., CLAYBERGER, C., DEKRUYFF, R., ROSENBLUM, D.S., and CANTOR, H. (1983). Sequential expression of new gene programs in inducer T-cell clones. Proc. Natl. Acad. Sci. USA **80**, 4094-4098.

FRICKER, L.D., DAS, B., and ANGELETTI, R.H. (1990). Identification of the pH-dependent membrane anchor of carboxypeptidase E (EC 3.4.17.10). J. Biol. Chem. **265**, 2476-2482.

FULLER, R.S., BRAKE, A.J., and THORNER, J. (1989). Intracellular targeting and structural conservation of a prohormone-processing endoprotease. Science **246**, 482--486.

GILBERT, W. (1985). Genes-in-pieces revisited. Science **228**, 823-824.

HATSUZAWA, K., HOSAKA, M., NAKAGAWA, T., NAGASE, M., SHODA, A., MURAKAMI, K., and NAKAYAMA, K. (1990). Structure and expression of mouse furin, a yeast KEX2-related protease. J. Biol. Chem. **265**, 22075-22078.

KIEFER, M.C., JOH, R.S., BAUER, D.M., and ZAPF, J. (1991). Molecular cloning of a new human insulin-like growth factor binding protein. Biochem. Biophys. Res. Comm. **176**, 219-225.

KIEFER, M.C., MASIARZ, F.R., BAUER, D.M., and ZAPF, J. (1991). Identification and molecular cloning of two new 30-kDa insulin-like growth factor binding proteins isolated from adult human serum. J. Biol. Chem. **266**, 9043-9049.

KORNER, J., CHUN, J., HARTER, D., and AXEL, R. (1991). Isolation and functional expression of a mammalian prohormone processing enzyme, murine prohormone convertase 1. Proc. Natl. Acad. Sci. USA **88**, 6834-6838.

HOSAKA, M., NAGAHAMA, M., KIM, W.-S., WATANABE, T., HATSUZAWA, K., IKEMIZU, J., MURAKAMI, K., and NAKAYAMA, K. (1991). Arg-X-Lys/Arg-Arg motif as a signal for precursor cleavage catalyzed by furin within the constitutive secretory pathway. J. Biol. Chem. **266**, 12127-12130.

LEHRACH, H., DIAMOND, D., WOZNEY, J.M., and BOEDTKER, H. (1977). RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical re-examination. Biochemistry **16**, 4743.

MISUMI, Y., ODA, K., FUJIWARA, T., TAKAMI, N., TASHIRO, K., and IKEHARA, Y. (1991). Functional expression of furin demonstrating its intracellular localization and endoprotease activity for processing of proalbumin and complement pro-C3. J. Biol. Chem. **266**, 16954-16959.

PINKEL, D., LANDEGENT, J., COLLINS, C., FUSCOE, J., SEGRAVES, R., LUCAS, J., and GRAY, J. (1988). Fluorescence in situ hybridization with human chromosome specific libraries: Detection of trisomy 21 and translocations of chromosome 4. Proc. Natl. Acad. Sci. USA **85**, 9138-9142.

ROEBROEK, A.J.M., SCHALKEN, J.A., BUSSEMAKERS, M.J.G., VAN HEERIKHUIZEN, H., ONNEKINK, C., DEBRUYNE, F.M.J., BLOEMERS, H.P.J., and VAN DE VEN, W.J.M. (1986). Characterization of human c-fes/fps reveals a new transcription unit (fur) in the immediately upstream region of the proto-oncogene. Mol. Biol. Rep. **11**, 117-125.

SAIKI, R.K., SCHARF, S., FALOONA, F., MULLIS, K.B., HORN, G.T., ERLICH, H.A., and ARNHEIM, N. (1985). Enzymatic amplification of b-Globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230**, 1350-1354.

SAMBROOK, J., MANIATIS, T., FRITSCH, E.F. (1989). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

SCHALKEN, J.A., ROEBROEK, A.J.M., OOMEN, P.P.C.A., WAGENAAR, S.SC., DEBRUYNE, F.M.J., BLOEMERS, H.P.J., and VAN DE VEN, W.J.M. (1987). fur Gene expression as a discriminating marker for small cell and nonsmall cell lung carcinomas. J. Clin. Invest. **80**, 1545-1549.

SEIDAH, N.G., GASPAR, L., MION, P., MARCINKIEWICZ, M., MBIKAY, M., and CHRETIEN, M. (1990). cDNA sequence of two distinct pituitary proteins homologous to Kex2 and furin gene products: Tissue-specific mRNAs encoding candidates for pro-hormone processing proteinases. DNA Cell Biol. **9**, 415-424.

SHENNAN, K.I.J., SEAL, A.J., SMEEKENS, S.P., STEINER, D.F., and DOCHERTY, K. (1991). Site-directed mutagenesis and expression of PC2 in microinjected xenopus oocytes. J. Biol. Chem.

SMEEKENS, S.P., AVRUCH, A.S., LAMENDOLA, J., CHAN, S.J., and STEINER, D.F. (1991). Identification of a cDNA encoding a second putative prohormone convertase related to PC2 in AtT20 cells and islets of Langerhans. Proc. Natl. Acad. Sci. USA **88**, 340-344.

SMEEKENS, S.P., and STEINER, D.F. (1990). Identification of a human insulinoma cDNA encoding a novel mammalian protein structurally related to the yeast dibasic processing protease Kex2. J. Biol. Chem. **265**, 2997-3000.

THOMAS, P.S., (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA 77, 5201.

THOMAS, L., LEDUC, R., THORNE, B.A., SMEEKENS, S.P., STEINER, D.F., and THOMAS, G. (1991). KEX2-like endoproteases PC2 and PC3 accurately cleave a model prohormone in mammalian cells: Evidence for a common core of neuroendocrine processing enzymes. Proc. Natl. Acad. Sci. USA **88**, 5297-5301.

TOMKINSON, B. and JONSSON, A.-K. (1990). Characterization of cDNA for human tripeptidyl peptidase II: The N-terminal part of the enzyme is similar to subtilisin. Biochem. **30**, 168-174.

VAN DE VEN, W.J.M., VOORBERG, J., FONTIJN, R., PANNEKOEK, H., VAN DEN OUWELAND, A.M.W., VAN DUIJNHOVEN, H.L.P., ROEBROEK, A.J.M., and SIEZEN, R.J. (1990). Furin is a subtilisin-like proprotein processing enzyme in higher eukaryotes. Mol. Biol. Rep. **14**, 265-275.

VON HEIJNE, G. (1986). A new method for predicting signal sequence cleavage sites. Nucl. Acids Res. **14**, 4683-4690.

WANG, P.H., DO, Y.S., MACAULAY, L, SHINAGAWA, T., ANDERSON, P.W., BAXTER, J.D., and HSUEH, W.A. (1991). Identification of renal cathepsin B as a human prorenin-processing enzyme. **266**, 12633-12638.

WISE, R.J., BARR, P.J., WONG, P.A., KIEFER, M.C., BRAKE, A.J., and KAUFMAN, R.J. (1990). Expression of a human proprotein processing enzyme: Correct cleavage of the von Willebrand precursor at a paired basic amino acid site. Proc. Natl. Acad. Sci. USA **87**, 9378-9382.

YANISCH-PERRON, C., VIEIRA, L., and MESSING, J. (1985). Improved M13 phage cloning vectors and host strains. Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33, 103-119.

ZAPF, J., KIEFER, M., MERRYWEATHER, J., MASIARZ, F., BAUER, D., BORN, W., FISCHER, J.A., and FROESCH, E.F. (1990). Isolation from adult human serum of four insulin-like growth factor (IGF) binding proteins and molecular cloning of one of them that is increased by IGF1 administration and in extra-pancreatic tumor hypoglycemia. J. Biol. Chem. **265**, 14892-14898.

SUMMARY OF THE INVENTION

Recombinant polynucleotides are provided for the expression of a novel mammalian endopeptidase, PACE 4, and its alternative form, PACE 4.1, which are involved in the production of mature polypeptides from precursor polypeptides by cleavage at pairs of basic amino acids (e.g., -LysArg-, -LysLys-, and -ArgArg-). These recombinant polynucleotides can be used for enhanced intracellular or extracellular production of PACE 4 and 4.1 in transformed cells. Additionally, efficient conversion of co-expressed heterologous polypeptides having processing sites recognized by PACE 4 or 4.1 can produce desired mature forms of those polypeptides.

Purified PACE 4 and 4.1 are also provided. These have utility for commercial and pharmaceutical purposes.

The present invention includes vectors, transformed mammalian, insect, and microorganism cells, methods for producing PACE 4 and 4.1, methods for expressing and secreting a desired mature polypeptide from a heterologous precursor polypeptide using PACE 4 or 4.1, purified PACE 4 and 4.1, a pharmaceutical preparation containing PACE 4 or 4.1, and therapeutic methods involving administration of PACE 4 or 4.1, or inhibitors of PACE 4 and 4.1.

DESCRIPTION OF THE DRAWING

Figure 1 shows the composite cDNA sequence encoding PACE 4 (SEQ ID NO: 1), and the amino acid sequence of the translation product encoded therein.

Figure 2 shows the composite cDNA sequence encoding the alternate splicing region of PACE 4.1 (SEQ ID NO: 3), and the amino acid sequence of the translation product encoded therein. Shown are the additional cDNA sequence encoding 16 amino acids, termination codon, and 3'-untranslated region which comprise the truncated cDNA and encoded protease amino acid sequence of PACE 4.1 (SEQ ID NO: 4). The

sequence of PACE 4.1 is identical to that of PACE 4 through Lys 471. The beginning of the divergence in cDNA sequence is indicated by an arrow.

Figure 3 shows the DNA sequence of the homolog of exon 12 of the furin/PACE gene (SEQ ID NO: 5) (Barr et al., 1991). This sequence contains the intron/exon splice junction for commitment to either PACE 4 (SEQ ID NO: 1) or PACE 4.1 (SEQ ID NO: 1) & (SEQ ID NO: 3) mRNA. This exon sequence corresponds to PACE 4 (SEQ ID NO: 1) residues Ser 473 to Pro 510.

Figure 4 provides a restriction map, clone diagrams, and composite diagram for PACE 4. The composite diagram indicates the amino terminal signal sequence by a shaded box, and the remainder of the coding region by the hatched box. Individual clones are identified above the composite diagram. Restriction endonucleases are abbreviated as follows: B, BamHI; G, BglII; R, EcoRI; Sm, SmaI; Sp, SphI. Each clone has synthetic EcoRI and BglII sites at the 5' and 3' termini.

Figure 5 provides a restriction map, clone diagrams, and composite diagram for PACE 4.1. The composite diagram indicates the amino terminal signal sequence by a shaded box, and the remainder of the coding region by the hatched box. Individual clones are identified above the composite diagram. Restriction endonucleases are abbreviated as follows: B, BamHI; G, BglII; R, EcoRI; Sm, SmaI; Sp, SphI. Each clone has synthetic EcoRI and BglII sites at the 5' and 3' termini.

Figure 6 is a domain map of PACE 4 and 4.1, together with comparative maps for other dibasic endoproteases KEX2, furin/PACE, PC2, and PCI/PC3. Active site Asp, His, Asn (Asp), and Ser residues are indicated as D, H, N (D), and S, respectively. Potential glycosylation sites are indicated by a diamond. Putative signal peptides (Sig), membrane-binding amphipathic helices (AH), and transmembrane domains (TM) are indicated by solid boxes. Cysteine-rich (CRR) and serine/threonine-rich (S/TRR) regions are indicated by hatched boxes. Proprotein regions are indicated by shaded boxes.

Figure 7 shows a comparison among the amino acid sequences of the catalytic domains of PACE 4 (SEQ ID NO: 11) and 4.1 (SEQ ID NO: 11) (human) and other dibasic endoproteases KEX2 (yeast) (SEQ ID NO: 7), furin/PACE (human) (SEQ ID NO: 8), PC2 (human) (SEQ ID NO: 9), and PC1/PC3 (mouse) (SEQ ID NO: 10). Amino acids that are identical among three or more proteases are indicated.

Figure 8 shows Northern blot analysis of PACE 4 and 4.1. The blots in panels A and B contain human poly(A)$^+$ RNAs from various tissues, while those in panels C and D contain poly(A)$^+$ RNAs from human or hamster cultured cell lines. Blots were hybridized with a PACE 4 specific probe (panels A and C) or a PACE 4.1 specific probe (panels B and D).

Figure 9 shows in situ hybridization of cosmid DNA to metaphase chromosomes as revealed by fluorescence of FITC labelled fur and Texas Red labelled PACE 4 gene probes.

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

As used herein, the term "PACE" is an acronym for paired basic amino acid converting enzyme. An enzyme belonging to the PACE family is a subtilisin-like endopeptidase, i.e., a propeptide-cleaving enzyme, which exhibits specificity for cleavage at basic residues of a polypeptide, e.g., -lys-arg-, -arg-arg, or -lys-lys-; is stimulated by calcium ions; and is inhibited by phenylmethyl sulfonyl fluoride (PMSF).

A cDNA that encodes a novel form of PACE, referred to herein as PACE 4, is presented in Figure 1. PACE 4 is derived from an animal cell, more specifically from a human cell. Previous applications, including U.S. Application Numbers 07/620,859, filed November 29, 1990; 07/621,443, filed November 29, 1990; and 07/621,457, filed November 30, 1990, herein incorporated by reference, disclose the isolation of a gene encoding a different form of PACE, also isolated from human cells. The present invention demonstrates that other forms of PACE exist and that they can be produced artifically. Thus, "PACE 4" refers to any of the naturally occurring forms, including the PACE 4 precursor protein shown in Figure 1, and various processed forms, including the mature PACE 4 polypeptide and modified proteins that have the properties of natural-sequence PACE 4. "Natural-sequence PACE 4" refers to the specific molecule having the sequence set forth in Figure 1 and the fragments and derivatives thereof defined in this specification. Polypeptide fragments that maintain the catalytic specificity of that enzyme are encompassed by the general term PACE 4. Additionally, analogs of PACE 4 are included within the definition of PACE 4, and

include truncated polypeptides (including fragments) and PACE 4-like polypeptides, e.g., mutants in which there are variations in the amino acid sequence that retain catalytic activity and preferably have a homology of at least 80%, more preferably 90%, and most preferably 95%, with the corresponding region of the PACE 4 sequence of Figure 1. Typically, such analogs differ by only 1, 2, 3, or 4 amino acids. Examples include polypeptides with minor amino acid variations from the natural amino acid sequence of PACE 4; in particular, conservative amino acid replacements. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are generally divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) non-polar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cystine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar conservative replacement of an amino acid with a structurally related amino acid will not have a major effect on the enzymatic activity, especially if the replacement does not involve an amino acid at the active site of the PACE-like polypeptide. As discussed elsewhere in this specification in detail, a PACE 4 molecule of a variant type, referred to as PACE 4.1, has a sequence partially identical with the sequence of PACE 4, and exhibits similar enzymatic properties. PACE 4.1 is produced by alternate splicing of mRNA, and encompasses the catalytic domain of PACE 4 truncated on the C-terminal side, to which is added an additional 16 amino acids comprising a novel C-terminus.

Utilizing the sequence data in Figure 1, as well as the denoted characteristics of a PACE molecule in general and those of PACE 4 and 4.1 in particular, it is within the skill of the art to obtain other PACE 4 or 4.1 polypeptides, or other DNA sequences encoding PACE 4 or 4.1. For example, the structural gene can be manipulated by varying individual nucleotides, while retaining the correct amino acid(s), or varying the nucleotides, so as to modify the amino acids, without loss of enzymatic activity. Nucleotides can be substituted, inserted, or deleted by known techniques, including, for example, in vitro mutagenesis and primer repair. The structural gene can be truncated

at its 3'-terminus and/or its 5'-terminus while retaining its endopeptidase activity. For example, PACE 4 as encoded in Figure 1 contains a cysteine-rich region (CRR), which it may be desirable to delete. It also may be desirable to remove the region encoding the signal sequence, and/or to replace it with a heterologous sequence. It may also be desirable to ligate a portion of the PACE 4 or 4.1 sequence (particularly that which includes the catalytic domain) to a heterologous coding sequence, and thus to create a fusion peptide with the enzymatic specificity of PACE 4.

In designing such modifications, it is expected that changes to nonconserved regions of the PACE-type structure will have relatively smaller effects on activity, whereas changes in the conserved regions, and particularly in or near the catalytic active site domains D, H, and S, are expected to produce larger effects.

The comparison among catalytic domain sequences of yeast KEX 2, human PACE, human PC2, mouse PC1/PC3, and human PACE 4, given in Figure 7, provides guidance on amino acid substitutions which are compatible with catalytic activity. Amino acid residues which are conserved among PACE 4 and at least three of the four related proteins are important for good catalytic activity and are not expected to be candidates for substitution. A residue which shows conservative variations among PACE 4 and at least two of the four related proteins is expected to be capable of similar conservative substitution of the PACE 4 sequence. Similarly, a residue which varies nonconservatively among PACE 4 and at least three of the four related proteins is expected to be capable of either conservative or nonconservative substitution. When designing substitutions to the PACE 4 or 4.1 sequence, replacement by an amino acid which is found in the comparable aligned position of one of the four related proteins is especially preferred.

In addition to the above, other open reading frames (ORFs) or structural genes encoding PACE 4 or 4.1 can be obtained and/or created from cDNA libraries from other animal cell sources.

As used herein, the term "polypeptide" refers to a polymer of amino acids and does not refer to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for

example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids) and polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The term "precursor polypeptide" denotes an expressed polypeptide which normally undergoes one or more posttranslational proteolytic cleavages to yield a biologically active mature polypeptide. Included within the term "precursor polypeptide" are "prepropolypeptides" and "propolypeptides."

A "prepeptide" is the portion of a precursor polypeptide which is removed by "signal peptidase" cleavage during translocation of the polypeptide into the endoplasmic reticulum. The "prepeptide" region is usually at or near the amino terminus.

A "propeptide" is the portion of a precursor polypeptide which is removed by a "propolypeptide convertase" or "endopeptidase" (for example, Kex2 and PACE) during the maturation process of the polypeptide. Many proteins, such as plasma proteins, hormones, neuropeptides, and growth factors, are translated with an additional "propeptide" region located to the carboxy side of the prepeptide region. After cleavage of the prepeptide, the "propeptide" segment is cleaved by a site-specific endopeptidase contributing to the maturation of the polypeptide. A "mature" form of a polypeptide has had a prepeptide and/or propeptide region removed.

A polypeptide or amino acid sequence "derived from" a designated nucleic acid sequence refers to a polypeptide having an amino acid sequence identical to that of a polypeptide encoded in the sequence, or a portion thereof wherein the portion consists of at least 3-5 amino acids, and more preferably at least 8-10 amino acids, and even more preferably at least 11-15 amino acids, or which is immunologically identifiable with a polypeptide encoded in the sequence. This terminology also includes a polypeptide expressed from a designated nucleic acid sequence.

A polypeptide of the invention is not necessarily translated from a designated nucleic acid sequence, for example, the sequence in Figure 1. It can be generated in any manner, including, for example, chemical synthesis, expression of a recombinant expression system, or isolation from a cell. A recombinant or derivative polypeptide of the invention can include one or more analogs of amino acids or unnatural amino acids in its sequence. Methods of inserting analogs of amino acids into a sequence are

known in the art. It also can include one or more labels, which are known to those of skill in the art.

The term "recombinant polynucleotide" as used herein intends a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation (1) is not associated with all or a portion of a polynucleotide with which it is associated in nature, (2) is linked to a polynucleotide other than that to which it is linked in nature, or (3) does not occur in nature.

The term "polynucleotide" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA and RNA. It also includes known types of modifications, for example, labels which are known in the art, methylation, "caps," substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example proteins (including for e.g., nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide.

A "replicon" is any genetic element, e.g., a plasmid, a chromosome, a virus, a cosmid, etc. that behaves as an autonomous unit of polynucleotide replication within a cell; i.e., capable of replication under its own control. This can include selectable markers.

A "vector" is a replicon in which a recombinant polynucleotide segment is attached, so as to bring about the replication and/or expression of the attached segment.

"Control sequence" refers to polynucleotide sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence; in eukaryotes, generally, such control sequences include

promoters and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is necessary for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

An "open reading frame" (ORF) is a region of a polynucleotide sequence which encodes a polypeptide; this region can represent a portion of a coding sequence or a total coding sequence.

A "coding sequence" is a polynucleotide sequence which is translated into a polypeptide, usually via mRNA, when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to, cDNA, and recombinant polynucleotide sequences.

"PCR" refers to the technique of polymerase chain reaction as described in Saiki, et al., Nature 324:163 (1986); U.S. Patent No. 4,683,195; and U.S. Patent No. 4,683,202.

As used herein, x is "heterologous" with respect to y if x is not naturally associated with y in the identical manner; i.e., x is not associated with y in nature or x is not associated with y in the same manner as is found in nature.

"Recombinant host cells," "host cells," "cells," "cell lines," "cell cultures," and other such terms denoting mammalian, insect, or microorganism cells that can be, or have been, used as recipients for a recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transformed. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

"Mammalian cells" are cells that are from a member of the Class Mammalia, and microorganism cells and insect cells are specifically excluded from this group.

Insect cells and compatible vectors which are useful as recombinant expression systems are known in the art, and include, for example, insect expression and transfer vectors derived from the baculovirus Autographa californica nuclear polyhedrosis virus (hereinafter "AcNPV" or "baculovirus"), which is a helper-independent, viral expression vector. Viral expression vectors derived from this system usually use the strong viral polyhedrin gene promoter to drive expression of heterologous genes.

As used herein, the term "microorganism" includes prokaryotic and eukaryotic microbial species such as bacteria and fungi, the latter including yeast and filamentous fungi; the term "microorganism" specifically excludes mammalian cells and insect cells.

"Transformation," as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, for example, direct uptake, transfection, transduction, infection, or electroporation. The exogenous polynucleotide can be maintained as a non-integrated vector, for example, a plasmid, or alternatively, can be integrated into the host genome.

"Substantially pure" as used herein means a level of purity which enables the practice of the appropriate function. With respect to PACE 4 or 4.1 polypeptides, these are in general substantially pure if the levels of competing and interfering proteolytic enzymes are low enough to permit the activity of PACE 4 or 4.1 to generate the major reaction products. Preferably this means that the PACE 4 or 4.1 will account for at least 90% of the endoprotease activity, more preferably at least 95% of endoprotease activity, and most preferably at least 99% of endoprotease activity.

As used herein, an "oligonucleotide" is a polynucleotide containing from 10 to approximately 200 bases, usually less than about 60 bases, and which is useful inter alia as a probe or diagnostic reagent for identifying the presence of a complementary nucleotide sequence.

II. General modes for carrying out the invention

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and

immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Sambrook, et al., MOLECULAR CLONING; A LABORATORY MANUAL, SECOND EDITION (1989); DNA CLONING, VOLUMES I AND II (D.N Glover ed. 1985); OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait ed, 1984); NUCLEIC ACID HYBRIDIZATION (B.D. Hames & S.J. Higgins eds. 1984); TRANSCRIPTION AND TRANSLATION (B.D. Hames & S.J. Higgins eds. 1984); ANIMAL CELL CULTURE (R.I. Freshney ed. 1986); IMMOBILIZED CELLS AND ENZYMES (IRL Press, 1986); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR CLONING (1984); the series, METHODS IN ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory), Methods in Enzymology Vol. 154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively), Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London), Scopes, (1987), PROTEIN PURIFICATION: PRINCIPLES AND PRACTICE, Second Edition (Springer-Verlag, N.Y.), and HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, VOLUMES I-IV (D.M. Weir and C. C. Blackwell eds 1986). All patents, patent applications, and publications mentioned herein, both supra and infra, are hereby incorporated herein by reference.

III. Specific modes for carrying out the invention

DNA sequences, constructs (vectors) containing the DNA sequences, and mammalian, insect, and microorganism hosts containing these constructs are provided for herein. These compositions have utility in processes for producing PACE 4 and 4.1 and for producing mature polypeptides from heterologous precursor polypeptides. Also provided are compositions comprising substantially pure PACE 4 or 4.1, and pharmaceutical compositions comprising PACE 4 or 4.1. Additionally, methods for producing PACE 4 and 4.1, methods for inhibiting blood coagulation and for reducing blood pressure, and also methods for producing a desired mature polypeptide are included in the present invention.

Methods for producing a desired mature polypeptide can include the following techniques. First, a vector coding for both PACE 4 or 4.1 and the heterologous

precursor polypeptide can be inserted into a host cell, or two vectors coding, respectively, for PACE 4 or 4.1 and the heterologous precursor polypeptide, can be inserted into a host. Furthermore, two transformed hosts can be employed wherein one host expresses PACE 4 or 4.1 and the other host expresses the heterologous precursor polypeptide; these hosts can be co-cultured whereby PACE processes the heterologous precursor into the mature polypeptide in the medium.

Another method employs a single transformed host cell expressing PACE 4 or 4.1, whereby PACE 4 or 4.1 is isolated from the cell and can be used in in vitro processing of a heterologous precursor polypeptide. The PACE 4 or 4.1 can be immobilized by, e.g., packing into or attaching to a column or other useful support configuration. Soluble or immobilized recombinant PACE 4 or 4.1 can be used as an added reagent to extracellular (or conditioned) media where a precursor product is secreted from the cell in which it is expressed.

In some instances, it may be desirable to have a plurality of copies, two or more, of the gene expressing the expression product precursor in relation to the PACE 4 or 4.1 gene, or vice versa. This can be achieved in a variety of ways. For example, one can use separate vectors, where the vector having either PACE 4/4.1 or the heterologous precursor polypeptide has a higher copy number than the other vector. In this situation, it would be desirable to have different markers on the two vectors, so as to ensure the continued maintenance of the vectors in the mammalian host. Alternatively, one could employ two transcriptional regulatory regions having different rates of transcriptional initiation, providing for enhanced expression of either the PACE 4 or 4.1 gene or the heterologous precursor polypeptide relative to the other gene. As another alternative, one can use different promoters, where one promoter provides for a low level of constitutive expression of either PACE or the heterologous precursor polypeptide, while the second promoter provides for a high level of induced expression of the other product.

Examples of precursor polypeptides include, but are not limited to, transforming growth factor (TGF) beta and its superfamily, including inhibin and activin; bone morphogenic proteins (BMP); insulin and relaxin; coagulation factors, such as von Willebrand factor; growth factors, such as platelet derived growth factor (PDGF) and nerve growth factor (NGF); and virus polypeptides, including those from

cytomegalovirus (CMV), hepatitis delta virus (HDV), hepatitis C virus (HCV), human immunodeficiency virus (HIV), and herpes simplex virus (HSV). Any precursor polypeptide with at least one dibasic cleavage site is a candidate for this aspect of the present invention.

The PACE 4 or 4.1 gene or a fragment thereof can be expressed in a mammalian, insect, or microorganism host. The polynucleotide encoding PACE is inserted into a suitable expression vector compatible with the type of host cell employed and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Site-specific DNA cleavage involved in such construction is performed by treating with suitable restriction enzymes under conditions which generally are specified by the manufacturer of these commercially available enzymes.

A suitable expression vector is one that is compatible with the desired function (e.g., transient expression, long term expression, integration, replication, amplification) and in which the control elements are compatible with the host cell.

Expression in mammalian cells

Vectors suitable for replication in mammalian cells are known in the art, and can include viral replicons, or sequences that ensure integration of the sequence encoding PACE into the host genome. Suitable vectors can include, for example, those derived from simian virus SV40, retroviruses, bovine papilloma virus, vaccinia virus, and adenovirus.

A suitable vector, for example, is one derived from vaccinia viruses. In this case, the heterologous DNA is inserted into the vaccinia genome. Techniques for the insertion of foreign DNA into the vaccinia virus genome are known in the art, and utilize, for example, homologous recombination. The insertion of the heterologous DNA is generally into a gene which is non-essential in nature, for example, the thymidine kinase gene (tk), which also provides a selectable marker. Plasmid shuttle vectors that greatly facilitate the construction of recombinant viruses have been described (see, for example, Mackett et al. (1984), Chakrabarti et al. (1985); Moss (1987)). Expression of the heterologous polypeptide then occurs in cells or individuals which are immunized with the live recombinant vaccinia virus.

Such suitable mammalian expression vectors usually contain one or more eukaryotic transcription units that are capable of expression in mammalian cells. The transcription unit is comprised of at least a promoter element to mediate transcription of foreign DNA sequences. Suitable promoters for mammalian cells are known in the art and include viral promters such as that from simian virus 40 (SV40), cytomegalovirus (CMV), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV).

The optional presence of an enhancer element (enhancer), combined with the promoter elements described above, will typically increase expression levels. An enhancer is any regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to endogenous or heterologous promoters, with synthesis beginning at the normal mRNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) Science 236:1237; Alberts et al. (1989) Molecular Biology of the Cell, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they typically have a broader host range. Examples useful in mammalian cells include the SV40 early gene enhancer [Dijkema et al (1985) EMBO J. 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) Proc. Natl. Acad. Sci. 79:6777] and from human cytomegalovirus [Boshart et al. (1985) Cell 41:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) Trends Genet. 2:215; Maniatis et al. (1987) Science 236:1237].

In addition, the transcription unit can also be comprised of a termination sequence and poly(A) addition sequences which are operably linked to the PACE 4 or 4.1 coding sequence. The transcription unit can also be comprised of an enhancer sequence which increases the expression of PACE 4 or 4.1.

Sequences that cause amplification of the gene may also be desirable, as are sequences which encode selectable markers. Selectable markers for mammalian cells are known in the art, and include for example, thymidine kinase, dihydrofolate reductase (together with methotraxate as a DHFR amplifier), aminoglycoside

phosphotransferase, hygromycin B phosphotransferase, asparagine synthetase, adenosine deaminase, metallothionien, and antibiotic resistant genes such as neomycin.

The vector that encodes PACE can be used for transformation of a suitable mammalian host cell. Transformation can be by any known method for introducing polynucleotides into a host cell, including, for example packaging the polynucleotide in a virus and transducing a host cell with the virus or by tranfection procedures known in the art, as exemplified by U.S. Patent Nos. 4,399,216; 4,912,040; 4,740,461; 4,959,455 (these patents are incorporated herein by reference). The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (e.g., Hep G2), and a number of other cell lines.

Expression in Insect Cells

In the case of expression in insect cells, generally the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site

32 basepairs downstream from the ATT; see Luckow and Summers, Virology (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) Ann. Rev. Microbiol., 42:177) and a procaryotic ampicillin-resistance (amp) gene and origin of replication for selection and propagation in E. coli.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region typically includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector can also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression can be either regulated or constitutive.

Expression in Microorganisms

Fungal expression systems can utilize both yeast and filamentous fungi hosts. Examples of filamentous fungi expression systems are Aspergillus, as described in EPO Pub. No. 357 127 (published March 7, 1990), and Acremonium chrysogenum, described in EPO Pub. No. 376 266 (published July 4, 1990).

A yeast expression system can typically include one or more of the following: a promoter sequence, fusion partner sequence, leader sequence, transcription termination sequence. These elements can be combined into an expression cassette, which can be maintained in a replicon, preferably with a selectable marker.

A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region typically includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter can also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible)

expression. Constitutive expression occurs in the absence of a UAS. Regulated expression can be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH)(E.P.O. Pub. No. 284044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK)(E.P.O. Pub. No. 329203). The yeast PHO5 gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara et al. (1983) Proc. Natl. Acad. Sci. USA 80:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter can be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (U.S. Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the ADH2, GAL4, GAL10, OR PHO5 genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK(E.P.O. Pub. No. 164556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, inter alia, [Cohen et al. (1980) Proc. Natl. Acad. Sci. USA 77:1078; Henikoff et al. (1981) Nature 283:835; Hollenberg et al. (1981) Curr. Topics Microbiol. Immunol. 96:119; Hollenberg et al. (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast Saccharomyces cerevisiae," in: Plasmids of Medical, Environmental and Commercial Importance (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon et al. (1980) Gene 11:163; Panthier et al. (1980) Curr. Genet. 2:109;].

The PACE 4 gene or a fragment thereof can be expressed intracellularly in yeast. A promoter sequence can be directly linked with the PACE 4 gene or fragment, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired,

methionine at the N-terminus can be cleaved from the protein by in vitro incubation with cyanogen bromide.

Intracellularly expressed fusion proteins provide an alternative to direct expression of the PACE gene or fragment. Typically, a DNA sequence encoding the N-terminal portion of a stable protein, a fusion partner, is fused to the 5' end of heterologous DNA encoding the desired polypeptide. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of the PACE 4 gene or fragment thereof and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See e.g., E.P.O. Pub. No. 196056. Another example is a ubiquitin fusion protein. Such a ubiquitin fusion protein preferably retains a site for a processing enzyme (e.g. ubiquitin-specific processing protease) to cleave the ubiquitin from the PACE 4 or 4.1 polypeptide. Through this method, therefore, a mature PACE 4 or 4.1 polypeptide can be isolated (P.C.T. WO 88/024066).

Alternatively, PACE 4 or 4.1 polypeptides can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion in yeast of the PACE 4 or 4.1 polypeptides. Preferably, there are processing sites encoded between the leader fragment and the PACE gene or fragment thereof that can be cleaved either in vivo or in vitro. The leader sequence fragment typically encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (E.P.O. Pub. No. 12873; J.P.O. Pub. No. 62,096,086) and the A-factor gene (U.S. Patent No. 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (E.P.O. Pub. No. 60057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated

alpha-factor leaders (typically about 25 to about 50 amino acid residues) (U.S. Patent Nos. 4,546,083 and 4,870,008; E.P.O. Pub. No. 324274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (See e.g., P.C.T. WO 89/02463.)

Typically, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Typically, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs or cassettes are often maintained in a replicon, such as an extrachromosomal element (e.g., plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon can have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a procaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein et al. (1979) Gene 8:17-24], pCl/1 [Brake et al. (1984) Proc. Natl. Acad. Sci USA 81:4642-4646], and YRp17 [Stinchcomb et al. (1982) J. Mol. Biol. 158:157]. In addition, a replicon can be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and typically about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Either a high or low copy number vector may be selected, depending upon the effect on the host of the vector and the PACE polypeptides. See e.g., Brake et al., supra.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors typically contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the

yeast chromosome [Orr-Weaver et al. (1983) Methods in Enzymol. 101:228-245]. An integrating vector can be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver et al., supra. One or more expression construct can integrate, possibly affecting levels of recombinant protein produced [Rine et al. (1983) Proc. Natl. Acad. Sci. USA 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Typically, extrachromosomal and integrating expression vectors can contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers can include biosynthetic genes that can be expressed in the yeast host, such as ADE2, HIS4, LEU2, TRP1, and ALG7, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker can also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of CUP1 allows yeast to grow in the presence of copper ions [Butt et al. (1987) Microbiol. Rev. 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are typically comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, inter alia, the following yeasts: Candida albicans [Kurtz, et al. (1986) Mol. Cell. Biol. 6:142], Candida maltosa [Kunze, et al. 91985) J. Basic Microbiol. 25:141]. Hansenula polymorpha [Gleeson, et al. (1986) J. Gen. Microbiol. 132:3459; Roggenkamp et al. (1986) Mol. Gen. Genet. 202:302], Kluyveromyces fragilis [Das, et al. (1984) J. Bacteriol. 158:1165], Kluyveromyces lactis [De Louvencourt et al. (1983) J. Bacteriol. 154:737; Van den Berg et al. (1990) Bio/Technology 8:135], Pichia guillerimondii [Kunze et al. (1985) J.

Basic Microbiol. 25:141], Pichia pastoris [Cregg, et al. (1985) Mol. Cell. Biol. 5:3376; U.S. Patent Nos. 4,837,148 and 4,929,555], Saccharomyces cerevisiae [Hinnen et al. (1978) Proc. Natl. Acad. Sci. USA 75:1929; Ito et al. (1983) J. Bacteriol. 153:163], Schizosaccharomyces pombe [Beach and Nurse (1981) Nature 300:706], and Yarrowia lipolytica [Davidow, et al. (1985) Curr. Genet. 10:380471 Gaillardin, et al. (1985) Curr. Genet. 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and typically include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See e.g., [Kurtz et al. (1986) Mol. Cell. Biol. 6:142; Kunze et al. (1985) J. Basic Microbiol. 25:141; Candida]; [Gleeson et al. (1986) J. Gen. Microbiol. 132:3459; Roggenkamp et al. (1986) Mol. Gen. Genet. 202:302; Hansenula]; [Das et al. (1984) J. Bacteriol. 158:1165; De Louvencourt et al. (1983) J. Bacteriol. 154:1165; Van den Berg et al. (1990) Bio/Technology 8:135; Kluyveromyces]; [Cregg et al. (1985) Mol. Cell. Biol. 5:3376; Kunze et al. (1985) J. Basic Microbiol. 25:141; U.S. Patent Nos. 4,837,148 and 4,929,555, Pichia]; [Hinnen et al. (1978) Proc. Natl. Acad. Sci. USA 75;1929; Ito et al. (1983) J. Bacteriol. 153:163 Saccharomyces]; [Beach and Nurse (1981) Nature 300:706; Schizosaccharomyces]; [Davidow et al. (1985) Curr. Genet. 10:39; Gaillardin et al. (1985) Curr. Genet. 10:49; Yarrowia].

Additionally, the PACE 4 gene or a fragment thereof can be expressed in a bacterial system. Therein, a bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3") transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region typically includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter can also have a second domain called an operator, that can overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein can bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression can occur in the absence of negative regulatory elements, such as the operator. In

addition, positive regulation can be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in Escherichia coli (E. coli) [Raibaud et al. (1984) Annu. Rev. Genet. 18:173]. Regulated expression can therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (lac) [Chang et al. (1977) Nature 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (trp) [Goeddel et al. (1980) Nuc. Acids Res. 8:4057; Yelverton et al. (1981) Nucl. Acids Res. 9:731; U.S. Patent No. 4,738,921; E.P.O. Pub. Nos. 36,776 and 121,775]. The ■-lactomase (bla) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In Interferon 3 (ed. I. Gresser)], bacteriophage lambda PL [Shimatake et al. (1981) Nature 292:128] and T5 [U.S. Patent No. 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter can be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [U.S. Patent No. 4,551,433]. For example, the tac promoter is a hybrid trp-lac promoter comprised of both trp promoter and lac operon sequences that is regulated by the lac repressor [Amann et al. (1983) Gene 25:167; de Boer et al. (1983) Proc. Natl. Acad. Sci. 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier et al. (1986) J. Mol. Biol. 189:113; Tabor et al. (1985) Proc Natl. Acad. Sci. 82:1074]. In addition, a hybrid

promoter can also be comprised of a bacteriophage promoter and an E. coli operator region (E.P.O. Pub. No. 267,851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of the PACE gene or fragment thereof in prokaryotes. In E. coli, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine et al. (1975) Nature 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' and of E. coli 16S rRNA [Steitz et al. (1979) "Genetic signals and nucleotide sequences in messenger RNA." In Biological Regulation and Development: Gene Expression (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook et al. (1989) "Expression of cloned genes in Escherichia coli." In Molecular Cloning: A Laboratory Manual].

PACE 4 or 4.1 can be expressed intracellularly. A promoter sequence can be directly linked with the PACE 4 gene or a fragment thereof, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus can be cleaved from the protein by in vitro incubation with cyanogen bromide or by either in vivo on in vitro incubation with a bacterial methionine N-terminal peptidase (E.P.O. Pub. No. 219,237).

Fusion proteins provide an alternative to direct expression. Typically, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous PACE 4 or 4.1 coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of the PACE gene or fragment thereof and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the PACE gene or fragment thereof [Nagai et al. (1984) Nature 309:810]. Fusion proteins can also be made with sequences from the lacZ [Jia et al. (1987) Gene 60:197], trpE [Allen et al. (1987) J. Biotechnol. 5:93; Makoff et al. (1989) J. Gen. Microbiol. 135:11], and Chey [E.P.O. Pub. No. 324,647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not

encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (e.g. ubiquitin specific processing-protease) to cleave the ubiquitin from the PACE polypeptide. Through this method, mature PACE polypeptides can be isolated [Miller et al. (1989) Bio/Technology 7:698].

Alternatively, PACE 4 or 4.1 polypeptides can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the PACE 4 or 4.1 polypeptides in bacteria [U.S. Patent No. 4,336,336]. The signal sequence fragment typically encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic spece, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either in vivo or in vitro encoded between the signal peptide fragment and the PACE 4 or 4.1 polypeptide.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the E. coli outer membrane protein gene (ompA) [Masui et al. (1983), in: Experimental Manipulation of Gene Expression; Ghrayeb et al. (1984) EMBO J. 3:2437] and the E. coli alkaline phosphatase signal sequence (phoA) [Oka et al. (1985) Proc. Natl. Acad. Sci. 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various Bacilus strains can be used to secrete heterologous proteins from B. subtilis [Palva et al. (1982) Proc. Natl. Acad. Sci. USA 79:5582; E.P.O. Pub. No. 244,042].

Typically, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the trp gene in E. coli as well as other biosynthetic genes.

Typically, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g., plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a procaryotic host either for expression or for cloning and amplification. In addition, a replicon can be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and typically about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector can be selected, depending upon the effect of the vector and the PACE 4 or 4.1 polypeptide on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors typically contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bactedrial chromosome. For example, integrating vectors constructed with DNA from various Bacillus strains integrate into the Bacillus chromosome (E.P.O. Pub. No. 127,328). Integrating vectors can also be comprised of bacteriophage or transposon sequences.

Typically, extrachromosomal and integrating expression constructs can contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and can include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies et al. (1978) Annu. Rev.Microbiol. 32:469]. Selectable markers can also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are typically comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, inter alia, the following bacteria: Bacillus subtilis [Palva et al. (1982) Proc. Natl. Acad. Sci. USA 79:5582; E.P.O. Pub. Nos. 36,259 and 63,953; P.C.T. WO 84/04541], Escherichia coli [Shimatake et al. (1981) Nature 292:128; Amann et al. (1985) Gene 40:183; Studier et al. (1986) J. Mol. Biol. 189:113; E.P.O. Pub. Nos. 36,776, 136,829 and 136,907; U.K. Patent Application Serial No. 8418273], Streptococcus cremoris [Powell et al. (1988) Appl. Environ. Microbiol. 54:655] Streptococcus lividans [Powell et al. (1988) Appl. Environ. Microbiol. 54:655], Streptomyces lividans [U.S. Patent No. 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and typically include either the transformation of bacteria treated with $CaCl_2$ or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See e.g., [Masson et al. (1989) FEMS Microbiol. Lett. 60:273; Palva et al. (1982) Proc. Natl. Acad. Sci. USA 79:5582; E.P.O. Pub. Nos. 36,259 and 63,953; P.C.T. WO 84/04541, Bacillus], [Miller et al. (1988) Proc. Natl. Acad. Sci. 85:856; Wang et al. (1990) J. Bacteriol. 172:949, Campylobacter], [Cohen et al. (1973) Proc. Natl. Acad. Sci. 69:2110; Dower et al. (1988) Nucleic Acids Res. 16:6127; Kushner (1978) "An improved method for transformation of Escherichia coli with ColE1-derived plasmids. In Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering (eds. H.W. Boyer and S. Nicosia); Mandel et al. (1970) J. Mol. Biol. 53:159; Taketo (1988) Biochim. Biophys. Acta 949:318; Escherichia], [Chassy et al. (1987) FEMS Microbiol. Lett. 44:173 Lactobacillus]; [Fiedler et al. (1988) Anal. Biochem 170:38, Pseudomonas]; [Augustin et al. (1990) FEMS Microbiol. Lett. 66:203, Staphylococcus], [Barany et al. (1980) J. Bacteriol. 144:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation," in: Streptococcal Genetics (ed. J. Ferretti and R. Curtiss III); Perry et al. (1981) Infec. Immun. 32:1295; Powell et al. (1988) Appl. Environ. Microbiol. 54:655; Somkuti et al. (1987) Proc. 4th Evr. Cong. Biotechnology 1:412, Streptococcus].

Expression and Detection of expressed PACE 4 or 4.1

In order to obtain PACE 4 or 4.1 expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant PACE encoding sequence. These conditions will vary, depending upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

Detection of PACE 4 or 4.1 expressed in the transformed host cell can be accomplished by several methods. For example, detection can be by enzymatic activity (or increased enzymatic activity or increased longevity of enzymatic activity) using fluorogenic substrates which are comprised of a dibasic cleavage site for which PACE is specific. PACE can also be detected by its immunological reactivity with anti-PACE antibodies.

Therapeutic aspects of PACE 4 or 4.1

PACE 4 and 4.1 mRNAs are expressed at relatively high levels in the kidney, as demonstrated by northern blots (Figure 8). Since physiological conversion of prorenin to renin occurs in the kidney, PACE 4 and 4.1 are candidates for in vivo prorenin processing enzymes. Furthermore, PACE 4 and 4.1 convert prorenin to renin, a process which results in elevation of blood pressure. Blockers of PACE 4 and 4.1 enzymatic activity therefore limit blood pressure increase. PACE 4 and 4.1 inhibitors have therapeutic utility in treating high blood pressure. PACE 4 and 4.1 appear likely to have physiological effects on endogenous proteins that possess dibasic cleavage sites. Any such effects can be exploited for therapeutic purposes.

Oligonucleotide probes

Probes based upon the PACE 4 gene sequence or its complementary sequence are useful for detecting the presence of nucleotide sequences encoding PACE 4 or related polypeptides. Such probes can be used for example to detect PACE 4 mRNA transcription in tissues, or to identify the PACE 4 encoding bands on gels, or to identify structurally related homologous proteins. Since PACE 4 and 4.1 are known to share regions of homology with other dibasic cleaving endoproteases such as KEX2, PC1/PC3, PC2, and furin/PACE, probes based upon regions of the PACE 4 sequence

which are not shared with these other endoproteases are especially useful in discriminating PACE 4 related sequences from the other endoproteases.

A comparison of the amino acid sequence of PACE 4 with these other endoproteases is given in Figure 7. Within the catalytic domain, the regions from amino acids 164 to 168, 172 to 189, 239 to 243, and 397 to 404 are seen to be unique to PACE 4. Other unique sequences exist in the noncatalytic domain of PACE 4 and 4.1. PACE 4 contains a large cysteine-rich region (CRR) within its C-terminal region which is unique to this protease, although a smaller CRR exists within furin/PACE. PACE 4 and 4.1 also have extended signal peptide sequences, compared to the other proteases in the family. In addition, PACE 4 and 4.1 have neither the classical transmembrane anchor domain of furin/PACE, nor the amphipathic helix (AH) surface membrane anchor proposed for PC 2 and PC1/PC3. Thus significant sequence differences from other PACE family proteases exist in the noncatalytic domains of PACE 4 and 4.1.

Probes based upon the nucleotide sequences corresponding to the above unique amino acid sequences are expected to be selective for PACE 4 and 4.1 and previously unknown homologues. A probe based on at least 15, and more preferably at least 20 consecutive nucleotides from the unique PACE 4 sequences and optionally in part upon the common endoprotease conserved regions, or alternatively the simultaneous or consecutive use of two probes based upon the unique and conserved regions respectively, permits identification of PACE 4 related endoproteases which are distinct from those identifiable by probes based only on sequences shared by PACE 4 and the known related protease sequences of Figure 7.

In light of the present disclosure, numerous embodiments within the scope of the claims will be apparent to those of ordinary skill in the art. The following experimental section is intended to be merely illustrative and does not limit the present scope in any way.

Example 1

PCR Identification of PACE 4$^+$ cDNA Clones

To identify and clone new members of the human subtilisin-like protease gene family, degenerate PCR primers were designed and synthesized corresponding to stretches of amino acids in the highly conserved catalytic domains of yeast KEX2 (SEQ ID NO: 7), human PACE (SEQ ID NO: 8), human PC2 (SEQ ID NO: 9), and mouse PC1/PC3 (SEQ ID NO: 10). These included regions containing the active site Asp (DDGI) and the active site His (HGTRC).

Oligonucleotide synthesis

Oligonucleotide adaptors, probes, and primers were synthesized as previously described (Kiefer et al., 1991). The two consensus PCR primers used to identify PACE 4 (SEQ ID NO: 1) were a sense primer consisting of a mixture of 16 23-mers (5'-AGATCTGAATTCGA$^{(C}/_{T)}$GA$^{(C}/_{T)}$GGXAT-3') (SEQ ID NO: 12) and an antisense primer consisting of a mixture of 64 27-mers (5'-AGATCTAAGCTTACAC-C$(^{G}/_{T)}$XGTXCC$(^{A}/_{G)}$TG-3') (SEQ ID NO: 13), where X denotes all four deoxynucleotides. The PACE 4 probes used to screen the Ost4 cDNA library were a 24mer, (5'-TATTCATTGCCGTTCACGTCGTAG-3') (SEQ ID NO: 14) and a 22mer (5'-GCTGGCATCATATCGTGGAGAT-3') (SEQ ID NO: 15) that were derived from the sequence complementary to the PACE 4 PCR product (see below).

Polymerase chain reactions (PCR)

PCR was performed as described (Saiki et al., 1985) with modifications (Kiefer et al., 1991). The reactions were performed using the PCR primers described above at 8 mM and human osteosarcoma, liver or kidney (293) cDNA at 0.1-1 ng/ml. PCR products migrating between 120-170 base pairs on a 7% acrylamide gel were excised, subcloned into M13 and sequenced.

Construction and screening of cDNA and genomic DNA libraries

A 293 cDNA library was constructed in ZAPII as described by Zapf et al. (1990). 6 x 10^7 independent recombinant clones were obtained. The HepG2 cDNA library (Zapf et al., 1990) and the human osteosarcoma (Ost4) cDNA library (Kiefer et

al., 1991) have been described previously. An oriented cDNA library was constructed in the yeast expression vector pAB23BXN using poly $(A)^+$ mRNA isolated from the human liver cell line HEPG2. pAB23BXN is a derivative of pAB23BX (D. Schild et al. (1990), Proc. Natl. Acad. Sci. USA 87:2916) into which a synthetic polylinker, that contained Bst X1 and Not 1 sites, was inserted for unidirectional cDNA cloning.

Approximately 300,000 recombinant phage from the Ost4 library were screened as described (Kiefer et al., 1991) using the two PACE 4 oligonucleotide probes. The HepG2 and 293 cDNA libraries were plated as above and screened with a ≈1.0 Kb BglII DNA fragment derived from the 5' end of the Ost 4 cDNA clone, Os-1.

The 293 cDNA library was subsequently further screened with a ≈500 bp EcoRI fragment derived from the 3' end of the 293 cDNA clone, K-15, and a synthetic oligonucleotide (PACE2.115) derived from a sequence complementary to the 5' end of PACE 4 cDNA (SEQ ID NO: 1) and including the initiation codon (5'-GGAGGCATAGCGGCGAC-3') (SEQ ID NO: 16). The cDNA probes were labeled as described (Feinberg and Vogelstein, 1984) and were hybridized to filters under conditions described above for the oligonucleotide probes except that the hybridization solution contained 40% formamide. The filters were washed twice for 20 min,. in 0.2 x SSC, 0.1% SDS at 65°C.

To obtain the human PACE 4 gene (SEQ ID NO: 1), 600,000 colonies from a human genomic DNA library cloned in the cosmid vector pWE15 (Stratagene, San Diego, CA) were plated and screened with a ≈1.0 Kb BglII DNA fragment from the 5' end of the Os-1 cDNA clone and subsequently with a ≈1.1 kb EcoRI DNA fragment from the 3' end of the L2-1 cDNA clone (Fig. 4). Two overlapping clones of ≈35-40 kb were obtained which hydridized to the 5' probe (PACE4-COS 3) and 3' probe (PACE4-COS 8.1).

Plasmid and cosmid isolation, subcloning and DNA sequencing

Standard procedures for the isolation and manipulation of DNA are from Sambrook et al. (1989). Plasmid DNA was propagated in E. coli strains HB101, D1210 or XL-1 Blue (Stratagene). DNA sequencing was performed by the dideoxy chain termination method (Sanger et al., 1977) using M13 primers as well as specific

internal primers. Ambiguous regions were resolved using 7-deaza-2'-deoxyguanosine-5'- triphosphate (Barr et al., 1986) and Sequenase (U.S. Biochemicals).

After PCR amplification of cDNA reverse transcribed from poly(A)$^+$RNA isolated from human cell lines HEPG2 and 293 and from human osteosarcoma cells, the expected products (circa 150 base pairs) were gel purified, pooled and subcloned into the M13 mp18 DNA sequencing vector. Sequence analysis of 10 PCR products revealed one sequence that was predicted to encode a stretch of 37 amino acids displaying 30-60% sequence identity to yeast KEX2 (SEQ ID NO: 7), human furin/PACE (SEQ ID NO: 8), human PC2 (SEQ ID NO: 9) and mouse PC1/PC3 (SEQ ID NO: 10).

In an attempt to isolate a full length cDNA encoding the new human subtilisin-like protease, an osteosarcoma (Ost4) cDNA library was screened with an oligonucleotide probe derived from the sequence of the 37 amino acid encoding PCR product. One clone (Os-1) was identified, among 300,000 recombinant plaques screened, and was found to contain a 3.6 kilobase (kb) cDNA insert (Fig. 4). Comparison of the deduced amino acid sequence of this clone with the sequence of furin/PACE (SEQ ID NO: 8) revealed that it was missing approximately 200 amino acids from the amino-terminus.

Molecular cloning of PACE 4 (SEQ ID NO: 1) and 4.1 (SEQ ID NO: 1) and (SEQ ID NO: 3) cDNAs and PACE 4 gene (SEQ ID NO: 1)

Subsequently, 300,000 recombinant plaques from a HepG2 cDNA library were probed with a 1.0 kb BglII DNA fragment derived from the 5' end of the Os-1 cDNA clone. This resulted in the isolation of two approximately full length clones with DNA inserts of 4.4 kb (clone L2-1) and 4.3 kb (clone L2-10).

Preliminary Northern blot analysis of several tissues and cell lines using the 1.0 kb Os-1 DNA probe identified a 4.4 kb mRNA transcript in most tissues (described below). In addition, the kidney cell line 293 revealed a mRNA of approximately 2.0 kb. The size of this transcript suggested that an alternative form of the PACE 4 protein (SEQ ID NO: 2) existed since the mRNA size required to encode PACE 4 is 2.9 kb (Fig. 1) (SEQ ID NO: 1). In an attempt to isolate this PACE 4 variant cDNA,

termed PACE 4.1 (SEQ ID NO: 1) and (SEQ ID NO: 3), a 293 cDNA library was constructed and 300,000 recombinant plaques were screened with the 1.0 kb Os-1 DNA probe. Several positive clones were identified including one that contained a cDNA insert of 1.7 kb (clone K-15). Preliminary DNA sequence analysis revealed that it was identical to PACE 4 (SEQ ID NO: 1) at the 5' end but was lacking 270bp encoding the amino-terminal 90 amino acids. DNA sequence analysis also revealed a unique 3' end on PACE 4.1 (SEQ ID NO: 1) and (SEQ ID NO: 3) that was contained almost entirely on a 0.5kb EcoRI DNA fragment (Fig. 5). Therefore, this unique PACE 4.1 EcoRI DNA fragment and an oligonucleotide probe complementary to the 5' end of PACE 4 were used to rescreen the 293 cDNA library. One clone was identified (K-1.1) that hybridized to both probes and was shown to contain a full-length cDNA insert of 1.9 kb.

Example 2

Structures of PACE 4 (SEQ ID NO: 1) and 4.1 (SEQ ID NO: 1) and (SEQ ID NO: 3) deduced from cDNA sequences

The full sequence of the PACE 4 cDNA and the encoded protein sequence (SEQ ID NO: 1) are shown in Figure 1. The encoded protein sequence is shown above that of the cDNA sequence; the numbering is based on the significant ORF in the cDNA. Likely active site residues are indicated by asterisks. Consensus sites for Asn-linked glycosylation are marked by diamonds and cysteine residues by bars. Potential dibasic proteolytic processing sites are indicated by arrows.

The coding sequence of PACE 4 (SEQ ID NO: 1) contains three consensus sites for N-linked glycosylation and fifty-six cysteine residues. The active site, shown in the box in Figure 1, includes a triad of amino acids: aspartate (Asp 205), histidine (His 246), and serine (Ser 420). A cysteine-rich region (CRR) is also present and, as shown in Figure 1, is located in the vicinity of amino acid Cys 695 to amino acid Cys 965.

The translation of PACE 4 (SEQ ID NO: 1) is most likely initiated at the ATG start codon shown (Fig. 1), since this is the only in-frame Met codon in the 5' region of the cDNA. In contrast to the amino-terminal sequence of furin/PACE (SEQ ID NO: 8), this Met is not followed immediately by a classical hydrophobic signal sequence.

However, amino acids 43-63 of the large open reading frame resemble closely the proposed signal sequence for furin/PACE (Barr et al., 1991) (SEQ ID NO: 8), including a predicted signal peptidase cleavage site following Ala63 (Fig. 1, arrowed) (von Heijne, 1986). The open reading frame encodes a PACE 4 precursor protein of 969 amino acids (SEQ ID NO: 1) with a calculated molecular weight of 106.4 kD. As with KEX2 (SEQ ID NO: 7) and its mammalian counterparts, PACE 4 (SEQ ID NO: 1) contains a region of clustered basic residues immediately preceding the catalytic domain (Fig 1). By direct comparison with the known cleavage sites for KEX2 (SEQ ID NO: 7), PC2 (SEQ ID NO: 9) and PC1/PC3 (SEQ ID NO: 10) (Christie et al., 1991; Shennan et al., 1991) the Arg-Val-Lys-Arg motif at amino acids 146-149 most likely represents a propeptide cleavage junction, thereby making Gln150 the likely amino terminus of mature PACE 4 (SEQ ID NO: 1).

The inferred sequence of PACE 4 (SEQ ID NO: 1) contains three consensus sites for N-linked glycosylation and 56 Cys residues, with 44 of these Cys residues clustered in a carboxyl-terminal Cys-rich region analogous to, although somewhat longer than, that of furin/PACE (SEQ ID NO: 8). Again, in contrast to furin/PACE (SEQ ID NO: 8), which has a classical transmembrane domain, PACE 4 (SEQ ID NO: 1) does not appear to possess such a region of hydrophobic amino acid residues for anchorage in cell membranes. Moreover, PACE 4 (SEQ ID NO: 1) does not have a carboxyl-terminal amphipathic helix type anchor sequence proposed for PC2 (SEQ ID NO: 9) and PC1/PC3 (SEQ ID NO: 10) (Smeekens et al., 1991) and carboxypeptidase E (Fricker et al., 1990), thereby suggesting an alternate mode of subcellular localization for this enzyme.

The PACE 4.1 cDNA (SEQ ID NO: 1) and (SEQ ID NO: 3), isolated from the human kidney cell line, encodes a much smaller subtilisin-like protease with a termination codon introduced immediately subsequent to the catalytic domain and leading to a calculated molecular weight for PACE 4.1 (SEQ ID NO: 1) and (SEQ ID NO: 3) of 53.3 kD (Figs. 2B, 3). Truncation at this point removes a region of the molecule referred to as the P-domain. This region is similar to one which was identified in KEX2 (SEQ ID NO: 7), and which has been shown to be essential for the activity of the yeast protease. The PACE 4.1 cDNA (SEQ ID NO: 1) and (SEQ ID NO: 3) also contains a 3-untranslated region that is distinct from that of PACE 4 (SEQ

ID NO: 1) (Fig. 2) (SEQ ID NO: 3), but is encoded by the 5'-cosmid clone PACE4-COS 3. That alternate RNA splicing is occurring in this region can be inferred from the sequences of genomic clones isolated from this cosmid. For example, the sequence of a 620bp PstI fragment from this cosmid encodes an exon (equivalent to exon 12 of furin/PACE (SEQ ID NO: 5)) that clearly illustrates the intron/exon splice junction leading to either PACE 4 (SEQ ID NO: 1) or PACE 4.1 (SEQ ID NO: 1) and (SEQ ID NO: 3) (Fig. 3) (SEQ ID NO: 5).

The unique CRR of PACE 4 (SEQ ID NO: 1) comprises the C-terminal portion of the protein, including 44 cystine residues occurring between Cys695 and Cys965. The CRR apparently provides for localized interactions between PACE 4 (SEQ ID NO: 1) and other cellular components, thereby modifying the activity of the enzyme.

The CRR domain is used to create novel fusion proteins by combining an oligonucleotide encoding the CRR with one encoding a functional domain of a heterologous protein. For example, a fusion protein having the catalytic domain of a different endoprotease and the CRR of PACE 4 (SEQ ID NO: 1) is constructed by joining the appropriate DNA sequence coding for the endoprotease catalytic domain in reading frame with a DNA sequence encoding the CRR, and operatively inserted into an expression vector. The catalytic domain of PC2 (SEQ ID NO: 9) provides proteolytic activity having altered specificity from that of natural PACE 4 (SEQ ID NO: 1). DNA encoding PC2 (SEQ ID NO: 9) is cleaved at a site 5' to the start codon and a site 3' to the region encoding the catalytic domain. PACE 4 clone L2-10 is cleaved with BglII and the C-terminal fragment including the CRR is isolated and ligated to the N-terminal DNA fragment from PC2. This construct is ligated into an expression vector. Expression of the construct in transformed cells yields a novel endoprotease having catalytic characteristics of PC2 and the cellular interactions of PACE 4.

Example 3

Similarity of PACE 4 to other dibasic endoproteases

Comparisons of the sequence and structure of PACE 4 (SEQ ID NO: 11) and PACE 4.1 (SEQ ID NO: 11) with KEX2 (SEQ ID NO: 7) and its mammalian homologs are illustrated in Figs. 6 and 7. In addition to the conservation of spacing of

the catalytic Asp, His, Asn (Asp), and Ser residues, also clearly apparent is the similarity in placement of the proposed processing site for removal of the propeptide region of each protease (Fig. 6). That this cleavage site is utilized has been demonstrated for KEX2 (SEQ ID NO: 7) and for PC2 (SEQ ID NO: 9) and PC1/PC3 (SEQ ID NO: 10) (Christie et al., 1991; Shennan et al., 1991).

Amino acid sequence identities between the catalytic domain of PACE 4 (SEQ ID NO: 11) and those of the previously described dibasic processing enzymes are shown in Fig. 7, and quantitated in Table 1.

TABLE 1. OVERALL % AMINO ACID SEQUENCE IDENTITIES BETWEEN THE CATALYTIC DOMAINS OF SACCHAROMYCES CEREVISIAE KEX2, AND THE MAMMALIAN PAIRED BASIC AMINO ACID RESIDUE CLEAVING ENZYMES

	yKEX2	furin/PACE	PC2	PC1/PC3	PACE4
yKEX2 (SEQ ID NO: 7)	--				
furin/PACE (SEQ ID NO: 8)	48	--			
PC2 (SEQ ID NO: 9)	47	56	--		
PC1/PC3 (SEQ ID NO: 10)	49	65	57	--	
PACE4 (SEQ ID NO: 11)	45	69	53	63	--

Example 4

Tissue distribution of PACE 4 mRNA

To determine the tissue distribution of PACE 4 (SEQ ID NO: 1) and PACE 4.1 (SEQ ID NO: 1) and (SEQ ID NO: 3), northern blot analysis of poly(A)+ RNA from a variety of tissues and cell lines was carried out using cDNA probes specific to each transcript (Fig. 8 A-D).

Northern blot analysis

Multiple tissue Northern (MTN) blots (panels A and B, Figure 8) containing poly(A)+ RNA (2 µg) from human heart (lane 1), brain (lane 2), placenta (lane 3), lung (lane 4), liver (lane 5), skeletal muscle (lane 6), kidney (lane 7) and pancreas (lane 8) were purchased from Clontech (Palo Alto, CA). Poly(A)+ RNA (2 µg) (panels C and D) from hamster insulinoma cell line HIT-15 (lane 9), human embryonic kidney cell line 293 (lane 10), human hepatoma cell line HEP G2 (lane 11), human umbilical vein endothelial cells (lane 12), human monocyte cell line U937 (lane 13), human lymphoblastoid cell line UC-729-6 (lane 14), and human osteosarcoma tissue (lane 15) were fractionated on 1.4% agarose gels in the presence of formaldehyde (Lehrach et al., 1977) transferred directly to nitrocellulose and processed as described (Thomas, 1980). Blots were hybridized with a PACE 4 specific probe (panels A and C) or a PACE 4.1 specific probe (panels B and D). Hybridization and washing conditions were as described above for the screening of cDNA libraries using cDNA probes.

A single PACE 4 4.4 kb transcript was seen at various levels in all tissues and cell lines tested (Fig. 8 A and C). This is similar to the widespread tissue distribution found with furin/PACE (Schalken et al., 1987; Barr et al., 1991) and is in contrast to the limited neuroendocrine tissue distribution of PC2 and PC3 transcripts (Smeekens and Steiner, 1990; Seidah et al., 1990; Smeekens et al., 1991). However, the 2.0 kb PACE 4.1 transcript was seen only in the 293 kidney cell line from which it was originally isolated (Fig. 8 B and D). The lack of the 2.0 kb PACE 4.1 transcript in all tissues tested suggests that it may represent an alternatively spliced PACE 4 transcript limited to certain cell types.

IV. Deposit of biological material

Escherichia coli strain HB101 host cells transformed with pAB23BXN-PACE 4 was deposited on October 11, 1991, with the American Type Culture Collection (ATCC), Rockville, MD, and designated as pAB23BXN-PACE 4 in E. coli. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for purposes of patent procedure. The accession number is available from the ATCC.

This deposit is provided merely as 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 nucleic acid sequence of this plasmid, as well as the amino acid sequence of the polypeptide encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with the description herein. A license may be required to make, use, or sell the deposited material, and no such license is hereby granted.

All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: BARR, PHILIP J
KIEFER, MICHAEL C

(ii) TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND PACE 4.1 GENE AND POLYPEPTIDES IN CELLS

(iii) NUMBER OF SEQUENCES: 16

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
(B) STREET: FIVE PALO ALTO SQUARE
(C) CITY: PALO ALTO
(D) STATE: CALIFORNIA
(E) COUNTRY: USA
(F) ZIP: 94306

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/848,629
(B) FILING DATE: 09-MAR-1992
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: NEELEY PH.D., RICHARD L
(B) REGISTRATION NUMBER: 30092
(C) REFERENCE/DOCKET NUMBER: CHIR-009/00US

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (415) 494-7622
(B) TELEFAX: (415) 857-0663
(C) TELEX: 380816 COOLEY PA

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4403 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (cDNA)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 170..3077

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CGGGAACGCG CCGCGGCCGC CTCCTCCTCC CCGGCTCCCG CCCGCGGCGG TGTTGGCGGC   60

GGCGGTGGCG GCGGCGGCGG CGCTTCCCCG GCGCGGAGCG GCTTTAAAAG GCGGCACTCC  120

ACCCCCCGGC GCACTCGCAG CTCGGGCGCC GCGCGAGCCT GTCGCCGCT ATG CCT      175
                                                          Met Pro
                                                            1

CCG CGC GCG CCG CCT GCG CCC GGG CCC CGG CCG CCG CCC CGG GCC GCC     223
Pro Arg Ala Pro Pro Ala Pro Gly Pro Arg Pro Pro Pro Arg Ala Ala
          5                  10                  15

GCC GCC ACC GAC ACC GCC GCG GGC GCG GGG GGC GCG GGG GGC GCG GGG     271
Ala Ala Thr Asp Thr Ala Ala Gly Ala Gly Gly Ala Gly Gly Ala Gly
     20                  25                  30

GGC GCC GGC GGG CCC GGG TTC CGG CCG CTC GCG CCG CGT CCC TGG CGC     319
Gly Ala Gly Gly Pro Gly Phe Arg Pro Leu Ala Pro Arg Pro Trp Arg
 35                  40                  45                  50

TGG CTG CTG CTG CTG GCG CTG CCT GCC GCC TGC TCC GCG CCC CCG CCG     367
Trp Leu Leu Leu Leu Ala Leu Pro Ala Ala Cys Ser Ala Pro Pro Pro
                55                  60                  65

CGC CCC GTC TAC ACC AAC CAC TGG GCG GTG CAA GTG CTG GGC GGC CCG     415
Arg Pro Val Tyr Thr Asn His Trp Ala Val Gln Val Leu Gly Gly Pro
            70                  75                  80

GCC GAG GCG GAC CGC GTG GCG GCG GCG CAC GGC TAC CTC AAC TTG GGC     463
Ala Glu Ala Asp Arg Val Ala Ala Ala His Gly Tyr Leu Asn Leu Gly
          85                  90                  95

CAG ATT GGA AAC CTG GAA GAT TAC TAC CAT TTT TAT CAC AGC AAA ACC     511
Gln Ile Gly Asn Leu Glu Asp Tyr Tyr His Phe Tyr His Ser Lys Thr
     100                 105                 110

TTT AAA AGA TCA ACC TTG AGT AGC AGA GGC CCT CAC ACC TTC CTC AGA     559
Phe Lys Arg Ser Thr Leu Ser Ser Arg Gly Pro His Thr Phe Leu Arg
115                 120                 125                 130

ATG GAC CCC CAG GTG AAA TGG CTC CAG CAA CAG GAA GTG AAA CGA AGG     607
Met Asp Pro Gln Val Lys Trp Leu Gln Gln Gln Glu Val Lys Arg Arg
                135                 140                 145

GTG AAG AGA CAG GTG CGA AGT GAC CCG CAG GCC CTT TAC TTC AAC GAC     655
Val Lys Arg Gln Val Arg Ser Asp Pro Gln Ala Leu Tyr Phe Asn Asp
            150                 155                 160

CCC ATT TGG TCC AAC ATG TGG TAC CTG CAT TGT GGC GAC AAG AAC AGT     703
Pro Ile Trp Ser Asn Met Trp Tyr Leu His Cys Gly Asp Lys Asn Ser
        165                 170                 175

CGC TGC CGG TCG GAA ATG AAT GTC CAG GCA GCG TGG AAG AGG GGC TAC     751
Arg Cys Arg Ser Glu Met Asn Val Gln Ala Ala Trp Lys Arg Gly Tyr
    180                 185                 190

ACA GGA AAA AAC GTG GTG GTC ACC ATC CTT GAT GAT GGC ATA GAG AGA     799
Thr Gly Lys Asn Val Val Val Thr Ile Leu Asp Asp Gly Ile Glu Arg
195                 200                 205                 210
```

```
AAT CAC CCT GAC CTG GCC CCA AAT TAT GAT TCC TAC GCC AGC TAC GAC     847
Asn His Pro Asp Leu Ala Pro Asn Tyr Asp Ser Tyr Ala Ser Tyr Asp
                215                 220                 225

GTG AAC GGC AAT GAT TAT GAC CCA TCT CCA CGA TAT GAT GCC AGC AAT     895
Val Asn Gly Asn Asp Tyr Asp Pro Ser Pro Arg Tyr Asp Ala Ser Asn
            230                 235                 240

GAA AAT AAA CAC GGC ACT CGT TGT GCG GGA GAA GTT GCT GCT TCA GCA     943
Glu Asn Lys His Gly Thr Arg Cys Ala Gly Glu Val Ala Ala Ser Ala
        245                 250                 255

AAC AAT TCC TAC TGC ATC GTG GGC ATA GCG TAC AAT GCC AAA ATA GGA     991
Asn Asn Ser Tyr Cys Ile Val Gly Ile Ala Tyr Asn Ala Lys Ile Gly
    260                 265                 270

GGC ATC CGC ATG CTG GAC GGC GAT GTC ACA GAT GTG GTC GAG GCA AAG    1039
Gly Ile Arg Met Leu Asp Gly Asp Val Thr Asp Val Val Glu Ala Lys
275                 280                 285                 290

TCG CTG GGC ATC AGA CCC AAC TAC ATC GAC ATT TAC AGT GCC AGC TGG    1087
Ser Leu Gly Ile Arg Pro Asn Tyr Ile Asp Ile Tyr Ser Ala Ser Trp
                295                 300                 305

GGG CCG GAC GAC GAC GGC AAG ACG GTG GAC GGG CCC GGC CGA CTG GCT    1135
Gly Pro Asp Asp Asp Gly Lys Thr Val Asp Gly Pro Gly Arg Leu Ala
            310                 315                 320

AAG CAG GCT TTC GAG TAT GGC ATT AAA AAG GGC CGG CAG GGC CTG GGC    1183
Lys Gln Ala Phe Glu Tyr Gly Ile Lys Lys Gly Arg Gln Gly Leu Gly
        325                 330                 335

TCC ATT TTC GTC TGG GCA TCT GGG AAT GGC GGG AGA GAG GGG GAC TAC    1231
Ser Ile Phe Val Trp Ala Ser Gly Asn Gly Gly Arg Glu Gly Asp Tyr
    340                 345                 350

TGC TCG TGC GAT GGC TAC ACC AAC AGC ATC TAC ACC ATC TCC GTC AGC    1279
Cys Ser Cys Asp Gly Tyr Thr Asn Ser Ile Tyr Thr Ile Ser Val Ser
355                 360                 365                 370

AGC GCC ACC GAG AAT GGC TAC AAG CCC TGG TAC CTG GAA GAG TGT GCC    1327
Ser Ala Thr Glu Asn Gly Tyr Lys Pro Trp Tyr Leu Glu Glu Cys Ala
                375                 380                 385

TCC ACC CTG GCC ACC ACC TAC AGC AGT GGG GCC TTT TAT GAG CGA AAA    1375
Ser Thr Leu Ala Thr Thr Tyr Ser Ser Gly Ala Phe Tyr Glu Arg Lys
            390                 395                 400

ATC GTC ACC ACG GAT CTG CGT CAG CGC TGT ACC GAT GGC CAC ACT GGG    1423
Ile Val Thr Thr Asp Leu Arg Gln Arg Cys Thr Asp Gly His Thr Gly
        405                 410                 415

ACC TCA GTC TCT GCC CCC ATG GTG GCG GGC ATC ATC GCC TTG GCT CTA    1471
Thr Ser Val Ser Ala Pro Met Val Ala Gly Ile Ile Ala Leu Ala Leu
    420                 425                 430

GAA GCA AAC AGC CAG TTA ACC TGG AGG GAC GTC CAG CAC CTG CTA GTG    1519
Glu Ala Asn Ser Gln Leu Thr Trp Arg Asp Val Gln His Leu Leu Val
435                 440                 445                 450
```

```
AAG ACA TCC CGG CCG GCC CAC CTG AAA GCG AGC GAC TGG AAA GTA AAC   1567
Lys Thr Ser Arg Pro Ala His Leu Lys Ala Ser Asp Trp Lys Val Asn
                455                 460                 465

GGC GCG GGT CAT AAA GTT AGC CAT TTC TAT GGA TTT GGT TTG GTG GAC   1615
Gly Ala Gly His Lys Val Ser His Phe Tyr Gly Phe Gly Leu Val Asp
            470                 475                 480

GCA GAA GCT CTC GTT GTG GAG GCA AAG AAG TGG ACA GCA GTG CCA TCG   1663
Ala Glu Ala Leu Val Val Glu Ala Lys Lys Trp Thr Ala Val Pro Ser
        485                 490                 495

CAG CAC ATG TGT GTG GCC GCC TCG GAC AAG AGA CCC AGG AGC ATC CCC   1711
Gln His Met Cys Val Ala Ala Ser Asp Lys Arg Pro Arg Ser Ile Pro
    500                 505                 510

TTA GTG CAG GTG CTG CGG ACT ACG GCC CTG ACC AGC GCC TGC GCG GAG   1759
Leu Val Gln Val Leu Arg Thr Thr Ala Leu Thr Ser Ala Cys Ala Glu
515                 520                 525                 530

CAC TCG GAC CAG CGG GTG GTC TAC TTG GAG CAC GTG GTG GTT CGC ACC   1807
His Ser Asp Gln Arg Val Val Tyr Leu Glu His Val Val Val Arg Thr
                535                 540                 545

TCC ATC TCA CAC CCA CGC CGA GGA GAC CTC CAG ATC TAC CTG GTT TCT   1855
Ser Ile Ser His Pro Arg Arg Gly Asp Leu Gln Ile Tyr Leu Val Ser
            550                 555                 560

CCC TCG GGA ACC AAG TCT CAA CTT TTG GCA AAG AGG TTG CTG GAT CTT   1903
Pro Ser Gly Thr Lys Ser Gln Leu Leu Ala Lys Arg Leu Leu Asp Leu
        565                 570                 575

TCC AAT GAA GGG TTT ACA AAC TGG GAA TTC ATG ACT GTC CAC TGC TGG   1951
Ser Asn Glu Gly Phe Thr Asn Trp Glu Phe Met Thr Val His Cys Trp
    580                 585                 590

GGA GAA AAG GCT GAA GGG CAG TGG ACC TTG GAA ATC CAA GAT CTG CCA   1999
Gly Glu Lys Ala Glu Gly Gln Trp Thr Leu Glu Ile Gln Asp Leu Pro
595                 600                 605                 610

TCC CAG GTC CGC AAC CCG GAG AAG CAA GGG AAG TTG AAA GAA TGG AGC   2047
Ser Gln Val Arg Asn Pro Glu Lys Gln Gly Lys Leu Lys Glu Trp Ser
                615                 620                 625

CTC ATA CTG TAT GGC ACA GCA GAG CAC CCG TAC CAC ACC TTC AGT GCC   2095
Leu Ile Leu Tyr Gly Thr Ala Glu His Pro Tyr His Thr Phe Ser Ala
            630                 635                 640

CAT CAG TCC CGC TCG CGG ATG CTG GAG CTC TCA GCC CCA GAG CTG GAG   2143
His Gln Ser Arg Ser Arg Met Leu Glu Leu Ser Ala Pro Glu Leu Glu
        645                 650                 655

CCA CCC AAG GCT GCC CTG TCA CCC TCC CAG GTG GAA GTT CCT GAA GAT   2191
Pro Pro Lys Ala Ala Leu Ser Pro Ser Gln Val Glu Val Pro Glu Asp
    660                 665                 670

GAG GAA GAT TAC ACA GCT CAA TCC ACC CCA GGC TCT GCT AAT ATT TTA   2239
Glu Glu Asp Tyr Thr Ala Gln Ser Thr Pro Gly Ser Ala Asn Ile Leu
675                 680                 685                 690
```

```
CAG ACC AGT GTG TGC CAT CCG GAG TGT GGT GAC AAA GGC TGT GAT GGC   2287
Gln Thr Ser Val Cys His Pro Glu Cys Gly Asp Lys Gly Cys Asp Gly
                695                 700                 705

CCC AAT GCA GAC CAG TGC TTG AAC TGC GTC CAC TTC AGC CTG GGG AGT   2335
Pro Asn Ala Asp Gln Cys Leu Asn Cys Val His Phe Ser Leu Gly Ser
            710                 715                 720

GTC AAG ACC AGC AGG AAG TGC GTG AGT GTG TGC CCC TTG GGC TAC TTT   2383
Val Lys Thr Ser Arg Lys Cys Val Ser Val Cys Pro Leu Gly Tyr Phe
        725                 730                 735

GGG GAC ACA GCA GCA AGA CGC TGT CGC CGG TGC CAC AAG GGG TGT GAG   2431
Gly Asp Thr Ala Ala Arg Arg Cys Arg Arg Cys His Lys Gly Cys Glu
    740                 745                 750

ACC TGC TCC AGC AGA GCT GCG ACG CAG TGC CTG TCT TGC CGC CGC GGG   2479
Thr Cys Ser Ser Arg Ala Ala Thr Gln Cys Leu Ser Cys Arg Arg Gly
755                 760                 765                 770

TTC TAT CAC CAC CAG GAG ATG AAC ACC TGT GTG ACC CTC TGT CCT GCA   2527
Phe Tyr His His Gln Glu Met Asn Thr Cys Val Thr Leu Cys Pro Ala
                775                 780                 785

GGA TTT TAT GCT GAT GAA AGT CAG AAA AAT TGC CTT AAA TGC CAC CCA   2575
Gly Phe Tyr Ala Asp Glu Ser Gln Lys Asn Cys Leu Lys Cys His Pro
            790                 795                 800

AGC TGT AAA AAG TGC GTG GAT GAA CCT GAG AAA TGT ACT GTC TGT AAA   2623
Ser Cys Lys Lys Cys Val Asp Glu Pro Glu Lys Cys Thr Val Cys Lys
        805                 810                 815

GAA GGA TTC AGC CTT GCA CGG GGC AGC TGC ATT CCT GAC TGT GAG CCA   2671
Glu Gly Phe Ser Leu Ala Arg Gly Ser Cys Ile Pro Asp Cys Glu Pro
    820                 825                 830

GGC ACC TAC TTT GAC TCA GAG CTG ATC AGA TGT GGG GAA TGC CAT CAC   2719
Gly Thr Tyr Phe Asp Ser Glu Leu Ile Arg Cys Gly Glu Cys His His
835                 840                 845                 850

ACC TGC GGA ACC TGC GTG GGG CCA GGC AGA GAA GAG TGC ATT CAC TGT   2767
Thr Cys Gly Thr Cys Val Gly Pro Gly Arg Glu Glu Cys Ile His Cys
                855                 860                 865

GCG AAA AAC TTC CAC TTC CAC GAC TGG AAG TGT GTG CCA GCC TGT GGT   2815
Ala Lys Asn Phe His Phe His Asp Trp Lys Cys Val Pro Ala Cys Gly
            870                 875                 880

GAG GGC TTC TAC CCA GAA GAG ATG CCG GGC TTG CCC CAC AAA GTG TGT   2863
Glu Gly Phe Tyr Pro Glu Glu Met Pro Gly Leu Pro His Lys Val Cys
        885                 890                 895

CGA AGG TGT GAC GAG AAC TGC TTG AGC TGT GCA GGC TCC AGC AGG AAC   2911
Arg Arg Cys Asp Glu Asn Cys Leu Ser Cys Ala Gly Ser Ser Arg Asn
    900                 905                 910

TGT AGC AGG TGT AAG ACG GGC TTC ACA CAG CTG GGG ACC TCC TGC ATC   2959
Cys Ser Arg Cys Lys Thr Gly Phe Thr Gln Leu Gly Thr Ser Cys Ile
915                 920                 925                 930
```

```
ACC AAC CAC ACG TGC AGC AAC GCT GAC GAG ACA TTC TGC GAG ATG GTG   3007
Thr Asn His Thr Cys Ser Asn Ala Asp Glu Thr Phe Cys Glu Met Val
                935                 940                 945

AAG TCC AAC CGG CTG TGC GAA CGG AAG CTC TTC ATT CAG TTC TGC TGC   3055
Lys Ser Asn Arg Leu Cys Glu Arg Lys Leu Phe Ile Gln Phe Cys Cys
            950                 955                 960

CGC ACG TGC CTC CTG GCC GGG T AAGGGTGCCT AGCTGCCCAC AGAGGGCAGG    3107
Arg Thr Cys Leu Leu Ala Gly
        965

CACTCCCATC CATCCATCCG TCCACCTTCC TCCAGACTGT CGGCCAGAGT CTGTTTCAGG 3167

AGCGGCGCCC TGCACCTGAC AGCTTTATCT CCCCAGGAGC AGCATCTCTG AGCACCCAAG 3227

CCAGGTGGGT GGTGGCTCTT AAGGAGGTGT TCCTAAAATG GTGATATCCT CTCAAATGCT 3287

GCTTGTTGGC TCCAGTCTTC CGACAAACTA ACAGGAACAA AATGAATTCT GGGAATCCAC 3347

AGCTCTGGCT TTGGAGCAGC TTCTGGGACC ATAAGTTTAC TGAATCTTCA AGACCAAAGC 3407

AGAAAAGAAA GGCGCTTGGC ATCACACATC ACTCTTCTCC CCGTGCTTTT CTGCGGCTGT 3467

GTAGTAAATC TCCCCGGCCC AGCTGGCGAA CCCTGGGCCA TCCTCACATG TGACAAAGGG 3527

CCAGCAGTCT ACCTGCTCGT TGCCTGCCAC TGAGCAGTCT GGGGACGGTT TGGTCAGACT 3587

ATAAATAAGA TAGGTTTGAG GGCATAAAAT GTATGACCAC TGGGGCCGGA GTATCTATTT 3647

CTACATAGTC AGCTACTTCT GAAACTGCAG CAGTGGCTTA GAAAGTCCAA TTCCAAAGCC 3707

AGACCAGAAG ATTCTATCCC CCGCAGCGCT CTCCTTTGAG CAAGCCGAGC TCTCCTTGTT 3767

ACCGTGTTCT GTCTGTGTCT TCAGGAGTCT CATGGCCTGA ACGACCACCT CGACCTGATG 3827

CAGAGCCTTC TGAGGAGAGG CAACAGGAGG CATTCTGTGG CCAGCCAAAA GGTACCCCGA 3887

TGGCCAAGCA ATTCCTCTGA ACAAAATGTA AAGCCAGCCA TGCATTGTTA ATCATCCATC 3947

ACTTCCCATT TTATGGAATT GCTTTTAAAA TACATTTGGC CTCTGCCCTT CAGAAGACTC 4007

GTTTTTAAGG TGGAAACTCC TGTGTCTGTG TATATTACAA GCCTACATGA CACAGTTGGA 4067

TTTATTCTGC CAAACCTGTG TAGGCATTTT ATAAGCTACA TGTTCTAATT TTTACCGATG 4127

TTAATTATTT TGACAAATAT TTCATATATT TTCATTGAAA TGCACAGATC TGCTTGATCA 4187

ATTCCCTTGA ATAGGGAAGT AACATTTGCC TTAAATTTTT TCGACCTCGT CTTTCTCCAT 4247

ATTGTCCTGC TCCCCTGTTT GACGACAGTG CATTTGCCTT GTCACCTGTG AGCTGGAGAG 4307

AACCCAGATG TTGTTTATTG AATCTACAAC TCTGAAAGAG AAATCAATGA AGCAAGTACA 4367

ATGTTAACCC TAAATTAATA AAAGAGTTAA CATCCC                           4403
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 969 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Pro Pro Arg Ala Pro Pro Ala Pro Gly Pro Arg Pro Pro Pro Arg
 1               5                  10                  15

Ala Ala Ala Ala Thr Asp Thr Ala Ala Gly Ala Gly Gly Ala Gly Gly
            20                  25                  30

Ala Gly Gly Ala Gly Gly Pro Gly Phe Arg Pro Leu Ala Pro Arg Pro
        35                  40                  45

Trp Arg Trp Leu Leu Leu Leu Ala Leu Pro Ala Ala Cys Ser Ala Pro
    50                  55                  60

Pro Pro Arg Pro Val Tyr Thr Asn His Trp Ala Val Gln Val Leu Gly
65                  70                  75                  80

Gly Pro Ala Glu Ala Asp Arg Val Ala Ala Ala His Gly Tyr Leu Asn
                85                  90                  95

Leu Gly Gln Ile Gly Asn Leu Glu Asp Tyr Tyr His Phe Tyr His Ser
            100                 105                 110

Lys Thr Phe Lys Arg Ser Thr Leu Ser Ser Arg Gly Pro His Thr Phe
        115                 120                 125

Leu Arg Met Asp Pro Gln Val Lys Trp Leu Gln Gln Gln Glu Val Lys
    130                 135                 140

Arg Arg Val Lys Arg Gln Val Arg Ser Asp Pro Gln Ala Leu Tyr Phe
145                 150                 155                 160

Asn Asp Pro Ile Trp Ser Asn Met Trp Tyr Leu His Cys Gly Asp Lys
                165                 170                 175

Asn Ser Arg Cys Arg Ser Glu Met Asn Val Gln Ala Ala Trp Lys Arg
            180                 185                 190

Gly Tyr Thr Gly Lys Asn Val Val Val Thr Ile Leu Asp Asp Gly Ile
        195                 200                 205

Glu Arg Asn His Pro Asp Leu Ala Pro Asn Tyr Asp Ser Tyr Ala Ser
    210                 215                 220

Tyr Asp Val Asn Gly Asn Asp Tyr Asp Pro Ser Pro Arg Tyr Asp Ala
225                 230                 235                 240

Ser Asn Glu Asn Lys His Gly Thr Arg Cys Ala Gly Glu Val Ala Ala
                245                 250                 255

Ser Ala Asn Asn Ser Tyr Cys Ile Val Gly Ile Ala Tyr Asn Ala Lys
            260                 265                 270

Ile Gly Gly Ile Arg Met Leu Asp Gly Asp Val Thr Asp Val Val Glu
        275                 280                 285
```

```
Ala Lys Ser Leu Gly Ile Arg Pro Asn Tyr Ile Asp Ile Tyr Ser Ala
    290                 295                 300

Ser Trp Gly Pro Asp Asp Asp Gly Lys Thr Val Asp Gly Pro Gly Arg
305                 310                 315                 320

Leu Ala Lys Gln Ala Phe Glu Tyr Gly Ile Lys Lys Gly Arg Gln Gly
                325                 330                 335

Leu Gly Ser Ile Phe Val Trp Ala Ser Gly Asn Gly Gly Arg Glu Gly
            340                 345                 350

Asp Tyr Cys Ser Cys Asp Gly Tyr Thr Asn Ser Ile Tyr Thr Ile Ser
        355                 360                 365

Val Ser Ser Ala Thr Glu Asn Gly Tyr Lys Pro Trp Tyr Leu Glu Glu
    370                 375                 380

Cys Ala Ser Thr Leu Ala Thr Thr Tyr Ser Ser Gly Ala Phe Tyr Glu
385                 390                 395                 400

Arg Lys Ile Val Thr Thr Asp Leu Arg Gln Arg Cys Thr Asp Gly His
                405                 410                 415

Thr Gly Thr Ser Val Ser Ala Pro Met Val Ala Gly Ile Ile Ala Leu
            420                 425                 430

Ala Leu Glu Ala Asn Ser Gln Leu Thr Trp Arg Asp Val Gln His Leu
        435                 440                 445

Leu Val Lys Thr Ser Arg Pro Ala His Leu Lys Ala Ser Asp Trp Lys
    450                 455                 460

Val Asn Gly Ala Gly His Lys Val Ser His Phe Tyr Gly Phe Gly Leu
465                 470                 475                 480

Val Asp Ala Glu Ala Leu Val Val Glu Ala Lys Lys Trp Thr Ala Val
                485                 490                 495

Pro Ser Gln His Met Cys Val Ala Ala Ser Asp Lys Arg Pro Arg Ser
            500                 505                 510

Ile Pro Leu Val Gln Val Leu Arg Thr Thr Ala Leu Thr Ser Ala Cys
        515                 520                 525

Ala Glu His Ser Asp Gln Arg Val Val Tyr Leu Glu His Val Val Val
    530                 535                 540

Arg Thr Ser Ile Ser His Pro Arg Arg Gly Asp Leu Gln Ile Tyr Leu
545                 550                 555                 560

Val Ser Pro Ser Gly Thr Lys Ser Gln Leu Leu Ala Lys Arg Leu Leu
                565                 570                 575

Asp Leu Ser Asn Glu Gly Phe Thr Asn Trp Glu Phe Met Thr Val His
            580                 585                 590

Cys Trp Gly Glu Lys Ala Glu Gly Gln Trp Thr Leu Glu Ile Gln Asp
        595                 600                 605

Leu Pro Ser Gln Val Arg Asn Pro Glu Lys Gln Gly Lys Leu Lys Glu
    610                 615                 620
```

Trp Ser Leu Ile Leu Tyr Gly Thr Ala Glu His Pro Tyr His Thr Phe
625 630 635 640

Ser Ala His Gln Ser Arg Ser Arg Met Leu Glu Leu Ser Ala Pro Glu
645 650 655

Leu Glu Pro Pro Lys Ala Ala Leu Ser Pro Ser Gln Val Glu Val Pro
660 665 670

Glu Asp Glu Glu Asp Tyr Thr Ala Gln Ser Thr Pro Gly Ser Ala Asn
675 680 685

Ile Leu Gln Thr Ser Val Cys His Pro Glu Cys Gly Asp Lys Gly Cys
690 695 700

Asp Gly Pro Asn Ala Asp Gln Cys Leu Asn Cys Val His Phe Ser Leu
705 710 715 720

Gly Ser Val Lys Thr Ser Arg Lys Cys Val Ser Val Cys Pro Leu Gly
725 730 735

Tyr Phe Gly Asp Thr Ala Ala Arg Arg Cys Arg Arg Cys His Lys Gly
740 745 750

Cys Glu Thr Cys Ser Ser Arg Ala Ala Thr Gln Cys Leu Ser Cys Arg
755 760 765

Arg Gly Phe Tyr His His Gln Glu Met Asn Thr Cys Val Thr Leu Cys
770 775 780

Pro Ala Gly Phe Tyr Ala Asp Glu Ser Gln Lys Asn Cys Leu Lys Cys
785 790 795 800

His Pro Ser Cys Lys Lys Cys Val Asp Glu Pro Glu Lys Cys Thr Val
805 810 815

Cys Lys Glu Gly Phe Ser Leu Ala Arg Gly Ser Cys Ile Pro Asp Cys
820 825 830

Glu Pro Gly Thr Tyr Phe Asp Ser Glu Leu Ile Arg Cys Gly Glu Cys
835 840 845

His His Thr Cys Gly Thr Cys Val Gly Pro Gly Arg Glu Glu Cys Ile
850 855 860

His Cys Ala Lys Asn Phe His Phe His Asp Trp Lys Cys Val Pro Ala
865 870 875 880

Cys Gly Glu Gly Phe Tyr Pro Glu Glu Met Pro Gly Leu Pro His Lys
885 890 895

Val Cys Arg Arg Cys Asp Glu Asn Cys Leu Ser Cys Ala Gly Ser Ser
900 905 910

Arg Asn Cys Ser Arg Cys Lys Thr Gly Phe Thr Gln Leu Gly Thr Ser
915 920 925

Cys Ile Thr Asn His Thr Cys Ser Asn Ala Asp Glu Thr Phe Cys Glu
930 935 940

Met Val Lys Ser Asn Arg Leu Cys Glu Arg Lys Leu Phe Ile Gln Phe
945 950 955 960

Cys Cys Arg Thr Cys Leu Leu Ala Gly
965

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 636 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (cDNA)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..58

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGT CAT AAA GGT GCG GCA GTG GCG TTC TGG TGG ACC ATT GGG TGG CCC 48
Gly His Lys Gly Ala Ala Val Ala Phe Trp Trp Thr Ile Gly Trp Pro
1 5 10 15

TGG AAT GTG T AGGAAGGGGT GTCATGAATT CCTTAAAAGG ACTCTCCAAA 98
Trp Asn Val

TAGCATTAGT TGTTATTATT AACTTAAAAG GACTCTCCAA ATAGCATTAG TTGTTATTAT 158

TAATTGTGTG TCACAAGAAT TTAAAACGCA TGTGCAGCTA TTTAAGAAAA GTATCCCGGA 218

AGCTCACAGT GACATTACGG AAGAACCCTC AGGTCACAAG AGTCTGGGGT CTCCTATACT 278

CTATAACTTT GGCCACACCG AGACACCACC TATACCAATA TTTACTCATA GTTCTCTTTA 338

AGCCAGGAGC AATGACGTGT GCCTATAGTC GCAGCTACTA GGGAAGTTGA GGCAGGAGGA 398

TTGCTTGAGC CCAGGAATTT GAGTCTAGCC TGGACAACAC AGCAGGACTC CATCTCTTAA 458

AAAAAAAATT ACTTCCCCCA CTACTTTTTT TGACATAAAA AAATGTATTT TAAAAGGAAA 518

CTGTACTACA TCTAGTTAAT CATAGGTTTG ATATGTAGTT ACGTATTTTT TCTAATGTGC 578

ATTAAAACAA ATCCATAATT ATTAAAATAA ATGTTGTTTG TGTGCCAAAA AAAAAAAA 636

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Gly His Lys Gly Ala Ala Val Ala Phe Trp Trp Thr Ile Gly Trp Pro
 1               5                  10                  15

Trp Asn Val
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 148 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 18..131

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
GGCCTGTCTT TTCAGTT AGC CAT TTC TAT GGA TTT GGT TTG GTG GAC GCA      50
                  Ser His Phe Tyr Gly Phe Gly Leu Val Asp Ala
                    1               5                  10

GAA GCT CTC GTT GTG GAG GCA AAG AAG TGG ACA GCA GTG CCA TCG CAG    98
Glu Ala Leu Val Val Glu Ala Lys Lys Trp Thr Ala Val Pro Ser Gln
            15                  20                  25

CAC ATG TGT GTG GCC GCC TCG GAC AAG AGA CCC AGGTAAGGCT CTGCTGT    148
His Met Cys Val Ala Ala Ser Asp Lys Arg Pro
        30                  35
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Ser His Phe Tyr Gly Phe Gly Leu Val Asp Ala Glu Ala Leu Val Val
 1               5                  10                  15

Glu Ala Lys Lys Trp Thr Ala Val Pro Ser Gln His Met Cys Val Ala
            20                  25                  30

Ala Ser Asp Lys Arg Pro
        35
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 280 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Asn Asp Pro Leu Phe Glu Arg Gln Trp His Leu Val Asn Pro Ser Phe
1               5                  10                  15

Pro Gly Ser Asp Ile Asn Val Leu Asp Leu Trp Tyr Asn Asn Ile Thr
            20                  25                  30

Gly Ala Gly Val Val Ala Ala Ile Val Asp Asp Gly Leu Asp Tyr Glu
        35                  40                  45

Asn Glu Asp Leu Lys Asp Asn Phe Cys Ala Glu Gly Ser Trp Asp Phe
    50                  55                  60

Asn Asp Asn Thr Asn Leu Pro Lys Pro Arg Leu Ser Asp Asp Tyr His
65                  70                  75                  80

Gly Thr Arg Cys Ala Gly Glu Ile Ala Ala Lys Lys Gly Asn Asn Phe
                85                  90                  95

Cys Gly Val Gly Val Gly Tyr Asn Ala Lys Ile Ser Gly Ile Arg Ile
            100                 105                 110

Leu Ser Gly Asp Ile Thr Thr Glu Asp Glu Ala Ala Ser Leu Ile Tyr
        115                 120                 125

Gly Leu Asp Val Asn Asp Ile Tyr Ser Cys Ser Trp Gly Pro Ala Asp
    130                 135                 140

Asp Gly Arg His Leu Gln Gly Pro Ser Asp Leu Val Lys Lys Ala Leu
145                 150                 155                 160

Val Lys Gly Val Thr Glu Gly Arg Asp Ser Lys Gly Ala Ile Tyr Val
                165                 170                 175

Phe Ala Ser Gly Asn Gly Gly Thr Arg Gly Asp Asn Cys Asn Tyr Asp
            180                 185                 190

Gly Tyr Thr Asn Ser Ile Tyr Ser Ile Thr Ile Gly Ala Ile Asp His
        195                 200                 205

Lys Asp Leu His Pro Pro Tyr Ser Glu Gly Cys Ser Ala Val Met Ala
    210                 215                 220

Val Thr Tyr Ser Ser Gly Ser Gly Glu Tyr Ile His Ser Ser Asp Ile
225                 230                 235                 240

Asn Gly Arg Cys Ser Asn Ser His Gly Gly Thr Ser Ala Ala Ala Pro
                245                 250                 255

Leu Ala Ala Gly Val Tyr Thr Leu Leu Leu Glu Ala Asn Pro Asn Leu
            260                 265                 270
```

Thr Trp Arg Asp Val Gln Tyr Leu
275 280

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 282 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Asp Pro Lys Phe Pro Gln Gln Trp Tyr Leu Ser Gly Val Thr Gln Arg
1 5 10 15

Asp Leu Asn Val Lys Ala Ala Trp Ala Gln Gly Tyr Thr Gly His Gly
20 25 30

Ile Val Val Ser Ile Leu Asp Asp Gly Ile Glu Lys Asn His Pro Asp
35 40 45

Leu Ala Gly Asn Tyr Asp Pro Gly Ala Ser Phe Asp Val Asn Asp Gln
50 55 60

Asp Pro Asp Pro Gln Pro Arg Tyr Thr Gln Met Asn Asp Asn Arg His
65 70 75 80

Gly Thr Arg Cys Ala Gly Glu Val Ala Ala Val Ala Asn Asn Gly Val
85 90 95

Cys Gly Val Gly Val Ala Tyr Asn Ala Arg Ile Gly Gly Val Arg Met
100 105 110

Leu Asp Gly Glu Val Thr Asp Ala Val Glu Ala Arg Ser Leu Gly Leu
115 120 125

Asn Pro Asn His Ile His Ile Tyr Ser Ala Ser Trp Gly Pro Glu Asp
130 135 140

Asp Gly Lys Thr Val Asp Gly Pro Ala Arg Leu Ala Glu Glu Ala Phe
145 150 155 160

Phe Arg Gly Val Ser Gln Gly Arg Gly Gly Leu Gly Ser Ile Phe Val
165 170 175

Trp Ala Ser Gly Asn Gly Gly Arg Glu His Asp Ser Cys Asn Cys Asp
180 185 190

Gly Tyr Thr Asn Ser Ile Tyr Thr Leu Ser Ile Ser Ser Ala Thr Gln
195 200 205

Phe Gly Asn Val Pro Trp Tyr Ser Glu Ala Cys Ser Ser Thr Leu Ala
210 215 220

Thr Thr Tyr Ser Ser Gly Asn Gln Asn Glu Lys Gln Ile Val Thr Thr
225 230 235 240

```
Asp Leu Arg Gln Lys Cys Thr Glu Ser His Thr Gly Thr Ser Ala Ser
                245                 250                 255

Ala Pro Leu Ala Ala Gly Ile Ile Ala Leu Thr Leu Glu Ala Asn Lys
            260                 265                 270

Asn Leu Thr Trp Arg Asp Met Gln His Leu
        275                 280
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 291 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear ,

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
Asn Asp Pro Leu Phe Thr Lys Gln Trp Tyr Leu Ile Asn Thr Gly Gln
1               5                   10                  15

Ala Asp Gly Thr Pro Gly Leu Asp Leu Asn Val Ala Glu Ala Trp Glu
            20                  25                  30

Leu Gly Tyr Thr Gly Lys Gly Val Thr Ile Gly Ile Met Asp Asp Gly
        35                  40                  45

Ile Asp Tyr Leu His Pro Asp Leu Ala Ser Asn Tyr Asn Ala Glu Ala
    50                  55                  60

Ser Tyr Asp Phe Ser Ser Asn Asp Pro Tyr Pro Tyr Pro Arg Tyr Thr
65                  70                  75                  80

Asp Asp Trp Phe Asn Ser His Gly Thr Arg Cys Ala Gly Glu Val Ser
                85                  90                  95

Ala Ala Ala Asn Asn Asn Ile Cys Gly Val Gly Val Ala Tyr Asn Ser
            100                 105                 110

Lys Val Ala Gly Ile Arg Met Leu Asp Gln Pro Phe Met Thr Asp Ile
        115                 120                 125

Ile Glu Ala Ser Ser Ile Ser His Met Pro Gln Leu Ile Asp Ile Tyr
    130                 135                 140

Ser Ala Ser Trp Gly Pro Thr Asp Asn Gly Lys Thr Val Asp Gly Pro
145                 150                 155                 160

Arg Asp Val Thr Leu Gln Ala Met Ala Asp Gly Val Asn Lys Gly Arg
                165                 170                 175

Gly Gly Lys Gly Ser Ile Tyr Val Trp Ala Ser Gly Asp Gly Gly Ser
            180                 185                 190

Tyr Asp Asp Cys Asn Cys Asp Gly Tyr Ala Ser Ser Met Trp Thr Ile
        195                 200                 205
```

Ser Ile Asn Ser Ala Ile Asn Asp Gly Arg Thr Ala Leu Tyr Asp Glu
210 215 220

Ser Cys Ser Ser Thr Leu Ala Ser Thr Phe Ser Asn Gly Arg Lys Arg
225 230 235 240

Asn Pro Glu Ala Gly Val Ala Thr Thr Asp Leu Tyr Gly Asn Cys Thr
245 250 255

Leu Arg His Ser Gly Thr Ser Ala Ala Ala Pro Glu Ala Ala Gly Val
260 265 270

Phe Ala Leu Ala Leu Glu Ala Asn Leu Gly Leu Thr Trp Arg Asp Met
275 280 285

Gln His Leu
290

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 289 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Asn Asp Pro Met Trp Asn Gln Gln Trp Tyr Leu Gln Asp Thr Arg Met
1 5 10 15

Thr Ala Ala Leu Pro Lys Leu Asp Leu His Val Ile Pro Val Trp Glu
20 25 30

Lys Gly Ile Thr Gly Lys Gly Val Val Ile Thr Val Leu Asp Asp Gly
35 40 45

Leu Glu Trp Asn His Thr Asp Ile Tyr Ala Asn Tyr Asp Pro Glu Ala
50 55 60

Ser Tyr Asp Phe Asn Asp Asn Asp His Asp Pro Phe Pro Arg Tyr Asp
65 70 75 80

Leu Thr Asn Glu Asn Lys His Gly Thr Arg Cys Ala Gly Glu Ile Ala
85 90 95

Met Gln Ala Asn Asn His Lys Cys Gly Val Gly Val Ala Tyr Asn Ser
100 105 110

Lys Val Gly Gly Ile Arg Met Leu Asp Gly Ile Val Thr Asp Ala Ile
115 120 125

Glu Ala Ser Ser Ile Gly Phe Asn Pro Gly His Val Asp Ile Tyr Ser
130 135 140

Ala Ser Trp Gly Pro Asn Asp Asp Gly Lys Thr Val Glu Gly Pro Gly
145 150 155 160

Arg Leu Ala Gln Lys Ala Phe Glu Tyr Gly Val Lys Gln Gly Arg Gln
165 170 175

Gly Lys Gly Ser Ile Phe Val Trp Ala Ser Gly Asn Gly Gly Arg Gln
180 185 190

Gly Asp Asn Cys Asp Cys Asp Gly Tyr Thr Asp Ser Ile Tyr Thr Ile
195 200 205

Ser Ile Ser Ser Ala Ser Gln Gln Gly Leu Ser Pro Trp Tyr Ala Glu
210 215 220

Lys Cys Ser Ser Thr Leu Ala Thr Ser Tyr Ser Ser Gly Asp Tyr Thr
225 230 235 240

Asp Gln Arg Ile Thr Ser Ala Asp Leu His Asn Asp Cys Thr Glu Thr
245 250 255

His Thr Gly Thr Ser Ala Ser Ala Pro Leu Ala Ala Gly Ile Phe Ala
260 265 270

Leu Ala Leu Glu Ala Asn Pro Asn Leu Thr Trp Arg Asp Met Gln His
275 280 285

Leu

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 287 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Asn Asp Pro Ile Trp Ser Asn Met Trp Tyr Leu His Cys Gly Asp Lys
1 5 10 15

Asn Ser Arg Cys Arg Ser Glu Met Asn Val Gln Ala Ala Trp Lys Arg
20 25 30

Gly Tyr Thr Gly Lys Asn Val Val Val Thr Ile Leu Asp Asp Gly Ile
35 40 45

Glu Arg Asn His Pro Asp Leu Ala Pro Asn Tyr Asp Ser Tyr Ala Ser
50 55 60

Tyr Asp Val Asn Gly Asn Asp Tyr Asp Pro Ser Pro Arg Tyr Asp Ala
65 70 75 80

Ser Asn Glu Asn Lys His Gly Thr Arg Cys Ala Gly Glu Val Ala Ala
85 90 95

Ser Ala Asn Asn Ser Tyr Cys Ile Val Gly Ile Ala Tyr Asn Ala Lys
100 105 110

Ile Gly Gly Ile Arg Met Leu Asp Gly Asp Val Thr Asp Val Val Glu
115 120 125

Ala Lys Ser Leu Gly Ile Arg Pro Asn Tyr Ile Asp Ile Tyr Ser Ala
130 135 140

Ser Trp Gly Pro Asp Asp Asp Gly Lys Thr Val Asp Gly Pro Gly Arg
145 150 155 160

Leu Ala Lys Gln Ala Phe Glu Tyr Gly Ile Lys Lys Gly Arg Gln Gly
165 170 175

Leu Gly Ser Ile Phe Val Trp Ala Ser Gly Asn Gly Gly Arg Glu Gly
180 185 190

Asp Tyr Cys Ser Cys Asp Gly Tyr Thr Asn Ser Ile Tyr Thr Ile Ser
195 200 205

Val Ser Ser Ala Thr Glu Asn Gly Tyr Lys Pro Trp Tyr Leu Glu Glu
210 215 220

Cys Ala Ser Thr Leu Ala Thr Thr Tyr Ser Ser Gly Ala Phe Tyr Glu
225 230 235 240

Arg Lys Ile Val Thr Thr Asp Leu Arg Gln Arg Cys Thr Asp Gly His
245 250 255

Thr Gly Thr Ser Val Ser Ala Pro Met Val Ala Gly Ile Ile Leu Ala
260 265 270

Leu Glu Ala Asn Ser Gln Leu Thr Trp Arg Asp Val Gln His Leu
275 280 285

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGATCTGAAT TCGAYGAYGG NAT 23

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGATCTAAGC TTACACCKNG TNCCRTG 27

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TATTCATTGC CGTTCACGTC GTAG 24

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GCTGGCATCA TATCGTGGAG AT 22

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GGAGGCATAG CGGCGAC 17

CLAIMS

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a sequence encoding PACE 4.

2. The polynucleotide according to Claim 1, wherein said sequence encodes a polypeptide having an amino acid sequence at least 90% homologous to an amino acid sequence encoded by the cDNA sequence of Figure 1 (SEQ ID NO: 1).

3. The polynucleotide according to Claim 2, wherein said polynucleotide sequence encodes a polypeptide which has an amino acid sequence identical to the amino acid sequence encoded by the cDNA sequence of Figure 1 (SEQ ID NO: 1).

4. The polynucleotide according to Claim 2, wherein said polynucleotide sequence does not encode a cysteine-rich region.

5. A transformed cell comprising a recombinant polynucleotide encoding PACE 4 or 4.1, wherein said cell is capable of expressing PACE 4 or 4.1.

6. The cell according to Claim 5, wherein said cell is a mammalian cell.

7. The mammalian cell according to claim 6, wherein the said mammalian cell is a CHO cell.

8. The cell according to Claim 5, wherein said cell is an insect cell.

9. The insect cell according to claim 8, wherein said insect cell is selected from the group consisting of Aedes aegypti, Autographa californica, Bombyx mori, Drosophila melanogaster, Spodoptera frugiperda, and Trichoplusia ni.

10. The cell according to Claim 5, wherein said cell is a microorganism selected from the group consisting of bacteria and fungi.

11. The microorganism cell according to Claim 10, wherein said microorganism is yeast.

12. A cell comprising a recombinant polynucleotide encoding a PACE 4 and a heterologous polynucleotide encoding a precursor polypeptide, wherein said precursor polypeptide is a substrate for the encoded PACE 4, and wherein said cell is capable of expressing the polynucleotides encoding said PACE 4 and said heterologous polypeptide.

13. The cell according to Claim 12, wherein said encoded PACE 4 and said encoded heterologous precursor polypeptide, when expressed, are secreted into extracellular medium.

14. The cell according to Claim 12, wherein said cell is a mammalian cell.

15. The mammalian cell according to claim 14, wherein said mammalian cell is a CHO cell.

16. The cell according to Claim 12, wherein said cell is an insect cell.

17. The cell according to claim 16, wherein said insect cell is selected from the group consisting of Aedes aegypti, Autographa californica, Bombyx mori, Drosophila melanogaster, Spodoptera frugiperda, and Trichoplusia ni.

18. The cell according to claim 12, wherein said cell is a microorganism selected from the group consisting of bacteria and fungi.

19. The microorganism cell according to claim 18, wherein said microorganism is yeast.

20. A recombinant expression vector suitable for expression in a cell comprising a polynucleotide sequence encoding PACE 4 or 4.1, wherein said PACE 4 or 4.1 encoding sequence is operably linked to a control sequence.

21. The expression vector according to claim 20, wherein said cell is a microorganism selected from the group consisting of bacteria and fungi.

22. The expression vector according to claim 21, wherein said microorganism is yeast.

23. The expression vector according to claim 20, wherein said cell is a mammalian cell.

24. The expression vector according to claim 23, wherein said vector is a recombinant vaccinia virus.

25. The expression vector according to claim 20, wherein said cell is an insect cell.

26. The expression vector according to claim 25, wherein said vector is a recombinant baculovirus.

25. A method for producing recombinant PACE 4 or 4.1 comprising incubating the cell of claim 5 under conditions that allow expression of the recombinant polynucleotide encoding PACE 4 or 4.1.

26. A method for producing recombinant PACE 4 or 4.1, comprising:

(a) providing a mammalian cell transformed by the recombinant expression vector of claim 23, wherein the vector contains control sequences for expression that are compatible with the mammalian cell; and

(b) incubating the mammalian cell under conditions that allow transformation with the expression vector.

27. A method for producing recombinant PACE 4 or 4.1, comprising:

(a) providing a microorganism transformed by the expression vector of claim 21, wherein the vector contains control sequences for expression which are compatible with the microorganism; and

(b) incubating the microorganism under conditions which allow transformation with the expression vector.

28. A method for producing a desired mature polypeptide comprising incubating the cell according to claim 12 under conditions that allow the expression of (1) recombinant PACE 4 or 4.1 and (2) the heterologous precursor polypeptide.

29. A method for producing a desired mature polypeptide comprising incubating the cell according to claim 13 under conditions that allow expression and secretion of (1) recombinant PACE 4 or 4.1 and (2) the heterologous precursor polypeptide.

30. A composition comprising substantially pure PACE 4 polypeptide.

31. The composition of Claim 30, wherein the sequence of said PACE 4 polypeptide is at least 90% homologous to the polypeptide sequence of Figure 1 (SEQ ID NO: 1).

32. The composition of Claim 30, wherein the sequence of said PACE 4 polypeptide is identical to the polypeptide sequence of Figure 1 (SEQ ID NO: 1).

33. A composition comprising substantially pure PACE 4.1 polypeptide.

34. The composition of Claim 33, wherein the sequence of said PACE 4.1 polypeptide is at least 90% homologous to the polypeptide sequence of Figure 2 (SEQ ID NO: 3) taken with Figure 1 (SEQ ID NO: 1).

35. The composition of Claim 33, wherein the sequence of said PACE 4.1 polypeptide is identical to the polypeptide sequence of Figure 2 (SEQ ID NO: 3) taken with Figure 1 (SEQ ID NO: 1).

36. A method for reducing blood pressure in a mammal comprising administering to said mammal an inhibiting compound in an amount effective to inhibit the conversion of prorenin to renin by PACE 4 or 4.1.

37. The method of Claim 36, wherein said PACE 4 or 4.1 inhibiting compound is α_1-antitrypsin or a mutant thereof.

38. A method of inhibiting coagulation of blood comprising administering to a mammal a coagulation inhibiting amount of a composition comprising PACE 4 or 4.1.

39. A pharmaceutical composition for inhibiting blood coagulation comprising a coagulation inhibiting amount of PACE 4 or 4.1 in a physiologically compatible matrix.

40. An isolated oligonucleotide comprising at least 15 consecutive nucleotides from the sequence of Figure 1, wherein said at least 15 consecutive nucleotides of said oligonucleotide are not consecutively present in the sequences encoding KEX2, PC1/PC3, furin/PACE, or PC2.

41. An isolated polypeptide comprising a domain having an amino acid sequence at least 90% homologous to the amino acid sequence from Cys695 to Cys965 of the cysteine rich region of Figure 1 (SEQ ID NO: 1).

42. A fusion protein comprising a domain having an amino acid sequence at least 90% homologous to the amino acid sequence from Cys695 to Cys965 of the cystine rich region of PACE 4 and a functional domain from a heterologous protein.

```
 541 His Val Val Arg Thr Ser Ile Ser His Pro Arg Arg Gly Asp Leu Gln Ile Tyr Leu
1621 CAC GTG GTT CGC ACC TCC ATC TCA CAC CCA CGC CGA GGA GAC CTC CAG ATC TAC CTG

 561 Val Ser Pro Ser Gly Thr Lys Ser Gln Leu Leu Ala Lys Arg Leu Leu Asp Leu Ser Asn
1681 GTT TCT CCC TCG GGA ACC AAG TCT CAA CTT TTG GCA AAG AGG TTG CTG GAT CTT TCC AAT

 581 Glu Gly Phe Thr Asn Trp Glu Phe Met Thr Val His Cys Trp Gly Glu Lys Ala Glu Gly
1741 GAA GGG TTT ACA AAC TGG GAA TTC ATG ACT GTC CAC TGC TGG GGA GAA AAG GCT GAA GGG

 601 Gln Trp Thr Leu Glu Ile Gln Asp Leu Pro Ser Gln Val Arg Asn Pro Glu Lys Gln Gly
1801 CAG TGG ACC TTG GAA ATC CAA GAT CTG CCA TCC CAG GTC CGC AAC CCG GAG AAG CAA GGG

 621 Lys Leu Lys Glu Trp Ser Leu Ile Leu Tyr Gly Thr Ala Glu His Pro Tyr His Thr Phe
1861 AAG TTG AAA GAA TGG AGC CTC ATA CTG TAT GGC ACA GCA GAG CAC CCG TAC CAC ACC TTC

 641 Ser Ala His Gln Ser Arg Ser Met Leu Glu Leu Ser Ala Pro Glu Leu Glu Pro Pro Pro
1921 AGT GCC CAT CAG TCC CGC TCG ATG CTG GAG CTC TCA GCC CCA GAG CTG GAG CCA CCC

 661 Lys Ala Ala Leu Ser Pro Ser Gln Val Glu Val Pro Glu Asp Glu Glu Asp Tyr Thr Ala
1981 AAG GCT GCC CTG TCA CCC TCC CAG GTG GAA GTT CCT GAA GAT GAG GAA GAT TAC ACA GCT

 681 Gln Ser Thr Pro Gly Ser Ala Asn Ile Leu Gln Thr Ser Val Cys His Pro Glu Cys Gly
2041 CAA TCC ACC CCA GGC TCT GCT AAT ATT TTA CAG ACC AGT GTG TGC CAT CCG GAG TGT GGT

 701 Asp Lys Gly Cys Asp Gly Pro Asn Ala Asp Gln Cys Leu Asn Cys Val His Phe Ser Leu
2101 GAC AAA GGC TGT GAT GGC CCC AAT GCA GAC CAG TGC TTG AAC TGC GTC CAC TTC AGC CTG

 721 Gly Ser Val Lys Thr Ser Arg Lys Cys Val Ser Cys Pro Leu Gly Tyr Phe Gly Asp
2161 GGG AGT GTC AAG ACC AGC AGG AAG TGC GTG AGT TGC CCC TTG GGC TAC TTT GGG GAC

 741 Thr Ala Ala Arg Arg Cys Arg Cys His Lys Gly Cys Glu Thr Cys Ser Ser Arg Ala
2221 ACA GCA GCA AGA CGC TGT CGC CAC AAG GGG TGT GAG ACC TGC TCC AGC AGA GCT

 761 Ala Thr Gln Cys Leu Ser Cys Arg Arg Gly Phe Tyr His His Gln Glu Met Asn Thr Cys
2281 GCG ACG CAG TGC CTG TCT TGC CGC CGG GGG TTC TAT CAC CAC CAG GAG ATG AAC ACC TGT

 781 Val Thr Leu Cys Pro Ala Gly Phe Tyr Ala Asp Glu Ser Gln Lys Asn Cys Leu Lys Cys
2341 GTG ACC CTC TGT CCT GCA GGA TTT TAT GCT GAT GAA AGT CAG AAA AAT TGC CTT AAA TGC

 801 His Pro Ser Cys Lys Lys Cys Val Asp Glu Pro Glu Lys Cys Thr Val Cys Lys Glu Gly
2401 CAC CCA AGC TGT AAA AAG TGC GTG GAT GAA CCT GAG AAA TGT ACT GTC TGT AAA GAA GGA

 821 Phe Ser Leu Ala Arg Gly Ser Cys Ile Pro Asp Cys Glu Pro Gly Thr Tyr Phe Asp Ser
2461 TTC AGC CTT GCA CGG GGC AGC TGC ATT CCT GAC TGT GAG CCA GGC ACC TAC TTT GAC TCA
```

FIG. 1-1

```
 841 Glu Leu Ile Arg Cys Gly Glu Cys His His Thr Cys Gly Thr Cys Val Gly Pro Gly Arg
2521 GAG CTG ATC AGA TGT GGG GAA TGC CAT CAC ACC TGC GGA ACC TGC GTG GGG CCA GGC AGA

 861 Glu Glu Cys Ile His Cys Ala Lys Asn Phe His Phe His Asp Trp Lys Cys Val Pro Ala
2581 GAA GAG TGC ATT CAC TGT GCG AAA AAC TTC CAC TTC CAC GAC TGG AAG TGT GTG CCA GCC

 881 Cys Gly Glu Gly Phe Tyr Pro Glu Glu Met Pro Gly Leu Pro His Lys Val Cys Arg Arg
2641 TGT GGT GAG GGC TTC TAC CCA GAA GAG ATG CCG GGC TTG CCC CAC AAA GTG TGT CGA AGG

 901 Cys Asp Glu Asn Cys Leu Ser Cys Ala Gly Ser Ser Arg Asn Cys Ser Arg Cys Lys Thr
2701 TGT GAC GAG AAC TGC TTG AGC TGT GCA GGC TCC AGC AGG AAC TGT AGC AGG TGT AAG ACG

 921 Gly Phe Thr Gln Leu Gly Thr Ser Cys Ile Thr Asn His Thr Cys Ser Asn Ala Asp Glu
2761 GGC TTC ACA CAG CTG GGG ACC TCC TGC ATC ACC AAC CAC ACG TGC AGC AAC GCT GAC GAG

 941 Thr Phe Cys Glu Met Val Lys Ser Asn Arg Leu Cys Glu Arg Lys Leu Phe Ile Gln Phe
2821 ACA TTC TGC GAG ATG GTG AAG TCC AAC CGG CTG TGC GAA CGG AAG CTC TTC ATT CAG TTC

 961 Cys Cys Arg Thr Cys Leu Leu Ala Gly OC
2881 TGC TGC CGC ACG TGC CTC CTG GCC GGG TAA GGGTGCCTAGCTGCCCACAGAGGGCAGGCACTCCCATCC
2950 ATCCATCCGTCCACCTTCCTCCAGACTGTCGGCCAGAGTCTGTTTCAGGAGCGGCGCCCTGCACCTGACAGCTTTATCT
3029 CCCCAGGAGCAGCATCTCTGAGCACCCAAGCCAGGTGGGTGGTGGCTCTTAAGGAGGTGTTCCTAAAATGGTGATATCC
3108 TCTCAAATGCTGCTTGTTGGCTCCAGTCTTCCGACAAACTAACAGGAACAAAATGAATTCTGGGAATCCACAGCTCTGG
3187 CTTTGGAGCAGCTTCTGGGACCATAAGTTTACTGAATCTTCAAGACCAAAGCAGAAAAGAAAGGCGCTTGGCATCACAC
3266 ATCACTCTTCTCCCCGTGCTTTTCTGCGGCTGTGTAGTAAATCTCCCCGGCCCAGCTGGCGAACCCTGGGCCATCCTCA
3345 CATGTGACAAAGGGCCAGCAGTCTACCTGCTCGTTGCCTGCCACTGAGCAGTCTGGGGACGGTTTGGTCAGACTATAAA
3424 TAAGATAGGTTTGAGGGCATAAAATGTATGACCACTGGGGCCGGAGTATCTATTTCTACATAGTCAGCTACTTCTGAAA
3503 CTGCAGCAGTGGCTTAGAAAGTCCAATTCCAAAGCCAGACCAGAAGATTCTATCCCCCGCAGCGCTCTCCTTTGAGCAA
3582 GCCGAGCTCTCCTTGTTACCGTGTTCTGTCTGTGTCTTCAGGAGTCTCATGGCCTGAACGACCACCTCGACCTGATGCA
3661 GAGCCTTCTGAGGAGAGGCAACAGGAGGCATTCTGTGGCCAGCCAAAAGGTACCCCGATGGCCAAGCAATTCCTCTGAA
3740 CAAAATGTAAAGCCAGCCATGCATTGTTAATCATCCATCACTTCCCATTTTATGGAATTGCTTTTAAAATACATTTGGC
3819 CTCTGCCCTTCAGAAGACTCGTTTTTAAGGTGGAAACTCCTGTGTCTGTGTATATTACAAGCCTACATGACACAGTTGG
3898 ATTTATTCTGCCAAACCTGTGTAGGCATTTTATAAGCTACATGTTCTAATTTTTACCGATGTTAATTATTTTGACAAAT
3977 ATTTCATATATTTTCATTGAAATGCACAGATCTGCTTGATCAATTCCCTTGAATAGGGAAGTAACATTTGCCTTAAATT
4056 TTTTCGACCTCGTCTTTCTCCATATTGTCCTGCTCCCCTGTTTGACGACAGTGCATTTGCCTTGTCACCTGTGAGCTGG
4135 AGAGAACCCAGATGTTGTTTATTGAATCTACAACTCTGAAAGAGAAATCAATGAAGCAAGTACAATGTTAACCCTAAAT
4214 TAATAAAAGAGTTAACATCCC
```

FIG. 1-2

```
-169                                                                                        CGGGAACGCGC
-158 CGCGGCCGCCTCCTCCTCCCCGGCTCCCGCCCGCGGCGGTGTTGGCGGCGGCGGTGGCGGCGGCGGCGGCGCTTCCCCG
 -79 GCGCGGAGCGGCTTTAAAAGGCGGCACTCCACCCCCCGGCGCACTCGCAGCTCGGGCGCCGCGCGAGCCTGTCGCCGCT

  1  Met Pro Pro Arg Ala Pro Pro Ala Pro Gly Pro Arg Pro Pro Pro Arg Ala Ala Ala Ala
  1  ATG CCT CCG CGC GCG CCG CCT GCG CCC GGG CCC CGG CCG CCG CCC CGG GCC GCC GCC GCC

 21  Thr Asp Thr Ala Ala Gly Ala Gly Gly Ala Gly Gly Ala Gly Gly Ala Gly Gly Pro Gly
 61  ACC GAC ACC GCC GCG GGC GCG GGG GGC GCG GGG GGC GCG GGG GGC GCC GGC GGG CCC GGG

 41  Phe Arg Pro Leu Ala Pro Arg Pro Trp Arg Trp Leu Leu Leu Leu Ala Leu Pro Ala Ala
121  TTC CGG CCG CTC GCG CCG CGT CCC TGG CGC TGG CTG CTG CTG CTG GCG CTG CCT GCC GCC

 61  Cys Ser Ala Pro Pro Pro Arg Pro Val Tyr Thr Asn His Trp Ala Val Gln Val Leu Gly
181  TGC TCC GCG CCC CCG CCG CGC CCC GTC TAC ACC AAC CAC TGG GCG GTG CAA GTG CTG GGC

 81  Gly Pro Ala Glu Ala Asp Arg Val Ala Ala Ala His Gly Tyr Leu Asn Leu Gly Gln Ile
241  GGC CCG GCC GAG GCG GAC CGC GTG GCG GCG GCG CAC GGC TAC CTC AAC TTG GGC CAG ATT

101  Gly Asn Leu Glu Asp Tyr Tyr His Phe Tyr His Ser Lys Thr Phe Lys Arg Ser Thr Leu
301  GGA AAC CTG GAA GAT TAC TAC CAT TTT TAT CAC AGC AAA ACC TTT AAA AGA TCA ACC TTG

121  Ser Ser Arg Gly Pro His Thr Phe Leu Arg Met Asp Pro Gln Val Lys Trp Leu Gln Gln
361  AGT AGC AGA GGC CCT CAC ACC TTC CTC AGA ATG GAC CCC CAG GTG AAA TGG CTC CAG CAA

141  Gln Glu Val Lys Arg Arg Val Lys Arg Gln Val Arg Ser Asp Pro Gln Ala Leu Tyr Phe
421  CAG GAA GTG AAA CGA AGG GTG AAG AGA CAG GTG CGA AGT GAC CCG CAG GCC CTT TAC TTC

161  Asn Asp Pro Ile Trp Ser Asn Met Trp Tyr Leu His Cys Gly Asp Lys Asn Ser Arg Cys
481  AAC GAC CCC ATT TGG TCC AAC ATG TGG TAC CTG CAT TGT GGC GAC AAG AAC AGT CGC TGC

181  Arg Ser Glu Met Asn Val Gln Ala Ala Trp Lys Arg Gly Tyr Thr Gly Lys Asn Val Val
541  CGG TCG GAA ATG AAT GTC CAG GCA GCG TGG AAG AGG GGC TAC ACA GGA AAA AAC GTG GTG

201  Val Thr Ile Leu Asp Asp Gly Ile Glu Arg Asn His Pro Asp Leu Ala Pro Asn Tyr Asp
601  GTC ACC ATC CTT GAT GAT GGC ATA GAG AGA AAT CAC CCT GAC CTG GCC CCA AAT TAT GAT

221  Ser Tyr Ala Ser Tyr Asp Val Asn Gly Asn Asp Tyr Asp Pro Ser Pro Arg Tyr Asp Ala
661  TCC TAC GCC AGC TAC GAC GTG AAC GGC AAT GAT TAT GAC CCA TCT CCA CGA TAT GAT GCC

241  Ser Asn Glu Asn Lys His Gly Thr Arg Cys Ala Gly Glu Val Ala Ala Ser Ala Asn Asn
721  AGC AAT GAA AAT AAA CAC GGC ACT CGT TGT GCG GGA GAA GTT GCT GCT TCA GCA AAC AAT
```

FIG. 1A-1

```
261  Ser Tyr Cys Ile Val Gly Ile Ala Tyr Asn Ala Lys Ile Gly Gly Ile Arg Met Leu Asp
781  TCC TAC TGC ATC GTG GGC ATA GCG TAC AAT GCC AAA ATA GGA GGC ATC CGC ATG CTG GAC

281  Gly Asp Val Thr Asp Val Val Glu Ala Lys Ser Leu Gly Ile Arg Pro Asn Tyr Ile Asp
841  GGC GAT GTC ACA GAT GTG GTC GAG GCA AAG TCG CTG GGC ATC AGA CCC AAC TAC ATC GAC

301  Ile Tyr Ser Ala Ser Trp Gly Pro Asp Asp Asp Gly Lys Thr Val Asp Gly Pro Gly Arg
901  ATT TAC AGT GCC AGC TGG GGG CCG GAC GAC GGC AAG ACG GTG GAC GGG CCC GGC CGA

321  Leu Ala Lys Gln Ala Phe Glu Tyr Gly Ile Lys Lys Gly Arg Gln Gly Leu Gly Ser Ile
961  CTG GCT AAG CAG GCT TTC GAG TAT GGC ATT AAA AAG GGC CGG CAG GGC CTG GGC TCC ATT

341  Phe Val Trp Ala Ser Gly Asn Gly Gly Arg Glu Gly Asp Tyr Cys Ser Cys Asp Gly Tyr
1021 TTC GTC TGG GCA TCT GGG AAT GGC GGG AGA GAG GGG GAC TAC TGC TCG TGC GAT GGC TAC

361  Thr Asn Ser Ile Tyr Thr Ile Ser Val Ser Ser Ala Thr Glu Asn Gly Tyr Lys Pro Trp
1081 ACC AAC AGC ATC TAC ACC ATC TCC GTC AGC GCC ACC GAG AAT GGC TAC AAG CCC TGG

381  Tyr Leu Glu Glu Cys Ala Ser Thr Leu Ala Thr Thr Tyr Ser Ser Gly Ala Phe Tyr Glu
1141 TAC CTG GAA GAG TGT GCC TCC ACC CTG GCC ACC ACC TAC AGC AGT GGG GCC TTT TAT GAG

401  Arg Lys Ile Val Thr Thr Asp Leu Arg Gln Arg Cys Thr Asp Gly His Thr Gly Thr Ser
1201 CGA AAA ATC GTC ACC ACG GAT CTG CAG CGC TGT ACC GAT GGC CAC ACT GGG ACC TCA

421  Val Ser Ala Pro Met Val Ala Gly Ile Ala Leu Ala Leu Glu Ala Asn Ser Gln Leu
1261 GTC TCT GCC CCC ATG GTG GCG GGC ATC GCC TTG GCT CTA GAA GCA AAC AGC CAG TTA

441  Thr Trp Arg Asp Val Gln His Leu Leu Val Lys Thr Ser Arg Pro Ala His Leu Lys Ala
1321 ACC TGG AGG GAC GTC CAG CAC CTG CTA GTG AAG ACA TCC CGG CCG GCC CAC CTG AAA GCG

461  Ser Asp Trp Lys Val Asn Gly Ala Gly His Lys Val Ser His Phe Tyr Gly Phe Gly Leu
1381 AGC GAC TGG AAA GTA AAC GGC GCG GGT CAT AAA GTT AGC CAT TTC TAT GGA TTT GGT TTG

481  Val Asp Ala Glu Ala Leu Val Glu Ala Lys Lys Trp Thr Ala Val Pro Ser Gln His
1441 GTG GAC GCA GAA GCT CTC GTT GAG GCA AAG AAG TGG ACA GCA GTG CCA TCG CAG CAC

501  Met Cys Val Ala Ala Ser Asp Lys Arg Pro Arg Ser Ile Pro Leu Val Gln Val Leu Arg
1501 ATG TGT GTG GCC GCC TCG GAC AAG AGA CCC AGG AGC ATC CCC TTA GTG CAG GTG CTG CGG

521  Thr Thr Ala Leu Thr Ser Ala Cys Ala Glu His Ser Asp Gln Arg Val Val Tyr Leu Glu
1561 ACT ACG GCC CTG ACC AGC TGC GCG GAG CAC TCG GAC CAG CGG GTG GTC TAC TTG GAG
```

FIG. 1A-2

GlyHisLysGlyAlaAlaValAlaPheTrpTrpThrIleGlyTrpProTrpAsnValAm
GGTCATAAAGGTGCGGCAGTGGCGTTCTGGTGGACCCATTGGGTGGCCCTGGAATGTGTAG
GAAGGGGTGTCATGAATTCCTTAAAGGACTCTCCAAATAGCATTAGTTGTTATTATTAA
CTTAAAGGACTCTCCAAATAGCATTAGTTGTTATTAATTGTGTCACAAGAATTT
AAAACGCATGTGCAGCTATTTAAGAAAAGTATCCCGGAAGCTCACAGTGACATTACGGAA
GAACCCTCAGGTCACAAGAGTCTGGGGTCTCCTATACTCTATAACTTTGGCCACACCGAG
ACACCACCCTATACCAATATTTACTCATAGTTCTCTTTAAGCCAGGAGCAATGACGTGTGC
CTATAGTCGCAGCTACTAGGGAAGTTGAGGCAGGAGGATTGCTTGAGCCCAGGAATTTGA
GTCTAGCCTGGACAACACAGCAGGACTCCATCTCTTAAAAAAAAAAATTACTTCCCCCACT
ACTTTTTTTGACATAAAAAAAATGTATTTTTAAAAGGAAACTGTACTACATCTAGTTAATCA
TAGGTTTGATATGTAGTTACGTATTTTTTTCTAATGTGCATTAAAACAAATCCATAATTAT
TAAAATAAATGTTGTTTGTGTGCCAAAAAAAAAAAA

FIG. 2

SerHisPheTyrGlyPheGlyLeuValAspAlaGluAla
5'...ggcctgtctttttcag TTAGCCATTTCTATGGATTTGGTTTGGTGGACGCAGAAGCT
LeuValGluAlaLysLysTrpThrAlaValProSer
CTCGTTGTGGAGGCAAAGAAGTGGACAGTGCCATCG
GlnHisMetCysValAlaAlaSerAspLysArgPro
CAGCACATGTGTGTGGCCGCCTCGGACAAGAGACCCAG gtaaggctctgctgt...3'

FIG. 3

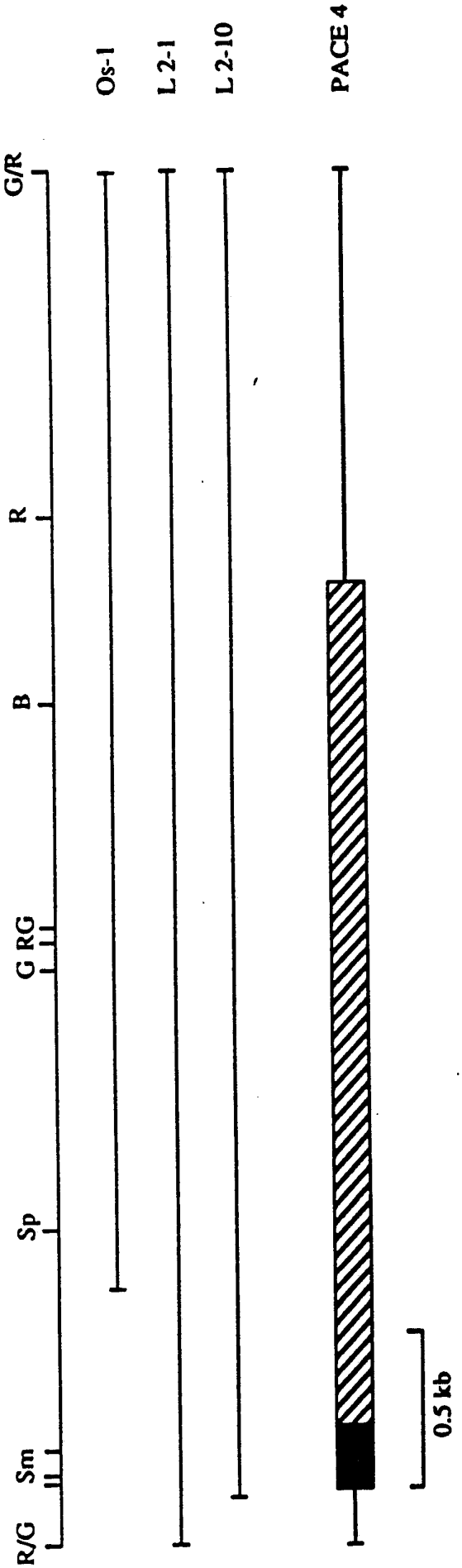


FIG. 4

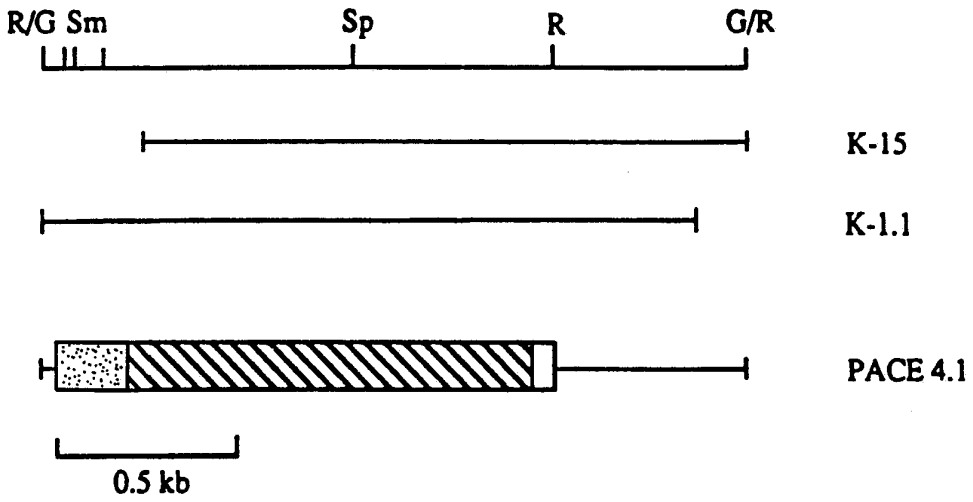


FIG. 5

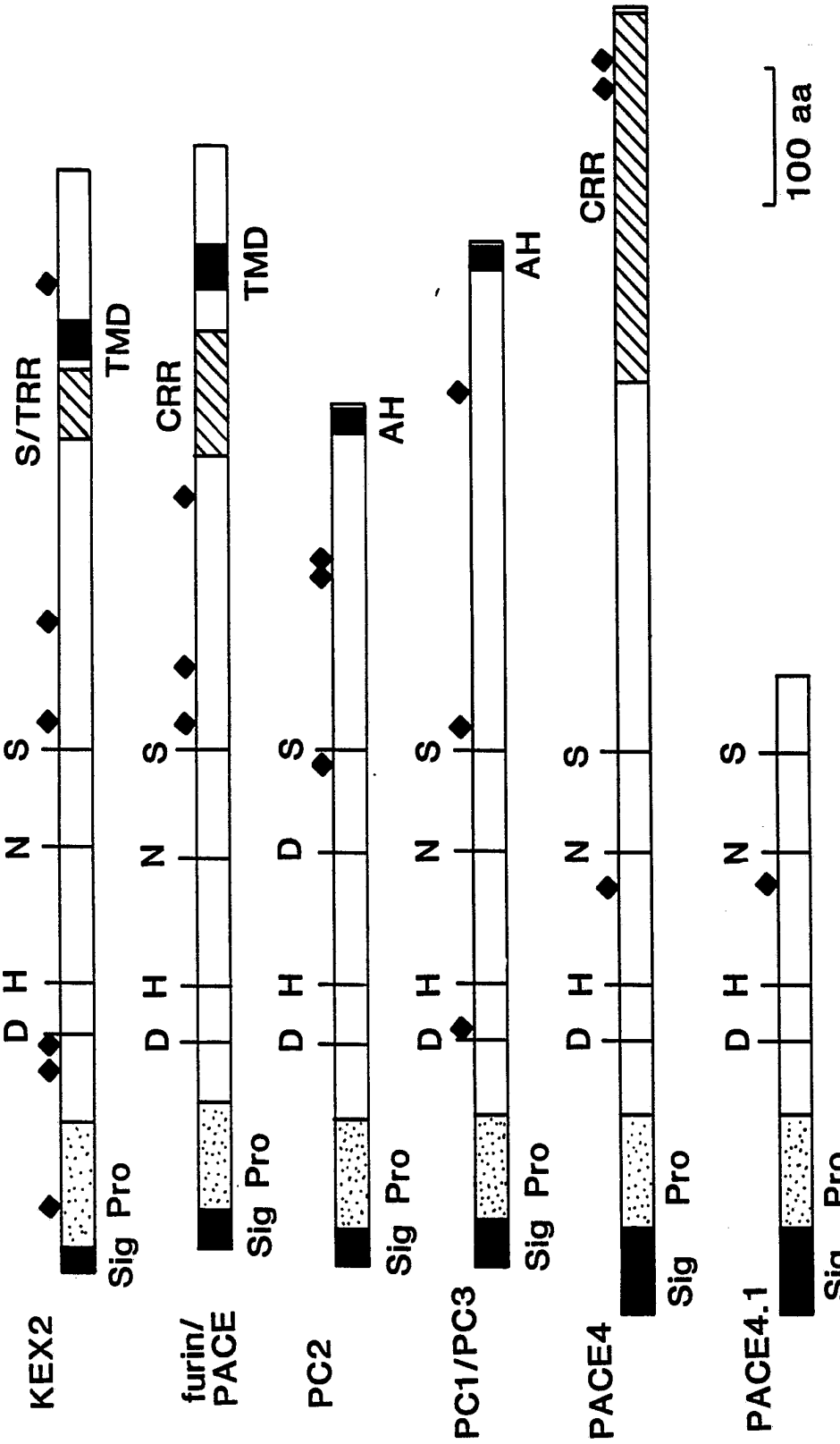

KEX2
furin/
PACE
PC2
PC1/PC3
PACE4
PACE4.1
Sig
Pro
D
H
N
S
S/TRR
TMD
CRR
AH
100 aa

FIG. 6

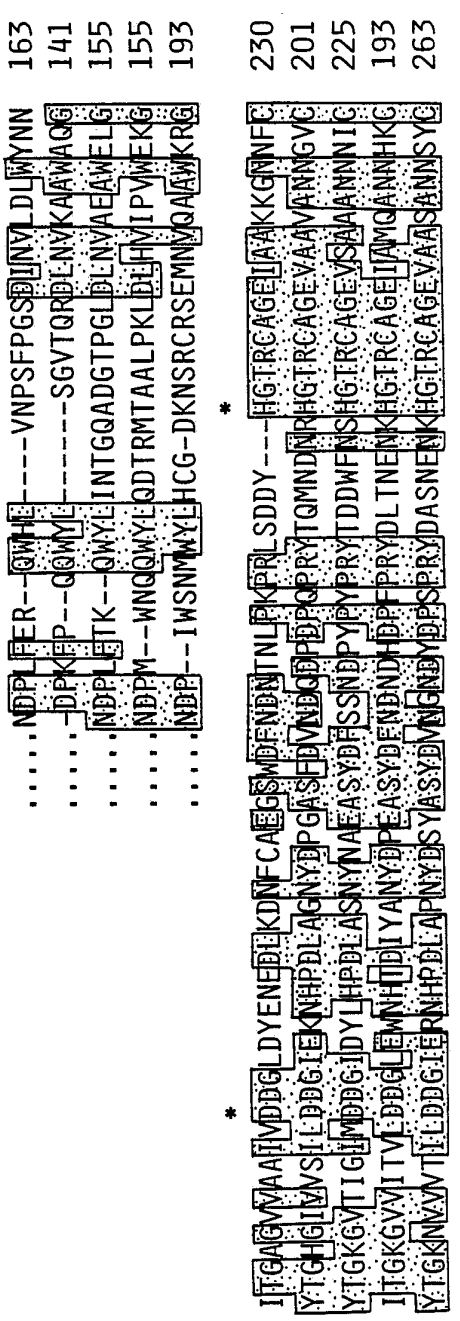


FIG. 7

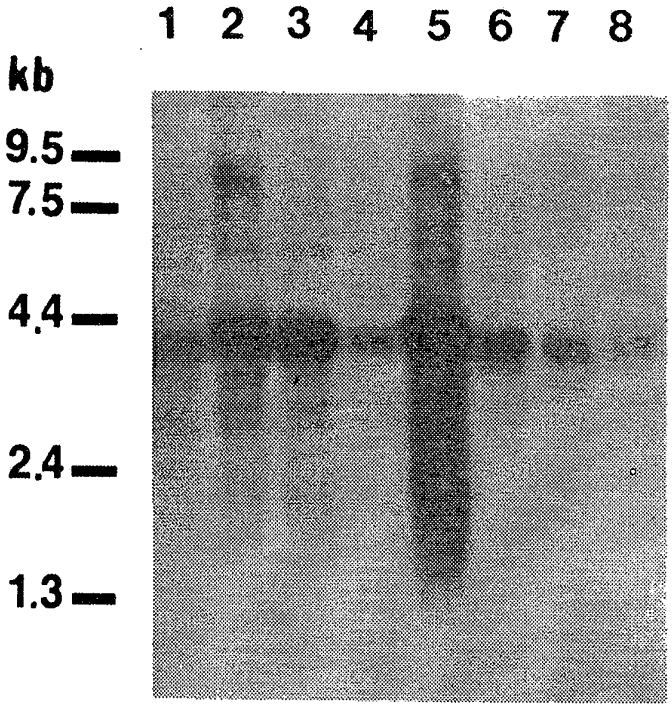


FIG. 8A

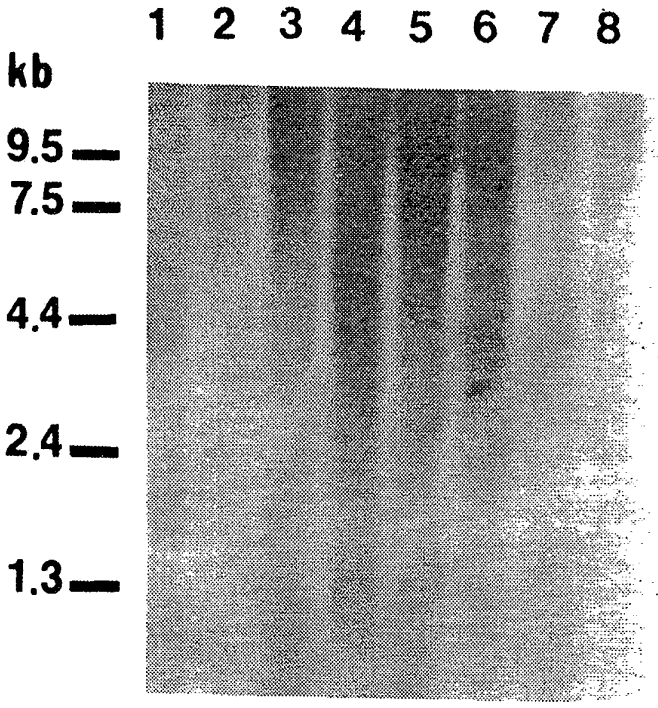


FIG. 8B

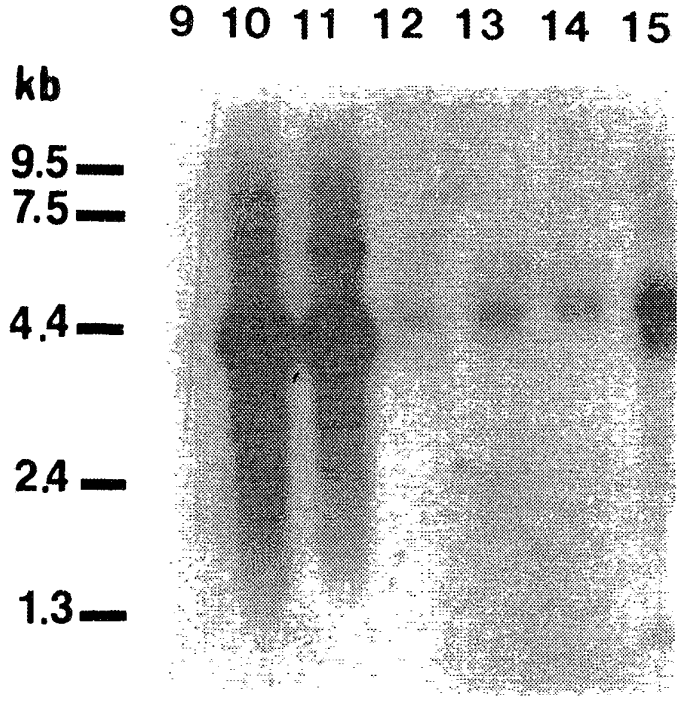

9 10 11 12 13 14 15
kb
9.5
7.5
4.4
2.4
1.3

FIG. 8C

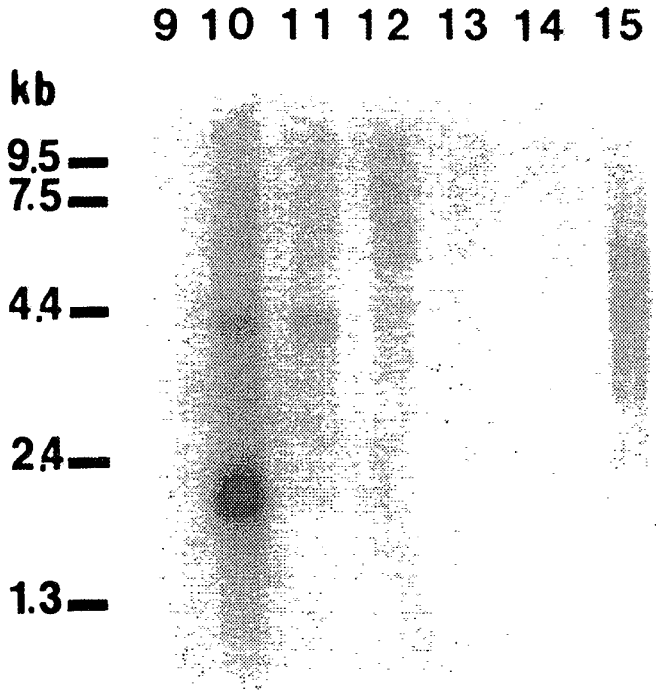

9 10 11 12 13 14 15
kb
9.5
7.5
4.4
2.4
1.3

FIG. 8D

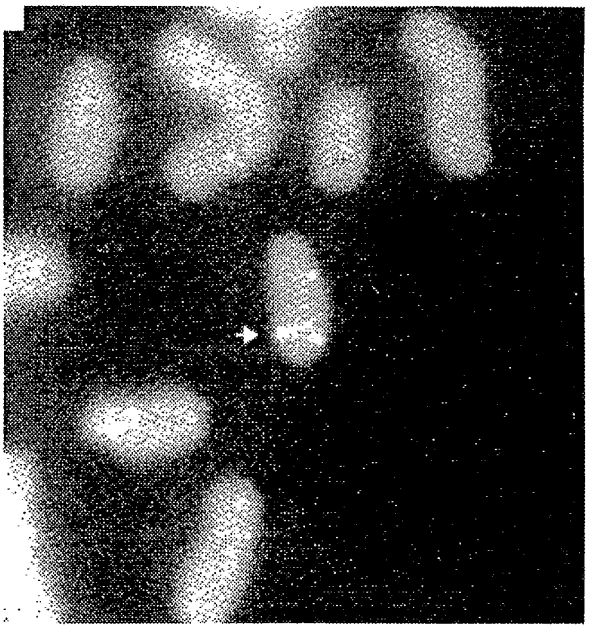


FIG. 9A

FIG. 9B

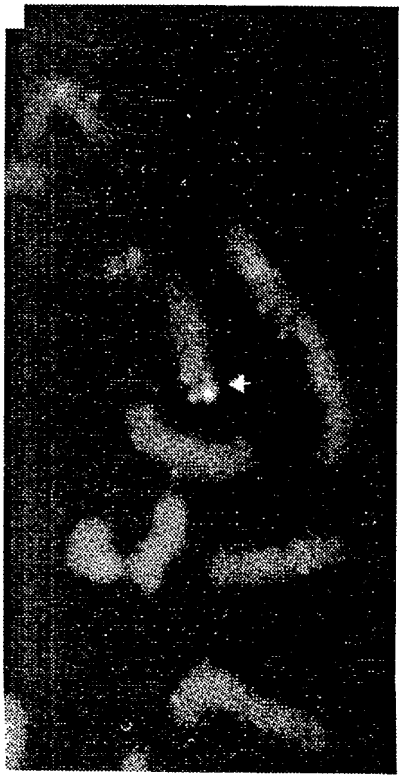


FIG. 9C

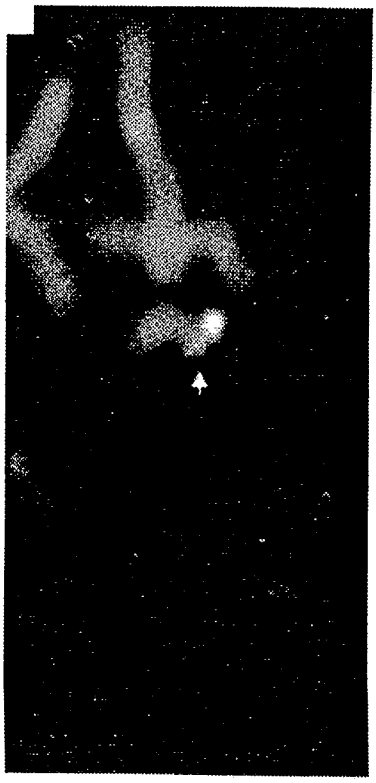


FIG. 9D

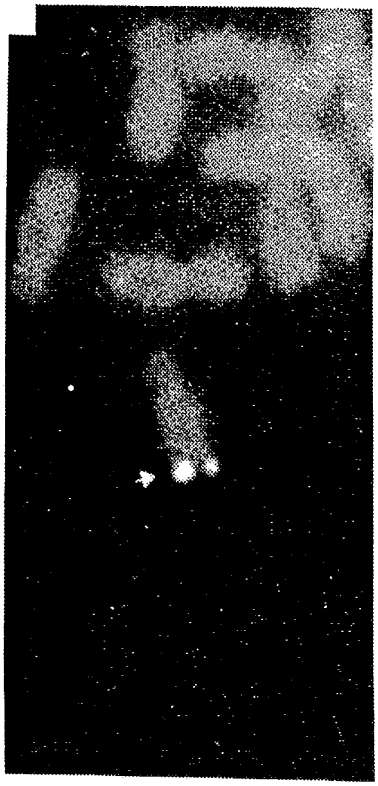


FIG. 9E