(54) Title: NARC-1, NOVEL SUBTILASE-LIKE HOMOLOGS

(57) Abstract: The present invention relates to a newly identified human and mouse programmed cell death (PCD) protein having homology to mammalian subtilases. The invention also relates to polynucleotides encoding the protein. The invention further relates to methods using the polypeptides and polynucleotides as a target for diagnosis and treatment in disorders mediated by or related to the protein. The invention further relates to drug-screening methods using the polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the polypeptides and polynucleotides. The invention further relates to procedures for producing the polypeptides and polynucleotides.
For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.
NARC-1, NOVEL SUBTILASE-LIKE HOMOLOGS

FIELD OF THE INVENTION

The present invention relates to newly identified human and mouse programmed cell death (PCD) proteins having homology to a mammalian subtilase-like protein family, including prohormone convertases. The invention also relates to polynucleotides encoding the protein. The invention further relates to methods using the polypeptides and polynucleotides as a target for diagnosis and treatment in disorders mediated by or related to the protein. The invention further relates to drug-screening methods using the polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the polypeptides and polynucleotides. The invention further relates to procedures for producing the polypeptides and polynucleotides.

BACKGROUND OF THE INVENTION

Programmed Cell Death

In multicellular organisms, homeostasis is maintained by balancing the rate of cell proliferation against the rate of cell death. Cell proliferation is influenced by numerous growth factors and the expression of proto-oncogenes, which typically encourage progression through the cell cycle. In contrast, numerous events, including the expression of tumor suppressor genes, can lead to an arrest of cellular proliferation.

In differentiated cells, a particular type of cell death called apoptosis occurs when an internal suicide program is activated. This program can be initiated by a variety of external signals as well as signals that are generated within the cell in response to, for example, genetic damage. Dying cells are eliminated by phagocytes, without an inflammatory response.
Programmed cell death is a highly regulated process (Wilson (1998) *Biochem. Cell. Biol.* 76:573-582). The death signal is then transduced through various signaling pathways that converge on caspase-mediated degradative cascades resulting in the activation of late effectors of morphological and physiological aspects of apoptosis, including DNA fragmentation and cytoplasmic condensation. In addition, regulation of programmed cell death may be integrated with regulation of energy, redox- and ion homeostasis in the mitochondria (reviewed by (Kroemer, 1998)), and/or cell-cycle control in the nucleus and cytoplasm (reviewed by (Choisy-Rossi and Yonish-Rouach, 1998; Dang, 1999; Kasten and Giordano, 1998)). Many mammalian genes regulating apoptosis have been identified as homologs of genes originally identified genetically in *Caenorhabditis elegans* or *Drosophila melanogaster*, or as human oncogenes. Other programmed cell death genes have been found by domain homology to known motifs, such as death domains, that mediate protein-protein interactions within the programmed cell death pathway.

The mechanisms that mediate apoptosis include, but are not limited to, the activation of endogenous proteases, loss of mitochondrial function, and structural changes, such as disruption of the cytoskeleton, cell shrinkage, membrane blebbing, and nuclear condensation due to degradation of DNA. The various signals that trigger apoptosis may bring about these events by converging on a common cell death pathway that is regulated by the expression of genes that are highly conserved.

Caspases (cysteine proteases having specificity for aspartate at the substrate cleavage site) are central to the apoptotic program. These proteases are responsible for degradation of cellular proteins that lead to the morphological changes seen in cells undergoing apoptosis. One of the human caspases was previously known as the interleukin-1β (IL-1β) converting enzyme (ICE), a cysteine protease responsible for the processing of pro-IL-1β to the active cytokine. Overexpression of ICE in Rat-1 fibroblasts induces apoptosis (Miura et al. (1993) *Cell* 75:653).

Many caspases and proteins that interact with caspases possess domains of about 60 amino acids called a caspase recruitment domain (CARD). Apoptotic proteins may bind to each other via their CARDs. Different subtypes of CARDs may confer binding specificity, regulating the activity of various caspases. (Hofmann et al. (1997) *TIBS* 22:155).
The functional significance of CARDs have been demonstrated in two recent publications. Duan et al. (1997) Nature 385:86 showed that deleting the CARD at the N-terminus of RAIDD, a newly identified protein involved in apoptosis, abolished the ability of RAIDD to bind to caspases. In addition, Li et al. (1997) Cell 91:479 showed that the N-terminal 97 amino acids of apoptotic protease activating factor-1 (Apaf-1) was sufficient to confer caspase-9-binding ability.

Thus, programmed cell death (apoptosis) is a normal physiological activity necessary to proper and differentiation in all vertebrates. Defects in apoptosis programs result in disorders including, but not limited to, neurodegenerative disorders, cancer, immunodeficiency, heart disease and autoimmune diseases (Thompson et al. (1995) Science 267:1456).

In vertebrate species, neuronal programmed cell death mechanisms have been associated with a variety of developmental roles, including the removal of neuronal precursors which fail to establish appropriate synaptic connections (Oppenheim et al. (1991) Annual Rev. Neuroscience 14:453-501), the quantitative matching of pre- and post-synaptic population sizes (Herrup et al. (1987) J. Neurosci. 7:829-836), and sculpting of neuronal circuits, both during development and in the adult (Bottjer et al. (1992) J. Neurobiol. 23:1172-1191).


In addition, improper expression of genes involved in apoptosis has been implicated in carcinogenesis. Thus, it has been shown that several “oncogenes” are in fact involved in apoptosis, such as in the Bcl family.

Accordingly, genes involved in apoptosis are important targets for therapeutic intervention. It is important, therefore, to identify novel genes involved in apoptosis or to discover whether known genes function in this process.
Subtilases/Proprotein Convertases

Subtilisin is an alkaline serine protease produced in various strains of Bacillus. Since it was first identified in B. subtilis, the enzyme was called "subtilisin." Numerous variants have been identified and studied. This exoenzyme was found to belong to a large family of proteins spanning both prokaryotes and eukaryotes, variously designated, including "subtilases," "subtilisin-related serine proteases," and further designated "proprotein" or "prohormone convertases."

Subtilisin is first produced as a precursor, pre- pro-subtilisin. The pre-sequence is a signal peptide which functions to export the protein across the membrane. The pro-sequence is essential for mediating proper folding of the mature catalytic region. Propeptide-mediated folding in subtilisin has been reviewed in Shinde et al. (In Subtilisin Enzymes; Practical Protein Engineering, 1996, Plenum Press, NY., pgs. 147-153). The function of the propeptide led to the concept of an intramolecular chaperone in subtilisin and subsequently in the subtilisin-related enzymes (see below).

The discovery of the fur locus led to the identification of the mammalian family of endoproteases, designated proprotein convertases (see above). These enzymes have a broad functional range and have been reviewed in Steiner, D.F. (1998) Current Opinion in Chemistry and Biology 2:31-39. Currently there are approximately seven members of the family. These catalyze the maturation of various peptide hormones and other precursor proteins and also are critical for virulence of pathogens, including bacterial and viral. Family members fall into two classes based on distribution, those expressed ubiquitously, such as furin and PACE4, and those with a more limited tissue distribution. Of the latter, prohormone convertase PC5/PC6 (differential nomenclature is a result of the naming by various groups) occurs mostly in gastrointestinal tissue. PC7/PC8/LPC is found in lymphoid tissue. PC1/sPC3 and PC2 are mainly restricted to tissues of neuroendocrine origin. PC4 is mainly localized to the testes.

The subtilases act within the secretory pathway to cleave polypeptide precursors at specific basic sites to generate their biologically active forms. Serum proteins, prohormones, receptors, zymogens, viral surface proteins, bacterial toxins, and others are activated by this route. Subtilisin-related serine proteases in the mammalian constitutive secretory pathway have been recently reviewed by Gensberg et al. (1998) Seminars in Cell and Dev. Biol. 9:11-17, summarized below.
These enzymes have also been referred to as Kex2-related serine proteases, or kexins because their discovery followed the characterization of Kex2, a calcium-dependent serine protease isolated from yeast. The subtilisin-related domain of Kex2 was found to be homologous to the human fur open reading frame. Moreover, the complete sequence of furin, the gene product of the fur gene, showed a more extensive similarity with Kex2. The preparation of degenerate PCR primers, designed to identify related sequences, led to the isolation of several other members of the family including PC2, PC1, PACE4, PC5/6, PC8/PC7/LPC and PC8. These family members share the same domain structure (shown in Figure 1 of Gensberg et al.). All the members have a signal peptide that targets the protein to the secretory pathway. All members contain a pro-peptide that provides for correct folding of the active polypeptide and correct secretion from the endoplasmic reticulum. Cleavage of the propeptide is essential for activation and occurs autocatalytically in the endoplasmic reticulum, at least in the case of furin. All members contain a catalytic domain related to the bacterial subtilisins that contains, in the active site, ASP, HIS, and SER, and the oxyanion hole residue ASN, except in PC2 where ASP is substituted. All members also contain the P or middle domain, also referred to as the homo B domain. This domain plays a role in folding. Mutants in the P domain are not autocatalytically processed and remain in the endoplasmic reticulum. Most of the variation is found in the C-terminus. Furin, PC5/6B and PC7/8/LPC have C terminal transmembrane domains. The C terminal domains of furin, PACE4 and PC5/6 include a cysteine rich region. Further, all known family members contain potential glycosylation sites. Inhibition of glycosylation causes rapid degradation of PC1/3 and PC2 in the endoplasmic reticulum. Further, gene regulation and cellular and tissue distribution are unique for each family member.

Furin, later referred to as PACE, is a ubiquitous housekeeping proprotein processing endopeptidase of the constitutive secretory pathway. The furin transcript is expressed in all cell types and encodes a type I membrane protein predominantly localized to the trans-Golgi network and immature secretory granules of neuroendocrine and endocrine cells. The cytosolic tail of furin contains two signals that mediate localization in the late secretory pathway. These include an acid casein kinase II site (CPSDSeEDEG) that retains furin in the trans-Golgi network and a tyrosine motif (YKGI) that serves as a retrieval signal for furin that has escaped to the cell surface,
cycling furin back to the trans-Golgi network via endosomes. A soluble form of furin, generated by cleavage N-terminal to the transmembrane domain, is shed from the cell.

The minimal recognition site for furin is R-XXR. However, the efficiency of cleavage may be modulated by the surrounding sequence. R-Q-P-R-G-W may be cleaved twice as efficiently as R-V-R-R-S-V, for example.

A wide range of protein precursors have been shown to be substrates for furin, such as parathyroid hormone-related peptide, pro-β-nerve growth factor, pro-albumin, complement pro-C3, semaphorins, pro-insulin-like growth factor 1A and integrin α-chain. Furin has been recently reviewed in detail in Nakayama (1997) *Biochem. J.* 327:625-635 and Molloy et al. (1999) *Trends in Cell Biology* 9:28-34, both of which are summarized below.

The expression of furin has been studied in rat development. mRNA is first detected in both endoderm and mesoderm in the primitive streak stage of embryogenesis. Subsequently a distinctly higher level of expression is observed in the heart and liver primordia. In mid and late gestational stages, furin is widely expressed in the peripheral tissues. The expression pattern of furin during embryogenesis is distinct from that of other ubiquitously expressed convertases and from neuroendocrine specific ones. This suggests that furin plays a role in processing various proproteins, such as growth factor precursors, during development. Furin knock-out mice die by e11-12.

The expression of furin is developmentally regulated and appears to control the growth and differentiation of cells such as pancreatic islet cells and gastric mucosal cells.

In furin, propeptide cleavage is not sufficient, although it is a prerequisite, for the activation of furin. After cleavage in the endoplasmic reticulum the propeptide remains associated with the mature furin moiety and functions as a potent autoinhibitor of the endoprotease. Upon transit through the endoplasmic reticulum, with a change of acidic conditions and calcium concentration, the propeptide is released, generating the active furin. This propeptide release requires a second cleavage at the ARG-GLY-VAL-THR-LYS-ARG site in the middle of the propeptide. Mutations in this sequence result in an endoprotease that cannot be activated by acid or calcium treatment in vitro.

Various substrates of furin and sequences around the cleavage sites of precursor proteins are shown in Nakayama, above. Furin is proposed to be responsible for processing precursors of constitutively secreted proteins rather than peptide hormones.
and neuropeptides. These include growth factors, their receptors, plasma proteins involved in blood clotting and complement systems, matrix metalloproteases, viral envelope glycoproteins, and bacterial exotoxins. Furin preferentially recognizes the cleavage sequence ARG-XAA-(LYS/ARG)-ARG. However, cleavage sites of some precursors cleaved by furin do not fully fit this consensus sequence. Accordingly, Nakayama has proposed the following sequence rules governing cleavage by furin: (1) An ARG residue is essential at the P₁ position; (2) In addition to the P₁ ARG, at least two out of the three residues at P₂, P₄ and P₆ are required to be basic for efficient cleavage; (3) At P₁ position an amino acid with a hydrophobic aliphatic side chain is not suitable. The cleavage site specificity determined by coexpression studies is in agreement with that determined by in vitro studies using purified recombinant soluble forms of furin.

Since furin cleavage is essential to produce a wide variety of biologically active proteins, it has been proposed that mutation of the cleavage site may result in genetic disorders. It has been reported that a severe form of hemophilia B is correlated with mutation of the P₄ ARG residue to GLN in pro-factor IX. There have also been many reports of hemophilia B cases with mutations of the P₄, P₂ or P₁ basic residue of pro-factor IX. Further, subjects with extreme insulin resistance were reported to have a mutation of the P₁ ARG residue to SER at the cleavage site of insulin proreceptor.

As discussed above, furin function is implicated in productive viral infection. Proteolytic activation of envelope glycoproteins is necessary for the entry of viruses into host cells. In some cases it has been shown that the cleavability of the envelope glycoproteins is an important determinant for viral pathogenicity. For example, proteins required for infectivity of mammalian influenza viruses and avirulent avian-influenza viruses, which can cause local infection, are susceptible to proteolytic cleavages only in specific cell types, such as in the respiratory and alimentary tract. In contrast, virulent avian-influenza viruses that cause systemic infection are cleaved in a variety of host cells. Similarly, avirulent and virulent Newcastle disease viruses cause local and systemic infections, respectively. A relationship has been suggested between viral pathogenicity and the cleavage site sequence of envelope glycoprotein precursors. In vitro experiments using purified furin have shown that furin is involved in cleavage of the glycoprotein precursors of virulent viruses. Accordingly, the widespread expression
of furin can account for systemic infections by virulent viruses. Furin has also been implicated in the activation of HIV-1 gp160. Furin is also expressed in CD4+ cell lines. However, other proteases may also be involved in gp160 cleavage.

The range of proproteins activated by furin (Table 1 in Molloy et al., above, incorporated herein by reference) is extensive and indicates an importance in fundamental biological processes. Furin is involved in cellular signaling at both the juxtacrine (for example, cell adhesion factors) and paracrine (for example, growth factors and receptors) levels. It can regulate the composition of the extracellular matrix (for example, processing of matrix components and activation of matrix metalloproteases) and contributes to the processes of embryonic induction.

As indicated above, proprotein convertases typically cleave their substrates on the C-terminal side of paired basic amino acids (for example, LYS-ARG\(^\dagger\), ARG-ARG\(^\dagger\). Furin generally requires an additional ARG at the P\(_4\) position for efficient cleavage of substrates (-ARG-X-LYS/ARG-ARG\(^\dagger\)-). The residues that are C-terminal to the cleavage site (P\(^\prime\)) also affect processing efficiency and possibly specificity.

As indicated, the propeptide is a multifunctional domain directing the compartment-specific activation of furin. Because of the role in directing the correct folding of the mature peptide, the propeptide functions as an intramolecular or steric chaperone. Autoproteolytic cleavage occurs at \(\text{\underline{\\underline{\ \ P} R-T-K-R}}_{107} \text{\underline{\underline{\ R-G-}}V-T-K-R}_{75}\). In addition, the prosequence contains an autoinhibitory domain. The catalytic domain is known to contain high and middle affinity calcium binding sites. The P domain is necessary for the activity of furin and other proprotein convertases. The P domain also functions in pH and calcium modulation. This domain also contains a conserved RGD integrin binding motif. Mutation of this site in PC1/3 disrupts proenzyme maturation and catalytic activity. In furin, this motif may function in matrix association.

The transmembrane domain (in furin) is followed by a cytosolic domain that contains sorting information, including multiple clathrin-coat-recruitment motifs that control internalization, budding from the \textit{trans}-Golgi network, and polarized sorting. This region also contains a cluster of acidic amino acids that directs phosphorylation-state-specific \textit{trans}-Golgi network localization and endosomal sorting, and a membrane proximal region that tethers furin to the cortical cytoskeleton.
The trafficking of furin between the trans-Golgi network, cell surface, and endosomes is directed by defined sequence motifs in the cytosolic domain. Localization to the trans-Golgi network and the endosomal routing of furin is dependent upon the phosphorylation state of the acidic cluster. Dephosphorylated furin is delivered to the trans-Golgi network, whereas the phosphorylated enzyme is recycled to the plasma membrane. Furin is retained in a bi-cycling loop by casein kinase II-mediated phosphorylation of serine residues in the acidic cluster motif. Dephosphorylation regulates movement between the loops. Furin molecules delivered to the cell surface can be tethered via binding to a component of the cortical actin cytoskeleton.

Localization of furin to the trans-Golgi network requires the cooperative effect of the phosphorylated acidic cluster motif together with one or more clathrin-coated pit recruitment signals. This bi-cycling has been described in Molloy et al., above.

The furin activation pathway is essentially as follows: (1) Initial synthesis as a zymogen within the neutral pH environment of the endoplasmic reticulum; (2) Rapid autoproteolytic cleavage of the propeptide at the consensus furin site ARG-THR-LYS-ARG_{107}; (3) Endoplasmic reticulum to Golgi transport, along with propeptide cleavage, where the cleaved propeptide remains associated with the enzyme and functions as a potent autoinhibitor; (4) Within the mildly acidic environment of the trans-Golgi network/endosomal system, the propeptide is cleaved autoproteolytically at a P1/P2/P6 ARG-containing furin site (ARG-GLY-VAL-THR-LYS-ARG_{751}), releasing the propeptide fragments and thus providing active furin; (5) Within the late secretory pathway, independent of the activation state, furin is cleaved upstream of its transmembrane domain, potentially functioning in the processing of extracellular substrates (for example, extracellular matrix components).

As indicated, furin is implicated in early development. Disruption of the mouse gene encoding furin results in embryonic lethality. This is associated with several defects, including failure of the heart tube to fuse or to undergo looping morphogenesis and failure of the embryos to undergo axial rotation. The results are consistent with a role for furin in maturation of members of the TGFβ family, particularly bone morphogenetic proteins and nodal-related proteins.

There is also a relationship between furin processing and regulation of the extracellular matrix. Specifically, the soluble (shed) furin is implicated in processing of
extracellular matrix proteins, for example fibrillin and zona pellucida proteins. It also has a role in the activation of matrix metalloproteases, such as BMP-1/procollagen, C-protease and stromelysin-3. Changes in furin-dependent matrix metalloprotease activation can contribute to the metastatic capacity of tumors. The observation that stromelysin-3 activity corresponds with tumor invasiveness supports this possibility.

Finally, as indicated, furin has been associated with pathogenic virulence. Numerous pathogens require cleavage by furin of viral envelope glycoproteins and bacterial toxins for their virulence. Processing is apparently a key determinant in viral tropism. For example, the cleavage site in Ebola virus glycoprotein GP is coupled to the lethality of the virus in humans. Also, a fatal respiratory illness has been traced to a specific strain of avian influenza A. Sequence analysis of genes encoding the HA gene showed a consistent alteration in the viral genomes, the generation of a second consensus furin site.

PC1 and PC2, although not as well characterized as furin, have also been analyzed in relative detail. PC1 and PC2 are primarily expressed in endocrine and neural cells, mostly localizing within the trans-Golgi network or dense core secretory granules. These molecules have been reviewed recently by Muller et al. (2000) *Progress in Nucleic Acid Research and Molecular Biology* 63:60-109. These enzymes participate in the regulated proteolysis of prohormones and are designated prohormone convertases (PC). A human patient exhibiting obesity, hyperproinsulinemia, and hypocortisolemia phenotypes was shown to have nucleotide mutations in each allele of the PC1 gene, resulting in an inactive convertase. In addition, homozygous PC2 null mice are viable but exhibit a hypoglycemia phenotype. PC1 substrates include but are not limited to POMC (ß-LPH, ACTH), proinsulin, proTRH, proENK, proDyn, proglucagon, prorenin, proMCH, proNT, and proCCK. PC2 substrates include but are not limited to POMC, proinsulin, proglucagon, proNT, proENK, proLHRH, proDyn, and proCCK.

Cellular expression and subcellular localization of prohormone convertases have also been reviewed in Seidah, et al. (1997) *Current Opinion in Biotechnology* 8:602-607, summarized herein.

PC4 is exclusively expressed in germ cells of the testes. Homozygous PC4 null mice are viable but have reduced male fertility.
PC5 and PACE4 are widely expressed and detected during early embryonic development. In the adult, PC5 is highly expressed in gut, endothelial and Sertoli cells and in the adrenal cortex.

Accordingly, subtilases/proprotein convertases ("subtilases") are a major target for drug action and development. Thus, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown subtilases or subtilase-like proteins. The present invention advances the state of the art by providing previously unidentified human, mouse, and rat subtilase-like proteins that are regulated in programmed cell death.

SUMMARY OF THE INVENTION

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

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

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

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

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

The invention is thus based on the identification of novel human, mouse, and rat subtilase-like proteins. The amino acid sequence is shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8. The nucleotide sequence is shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and SEQ ID NO:7.

The invention provides isolated subtilase-like polypeptides, including a polypeptide having an amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8.
The invention also provides isolated subtilase-like nucleic acid molecules having a sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7.

The invention also provides variant polypeptides having an amino acid sequence that is substantially homologous to an amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8.

The invention also provides variant nucleic acid sequences that are substantially homologous to a nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7.

The invention also provides fragments of a polypeptide shown in SEQ ID NOS:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8 and nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7, as well as substantially homologous fragments of the polypeptide or nucleic acid.

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

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

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

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

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

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

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

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

DESCRIPTION OF THE DRAWINGS

**Figure 1** shows the human nucleotide sequence NARC1A (SEQ ID NO:1) and the deduced amino acid sequence (SEQ ID NO:2).

**Figure 2** shows an analysis of the human NARC1A amino acid sequence: αβturn and coil regions; hydrophilicity; amphipathic regions; flexible regions; antigenic index; and surface probability plot.

**Figure 3** shows a hydrophobicity plot of the human NARC1A protein.

**Figure 4** shows an analysis of the human NARC1A protein open reading frame for amino acids corresponding to specific functional sites. For the phosphorylation and myristoylation sites, the actual modified residue is the first amino acid.

**Figure 5** shows the rat NARC1 nucleotide sequence (SEQ ID NO:7) and the deduced amino acid sequence (SEQ ID NO:8). Note that the numbers on the left refer to the number of amino acids or nucleotides in the preceding line.

**Figure 6** shows the mouse NARC1 nucleotide sequence (SEQ ID NO:5) and the deduced amino acid sequence (SEQ ID NO:6) designated mouse NARC1. This gene is a murine ortholog of the rat and human NARC1 sequences above. Note that the numbers on the left refer to the number of amino acids or nucleotides in the preceding line.
Figure 7 shows the human NARC1C nucleotide sequence (SEQ ID NO:3) and the deduced amino acid sequence (SEQ ID NO:4). Note that the numbers on the left refer to the number of amino acids or nucleotides in the preceding line.

Figure 8 shows the result of experiments designed to characterize transcriptional characteristics for the rat NARC1 gene. The top panel summarizes results from transcription profiling experiments performed on Smart Chip I for the rat NARC1. Hybridization signals (gene expression intensities) are plotted on the Y-axis. Experiments are listed along the X-axis in the following order from left to right: 1, 3, 6, 12, 24 h post serum add-back (SA), 1, 3, 6, 12 h post KCl plus serum withdrawal (KCl/S), 1, 3, 6, 12 h post sham KCl treatment (KCl C), 1, 3, 6, 12 h KCl withdrawal (KCl), 2, 4, 8, 12 h post sham kainic acid treatment (KT C), and 2, 4, 8, 12 h post kainic acid treatment (KT). The RT-PCR panel shows confirmation of the transcription profiling result at 3 h post KCl plus serum withdrawal (>4X upregulation) (See Materials and Methods section of U.S. Provisional Patent Application 60/161,188). Rat NARC1 was originally cloned by differential display (RADE) (U.S. Patent Application 09/393,174). The transcript size of rat NARC1 measured by a multiple tissue Northern (bottom panel) was 3.4 kb. The result of multiple tissue Northern indicated high levels of rat NARC1 expression in the liver, and less expression in the kidney and testes. The signal in testes indicates the presence of a shorter isoform consistent with the size of human NARC1A (Figure 1). Human NARC1C (Figure 7) is an ortholog of the larger rat splice variant.

DETAILED DESCRIPTION OF THE INVENTION

Polypeptides

The invention is based on the identification of a novel human subtilase-like protein which is regulated in programmed cell death (apoptosis).

Programmed cell death (PCD) in rat cerebellar granule neurons (CGNs) induced by potassium (K⁺) withdrawal has been shown to depend on de novo RNA synthesis. This transcriptional component of CGN programmed cell death was characterized using a custom-built brain-biased cDNA array representing over 7000
different rat genes. Consistent with carefully orchestrated mRNA regulation, the profiles of 234 differentially expressed genes segregated into distinct temporal groups (immediate early, early, middle, and late) encompassing genes involved in distinct physiological responses including cell-cell signaling, nuclear reorganization, apoptosis, and differentiation. A set of 64 genes, including 22 novel genes, were regulated by both K\textsuperscript+ withdrawal and kainate treatment. Thus, by using array technology, physiological responses at the transcriptional level were characterized and novel genes induced by multiple models of programmed cell death were identified. The rat NARC1 was among these genes.

The above experimental subject matter was disclosed in U.S. Provisional Patent Application No. 60/161,188, filed October 22, 1999, entitled "Nucleic Acid Molecules Derived From Rat Brain and Programmed Cell Death Models." The novel genes included the gene designated “NARC1.” This gene is further disclosed in U.S. Provisional Patent Application No. 60/099,616, entitled "Neuronal Cell Death Associated Molecules and Uses Therefor", now U.S. Patent Application No. 09/393,174. In the present application, a human NARC1 ortholog has been identified, designated NARC1A.

Accordingly, the human ortholog was cloned from a cDNA library of human keratinocytes treated with KGF, GF and cycloheximide.

The invention thus relates to novel human, mouse, and rat subtilase-like proteins having a deduced amino acid sequence shown in Figures 1, and 5-7 (SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8).

The polypeptide encoded by the sequence, is incorporated herein by reference and controls in the event of any conflict, such as a sequencing error, with description in this application.

“Subtilase-like polypeptide” or “subtilase-like protein” refers to a polypeptide in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8. The term “subtilase-like protein” or “subtilase-like polypeptide,” however, further includes the numerous variants described herein, as well as fragments derived from the full-length subtilase-like proteins and their variants.

The present invention thus provides an isolated or purified subtilase-like polypeptide and variants and fragments thereof.
Based on a BLAST search, highest homology of the human protein in Figure 1 was shown to an aqualysin precursor. Homology was also shown to subtilase-like proteins in other organisms and to prohormone convertases.

The human protein in Figure 1 is expressed in tissues that include but are not limited to testes and liver. High relative expression occurs in liver.

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

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

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

A subtilase-like polypeptide is also considered to be isolated when it is part of a membrane preparation or is purified and then reconstituted with membrane vesicles or liposomes.

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

In one embodiment, the subtilase-like polypeptide comprises an amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8. However, the invention also encompasses sequence variants. Variants include a substantially homologous protein encoded by the same genetic locus in an organism, i.e., an allelic variant. The human subtilase-like protein in Figure 1 has been mapped to human chromosome 1p32. Two diseases are known to map at this locus. These include muscle-eye-brain disease at 1p34-p32 (MEB) and Bartter Syndrome, infantile, with sensorineural deafness (BSND), at 1p31, both of which are discussed in more detail herein below.

Variants also encompass proteins derived from other genetic loci in an organism, but having substantial homology to a subtilase-like protein of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8. Variants also include proteins substantially homologous to the subtilase-like protein but derived from another organism, i.e., an ortholog. Variants also include proteins that are substantially homologous to the subtilase-like protein that are produced by chemical synthesis. Variants also include proteins that are substantially homologous to the subtilase-like protein that are produced by recombinant methods. It is understood, however, that variants exclude any amino acid sequence disclosed prior to the invention.

As used herein, two proteins (or a region of the proteins) are substantially homologous when the amino acid sequences are at least about 50-55%, 55-60%, 60-65%, 65-70%, 70-75%, typically at least about 80-85%, and most typically at least about 90-95% or more homologous. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence hybridizing to the nucleic acid sequence, or portion thereof, of a sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 under stringent conditions as more fully described below.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid
sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid “identity” is equivalent to amino acid or nucleic acid “homology”). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by the subtilase-like protein. Similarity is determined by conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie et al. (1990) Science 247:1306-1310.

TABLE 1. Conservative Amino Acid Substitutions.
| Aromatic       | Phenylalanine |
|               | Tryptophan    |
|               | Tyrosine      |
| Hydrophobic   | Leucine       |
|               | Isoleucine    |
|               | Valine        |
| Polar         | Glutamine     |
|               | Asparagine    |
| Basic         | Arginine      |
|               | Lysine        |
|               | Histidine     |
| Acidic        | Aspartic Acid |
|               | Glutamic Acid |
| Small         | Alanine       |
|               | Serine        |
|               | Threonine     |
|               | Methionine    |
|               | Glycine       |


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

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

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

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

Variant polypeptides can be fully functional or can lack function in one or more activities. Thus, in the present case, variations can affect the function, for example, of one or more of the regions corresponding to the prodomain, catalytic domain, P domain, cysteine-rich domain, transmembrane domain, and cytosolic domain. Functions that can
be affected include but are not limited to autoproteolysis, intracellular chaperone function, propeptide processing, and autoinhibitory function in the prodomain, the ability to be modulated by pH and calcium, cell adhesion/integrin-binding, and collateral catalytic activity in the P domain, and cell surface tethering, TGN localization, and casein kinase II phosphorylation in the cytosolic domain.

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

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

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

Useful variations further include alteration of catalytic activity. For example, one embodiment involves a variation in the catalytic domain that results in more or less affinity for the substrate propeptide. Another variation in this domain would result in greater or lesser rates of hydrolysis of propeptide substrate. A further variation in the catalytic domain results in altered specificity for the substrate propeptide, for example affinity for another (different) substrate which can include affinity for additional substrates or loss of specificity for the native substrate. Another variation is alteration of autocatalytic activity. This in turn would affect intramolecular chaperone functions. A further variation is one that affects the ability to be activated, for example by pH or calcium. A further variation includes a variation in the targeting potential. For example, a variation in the ability to be phosphorylated by casein kinase II could affect intracellular trafficking. Another variation involves an alteration in the acidic cluster motif in the cytosolic domain which results in changes in intracellular localization. A further variation includes one that prevents truncation of the molecule and hence affects extracellular matrix-associated functions. Another useful variation provides a fusion
protein in which one or more domains or subregions are operationally fused to one or more domains or subregions from a different subtilase, subtilase-like protein, prohormone convertase, or proprotein convertase.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham et al. (1985) Science 244:1081-1085). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity, such as autoproteolysis or propeptide proteolysis in vitro, or in vitro or in vivo activity that is dependent upon autoproteolytic cleavage or propeptide proteolytic cleavage, such as cell proliferation, development, V-ATPase function, extracellular matrix formation, inflammation, apoptosis/programmed cell death, and viral and bacterial pathogenesis and toxicity, as well as other effects disclosed herein. Sites that are critical, for example, for propeptide binding, can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith et al. (1992) J. Mol. Biol. 224:899-904; de Vos et al. (1992) Science 255:306-312).

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

The invention thus also includes polypeptide fragments of the subtilase-like protein. Fragments can be derived from an amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8. However, the invention also encompasses fragments of the variants of the subtilase-like proteins as described herein.

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

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

Biologically active fragments (peptides which are, for example, 5, 7, 10, 12, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) can comprise a domain or motif, e.g., as discussed above, as well as functional sites shown in Figure 4 herein.
Such domains or motifs can be identified by means of routine computerized homology searching procedures or by routine assays, such as those disclosed herein.

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

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

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

Non-limiting examples of antigenic polypeptides that can be used to generate antibodies include but are not limited to peptides derived from an extracellular site. Regions having a high antigenicity index are shown in Figure 2. However, intracellularly-made antibodies ("intrabodies") are also encompassed, which would recognize intracellular peptide regions.


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

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

In one embodiment the fusion protein does not affect the subtilase-like protein function per se. For example, the fusion protein can be a GST-fusion protein in which the subtilase-like protein sequences are fused to the C-terminus of the GST sequences. Other types of fusion proteins include, but are not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL-4 fusions, poly-His fusions and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant subtilase-like protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence. Therefore, in another embodiment, the fusion protein contains a heterologous signal sequence at its N-terminus.

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

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

Another form of fusion protein is one that directly affects the functions of the protein of the invention. Accordingly, a subtilase-like polypeptide is encompassed by the present invention in which one or more of the native protein domains (or parts thereof) has been replaced by homologous domains (or parts thereof) from another subtilase or subtilase-like protein. Accordingly, various permutations are possible. For example, the propeptide or subregion thereof can be replaced with the propeptide or subregion thereof from another subtilase or subtilase-like protein. As a further example, the catalytic domain or subregions thereof can be replaced; the P domain or subregion thereof can be replaced; the carboxyterminal region or parts thereof can be replaced; the transmembrane domain or parts thereof can be replaced; furthermore, domains not present in the native molecule could be added. These might include a cysteine-rich region, transmembrane region, or other carboxyterminal region if not present in the subtilase-like protein of the invention. Thus, chimeric proteins can be formed in which one or more of the native domains or subregions has been replaced by another.

Additionally, chimeric proteins can be produced in which one or more functional sites is derived from a different isoform, or from another subtilase or subtilase-like protein. It is understood however that sites could be derived from subtilases or subtilase-like proteins that occur in the mammalian genome but which have not yet been discovered or characterized. Such sites include, but are not limited to, those discussed above that affect such functions as autoproteolysis, substrate processing, secretion, subcellular localization, and specific membrane association, such as with the plasma membrane.

The isolated subtilase-like protein can be purified from cells that naturally express it, such as liver and testes, especially purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods.
In one embodiment, the protein is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the subtilase-like polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques.

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

Accordingly, the polypeptides also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature polypeptide or a proprotein sequence.

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

Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of

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

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. Blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally-occurring and synthetic polypeptides. For instance, the aminoterminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

The modifications can be a function of how the protein is made. For recombinant polypeptides, for example, the modifications will be determined by the host cell posttranslational modification capacity and the modification signals in the polypeptide amino acid sequence. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell.

Insect cells often carry out the same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to efficiently express mammalian proteins having native patterns of glycosylation. Similar considerations apply to other modifications.

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

Cell-based and cell-free assays directed to expression or function of the NARC subtilase-like proteins are applicable to the uses disclosed herein. Cell-free assays include but are not limited to cleavage of substrate precursors and analogs, for example as disclosed in Nakayama et al., above. Cellular assays include recombinant cells coexpressing substrate precursor and the subtilase protein, such as disclosed in Nakayama et al., above. See also the assays disclosed in Wise et al. (1990) Proc. Natl. Acad. Sci. USA 87:9378-9382, Bresnahan et al. (1990) J. Cell Biol. 111:2851-2859, Van de Ven et al. (1990) Mol. Biol. Rep. 14:265-275, and Misumi et al. (1991) J. Biol. Chem. 266:16954-16959. Assays related to cellular toxin sensitivity include assays in RPE 40 cells, for example, as in Nakayama et al., above. Assays for pathogenic virulence can also be performed in transgenic animals as disclosed in Nakayama et al., above. Coexpression of precursor substrates and subtilases are also disclosed in Creemers et al., above and in Jutras et al. (1997) J. Biol. Chem. 272:15184-15188. Recombinant production of subtilases is disclosed in Seidah et al., above, and also in references cited therein (21, 30, 35-40). Moreover, recombinant production in milk of subtilase enzymes is disclosed in Seidah et al., above, and also in Lamango et al. (1996) Arch. Biochem. Biophys. 330:238-250. Coexpression of substrate precursors and subtilase enzymes are also disclosed in Seidah et al., above, and in references cited therein (35, 36, 37 and 42). Further, transgenic coexpression is also disclosed in Seidah et al., above, in Velander et al. (1997) Scientific American 276:70-74 and in Subramaniam et al. (1996) Ann. NY Acad. Sci. 782:87-96. All of these references are incorporated herein by reference for disclosure of these assays. It is also understood that these assays apply not only to polypeptide uses but also to uses of any of the nucleic acids or antibodies disclosed herein.

Further, apoptosis-specific assays may be used to identify modulators of any of the target nucleic acids or proteins of the present invention, which proteins and/or nucleic acids are related to apoptosis. Accordingly, an agent that modulates the level or activity of any of these nucleic acids or proteins can be identified by means of apoptosis-specific assays. For example, high throughput screens exist to identify apoptotic cells by the use of chromatin or cytoplasmic-specific dyes. Thus, hallmarks of apoptosis, cytoplasmic condensation and chromosome fragmentation, can be used
as a marker to identify modulators of any of the genes related to programmed-cell death described herein. Other assays include, but are not limited to, the activation of specific endogenous proteases, loss of mitochondrial function, cytoskeletal disruption, cell shrinkage, membrane blebbing, and nuclear condensation due to degradation of DNA.


Apoptosis can be triggered by the addition of apoptosis-promoting ligands to a cell in culture or in vivo. Apoptosis can also be triggered by decreasing or removing an apoptosis-inhibiting or survival-promoting ligand. Accordingly, apoptosis is triggered in view of the fact that the cell lacks a signal from a cell surface survival factor receptor. Ligands include, but are not limited to, FasL. Death-inhibiting ligands include, but are not limited to, IL-2. See Hetts et al. (1998) JAMA 279:300-307 (incorporated by reference in its entirety for teaching of ligands involved in active and passive apoptosis pathways). Central in the pathway, and also serving as potential molecules for inducing (or releasing from inhibition) apoptosis pathways include FADD, caspases, human CED4 homolog (also called apoptotic protease activating factor 1), the Bcl-2 family of genes including, but not limited to, apoptosis promoting (for example, Bax and Bad) and apoptosis inhibiting (for example, Bcl-2 and Bcl-x) molecules. See Hetts et al., above.

Multiple caspases upstream of caspase-3 can be inhibited by viral proteins such as cowpox, CrmA, and baculovirus, p35. Synthetic tripeptides and tetrapeptides inhibit caspase-3 specifically (Hetts, above).

Accordingly, cellular and animal models also exist for studying expression or function of the subtilase-like protein sequences in apoptosis and with regard to their effect on apoptosis. Such model systems can be applied in the context of the assays described herein below, for example the effect of specific mutations in the subtilase-like protein, the effect of compounds on the subtilase-like protein, and any of the other assays in which the effect of altered expression or activity of the subtilase-like protein is within the context of effects on apoptosis.
The protein sequences of the present invention can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul et al. (1990) J. Mol. Biol. 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) Nucleic Acids Res. 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The subtilase-like polypeptides are useful for producing antibodies specific for the subtilase-like protein, regions, or fragments. Regions having a high antigenicity index score are shown in Figure 2.

A polypeptide and fragments and sequences thereof and antibodies specific thereto can be used to map the location of the gene encoding the polypeptide on a chromosome. This mapping can be carried out by specifically detecting the presence of the polypeptide in members of a panel of somatic cell hybrids between cells of a first species of animal from which the protein originates and cells from a second species of animal and then determining which somatic cell hybrid(s) expresses the polypeptide and noting the chromosome(s) from the first species of animal that it contains. For examples of this technique, see Pajunen et al. (1988) Cytogenet. Cell Genet. 47:37-41 and Van Keuren et al. (1986) Hum. Genet. 74:34-40. Alternatively, the presence of the polypeptide in the somatic cell hybrids can be determined by assaying an activity or property of the polypeptide, for example, enzymatic activity, as described in Bordelon-Riser et al. (1979) Somatic Cell Genetics 5:597-613 and Owerbach et al. (1978) Proc. Natl. Acad. Sci. USA 75:5640-5644.

The subtilase-like polypeptides are useful for providing desired amounts, including commercially valuable amounts, of a mature protein from a proprotein precursor. Accordingly, in one embodiment the present invention is valuable in that
large amounts of a precursor protein can be produced in a recombinant cell in which the
subtilase-like protein of the present invention is also overexpressed. This allows for the
production of relatively large amounts of a mature protein produced by subtilase
cleavage.

The subtilase-like polypeptides are also useful for producing reagents that inhibit
viral or bacterial infection. Accordingly, the production of a propeptide that acts as a
potent competitive inhibitor of the natural subtilase-like protein can be used to prevent
the processing of bacterial endotoxins and viral envelope glycoproteins and hence to
prevent infection.

The subtilase-like polypeptides of the invention are also useful as a screen for
developing inhibitors of protein activation. Such inhibitors are useful, among other uses,
for preventing pathogenic infection. In this regard, the polypeptides are useful in drug
screening assays as described further herein below.

The subtilase-like polypeptides are useful for biological assays related to the
subtilase-like proteins. Such assays involve any of the known subtilase functions or
activities or properties useful for diagnosis and treatment of subtilase-like protein-related
conditions, such as those disclosed herein.

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

Determining the ability of the test compound to interact with the subtilase-like
protein can also comprise determining the ability of the test compound to preferentially
bind to the polypeptide as compared to the ability of a known binding molecule to bind
to the polypeptide. Such molecules include but are not limited to glycosylation
enzymes, phosphorylation enzymes such as casein kinase II, substrate precursor proteins,
cleaved propeptides, and membrane components, for example those that interact with a
transmembrane domain. Substrates include any of those disclosed herein known to be
processed by subtilases, that include but are not limited to growth factors and hormones,
including mouse pro-β-nerve growth factor, porcine pro-brain-derived neurotrophic
factor, human pro-neurotrophin-3, human pro-transforming growth factor β1, rat pro-
Müllerian inhibiting substance, human pro-insulin-like growth factor I, human proendothelin-1, human pro-parathyroid hormone-related peptide, human pro-parathyroid hormone; receptors, including human insulin pro-receptor, human hepatocyte growth factor pro-receptor, human pro-LRP, human integrin α3-chain, human integrin α6-chain; plasma proteins, including human proalbumin, rat complement pro-C3, human pro-factor IX, human pro-factor X, human pro-von Willebrand Factor, human proprotein C; matrix metalloproteinases, including human stromelysin-3, human MT-MMP1; viral envelope glycoproteins, including human immunodeficiency virus gp160, human cytomegalovirus glycoprotein B, mouse mammary tumor virus-7 superantigen, avian influenza virus A hemagglutinin, measles virus F₀, Newcastle disease virus F₀, Sindbis virus gpE2, human parainfluenza virus type 3 F₀; bacterial exotoxins, including anthrax toxin protective antigen, diphtheria toxin, Pseudomonas exotoxin A, Shiga toxin; and others, including human pro-furin, rat pro-endopeptidase 3.4.24. and 18, and mouse pro-7B2.

The polypeptides can be used to identify compounds that can modulate the subtilase-like protein activity. Such compounds, for example, can increase or decrease affinity or rate of binding to substrate, compete with substrate for binding to the protein, or displace substrate bound to the protein. Both subtilase-like protein and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the protein. These compounds can be further screened against a functional subtilase-like protein to determine the effect of the compound on the protein activity. Compounds can be identified that activate (agonist) or inactivate (antagonist) the protein to a desired degree. Modulatory methods can be performed in vitro (e.g., by culturing the cell with the agent) or, alternatively, in vivo (e.g., by administering the agent to a subject.

The subtilase-like polypeptides can be used to screen a compound for the ability to stimulate or inhibit interaction between the protein and a target molecule that normally interacts with the protein. The target can be any of the molecules with which the protein interacts as described herein. The assay includes the steps of combining the protein with a candidate compound under conditions that allow the protein or fragment to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the
protein and the target. Such consequences include production of a mature substrate molecule, for example mature insulin from pro-insulin, or include the biological consequences of cleavage (or lack thereof), such as effects on embryogenesis, formation of extracellular matrix, pathogen virulence, cell proliferation, inflammation, apoptosis, blood clotting and complement function, cellular differentiation, metabolic activity, cell adhesion, cell signaling, and tumor formation. Moreover, such end results can also be assayed at the level of the organism to further include symptoms such as obesity, tumor formation, endocrine disorders, embryonic induction, bleeding time, and other effects of abnormal processing, including but not limited to those abnormal processing events disclosed herein.

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

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


Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam et al. (1991) Nature 354:82-84; Houghten et al. (1991) Nature 354:84-86) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang et al. (1993) Cell 72:767-778); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')2, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

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

Another candidate compound is a propeptide that acts as a competitive inhibitor of the mature subtilase-like protein.

WO 98/37910 discloses peptide inhibitors of propeptide/prohormone convertases. These inhibitors can be used to inhibit propeptide/prohormone convertases and to treat such disorders as cancer, endocrine disorders, and viral infections, including AIDS. Accordingly, the disclosure provides various peptides useful for inhibition, and longer peptides containing those peptides. The disclosure of those peptide sequences is incorporated herein by reference. In particular, see pages 6-16 or the sequence listing in
the disclosure. With respect to the present invention, accordingly, such inhibitors are useful for treating the disorders, such as those disclosed herein, by inhibiting the subtilase-like protein of the present invention. Inhibition of conversion has uses that include, but are not limited to, reducing malignant transformation and tumorigenesis, reducing the physiological consequences of tumor production and release of bioactive peptides, such as those derived from insulinomas, gastrinomas, or lung cancer cells that may hypersecrete hormonally active peptides, inhibiting neoplasia by blocking subtilase-like protein-mediated processing of growth factors that are produced in many types of tumor cells, reducing or preventing HIV infection via inhibition of processing of gp160, thereby blocking formation of gp120, and diminishing the infectivity of newly synthesized virions. Inhibition of conversion is also useful for inhibiting overproduction of endocrine or neuroendocrine hormones that result in pathophysiology.

Other inhibitors include acylated peptidyl chloromethanes containing a consensus furin cleavage sequence, such as decanoyl-ARG-GLU-LYS-ARG-CH₂Cl. See, for example Stieneke-Grober (1992) EMBO Journal 11:2407-2414. Further candidates include reversible peptide inhibitors in which the -NH-group of the scissile P₁-P₁' bond has been replaced with a methylene group or a methylene group has been inserted between the -CO- and -NH- of the scissile bond. See Angliker (1995) J. Med. Chem. 38:4014-4018. Protein-based furin inhibitors have also been developed, such as a variant of α₁-antitrypsin that has a replacement of the reactive-site MET residue by ARG. This has been shown to inhibit the in vitro conversion of proalbumin. See Bathurst (1987) Science 235:348-350. Other α₁-antitrypsin variants have been constructed, such as α₁-PDX in which the reactive center ALA²⁴-ILE-PRO-MET²¹ sequence has been replaced by ARG-ILE-PRO-ARG. See Mizuno et al. (1988) Biochem. Biophys. Resp. Commun. 156:246-254. This particular candidate has been shown to inhibit the cleavage of viral envelope glycoproteins, including HIV gp160.

The invention provides other end points to identify compounds that modulate (stimulate or inhibit) the subtilase-like protein activity. The assays typically involve an assay of molecular, subcellular, cellular, or in vivo events that indicate the subtilase-like protein activity. These include but are not limited to those that have been discussed above, including the production of mature substrate peptide, association with specific
subcellular locations, effects on cell growth or differentiation, including apoptosis, pathogen virulence, obesity, and the like.

Thus, the expression of genes that are up- or down-regulated in response to the subtilase-like protein activity pathway can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase. Alternatively, phosphorylation of the subtilase-like protein or target could also be measured.

Any of the biological or biochemical functions mediated by the subtilase-like protein can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art.

Binding and/or activating compounds can also be screened by using chimeric subtilase-like proteins in which one or more domains, sites, and the like, as disclosed herein, or parts thereof, can be replaced by their heterologous counterparts derived from other subtilases or subtilase-like proteins or from other subtilase or subtilase-like isoforms. For example, a catalytic region can be used that interacts with a different substrate specificity and/or affinity than the native subtilase-like protein of the invention. Accordingly, a different set of components is available as an end-point assay for activation. Alternatively, a heterologous COOH sequence can replace a native COOH sequence or can be added where no COOH sequence existed. This will result in different subcellular or cellular localization and accordingly can result in having an effect on a different set of components or pathway. Accordingly, a different set of components or pathway is available as an endpoint assay for activation. As a further alternative, the site of modification by an effector protein, for example phosphorylation by casein kinase II, can be replaced with the site from a different effector protein. Activation can also be detected by a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native signal transduction pathway.

The subtilase-like polypeptides are also useful in competition binding assays in methods designed to discover compounds that interact with the polypeptide. Thus, a compound is exposed to the polypeptide under conditions that allow the compound to
bind or to otherwise interact with the polypeptide. Soluble subtilase-like polypeptide is also added to the mixture. If the test compound interacts with the soluble subtilase-like polypeptide, it decreases the amount of complex formed with or activity from the subtilase-like polypeptide target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the subtilase-like protein. Thus, the soluble polypeptide that competes with the target protein region is designed to contain peptide sequences corresponding to the region of interest.

Another type of competition-binding assay can be used to discover compounds that interact with specific functional sites. As an example, substrate and a candidate compound can be added to a sample of the subtilase-like protein. Compounds that interact with the subtilase-like protein at the same site as the substrate will reduce the amount of complex formed between the subtilase-like protein and substrate. Accordingly, it is possible to discover a compound that specifically prevents or alters interaction between the subtilase-like protein and substrate. Another example involves adding a candidate compound to a sample of subtilase-like protein and propeptide. A compound that competes with the propeptide will reduce the amount of binding of the propeptide to the subtilase-like protein. Accordingly, compounds can be discovered that directly interact with the subtilase-like protein and compete with the propeptide. Such assays can involve any other component that interacts with the subtilase-like protein.

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

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/subtilase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., 35S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined.
directly, or in the supernatant after the complexes is dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of subtilase-like-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a subtilase-like-binding target component and a candidate compound are incubated in the subtilase-like protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the subtilase-like target molecule, or which are reactive with the subtilase-like protein and compete with the target molecule; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Modulators of the subtilase-like protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the subtilase-like protein pathway, by treating cells that express the subtilase-like protein. These methods of treatment include the steps of administering the modulators of the subtilase-like protein activity in a pharmaceutical composition as described herein, to a subject in need of such treatment.

Treatment is of disorders related to improper proprotein processing. Disorders result from events including but not limited to processing of the extracellular matrix, growth factors, including during early embryogenesis, serum proteins, including proteases of blood clotting and complement systems, matrix metalloproteinases, receptors, enzymes, adhesion molecules, hormones, cell surface signaling components, and endocrine and neural polypeptide hormones. Accordingly, treatment is of the consequences of such abnormal processing of these components, including defects in embryogenesis, tumor formation, inflammation, apoptosis, defects in differentiation, improper metabolic activity, defects in cell signaling, defects in programmed cell death, and endocrine disorders and endocrine tumors resulting from improper prohormone
processing. On another level, treatment can be of such disorders as obesity. Further, since pathogenic virulence is related to processing, disorders also include increased virulence as a result of over-expression or increased activity of the subtilase-like protein, resulting in relatively high viral and bacterial virulence.

In the present case, a relevant disorder that maps to chromosome 1 p34-p32 is the muscle-eye brain disease (MEB). See http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispimim253280. This text describes a disorder comprising congenital muscular dystrophy with high serum CPK, severe congenital myopia, congenital glaucoma, pallor of the optic discs, retinal hypoplasia, mental retardation, hydrocephalus, abnormal EEG, and myoclonic jerks. Characteristics are severe early-onset muscle weakness, mental retardation and pathologic eye findings, usually congenital myopia. A further study showed the combination of congenital muscular dystrophy and involvement of the central nervous system and eyes. This disease has phenotypic similarities with the Walker-Warburg syndrome.

A further relevant disorder is described in http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispimim602522, designated Bartter Syndrome, infantile, with sensorineural deafness. Bartter syndrome is an autosomal recessive disorder defined by hypokalemic metabolic alkalosis. Affected individuals have elevated plasma renin activity and hyperaldosteronism, with normal blood pressure, altered prostaglandin metabolism (with increased levels of urinary prostaglandins), and increased urinary chloride excretion. One form of Bartter syndrome is due to mutation in the kidney chloride channel B and maps to 1p36.

Other especially relevant disorders include those that are associated with programmed cell death, and particularly with neuronal programmed cell death. These include but are not limited to those described herein and also in the cross-referenced applications above, that are incorporated herein by reference for disclosure of disorders associated with neuronal programmed cell death.

As used herein, "programmed cell death" refers to a genetically regulated process involved in the normal development of multicellular organisms. This process occurs in cells destined for removal in a variety of normal situations, including larval development of the nematode *C. elegans*, insect metamorphosis, development in mammalian embryos, including the nephrogenic zone in the developing kidney, and
regression or atrophy (e.g., in the prostate after castration). Programmed cell death can occur following the withdrawal of growth and trophic factors in many cells, nutritional deprivation, hormone treatment, ultraviolet irradiation, and exposure to toxic and infectious agents including reactive oxygen species and phosphatase inhibitors, e.g., okadaic acid, calcium ionophores, and a number of cancer chemotherapeutic agents. See Wilson (1998) *Biochem. Cell Biol.* 76:573-582 and Hetts (1998) *JAMA* 279:300-307, the contents of which are incorporated herein by reference. Thus, the proteins of the invention, by being differentially expressed during programmed cell death, e.g., neuronal programmed cell death, can modulate a programmed cell death pathway activity and provide novel diagnostic targets and therapeutic agents for disorders characterized by deregulated programmed cell death, particularly in cells that express the protein.

As used herein, a "disorder characterized by deregulated programmed cell death" refers to a disorder, disease or condition which is characterized by a deregulation, e.g., an upregulation or a downregulation, of programmed cell death. Programmed cell death deregulation can lead to deregulation of cellular proliferation and/or cell cycle progression. Examples of disorders characterized by deregulated programmed cell death include, but are not limited to, neurodegenerative disorders, e.g., Alzheimer's disease, dementias related to Alzheimer's disease (such as Pick's disease), Parkinson's and other Lewy diffuse body diseases, multiple sclerosis, amyotrophic lateral sclerosis, progressive supranuclear palsy, epilepsy, Jakob-Creutzfeldt disease, or AIDS related dementias; myelodysplastic syndromes, e.g., aplastic anemia; ischemic injury, e.g., myocardial infarction, stroke, or reperfusion injury; autoimmune disorders, e.g., systemic lupus erythematosus, or immune-mediated glomerulonephritis; or proliferative disorders, e.g., cancer, such as follicular lymphomas, carcinomas with p53 mutations, or hormone-dependent tumors, e.g., breast cancer, prostate cancer, or ovarian cancer). Clinical manifestations of faulty apoptosis are also seen in stroke and in rheumatoid arthritis. Wilson (1998) *Biochem. Cell Biol.* 76:573-582.

Failure to remove autoimmune cells that arise during development or that develop as a result of somatic mutation during an immune response can result in
autoimmune disease. One of the molecules that plays a critical role in regulating cell
dehath in lymphocytes is the cell surface receptor for Fas.

Viral infections, such as those caused by herpesviruses, poxviruses, and
adenoviruses, may result in aberrant apoptosis. Populations of cells are often depleted
in the event of viral infection, with perhaps the most dramatic example being the cell
depletion caused by the human immunodeficiency virus (HIV). Most T cells that die
during HIV infections do not appear to be infected with HIV. Stimulation of the CD4
receptor may result in the enhanced susceptibility of uninfected T cells to undergo
apoptosis.

Many disorders can be classified based on whether they are associated with
abnormally high or abnormally low apoptosis. Thompson (1995) Science 267:1456-
1462. Apoptosis may be involved in acute trauma, myocardial infarction, stroke, and
infectious diseases, such as viral hepatitis and acquired immunodeficiency syndrome.

Primary apoptosis deficiencies include graft rejection. Accordingly, the
invention is relevant to the identification of genes useful in inhibiting graft rejection.

Primary apoptosis deficiencies also include autoimmune diabetes.
Accordingly, the invention is relevant to the identification of genes involved in
autoimmune diabetes and accordingly, to the identification of agents that act on these
targets to modulate the expression of these genes and hence, to treat or diagnose this
disorder. Further, it has been suggested that all autoimmune disorders can be viewed
as primary deficiencies of apoptosis (Hetts, above). Accordingly, the invention is
relevant for screening for gene expression and transcriptional profiling in any
autoimmune disorder and for screening for agents that affect the expression or
transcriptional profile of these genes.

Primary apoptosis deficiencies also include local self reactive disorder. This
includes Hashimoto thyroiditis.

Primary apoptosis deficiencies also include lymphoproliferation and
autoimmunity. This includes, but is not limited to, Canale-Smith syndrome.

Primary apoptosis deficiencies also include cancer. For example, p53 induces
apoptosis by acting as a transcription factor that activates expression of various
apoptosis-mediating genes or by upregulating apoptosis-mediating genes such as Bax.
Primary apoptosis excesses are associated with neurodegenerative disorders including Alzheimer's disease, Parkinson's disease, spinal muscular atrophy, and amyotrophic lateral sclerosis.

Primary apoptosis excesses are also associated with heart disease including idiopathic dilated cardiomyopathy, ischemic cardiomyopathy, and valvular heart disease. Evidence has also been shown of apoptosis in heart failure resulting from arrhythmogenic right ventricular dysplasia. For all these disorders, see Hetts, above.

Death receptors also include the TNF receptor-1 and hence, TNF acts as a death ligand.

A wide variety of neurological diseases are characterized by the gradual loss of specific sets of neurons. Such disorders include Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis (ALS) retinitis pigmentosa, spinal muscular atrophy, and various forms of cerebellar degeneration. The cell loss in these diseases does not induce an inflammatory response, and apoptosis appears to be the mechanism of cell death.

In addition, a number of hematologic diseases are associated with a decreased production of blood cells. These disorders include anemia associated with chronic disease, aplastic anemia, chronic neutropenia, and the myelodysplastic syndromes. Disorders of blood cell production, such as myelodysplastic syndrome and some forms of aplastic anemia, are associated with increased apoptotic cell death within the bone marrow.

These disorders could result from the activation of genes that promote apoptosis, acquired deficiencies in stromal cells or hematopoietic survival factors, or the direct effects of toxins and mediators of immune responses.

Two common disorders associated with cell death are myocardial infarctions and stroke. In both disorders, cells within the central area of ischemia, which is produced in the event of acute loss of blood flow, appear to die rapidly as a result of necrosis. However, outside the central ischemic zone, cells die over a more protracted time period and morphologically appear to die by apoptosis.

The invention also pertains to disorders of the central nervous system (CNS). These disorders include, but are not limited to cognitive and neurodegenerative disorders such as Alzheimer's disease, senile dementia, Huntington's disease,
amyotrophic lateral sclerosis, and Parkinson’s disease, as well as Gilles de la Tourette’s syndrome, autonomic function disorders such as hypertension and sleep disorders, and neuropsychiatric disorders that include, but are not limited to schizophrenia, schizoaffective disorder, attention deficit disorder, dysthymic disorder, major depressive disorder, mania, obsessive-compulsive disorder, psychoactive substance use disorders, anxiety, panic disorder, as well as bipolar affective disorder, e.g., severe bipolar affective (mood) disorder (BP-I), bipolar affective (mood) disorder with hypomania and major depression (BP-II). Further CNS-related disorders include, for example, those listed in the American Psychiatric Association’s Diagnostic and Statistical manual of Mental Disorders (DSM), the most current version of which is incorporated herein by reference in its entirety.

As used herein, "differential expression" or differentially expressed" includes both quantitative and qualitative differences in the temporal and/or cellular expression pattern of a gene, e.g., the programmed cell death genes disclosed herein, among, for example, normal cells and cells undergoing programmed cell death. Genes which are differentially expressed can be used as part of a prognostic or diagnostic marker for the evaluation of subjects at risk for developing a disorder characterized by deregulated programmed cell death. Depending on the expression level of the gene, the progression state of the disorder can also be evaluated.

Further relevant disorders include those associated with aberrant mitochondrial function. Open reading frame analysis of the human subtilase-like protein of the present invention indicates that the enzyme is localized in mitochondria.

The yeast V-ATPase is similar to the V-ATPases of higher organisms and has shown to be an accessible model for many aspects of V-ATPase function. See Kane, J. (1999) Bioenergetics Biomembranes 31:40-56. In yeast this ATPase acidifies the vacuole to a pH of approximately 6 and drives secondary transport of calcium, amino acids and other nutrients. V-ATPases also reside in other intracellular compartments. Accordingly, the yeast ATPase is analogous to the role of V-ATPases in intracellular compartments of all eukaryotic cells. The V-ATPase is a substrate for subtilase-related enzymes. Accordingly, with regard to the present invention, further relevant disorders include those that result from defective V-ATPase processing.
Since the gene is expressed in (among others) liver, kidney, and testes, further relevant disorders are those involving these tissues, especially liver, where the gene is relatively highly expressed, and particularly apoptosis-related liver disorders.

Disorders involving the liver include, but are not limited to, hepatic injury; jaundice and cholestasis, such as bilirubin and bile formation; hepatic failure and cirrhosis, such as cirrhosis, portal hypertension, including ascites, portosystemic shunts, and splenomegaly; infectious disorders, such as viral hepatitis, including hepatitis A-E infection and infection by other hepatitis viruses, clinicopathologic syndromes, such as the carrier state, asymptomatic infection, acute viral hepatitis, chronic viral hepatitis, and fulminant hepatitis; autoimmune hepatitis; drug- and toxin-induced liver disease, such as alcoholic liver disease; inborn errors of metabolism and pediatric liver disease, such as hemochromatosis, Wilson disease, α1-antitrypsin deficiency, and neonatal hepatitis; intrahepatic biliary tract disease, such as secondary biliary cirrhosis, primary biliary cirrhosis, primary sclerosing cholangitis, and anomalies of the biliary tree; circulatory disorders, such as impaired blood flow into the liver, including hepatic artery compromise and portal vein obstruction and thrombosis, impaired blood flow through the liver, including passive congestion and centrilobular necrosis and peliosis hepatis, hepatic vein outflow obstruction, including hepatic vein thrombosis (Budd-Chiari syndrome) and veno-occlusive disease; hepatic disease associated with pregnancy, such as preeclampsia and eclampsia, acute fatty liver of pregnancy, and intrahepatic cholestasis of pregnancy; hepatic complications of organ or bone marrow transplantation, such as drug toxicity after bone marrow transplantation, graft-versus-host disease and liver rejection, and nonimmunologic damage to liver allografts; tumors and tumorous conditions, such as nodular hyperplasias, adenomas, and malignant tumors, including primary carcinoma of the liver and metastatic tumors.

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

Disorders involving the kidney include, but are not limited to, congenital
anomalies including, but not limited to, cystic diseases of the kidney, that include but are
not limited to, cystic renal dysplasia, autosomal dominant (adult) polycystic kidney
disease, autosomal recessive (childhood) polycystic kidney disease, and cystic diseases
of renal medulla, which include, but are not limited to, medullary sponge kidney, and
nephronophthisis-uremic medullary cystic disease complex, acquired (dialysis-
associated) cystic disease, such as simple cysts; glomerular diseases including
pathologies of glomerular injury that include, but are not limited to, in situ immune
complex deposition, that includes, but is not limited to, anti-GBM nephritis, Heymann
nephritis, and antibodies against planted antigens, circulating immune complex nephritis,
antibodies to glomerular cells, cell-mediated immunity in glomerulonephritis, activation
of alternative complement pathway, epithelial cell injury, and pathologies involving
mediators of glomerular injury including cellular and soluble mediators, acute
glomerulonephritis, such as acute proliferative (poststreptococcal, postinfectious)
glomerulonephritis, including but not limited to, poststreptococcal glomerulonephritis
and nonstreptococcal acute glomerulonephritis, rapidly progressive (crescentic)
glomerulonephritis, nephrotic syndrome, membranous glomerulonephritis (membranous
nephropathy), minimal change disease (lipoid nephrosis), focal segmental
glomerulosclerosis, membranoproliferative glomerulonephritis, IgA nephropathy
(Berger disease), focal proliferative and necrotizing glomerulonephritis (focal
glomerulonephritis), hereditary nephritis, including but not limited to, Alport syndrome
and thin membrane disease (benign familial hematuria), chronic glomerulonephritis,
glomerular lesions associated with systemic disease, including but not limited to,
systemic lupus erythematosus, Henoch-Schönlein purpura, bacterial endocarditis,
diabetic glomerulosclerosis, amyloidosis, fibrillary and immunotactoid
glomerulonephritis, and other systemic disorders; diseases affecting tubules and
interstitium, including acute tubular necrosis and tubulointerstitial nephritis, including
but not limited to, pyelonephritis and urinary tract infection, acute pyelonephritis, chronic pyelonephritis and reflux nephropathy, and tubulointerstitial nephritis induced by drugs and toxins, including but not limited to, acute drug-induced interstitial nephritis, analgesic abuse nephropathy, nephropathy associated with nonsteroidal anti-inflammatory drugs, and other tubulointerstitial diseases including, but not limited to, urate nephropathy, hypercalcemia and nephrocalcinosis, and multiple myeloma; diseases of blood vessels including benign nephrosclerosis, malignant hypertension and accelerated nephrosclerosis, renal artery stenosis, and thrombotic microangiopathies including, but not limited to, classic (childhood) hemolytic-uremic syndrome, adult hemolytic-uremic syndrome/thrombotic thrombocytopenic purpura, idiopathic HUS/TTP, and other vascular disorders including, but not limited to, atherosclerotic ischemic renal disease, atheroembolic renal disease, sickle cell disease nephropathy, diffuse cortical necrosis, and renal infarcts; urinary tract obstruction (obstructive uropathy); urolithiasis (renal calculi, stones); and tumors of the kidney including, but not limited to, benign tumors, such as renal papillary adenoma, renal fibroma or hamartoma (renomedullary interstitial cell tumor), angiomyolipoma, and oncocytooma, and malignant tumors, including renal cell carcinoma (hypernephroma, adenocarcinoma of kidney), which includes urothelial carcinomas of renal pelvis.

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

Methods for treatment include but are not limited to the use of soluble subtilase-like protein or fragments of the subtilase-like protein that compete for substrate or propeptide. These proteins or fragments can have a higher affinity for the target so as to provide effective competition. Methods of treatment also include the use of candidate compounds as described hereinabove.

Stimulation of activity is desirable in situations in which the protein is abnormally downregulated and/or in which increased activity is likely to have a
beneficial effect. Likewise, inhibition of activity is desirable in situations in which the protein is abnormally upregulated and/or in which decreased activity is likely to have a beneficial effect. In one example of such a situation, a subject has a disorder characterized by aberrant development or cellular differentiation. In another example, the subject has a proliferative disease (e.g., cancer) or a disorder characterized by an aberrant hematopoietic response. In another example, it is desirable to achieve tissue regeneration in a subject (e.g., where a subject has undergone brain or spinal cord injury and it is desirable to regenerate neuronal tissue in a regulated manner).

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

The subtilase-like polypeptides also are useful to provide a target for diagnosing a disease or predisposition to disease mediated by the subtilase-like protein, and particularly in obesity, liver disorders, and disorders related to neuronal programmed cell death. Accordingly, methods are provided for detecting the presence, or levels of, the subtilase-like protein in a cell, tissue, or organism. The method involves contacting a biological sample with a compound capable of interacting with the protein such that the interaction can be detected.

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

The subtilase-like protein also provides a target for diagnosing active disease, or predisposition to disease, in a patient having a variant of the subtilase-like protein. Thus, the subtilase-like protein can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in an aberrant protein. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered protein activity in
cell-based or cell-free assay, alteration in substrate binding or degradation, propeptide binding or phosphorylation, or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein in general or in a subtilase-like protein specifically.

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

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

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

Antibodies

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

To generate antibodies, an isolated subtilase-like polypeptide is used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. Either the full-length protein or antigenic peptide fragment can be used. Regions having a high antigenicity index are shown in Figure 2.
Antibodies are preferably prepared from these regions or from discrete fragments in these regions. However, antibodies can be prepared from any region of the peptide as described herein. A preferred fragment produces an antibody that diminishes or completely prevents substrate hydrolysis or binding. Antibodies can be developed against the entire protein or domains of the protein as described herein. Antibodies can also be developed against specific functional sites as disclosed herein.

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

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

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

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

**Antibody Uses**

The antibodies can be used to isolate the subtilase-like proteins of the invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural subtilase-like protein from cells and a recombinantly produced subtilase-like protein expressed in host cells.
The antibodies are useful to detect the presence of the subtilase-like protein in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development.

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

The antibodies can be used to assess abnormal tissue distribution or abnormal expression during development. Antibody detection of circulating fragments of the full length protein can be used to identify protein turnover.

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

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Antibodies can be developed against the whole protein or portions of the protein.

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

Antibodies accordingly can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen.
Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against a polymorphic subtilase-like protein can be used to identify individuals that require modified treatment modalities.

The antibodies are also useful as diagnostic tools as an immunological marker for aberrant subtilase-like protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

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

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

The antibodies are also useful for inhibiting protein function, for example, blocking substrate, propeptide, or the subcellular localization site(s).

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


Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response to protect the animal from the diseases herein mentioned, among others. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises delivering a polypeptide of the present invention via a
vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect the animal from diseases.

A further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces an immunological response in that mammal to a polypeptide of the present invention where the composition comprises a polypeptide or polynucleotide of the present invention. The vaccine formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain antioxidants, buffers, bacteriostats and solutes which render the formulation instonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

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

The specifically disclosed cDNAs comprise the coding region and 5' and 3' untranslated sequences in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and SEQ ID NO:7.

The invention provides isolated polynucleotides encoding the novel subtilase-like proteins. The term “subtilase-like polynucleotide” or “subtilase-like nucleic acid” refers to the sequences shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7. The term “subtilase-like polynucleotide” or “subtilase-like nucleic acid” further includes variants and fragments of the subtilase-like polynucleotides.

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

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

In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC.

Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 % (on a molar basis) of all macromolecular species present.
For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include in vivo or in vitro RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

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

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

The subtilase-like polynucleotides include, but are not limited to, the sequence encoding the mature polypeptide alone, the sequence encoding the mature polypeptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or proprotein sequence), the sequence encoding the mature polypeptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the polynucleotide may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.
Subtilase-like polynucleotides can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

Subtilase-like nucleic acid can comprise a nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7, corresponding to human, mouse, or rat cDNA.

In one embodiment, the subtilase-like nucleic acid comprises only the coding region.

The invention further provides variant subtilase-like polynucleotides, and fragments thereof, that differ from a nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequences.

The invention also provides subtilase-like nucleic acid molecules encoding the variant polypeptides described herein. Such polynucleotides may be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions.

Typically, variants have a substantial identity with a nucleic acid molecule of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 and the complements thereof. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. These variants comprise a nucleotide sequence encoding a subtilase-like protein that is at least about 60-65%, 65-70%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more homologous to a nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 or a fragment of this sequence. Such nucleic acid molecules
can readily be identified as being able to hybridize under stringent conditions, to a nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7, or a fragment of the sequence. It is understood that stringent hybridization does not indicate substantial homology where it is due to general homology, such as poly A sequences, or sequences common to all or most proteins, all subtilases, or common to a known subtilase family. Moreover, it is understood that variants do not include any of the nucleic acid sequences that may have been disclosed prior to the invention.

As used herein, the term “hybridizes under stringent conditions” is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a polypeptide at least about 60-65% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 90%, at least about 95% or more identical to each other remain hybridized to one another. Such stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6, incorporated by reference. One example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by

one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. In another non-limiting example, nucleic acid molecules are allowed to hybridize in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more low stringency washes in 0.2 X SSC/0.1% SDS at room temperature, or by one or more moderate stringency washes in 0.2 X SSC/0.1% SDS at 42°C, or washed in 0.2 X SSC/0.1% SDS at 65°C for high stringency. In another embodiment, hybridization is in 3 X SSC, at about 65°C, followed by washes that include 0.2 X SSC at 65°C for at least 30 minutes. In one embodiment, an isolated nucleic acid molecule that hybridizes under stringent conditions to a sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a “naturally-occurring” nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

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

The present invention also provides isolated nucleic acids that contain a single or double stranded fragment or portion that hybridizes under stringent conditions to a nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 or the complement. In one embodiment, the nucleic acid consists of a portion of a nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 and the complement. The nucleic acid fragments of the invention are at least about 15, preferably at least about 18, 20, 23 or 25 nucleotides, and can be 30, 40, 50, 100, 200, 500 or more nucleotides in length. Longer fragments, for example, 30 or more nucleotides in length, which encode antigenic proteins or polypeptides described herein are useful.

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

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

Thus, nucleic acid fragments further include sequences corresponding to the domains described herein, subregions also described, and specific functional sites. Nucleic acid fragments also include combinations of the domains, segments, and other functional sites described above. A person of ordinary skill in the art would be aware of the many permutations that are possible.
Where the location of the domains or sites have been predicted by computer analysis, one of ordinary skill would appreciate that the amino acid residues constituting these domains can vary depending on the criteria used to define the domains.

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

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

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

Polynucleotide Uses

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

The nucleic acid fragments of the invention provide probes or primers in assays such as those described below. "Probes" are oligonucleotides that hybridize in a base-specific manner to a complementary strand of nucleic acid. Such probes include polypeptide nucleic acids, as described in Nielsen et al. (1991) Science 254:1497-1500. Typically, a probe comprises a region of nucleotide sequence that hybridizes under highly stringent conditions to at least about 15, typically about 20-25, and more typically about 40, 50 or 75 consecutive nucleotides of a nucleic acid sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7.
and the complements thereof. More typically, the probe further comprises a label, e.g., radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

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

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

The polynucleotides are useful as a hybridization probe for cDNA and genomic DNA to isolate a full-length cDNA and genomic clones encoding a polypeptide described in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8 and to isolate cDNA and genomic clones that correspond to variants producing the same polypeptides shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8 or the other variants described herein. Variants can be isolated from the same tissue and organism from which a polypeptide shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8 were isolated, different tissues from the same organism, or from different organisms. This method is useful for isolating genes and cDNA that are developmentally-controlled and therefore may be expressed in the same tissue or different tissues at different points in the development of an organism.
The probe can correspond to any sequence along the entire length of the gene encoding the subtilase-like protein. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions.

The nucleic acid probe can be, for example, a full-length cDNA of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7, or a fragment thereof, such as an oligonucleotide of at least 12, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to mRNA or DNA.

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

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

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

Additionally, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup et al. (1996) Bioorganic & Medicinal Chemistry 4:5). As used herein, the terms “peptide nucleic acids” or “PNAs” refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNA has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al. (1996), supra; Perry-O'Keefe et al. (1996) Proc. Natl. Acad. Sci. USA 93:14670. PNAs can be further modified, e.g., to enhance their stability, specificity or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), supra, Finn et al. (1996) Nucleic Acids Res. 24(17):3357-63, Mag et al. (1989) Nucleic Acids Res. 17:5973, and Peterser et al. (1975) Bioorganic Med. Chem. Lett. 5:1119.

The nucleic acid molecules and fragments of the invention can also include other appended groups such as peptides (e.g., for targeting host cell subtilase-like proteins in vivo), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al. (1989) Proc. Natl. Acad. Sci. USA 86:6553-6556; Lemaitre et al. (1987) Proc. Natl. Acad. Sci. USA 84:648-652; PCT Publication No. WO 88/0918) or the blood brain barrier (see, e.g., PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see,

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

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

The polynucleotides are also useful for expressing antigenic portions of the subtilase-like proteins.

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

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

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, *Mendelian Inheritance in Man*, available on-line through Johns Hopkins University Welch Medical Library.) The relationship between a gene and a disease mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland *et al.* ((1987) *Nature* 325:783-787).
Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a specified gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease.

Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations, that are visible from chromosome spreads, or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

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

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

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

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

The polynucleotides are also useful for making vectors that express part, or all, of the subtilase-like polypeptides.

The polynucleotides are also useful as hybridization probes for determining the level of nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, subtilase-like nucleic acid in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the polypeptides described herein can be used to assess gene copy number in a given cell, tissue, or organism. This is particularly relevant in cases in which there has been an amplification of the gene of the invention.
Alternatively, the probe can be used in an *in situ* hybridization context to assess the position of extra copies of the gene, as on extrachromosomal elements or as integrated into chromosomes in which the gene is not normally found, for example as a homogeneously staining region.

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

As such, the gene is particularly relevant for the treatment of disorders including but not limited to those disclosed herein.

Disorders in which subtilase-like protein expression is particularly relevant also include, but are not limited to, disorders involving programmed cell death, such as those disclosed herein, disorders involving obesity, liver disorders, and disorders associated with mitochondrial dysfunction as a result of defects in proprotein processing.

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

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

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

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

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

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

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

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

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

The assay for nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds such as free propeptide, mature substrate, and any of the downstream components or cellular events that result from subtilase-like protein expression, including but not limited to those disclosed hereinabove. Further, the expression of genes that are up- or down-regulated in response
to the subtilase-like protein expression can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

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

Accordingly, the invention provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate the nucleic acid expression. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or effects on nucleic acid activity (e.g. when nucleic acid is mutated or improperly modified). Treatment is of disorders characterized by aberrant expression or activity of the nucleic acid.

The gene is particularly relevant for the treatment of disorders involving obesity, liver function, mitochondrial dysfunction, and programmed cell death, and in particular, neuronal cell death, especially in brain.

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

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

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

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

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

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

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

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

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.
Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature. Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method.


Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers et al. (1985) Science 230:1242; Cotton et al. (1988) PNAS 85:4397; Saleeba et al. (1992) Meth. Enzymol. 217:286-295), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita et al. (1989) PNAS 86:2766; Cotton et al. (1993) Mutat. Res. 285:125-144; and Hayashi et al. (1992) Genet. Anal. Tech. Appl. 9:73-79), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers et al. (1985) Nature 313:495). The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) Trends Genet. 7:5). Examples of other techniques for detecting point mutations include, selective oligonucleotide hybridization, selective amplification, and selective primer extension.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Examples of antisense molecules useful to inhibit nucleic acid expression include antisense molecules complementary to a fragment of the 5' untranslated region of SEQ ID NOS:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 which also includes the start codon and antisense molecules which are complementary to a fragment of the 3' untranslated region of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7. Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of subtilase-like nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired expression of the nucleic acid of the invention. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the subtilase-like protein.

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

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

Computer Readable Means

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

In one application of this embodiment, a nucleotide or amino acid sequence of the present invention can be recorded on computer readable media. As used herein, “computer readable media” refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. The skilled artisan will readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention.

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

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

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

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).
Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI).

For example, software which implements the BLAST (Altschul et al. (1990) J. Mol. Biol. 215:403-410) and BLAZE (Brutlag et al. (1993) Comp. Chem. 17:203-207) search algorithms on a Sybase system can be used to identify open reading frames (ORFs) of the sequences of the invention which contain homology to ORFs or proteins from other libraries. Such ORFs are protein encoding fragments and are useful in producing commercially important proteins such as enzymes used in various reactions and in the production of commercially useful metabolites.

Vectors/Host Cells

The invention also provides vectors containing the polynucleotides of the invention. The term “vector” refers to a vehicle, preferably a nucleic acid molecule that can transport the polynucleotides. When the vector is a nucleic acid molecule, the polynucleotides are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

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

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

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

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

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

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

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

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

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

The vector containing the appropriate polynucleotide can be introduced into an appropriate host cell for propagation or expression using well-known techniques.

Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

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

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

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

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


The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the polynucleotides of the invention. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the polynucleotides described herein. These are found for example in Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the polynucleotide sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

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

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

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the polynucleotides of the invention can be introduced either alone or with other polynucleotides that are not related to the polynucleotides of the invention such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the subtilase-like polynucleotide vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.
Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the polynucleotides described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

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

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

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

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

Uses of Vectors and Host Cells

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

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

Host cells are also useful for conducting cell-based assays involving the subtilase-like protein or subtilase-like protein fragments. Thus, a recombinant host cell expressing a native subtilase-like protein is useful to assay for compounds that stimulate or inhibit the subtilase-like protein function. This includes substrate binding, gene expression at the level of transcription or translation, propeptide interaction, and downstream components of pathways affected by subtilase-like protein activation.

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

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

Further, mutant subtilase-like proteins can be designed in which one or more of the various functions is engineered to be increased or decreased and used to augment or replace subtilase-like proteins in an individual. Thus, host cells can provide a therapeutic benefit by replacing an aberrant subtilase-like protein or providing an aberrant subtilase-like protein that provides a therapeutic result. In one embodiment, the cells provide a subtilase-like protein that is abnormally active.

In another embodiment, the cells provide subtilase-like proteins that are abnormally inactive. These can compete with the endogenous subtilase-like proteins in the individual.
In another embodiment, cells expressing a subtilase-like protein that cannot be activated are introduced into an individual in order to compete with the endogenous one.

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

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


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

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

In one embodiment, a host cell is a fertilized oocyte or an embryonic stem cell into which a polynucleotide sequence of the invention has been introduced.

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

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

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

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

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

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

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

**Pharmaceutical Compositions**

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

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

As used herein the language “pharmaceutically acceptable carrier” is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, such media can be used in the compositions of the invention. Supplementary active compounds can also be incorporated into the compositions. A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration.

Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous
administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyl (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a subtilase-like protein or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

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

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

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

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

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova
Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected
cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically
acceptable carriers. These can be prepared according to methods known to those skilled
in the art, for example, as described in U.S. Patent No. 4,522,811.

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

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

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

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

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

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

It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per
kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

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

EXAMPLE

In neurons, programmed cell death is an essential component of neuronal development (Jacobson et al. 1997; Pettmann and Henderson (1998)) and has been associated with many forms of neurodegeneration (Hetts, 1998)). In the cerebellum, granule cell development occurs postnatally. The final number of neurons represents the combined effects of additive processes such as cell division and subtractive processes such as target-related programmed cell death. Depolarization due to high concentrations (25 mM) of extracellular potassium (K\(^{+}\)) promotes the survival of
cerebellar granule neurons (CGNs) \textit{in vitro}. CGNs maintained in serum containing medium with high K$^+$ will undergo programmed cell death when switched to serum-free medium with low K$^+$ (5 mM) (D'Mello \textit{et al.} 1993; Miller and Johnson Jr. 1996). The resulting programmed cell death has a transcriptional component that can be blocked by inhibitors of new RNA synthesis (Galli \textit{et al.} 1995; Schulz \textit{et al.} 1996).

As previously disclosed in U.S. Provisional Patent application No. 60/161/188, the inventors in that application constructed a brain-biased and programmed cell death-enriched clone set by arraying ~7300 consolidated ESTs from two cDNA libraries cloned from rat frontal cortex and differentiated PC12 cells deprived of nerve growth factor (NGF), and >300 genes that are known markers for the central nervous system and/or programmed cell death. They reproducibly and simultaneously monitored the expression of the genes at 1, 3, 6, 12, and 24 hours after K$^+$ withdrawal. They then categorized the regulated genes by time course expression pattern to identify cellular processes mobilized by CGN programmed cell death at the RNA level. In particular they focused on the expression profiles of many known pro- and anti-apoptotic regulatory proteins, including transcription factors, Bcl-2 family members, caspases, cyclins, heat shock proteins (HSPs), inhibitors of apoptosis (IAPs), growth factors and receptors, other signal transduction molecules, p53, superoxide dismutases (SODs), and other stress response genes. Finally, they compared the time courses of regulated genes induced by K$^+$ withdrawal in the presence or absence of serum to those induced by glutamate toxicity. Thus, they identified a restricted set of relevant genes regulated by multiple models of programmed cell death in CGNs. These genes included the rat NARCl.

Using the brain-biased, programmed cell death nucleic acid-enriched Smart Chip, global mRNA expression was profiled throughout a time course of KCl/serum-withdrawal-induced cell death in primary cultures of CGNs. The transcription-dependent CGN programmed cell death was coordinated, resulting in less than 30% survival at 24 hours post-withdrawal as quantified by cell counting (data not shown). RNA samples, designated “treated”, were isolated at 1, 3, 6, 12, and 24 hours after switching post-natal day eight CGNs from medium containing 5% serum and 25 mM KCl to serum-free medium with 5 mM KCl. For controls, the 5% serum/25 mM KCl medium was replaced, and “sham” RNA at 1, 3, 6, 12, and 24 hours was isolated.
A third model of programmed cell death used to assay NARC expression involves kainate treatment. See, for example, Figure 8 herein, and U.S. Provisional Application No. 60/161,188, incorporated herein by reference for teaching (among other things) this model.

Figure 8 shows the result of experiments designed to characterize transcriptional characteristics for the rat NARC1 gene. For the top panel, the results show that, for NARC1, gene expression peaked at 3 hr both in the potassium/serum withdrawal paradigm and also in the potassium alone withdrawal paradigm. The conclusion from these results is that NARC1 is a gene that is regulated by the transcriptionally-dependent models of programmed cell death in the cerebellar granular neurons.

RT-PCR performed with sequence-specific primers for NARC1, confirm, for one of the time points in the top panel, the results in the top panel. That is, these RT-PCR results show that at 3 hr after potassium/serum withdrawal, the same regulation was observed using an independent technique. Accordingly, RT-PCR results confirm an upregulation of greater than four-fold for the NARC1 gene 3 hours after withdrawal.

The transcript size is 3.4 kb, which is the length of the sequence determined for NARC1.

The tissue distribution shows high expression in the liver with lower levels of expression in the testes and in the kidney. There is little expression in brain. This fits the disease model which is that this gene is expressed only when neurons are undergoing cellular distress and cell death. Accordingly, the gene provides a drug target for apoptosis/programmed cell death.
THAT WHICH IS CLAIMED:

1. An isolated polypeptide having an amino acid sequence selected from the group consisting of:
   (a) The amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8;
   (b) The amino acid sequence of an allelic variant of the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8;
   (c) The amino acid sequence of a sequence variant of the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8, wherein the sequence variant is encoded by a nucleic acid molecule hybridizing to the nucleic acid molecule shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 under stringent conditions;
   (d) A fragment of the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8, wherein the fragment comprises at least 10 contiguous amino acids;
   (e) The amino acid sequence of an epitope bearing region of any one of the polypeptides of (a)-(e).

2. An isolated antibody that selectively binds to a polypeptide of claim 1, (a)-(e).

3. An isolated nucleic acid molecule having a nucleotide sequence selected from the group consisting of:
   (a) The nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7;
   (b) A nucleotide sequence encoding the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8; and
   (c) A nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b).
4. An isolated nucleic acid molecule having a nucleotide sequence selected from the group consisting of:
   (a) A nucleotide sequence encoding an amino acid sequence of a sequence variant of the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8 that hybridizes to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 under stringent conditions; and
   (b) A nucleotide sequence complementary to the nucleotide sequences in (a).

5. An isolated nucleic acid molecule having a nucleotide sequence selected from the group consisting of:
   (a) A nucleotide sequence encoding a fragment of the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8, wherein the fragment comprises at least 10 contiguous amino acids;
   (b) A nucleotide sequence complementary to the nucleotide sequences in (a).

6. A nucleic acid vector comprising the nucleic acid sequences in any of claims 3-5.

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

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

9. A method for detecting the presence of any of the polypeptides in claim 1 in a sample, said method comprising contacting said sample with an agent that specifically allows detection of the presence of the polypeptide in the sample and then detecting the presence of the polypeptide.
10. The method of claim 9, wherein said agent is capable of selective physical association with said polypeptide.

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

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

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

14. A kit comprising reagents used for the method of claim 9, wherein the reagents comprise an agent that specifically binds to said polypeptide.

15. A method for detecting the presence of any of the nucleic acid sequences in any of claims 3-5 in a sample, the method comprising contacting the sample with an oligonucleotide that hybridizes to the nucleic acid sequences under stringent conditions and determining whether the oligonucleotide binds to the nucleic acid sequence in the sample.

16. The method of claim 15, wherein the nucleic acid, whose presence is detected, is mRNA.

17. A kit comprising reagents used for the method of claim 15, wherein the reagents comprise a compound that hybridizes under stringent conditions to any of the nucleic acid molecules.

18. The method of claim 17 wherein a fragment of the polypeptide is contacted.

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

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

21. A method for treating obesity, a liver disorder, or a disorder involving aberrant programmed cell death comprising administering the polypeptides of claim 1 to a subject having or at risk of developing obesity, a liver disorder, or a disorder involving aberrant programmed cell death.

22. The method of claim 20 wherein said modulation is in a subject having or predisposed to having obesity, a liver disorder, or a disorder involving aberrant programmed cell death.
FIG. 1A.
CAG CTG GTC CAG CCT GTG GGG CCA CTG GTG CTG CTG CCC CTG GGG GGT GGG TAC AGC
R V L N A A A C Q R L A R V G V V L V T A
GTC CTG CTC AAC GCC GCC TGC CAG CCG CTG GGG AGG GTT GGG GTG CTG CTG ACC GCT
A G N F R D D A C L Y S P A S A P E V I
GCC GCC AAC TTC CGG GAC GAT GCC TGC CTC TAC TCC CCA GCC TCA GCC TCC GAG GTC ATC
T V G A T N A Q D Q P V T L G T L G T N
ACG GTT GGG GCC ACC AAT GCC CAG GAC CAG CGG GTG ACC CTG GGG ACT TTG GGG ACC AAC
F G R C V D L F A P G E D I I G A S D
TTT GCC CGC TGG GTG GAC CTC TTT GCC CCA GGG GAG GAC ATC ATT GGT GCC TCC AGC GAC
C S T C F V S Q S G T S Q A A A H V A G
GTC ACC TGC TTT GTG TCA CAG AGT GGG ACA TCA CAG GCT GCC CAC GTG GCT GCC
I A A A M M L S A E P E L T L A E L R Q R
ATT GCA GCC ATG ATG CTG CTG GCC GAG CGG GAC CTC ACC CTG GCC GAG TTG AGG CAG AGA
L I H F S A K D V I N E A W F P E D Q R
CTG ATC CAC TTC TCT GCC AAA GAT GTC ATC AAT GAG GCC TGG TTC CCT GAG GAC CAG CGG
V L T P N L V A A L P S T H G A G W Q
GTA CTG ACC CCC AAC CTG GTG GCC GCC CTG CCC CCC AGC ACC CAT GGG GCA GGT TGG CAG
L F C R T V W S A H S G P T R M A T A I
CTG TTT GGC AGT GCT GTG TGG TCA GCA CAC TCG GGG CCT ACA CGG ATG GCC ACA GCC ATC
A R C A P D E E L L S C S S F S R G K
GCC CGC TGC GCC CCA GAT GAG GAG CTG CTG ATC GCC TGC TCC AGT TTT TCC AGG AGT GGG AGG
R R G E R M E A Q G G K L V C R A H N A
CGG CGG GCC GAG CGG GCC ATG GAG GCC CAA GGG GGG CAG ATG CTG TGG GCC GCC CAC AAC GCT
F G G E G V Y A I A R C C L L P Q A N C
TGG GGT GAG GGT GTG GTC GCC ATT GGC AGG TGC TGC CTG CTA CCC CAG GCC AAC TGC
S V H T A P P A E A S G M T R V H C H Q
AGC GTC CAC ACA GCT CCA CCA GCT GAG GCC AGC ATG GGG ACC CGT GTC CAC TGC CAC CAA
Q G H V L T G F L A L A S D L K E R G S
CAG GCC CAC GTC CTC ACA GGT TTC CTA GCT CTT GCC TCA GAC CTG AAA GAG AGA GGG TCT
D G D G H W R R S I P A F H I *
GAT GGG GAT GGG CAC TGG AGA CGG AGC ATC CCA GCA TTT CAC ATC TGA

FIG. 1B.

SUBSTITUTE SHEET (RULE 26)
FIG. 1C.
FIG. 2.
Transmembrane Segments Predicted by MEMSAT

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Orient</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>28</td>
<td>out-&gt;ins</td>
<td>4.6</td>
</tr>
<tr>
<td>107</td>
<td>124</td>
<td>ins-&gt;out</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Transmembrane segments for presumed mature peptide

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Orient</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>246</td>
<td>263</td>
<td>ins-&gt;out</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Prosite Pattern Matches for hNARC1

Prosite version: Release 12.2 of February

>PS00001|PD000001|ASN_GLYCOSYLATION N-glycosylation site.

Query: 533 NCSV 536

>PS00005|PD000005|PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 5 SSR 7
Query: 91 SER 93
Query: 94 TAR 96
Query: 235 SGR 237
Query: 246 SMR 248
Query: 249 SLR 251
Query: 419 SAK 421
Query: 492 SGK 494

>PS00006|PD000006|CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query: 47 SEED 50
Query: 89 SWSE 92
Query: 178 SLVE 181
Query: 203 TDFE 206
Query: 221 SKCD 224
Query: 235 SGRD 238
Query: 329 SAPE 332
Query: 407 TLAE 410
Query: 419 SAKD 422
Query: 574 SDGD 577

FIG. 4A.
>PS000008|PDOC00008|MYRISTYL N-myristoylation site.

<table>
<thead>
<tr>
<th>Query</th>
<th>2</th>
<th>GTVSSR</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query</td>
<td>59</td>
<td>GTTATF</td>
<td>64</td>
</tr>
<tr>
<td>Query</td>
<td>117</td>
<td>GLLPGF</td>
<td>122</td>
</tr>
<tr>
<td>Query</td>
<td>227</td>
<td>GTHLAG</td>
<td>232</td>
</tr>
<tr>
<td>Query</td>
<td>240</td>
<td>GVAKGA</td>
<td>245</td>
</tr>
<tr>
<td>Query</td>
<td>259</td>
<td>GTVSGT</td>
<td>264</td>
</tr>
<tr>
<td>Query</td>
<td>337</td>
<td>GATNAQ</td>
<td>342</td>
</tr>
<tr>
<td>Query</td>
<td>349</td>
<td>GTLGTN</td>
<td>354</td>
</tr>
<tr>
<td>Query</td>
<td>384</td>
<td>GTSQAA</td>
<td>389</td>
</tr>
</tbody>
</table>

>PS000009|PDOC00009|AMIDATION Amidation site.

| Query | 492 | SGKR | 495 |

**FIG. 4B.**
Input file RatNARCl.seq  Output File RatNARCl.pat  
Sequence length 3372

MWT RSLPLGS
GGGTCCAGCAGCAGCCACACTTAAAGGCTCTTTTATGTGAGCTGCAGCTCTGCGCT
69
RSDLDRDLRTEPVLGSRPRDI
CGTTCTCCTGACGAGACTGAGATGGACTGCCCAGTGGTCCAGAAGAGCTG
129
TACSPRAQCPAFTSFPRPRACA
AGCTGCTCAGAGGCAGCTGGTGCCCAGGCTTCAGGCTTTAGGTGCTG
189
PLLMGIRCRSTWLWPPLS
CTCTGCCTGCGATGGTGCATACGTCATCTAGTGGTCGTCGGTCGGTGGTGG
249
QLLLLLLCPTGSRADEDG
CAGCTGGTGTTGCTACGTGCTACGTCCTGACAGCTGCCAGGAGGAGGAG
309
DYEELMLALPSQEDSLVEDA

FIG. 5A.
FIG. 5B.
10/22

L N T A C Q R L A R T G V V L V A A A G
370
1149
N F R D D A C L Y S P A S A P E V I T V
390
1289
G A T N A Q D Q P V T L G T L G T N F G
410
1289
R C V D L F A P G K D I I G A S S D C S
430
1289
T C Y M S Q S G T S Q A A A H V A G I V
450
1389
A M M L N R D P A L T L A E L R Q R L I
470
1449
L F S T K D V I N M A W F P E D Q R V L
490
1599
T P N R V A T L P P S T Q E T G G Q L L
510
1599
C R T V M S A H S G P T R T A T A R
530
1629
C A P E E E L L S C S S F S R S G R R R
550
1669
G D R I E A I G G Q Q V C K A L N A F G
570
1749
G E G V Y A V A R C C L L L P R V N C S I
590
1809
H N T P A A R A G P Q T P V H C H Q K D
610

FIG. 5C.

SUBSTITUTE SHEET (RULE 26)
FIG. 5D.
TCCAGAGTGCTGCGCTGCAAGCATTGCCCTCTTG6CAATGCAAGCTGGGACATCCATTATTTTGGAAGCAAAAAAATG
3070
AGAGGACAGGGCTGCTCCTACGCGGTGGTTCTCCTGAGGACCTATGCTCCCTTAGCCCAGGTCCCCAGGATATTATGTGGAGAT
3149
GGAGGAAGCAACAGAGGGTCGACTCTAAGGCCACAAAGTGCAAGACACCTATTACCTCTTCCTGCT
3228
TTTATAGTGAAGTTCCTCTACCTGGTGGTTTTCTGGATTTATCTGGGTTTTATAGATTTTATATGT
3307
AGTGAAGTTTCTGAGAAAAGAAGCTGATGATCTGACCTGAAAAAAAAGGCGCGCGC
3372

FIG. 5E.
Input file MouseNARC1.seq; Output File MouseNARC1.pat
Sequence length 2259

GGACCTTCAGTGAGCGCGAGCTGCCGGTGGCCTCCGGTTCTCCTCTCTTCTTCTTCTGAAGCTAGAGACTGAGCAGCTC
79

MSF

CTTGGCTCCCCAGAGCAGCTACGCGCGCCGAGCCCCGGAGCCAGTGCCCCGATCCGCGTCCTG
189

P R P R A P L L A P M G T H C S A W L R
23

CCG AGG CGG CGG GCA CCT CTC CCT GCC CCG ATG GCC ACC TGC TCT GCG TGG CTG CCG
213

W P L L P L L L L L L L L L L L L C P T G A G
43

TGG CCG CTG TTG CCG CTG CTG CTG CTG CTG TTA CTG TGC CCC ACG GCC GGT GGT
273

T Q D E D D G D Y E E L M L A L P S Q E D
63

ACC CAG GAC GAG GAC GGA GAT TAT GAA GAG CTG ATC TCT GCC CTC CCG TCC CAG GAG GAT
333

G L A D E A A H V A T A T F R R C S K E
83

GCC CTG GCT GAT GAG GCC GCC CAT GTG GCC ACC GCC ACC TCC GCC GCT TGC TCC AAG GAG
393

A W R L P G T Y I V V L M E E T Q R L Q
108

GCC TGG AGG CTG CCA GGA ACC TAC ATT GTG GTG CTG ATG GAG GAG ACC CAG AGG CTA CAG
453

FIG. 6A.
FIG. 6D.
Input file HumanNARCIC.seq  Output File HumanNARCIC.pat
Sequence length 3617

CCCACGCGTCGCCGCCGGCTGGAGGGTGAGCGGAGATTGATGGACTGCCTGGGGGTCGCCGGAGCGTGCCTGTGTGACAGCAGCGG
CTCCAGCTCCAGCGGAGATTCCCGGGCCCTTACGGCACCCTGCTGAACGAGCTCTCAGGCTCTGACAGTCTCTCCCC
ACCAGAAGGTCAGGCGGCCGCGGCGGCTTGAGCCGGACCCGGCCTCAGGCTCGCCAGGACAGAAACCTCTCCCC

MGTVSSRSWWPLL
GGCCCTC ATG GCC ACC GTC AGC TCC AGG CGG TCC TGG TGG CCG CTG CCA CTG CTG CTG CTG
LLLLGGPAGARAEDEDEGDY
CTG CTG CTG TCT CTG GGT CCC GGC GCC GGC CTG CGG GAG GAC GAC GCC GAC TAC
EEELVLALESDEGLAEAPEH
GAG GAG CTG CTG CTA GCC TTG CTG TCC GAG GAG GAC GCC CTG GCC GAA GCA CCC GAG CAC

FIG. 7A.
19/22
GCC GCC TGC CAG CGC CTG GGG AGG ACT GCT GGG GTG GTG GTC ACC GCT GCC GGC AAC TTC
1198
R D D A C L Y S P A S A P E V I T V G A
338
GGG GAC GAT GCC TGC CTC TAC TCC CCA GCC TCA GCT CCC GAG GTC ATC ACA GTT GGG GCC
1268
T N A Q D O P V T L G T L G T N F G R C
358
ACC AAT GEC CAG GAC CAG CGG GTG ACC CTG GGG ACT TTG GGG ACC AAC TTT GEC CGC TGT
1318
V D L F A P G E D I I G A S S D C S T C
378
GTG GAC CTC TTT GCC CCA GGG GAG GAC ATC ATT GCT GCC TC AGC GAC TGC AGC ACC TGC
1378
F V S Q S G T S Q A A A H V A G I A A M
398
TTT GTG TCA CAG AGT GGG ACA TCA CAG GCT GCT GCC CAC GTG GCT GCC ATT GCA GCC ATG
1438
M L S A E P E L T L A E L R Q R L I H F
418
ATG CTG TCT GCC GAG CCC GAG CTC ACC CTG GCC GAG TTG GAG CAG AGA CTG ATC CAC TTC
1498
S A K D V I N E A W F P E D Q R V L T P
438
TCT GCC AAA GAT GTC ATC AAT GAG GCC TGG TTC CCT GAG GAC CAG CGG GTA CTG ACC CCC
1558
N L V A A L P P S T H G A G W Q L F C R
458
AAC CTG GTG GCC CCC GTC CCC CCC ACC CAT GGG GCA GGT TGG CAG CTG TTC TGG AGG
1618
I V W S A H S G P T R M A T A I A R C A
478
ACT CTG TGG TCA GCA CAC TCG GGG CCT ACA CCG ATG GCC ACA GCC ATC GCC CCC TGC GCC
1678
P D E E L L S C S S F S S R S G K R R G E
498
CCA GAT GAG GAG CTG CGT AGC TGC TCC ACT TTG TCC ACG GGT AGT GGG AAG CGG CGG GCC GAG
1738
R M E A Q G G K L V C R A H N A F G G E
518
CGC ATG GAG GCC CAA GGG GGC AAG CTG GTC TGC CGG GCC CAC ACG GCT TTG GGG GGT GAG
1798
G V Y A I A R C C L L P Q A N C S V H T
538
GGT GTC TAC GCC ATT GGG TGC TGC CTG CTA CCC CAG GCC AAC TGC AGC GTC CAC ACA
1858
A P P A E A S M G T R V H C H Q Q G H V
558
GCT CCA CCA GCT GAG GCC ACC ATG GGG ACC CGT GTC CAC TGC CAC CAA CAG GGC CAC GTC
1918

FIG. 7C.
FIG. 7D.

SUBSTITUTE SHEET (RULE 26)
FIG. 7E.
RT-PCR: Up >4x 3 hr S

RADE Validated
Transcript Size: 3.4kb
Tissue Distribution: Liver, Kidney, Testes

FIG. 8.
SEQUENCE LISTING

<110> Chiang, Lillian W.
<120> NARC-1, Novel Subtilase-Like Homolog
<130> 35800/207880
<160> 8
<170> FastSEQ for Windows Version 4.0
<210> 1
<211> 3585
<212> DNA
<213> Homo sapiens
<220>
<221> CDS
<222> (99)...(1863)
<400> 1
cacgagccgcg tgggccagcg ccgcccggcgt ggcgcggcgc gggcctctag 60
gtgctctgc gcagcagcgc aacctcccccc tggcctct atg gcc acc gtc agc tcc 116
Met Gly Thr Val Ser Ser
1 5

agg cgg tcc tgg tgg cgg ctc ctc ctc ctc ctc ctc ctc 164
Arg Arg Ser Trp Trp Pro Leu Pro Leu Leu Leu Leu Leu
10 15 20

cgg ggt ccc ggc ggc cgt cgc cag gag gag gac ggc gac tac 212
Leu Gly Pro Ala Gly Ala Arg Ala Glu Asp Glu Asp Tyr
25 30 35

gag gag ctc tgt cta gcc ttg tgt tgt tcc gag gag gag gcc ctc gaa 260
Glu Glu Leu Val Leu Ala Leu Arg Ser Glu Asp Gly Leu Ala Glu
40 45 50

gca ccc gag cac gga acc acc gcc acc ttc cac gcc gcc gac cag cgc gtc aag gat 308
Ala Pro Glu His Gly Thr Thr Ala Thr Phe His Arg Cys Ala Lys Asp
55 60 65 70

cgc tgg aag tgt ctc gcc acc tac tgt tgt tgt ctc aag gag gag acc 356
Pro Trp Arg Leu Pro Gly Thr Tyr Val Val Val Lys Glu Gly Thr
75 80 85

cac ctc tgc cag tca gag cgc act gcc cgc ctc cag gcc cgc cag cgc ctc 404
His Leu Ser Glu Ser Glu Arg Thr Arg Arg Leu Glu Glu Ala
90 95 100

gcc cgc cgg tac ctc acc aag atc ctc cat gtc ttc cat ggc ctt 452
Ala Arg Arg Gly Tyr Leu Thr Lys Ile Leu His Val Phe His Gly Leu
105 110 115

ctt ctc ctc ctc ctc ctc ctc ctc ctc ctc ctc 500
Leu Pro Gly Phe Leu Val Lys Met Ser Gly Asp Leu Leu Leu Ala
120 125 130

ttg aag tgt ccc cat gtc gac tac atc gag gac gcc ttc tct gtc ttt 548
Leu Lys Leu Pro His Val Asp Tyr Ile Glu Glu Asp Ser Ser Val Phe
135 140 145 150

gcc cag ago atc ccc tgg aac ctc cag cgg att acc ctt cca cgg tac 596
Ala Gin Ser Ile Pro Trp Asn Leu Glu Arg Ile Thr Pro Pro Arg Tyr
CGG GCG GAT GAA TAC CAG CCC CCC GAC GGA GGC AGC CTG GTG GAG GTG
Arg Ala Asp Glu Tyr Gln Pro Pro Asp Gly Gly Ser Leu Val Glu Val
170 175 180

TAT CTC CTA GAC ACC AGC ATA CAG AGT GAC CAC CGG GEA ATC GAG GGC
Tyr Leu Leu Asp Thr Ser Ile Gln Ser Asp His Arg Glu Ile Gln Gly
185 190 195

AGG GTG ATG GTG ACC GAC TCC GAG AAT GTG CCC GAG GAC GGC GGG ACC
Arg Val Met Val Thr Asp Phe Glu Asn Val Pro Glu Asp Gly Thr Thr
200 205 210

CGC TTC CAC AGA CAG GCC AGC AAG GTT GAC AGT CAT GGC ACC CAC CTG
Arg Phe His Arg Gin Ala Ser Lys Cys Asp Ser His Gly Thr His Leu
215 220 225 230

GCA GGG GTG TGC AGC GGC CCG GAT GCC GGC GTG GCC AAG GGT GCC AGC
 Ala Gly Val Val Val Arg Ala Gin Ala Tyr Lys Ala Lys Gly Ala Ser
235 240 245

ATG CGC AGC CTG CGC GTG CTC ACC TGC CAA GGG AAG GGC ACG GTC AGC
Met Arg Ser Leu Arg Val Leu Gin Gly Gln Gys Gly Thr Val Ser
250 255 260

GGC ACC CTC ATA GGC GTG GAG TTT ATT CCG AAA AGC CAG CTG GTC CAG
Gly Thr Leu Ile Gly Leu Glu Phe Arg Lys Ser Gin Leu Val Gln
265 270 275

CCT GTG GGG CCA CTG GTG GTG CTC CCG CCG GTG GGT GGG TAC AGC
Pro Val Gly Pro Leu Val Val Leu Leu Pro Ala Gly Val Tyr Ser
280 320 285 290

CGC GTG CTC AAC GCC TGC CAG CGC GTG GCG AGG GTT GGG GTG CTG
Arg Val Leu Asn Ala Ala Cys Gln Arg Leu Ala Arg Val Gly Val
295 300 305 310

CTG GTG ACC GGT GCC ACC TTC CCG GAC GAT GCC TGC CTC TAC TTC
Leu Val Thr Ala Gly Gin Ala Pro Gin Phe Arg Asp Asp Ala Cys Leu Tyr Ser
315 320 325

CCA GCC TCA GCT CCC GAG GTC ATC ACA GTT GGG GCC ACC AAT GCC CAG
Pro Ala Ser Ala Pro Glu Val Leu Thr Val Gly Thr Asn Ala Gin
330 335 340

GAC CAG CCG GTG ACC CTG GGG ACT TTG GGG ACC AAT CCC TGT GCC
Asp Gin Pro Val Thr Leu Gly Thr Leu Gly Thr Asn Phe Arg Cys
345 350 355

GTG GAC CTC TTT GCC CCA GGG GAG GAC ATC ATT GTT GCC TCC AGC GAC
Val Asp Leu Phe Ala Pro Gly Asp Ile Ile Gly Ala Ser Ser Asp
360 365 370

TGC AGC ACC TGC TTT GTG TCA CAG AGT GGG ACA TCA CAG GTT GCT GCC
Cys Ser Thr Cys Phe Val Ser Gin Ser Gly Thr Ser Gin Ala Ala
375 380 385 390

CAC GTG GCT ATT GCA GCC ATG GTG CTC GCC GAG CCG GAG CTC
His Val Ala Gly Ile Ala Ala Met Met Leu Ser Ala Glu Pro Glu Leu
395 400 405

ACC CTC GCC GAG TGG AGG CAG AGA CTG ATC ACC TAC TTC TCT GCC AAA GAT
Thr Leu Ala Glu Leu Arg Gin Arg Leu Ile His Phe Ser Ala Lys Asp
410 415 420

GTC ATC AAT GAG GCC TGG TTC CTC GAG GAC CAG CCG GTC ATG ACC CCC
1412


<210> 2

<211> 589

<212> PRT

<213> Homo sapiens

<400> 2

Met Gly Thr Val Ser Ser Arg Arg Ser Trp Trp Pro Leu Pro Leu
1  5  10  15
Leu Leu Leu Leu Leu Leu Leu Gly Pro Ala Gly Ala Arg Ala Gin Glu
20  25  30
Asp Glu Asp Gly Asp Tyr Glu Glu Leu Val Leu Ala Leu Arg Ser Glu
35  40  45
Glu Asp Gly Leu Ala Glu Ala Pro Glu His Gly Thr Thr Ala Thr Phe
50  55  60
His Arg Cys Ala Lys Asp Pro Trp Arg Leu Pro Gly Thr Tyr Val Val
65  70  75  80
Val Leu Lys Gly Thr His Leu Ser Gin Ser Glu Arg Thr Ala Arg
85  90  95
Arg Leu Gin Ala Gin Ala Ala Arg Arg Gly Tyr Leu Thr Lys Ile Leu
100  105  110
His Val Phe His Gly Leu Leu Pro Gly Phe Leu Val Lys Met Ser Gly
115  120  125
Asp Leu Gly Leu Leu Leu Pro His Val Asp Tyr Ile Glu
130  135  140
Glu Asp Ser Ser Val Phe Ala Gin Ser Ile Pro Trp Asn Leu Gly Arg
145  150  155  160
Ile Thr Pro Pro Arg Tyr Arg Ala Asp Gly Thr Gin Pro Pro Asp Gly
165  170  175
Gly Ser Leu Val Glu Val Tyr Leu Asp Thr Ser Ile Gin Ser Asp
180  185  190
His Arg Glu Ile Glu Gly Arg Val Met Val Thr Asp Phe Glu Asn Val
195  200  205
Pro Glu Asp Gly Thr Arg Phe His Arg Gin Ala Ser Lys Cys Asp
210  215  220
Ser His Gly Thr His Thr His Ala Gly Val Val Ser Gly Arg Asp Ala Gin
225  230  235  240
Val Ala Lys Gly Ala Ser Met Arg Ser Leu Arg Val Leu Asn Cys Gin
245  250  255
Gly Lys Gly Thr Val Ser Gly Thr Leu Ile Gly Leu Phe Ile Arg
260  265  270
Lys Ser Gin Leu Val Gin Pro Val Gly Pro Leu Val Val Leu Leu Pro
275  280  285
Leu Ala Gly Tyr Ser Leu Leu Asn Ala Ala Cys Gin Arg Leu
290  295  300
Ala Arg Val Gly Val Leu Val Thr Ala Gly Asn Fhe Asp Arg
305  310  315  320
Asp Ala Cys Leu Tyr Ser Pro Ala Ser Ala Pro Glu Val Ile Thr Val
325  330  335
Gly Ala Thr Asn Ala Gin Asp Gin Pro Val Thr Leu Gly Thr Leu Gly
340  345  350
Thr Asn Phe Gly Arg Cys Val Asp Leu Phe Ala Pro Gly Glu Asp Ile
355  360  365
Ile Gly Ala Ser Ser Asp Cys Ser Thr Cys Phe Val Ser Gin Ser Gly
370  375  380
Thr Ser Gin Ala Ala His Val Ala Gly Ile Ala Ala Met Met Leu
385  390  395  400
Ser Ala Glu Pro Glu Leu Thr Leu Ala Glu Leu Arg Gin Arg Leu Ile
405  410  415
His Phe Ser Ala Lys Asp Val Ile Asn Gin Ala Trp Phe Pro Glu Asp
420  425  430
Gln Arg Val Leu Thr Pro Asn Leu Val Ala Ala Leu Pro Pro Ser Thr
<table>
<thead>
<tr>
<th>Amino Acid Sequence</th>
<th>DNA Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>His Gly Ala Gly Trp Glu Leu Phe Cys Arg Thr Val Trp Ser Ala His</td>
<td>cccacaggttc ggccttgag gatgtaagca ggcagtctga gttgcctcggt ggcccgggga</td>
</tr>
<tr>
<td>Ser Gly Pro Thr Arg Met Ala Thr Ala Ile Ala Arg Cys Ala Pro Asp</td>
<td>gcgcgtctgg gcacagggcc tcctcagtc gcacgaggtc gcggtcgcttc cctctacgct</td>
</tr>
<tr>
<td>Glu Glu Leu Leu Ser Cys Ser Ser Phe Ser Arg Ser Gly Lys Arg Arg</td>
<td>cctgtcctct gcacactcag tctgctcagc tcctctcctc cagctaagctc aagggccgcc</td>
</tr>
<tr>
<td>Gly Glu Arg Met Glu Ala Gln Gly Gly Lys Leu Val Cys Arg Ala His</td>
<td>cgccgctggac cgccacgacgc ctcttagctt cctgctcagc acagcaacct ccctcctgctt</td>
</tr>
<tr>
<td>Asn Ala Phe Gly Gly Glu Gly Val Tyr Ala Ile Ala Arg Cys Cys Leu</td>
<td>cctc atg gcc acc gtc agc tcc agg cgg tcc tgg tgg ccc ctc ctc</td>
</tr>
<tr>
<td>Leu Pro Gln Ala Asn Cys Ser Val His Thr Ala Pro Ala Glu Ala Glu</td>
<td>Met Gly Thr Val Ser Ser Arg Arg Ser Trp Trp Pro Leu Pro Leu</td>
</tr>
<tr>
<td>Ser Met Gly Thr Arg Val His Cys His Gln Gln Gly His Val Leu Thr</td>
<td>ctg ctg ctg ctg ctg ctc ctg gtt ccc ggc ggc ggc gct ggc gag</td>
</tr>
<tr>
<td>Gly Phe Leu Ala Ala Ser Asp Leu Lys Glu Arg Gly Ser Asp Gly</td>
<td>gag gcg gac gag gcc gac tac gag gag ctg tga ctc gtt gtc tcc</td>
</tr>
<tr>
<td>Asp Gly His Trp Arg Arg Ser Ile Pro Ala Phe His Ile</td>
<td>gag gag gcc gtc gcc ggc aa gca gca gcc gag cag cag cgc ctc</td>
</tr>
</tbody>
</table>

**DNA Sequences:**

- 3617: cccacaggttc ggccttgag gatgtaagca ggcagtctga gttgcctcggt ggcccgggga
- 3618: gcgcgtctgg gcacagggcc tcctcagtc gcacgaggtc gcggtcgcttc cctctacgct
- 3619: cctgtcctct gcacactcag tctgctcagc tcctctcctc cagctaagctc aagggccgcc
- 3620: cgccgctggac cgccacgacgc ctcttagctt cctgctcagc acagcaacct ccctcctgctt
- 3621: Met Gly Thr Val Ser Ser Arg Arg Ser Trp Trp Pro Leu Pro Leu

**Codon Table:**

- 3622: ctg ctg ctg ctg ctg ctc ctg gtt ccc ggc ggc ggc gct ggc gag
- 3623: gag gcg gac gag gcc gac tac gag gag ctg tga ctc gtt gtc tcc
- 3624: gag gag gcc gtc gcc ggc aa gca gca gcc gag cag cag cgc ctc
- 3625: ttc cac cgc tgc gcc aag gat ccg tgg agg tgt ccc ggc acc tac gtc
- 3626: gtc gtc ctg aag gag gcc cac tct tcg gac tca gac gcc gtc tct
- 3627: cgc cgctg cag gcc cag gct gcc cgcc ccgg gga tac tcc acc aag atc
- 3628: ctg cat gtc ttc cat gcc ctt ctt cct gcc ttc ctg aag atg atg
- 3629: gcc gcg ctg gtc ggt ggc cgg tgg aag tgt ccc cat gtc gac tac atc

**Protein Sequences:**

- 3630: His Gly Ala Gly Trp Glu Leu Phe Cys Arg Thr Val Trp Ser Ala His
- 3631: Ser Gly Pro Thr Arg Met Ala Thr Ala Ile Ala Arg Cys Ala Pro Asp
- 3632: Glu Glu Leu Leu Ser Cys Ser Ser Phe Ser Arg Ser Gly Lys Arg Arg
- 3633: Gly Glu Arg Met Glu Ala Gln Gly Gly Lys Leu Val Cys Arg Ala His
- 3634: Asn Ala Phe Gly Gly Glu Gly Val Tyr Ala Ile Ala Arg Cys Cys Leu
- 3635: Leu Pro Gln Ala Asn Cys Ser Val His Thr Ala Pro Ala Glu Ala Glu
- 3636: Ser Met Gly Thr Arg Val His Cys His Gln Gln Gly His Val Leu Thr
- 3637: Gly Phe Leu Ala Ala Ser Asp Leu Lys Glu Arg Gly Ser Asp Gly
- 3638: Asp Gly His Trp Arg Arg Ser Ile Pro Ala Phe His Ile

**Codon Count:**

- 3617: 60
- 3618: 120
- 3619: 180
- 3620: 240
- 3621: 289
- 3622: 337
- 3623: 385
- 3624: 433
- 3625: 481
- 3626: 529
- 3627: 577
- 3628: 625
- 3629: 673
Gly Asp Leu Leu Glu Leu Ala Leu Lys Leu Pro His Val Asp Tyr Ile
130 135 140

gag gag gac tcc tct gtc ttt gcc cag agc atc ccg tgg aac ctg gag
Glu Glu Asp Ser Ser Val Phe Ala Gln Ser Ile Pro Trp Asn Leu Glu
145 150 155

cgg att acc cct cca cgg tac cgg ggc gat gaa tac cag ccc ccc gac
Arg Ile Thr Pro Pro Arg Tyr Ala Asp Gly Tyr Gln Pro Pro Asp
160 165 170 175

gga ggc agc ctg gtg gag gtg tat ctc cta gac acc agc ata cag agt
Gly Gly Ser Leu Val Val Tyr Leu Leu Asp Thr Ser Ile Gln Ser
180 185 190

gac cac cgg gaa atc gag ggc agg gtg atg gtc acc gac ttc gag aat
Asp His Arg Glu Ile Glu Gly Arg Val Met Val Thr Asp Phe Asn
195 200 205

gtg ccc gag gag gac ggg acc cgc ttc cac aga cag gcc acc aag tgt
Val Pro Glu Glu Asp Gly Thr Arg Phe His Arg Gln Ala Ser Lys Cys
210 215 220

gac agt cat ggc acc cac ctg gca ggg gtg gtc agc ggc cgg gat gcc
Asp Ser His Gly Thr His Leu Ala Gly Val Ser Gly Arg Asp Ala
225 230 235

ggc gtg gcc aag ggt gcc agc atg cgc agc ctg cgc gtg ctc aac tgt
Gly Val Ala Lys Gly Ala Ser Met Arg Ser Leu Arg Val Leu Asn Cys
240 245 250 255

caa ggg aag ggc acg gtt agc ggc acc ctc ata ggc ctg gag ttt att
Gln Gly Lys Gly Thr Val Leu Ser Gly Thr Leu Ile Gly Glu Phe Ile
260 265 270

cgg aca agc cag ctg gtc cag cct gtg ggg cca ctg gtg gtg ctg ctg
Arg Lys Ser Glu Leu Val Gln Pro Val Gly Pro Leu Val Leu Leu
275 280 285

ccc ctg gcc ggt ggg tac agc cgc gtc ctc aac gcc gcc tgc cag cgc
Pro Leu Ala Gly Gly Tyr Ser Arg Val Leu Asn Ala Ala Cys Gln Arg
290 295 300

cgt gcc agg gct ggg gtg ctc gtc acc gct gcc ggc aac ttc cgg
Leu Ala Arg Ala Gly Val Val Leu Val Thr Ala Ala Gly Asn Phe Arg
305 310 315

gac gat gcc tgc ctc tac tcc cca gcc tca gct ccc gac ggc atc aca
Asp Asp Ala Cys Leu Tyr Ser Pro Ala Ser Ala Pro Glu Val Ile Thr
320 325 330 335

gtt ggg gcc acc aat gcc cag gac cag cgg gtc acc ctg ggg act tgt
Val Gly Ala Thr Asn Ala Glu Asp Glu Pro Val Thr Leu Gly Thr Leu
340 345 350

ggg acc aac ttt ggc cgc tgt gtt gac ctc ttt gcc cca ggg gag gac
Gly Thr Asn Phe Gly Arg Cys Val Asp Phe Ala Pro Gly Glu Asp
355 360 365

atc att gtg gcc tcc agc gac tgc agc acc tgc ttt gtg tca cag agt
Ile Ile Gly Ala Ser Ser Asp Cys Ser Thr Cys Phe Val Ser Gln Ser
370 375 380

ggg tca cag gct gcc cac gtg gct ggc att gca gcc atg atg
Gly Thr Ser Gln Ala Ala Ala His Val Ala Gly Ile Ala Ala Met Met
385 390 395
gtg aca gcc gtt gcc atc tgc tgc cgg agc cgg cac ctg ggc cag gcc
Val Thr Ala Val Ala Ile Cys Cys Arg Ser Arg His Leu Ala Gln Ala
675
680
685
tcc cag gag ctc cag tgcacgccccc atcccaggggt ggtgtctgg ggagggccta
Ser Gln Glu Leu Glu
690
2305
2360
gggctgggcc tcagctttta aatgcttcc cgcttctcct cttcctgaccc tccatgctcc
2420
gcggcaggg gtatggtgtg ttctcctgcc tcctgctggt tcggtctgtgc ctttgctgtg
2480
gcgcctgcc tggctctgaa cattcctca ctggctctggt tcgcttcagct gccgtttctgc
2540
ggagagac gccgtctctct ctgggtcatt tccaggtcag caagagcgtgg ctggtgctg
2600
gggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
2660
tctgttctcg gctgctgtcct caacctcttg cgggtctctct tcggtgctgg gcagagcgtg
2720
tggataggg ggtcggcctg tgggtcctgt ttcggcctgg tgggtcctgt ttcggcctgg
2780
2840
taatgaggtgc ctcggtctct ctgggtcatt tccaggtcag caagagcgtgg ctggtgctg
2880
tggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
2940
tggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
2980
tggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
3020
tggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
3080
tggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
3140
tggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
3200
tggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
3260
tggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
3320
tggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
3380
tggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
3440
tggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
3500
tggtgctggca ggtcctcccag tagctggcag tgcctgctgg gcagagcgtgg ctggtgctg
3560
<210> 4
<211> 692
<212> PRT
<213> Homo sapiens

<400> 4
Met Gly Thr Val Ser Ser Arg Arg Ser Thr Trp Pro Leu Pro Leu Leu
1     5     10     15
Leu Leu Leu Leu Leu Leu Leu Gly Pro Ala Gly Ala Arg Arg Ser Ser
20    25    30
Asp Glu Asp Gly Asp Tyr Glu Glu Leu Val Leu Ala Leu Arg Ser Glu
35    40    45
Glu Asp Gly Leu Ala Glu Ala Pro Glu His Gly Thr Thr Ala Thr Phe
50    55    60
His Arg Cys Ala Lys Asp Pro Trp Arg Leu Pro Gly Thr Thr Tyr Val Val
65    70    75    80
Val Leu Lys Glu Glu Thr His Leu Ser Gln Ser Glu Arg Thr Ala Arg
85    90    95
Arg Leu Gln Ala Glu Ala Ala Arg Arg Gly Tyr Leu Thr Lys Ile Leu
100   105   110
Arg Val Phe His Gly Leu Leu Pro Gly Phe Val Leu Val Lys Met Ser Gly
115   120   125
130   135   140
Asp Leu Leu Glu Leu Ala Lys Leu Pro His Val Asp Tyr Ile Glu
145   150   155   160
165   170   175
Glu Ser Val Leu Val Tyr Leu Leu Asp Ser Thr Ser Ile Glu Ser Asp
180   185   190
195   200   205
His Arg Glu Ile Glu Gly Arg Val Met Val Thr Asp Phe Gln Asn Val
210   215   220
225   230   235   240
Val Ala Lys Gly Ala Ser Met Arg Ser Leu Arg Val Leu Asn Cys Gln 245 250 255
Gly Lys Gly Thr Val Ser Gly Thr Leu Ile Gly Leu Glu Phe Ile Arg 260 265 270
Lys Ser Gln Leu Val Gln Pro Val Gly Pro Leu Val Val Leu Leu Pro 275 280 285
Leu Ala Gly Gly Tyr Ser Arg Val Leu Asn Ala Ala Ala Cys Gln Arg Leu 290 295 300
 Ala Arg Ala Gly Val Leu Val Thr Ala Ala Gly Asn Phe Arg Asp 305 310 315 320
Asp Ala Cys Leu Tyr Ser Pro Ala Ser Pro Glu Val Ile Thr Val 325 330 335
Gly Ala Thr Asn Ala Gln Asp Gln Pro Val Thr Leu Thr Leu Gly 340 345 350
Thr Asn Phe Gly Arg Cys Val Asp Leu Phe Ala Pro Gly Glu Asp Ile 355 360 365
Ile Gly Ala Ser Ser Asp Cys Ser Thr Cys Phe Val Ser Gln Ser Gly 370 375 380
Thr Ser Gln Ala Ala His Val Ala Gly Ile Ala Ala Met Met Leu 385 390 395 400
Ser Ala Glu Pro Glu Leu Thr Leu Ala Glu Leu Arg Gln Arg Leu Ile 405 410 415
His Phe Ser Ala Lys Asp Val Ile Asn Glu Ala Trp Phe Pro Glu Asp 420 425 430
Gln Arg Val Leu Thr Pro Asn Leu Val Ala Ala Leu Pro Pro Ser Thr 435 440 445
His Gly Ala Gly Trp Gln Leu Phe Cys Arg Thr Val Trp Ser Ala His 450 455 460
Ser Gly Pro Thr Arg Met Ala Ala Ala Ala Arg Cys Ala Pro Asp 465 470 475 480
Glu Glu Leu Leu Ser Cys Ser Ser Phe Ser Arg Gly Ser Lys Arg Arg 485 490 495
Gly Glu Arg Met Glu Ala Gln Gly Gly Gly Lys Leu Val Cys Arg Ala His 500 505 510
Asn Ala Phe Gly Gly Glu Gly Val Tyr Ala Ile Ala Arg Cys Cys Leu 515 520 525
Leu Pro Gln Ala Asn Cys Ser Val His Thr Ala Pro Pro Ala Glu Ala 530 535 540
 Ser Met Gly Thr Arg Val His Cys His Gln Gln Gly His Val Leu Thr 545 550 555 560
Gly Cys Ser Ser His Trp Glu Val Glu Asp Leu Gly Thr His Lys Pro 565 570 575
Pro Val Leu Arg Pro Arg Gly Gln Pro Asn Gln Cys Val Gly His Arg 580 585 590
Glu Ala Ser Ile His Ala Ser Cys His Ala Pro Gly Leu Glu Cys 595 600 605
Lys Val Lys Glu His Gly Ile Pro Ala Pro Glu Gln Glu Val Thr Val 610 615 620
Ala Cys Glu Gly Gly Trp Thr Leu Thr Gly Cys Ser Ala Leu Pro Gly 625 630 635 640
Thr Ser His Val Leu Gly Ala Tyr Ala Val Asp Asn Thr Cys Val Val 645 650 655
Arg Ser Arg Asp Val Ser Thr Gly Ser Thr Ser Glu Ala Val 660 665 670
Thr Ala Val Ala Ile Cys Cys Arg Ser Arg His Leu Ala Gln Ala Ser 675 680 685
Gln Glu Leu Gln 690

<210> 5
<211> 2259
<212> DNA
<213> Mus musculus

<220>
<221> CDS
<222> (145)...(2256)
<400> 5
ggacctccac gttgacgcgcg aggcgtcggccg tgggtctccccg ttcttctctct tttctgaggg 60
cctagaggact ggcacgacgct tgggtctccccg agacataccg ccggcggcag cccggagcc 120
agggccggc agctccgagcc gttc atgc tcct ccag cgc gca cct 171
Met Ser Phe Pro Arg Pro Arg Ala Pro 1 5

tct ctc gcc ccg atg gcc acc cac tgc tct ggc tgg ctg cgg tgg ccg 219
Leu Leu Ala Pro Met Gly Thr His Cys Ser Ala Trp Leu Arg Trp Pro 10 15 20 25

tgg tct ccg ctg ctg ctg ctg tgg ctg ctg ctg ctg tgc ccc aac ggc 267
Leu Leu Pro Leu Leu Leu Leu Leu Leu Leu Leu Leu Cys Pro Thr Gly 30 35 40

gct ggt acc cag cag gag gac gga gat tat gaa gac ctg atg ccct gcc 315
Ala Gly Thr Gln Asp Glu Asp Gly Asp Tyr Glu Leu Met Leu Ala 45 50 55

tct ccg tctc gag gat ggc ctg gct gat gac gac gca cat ctg gcc 363
Leu Pro Ser Gln Glu Asp Gly Asp Ala Ala His Val Ala 60 65 70

acc gcc acc ttc cgc cgt tgc tcc aag gag gcc tgg agg ctg cca gga 411
Thr Ala Thr Phe Arg Arg Cys Ser Lys Ala Trp Arg Leu Pro Gly 75 80 85

acc tac att tgt tgt tgt tgt gag gac gac cag agg cta cag att gaa 459
Thr Tyr Ile Val Val Leu Met Glu Glu Thr Gln Arg Leu Gln Ile Glu 90 95 100 105

caa act gcc cac cgc ctg cag acc cgg gct gcc cgc cgg ggc tat gtc 507
Gln Thr Ala His Arg Leu Gln Thr Arg Ala Ala Arg Gly Tyr Val 110 115 120

atc aag gct cta cat atc ttt tat gac ctc ttc cct gcc ttc tgg ctg 555
Ile Lys Val Leu His Ile Phe Tyr Asp Leu Phe Pro Gly Phe Leu Val 125 130 135

aag atg agc agt gac ctg tgt gcc ctg gcc ctg aag tgt ccc cat ctg 603
Lys Met Ser Ser Asp Leu Leu Leu Leu Lys Pro His Leu Pro His 140 145 150

gag tac att gag gaa gag gcc ccc ttt gtc ttc gcc cag agc atc cca tgg 651
Glu Tyr Ile Glu Glu Asp Ser Phe Val Phe Ala Gin Ser Phe Pro Trp 155 160 165

aac ctg gag cga att atc cca gca tgt cac cag cag aca gag gaa gac cgc 699
Asn Leu Glu Arg Ile Ile Pro Ala Trp His Glu Thr Glu Asp Arg 170 175 180 185

tcc cct gat gga aagc agc cag gtt gac ggg gag gtt tac acc atgc 747
Ser Pro Asp Gly Ser Ser Gin Val Glu Val Tyr Leu Leu Gly Thr Ser 190 195 200

atc cag ggt gcc tat cgg gag att gcc ggc aag gcct acc aag ctc ccg 795
Ile Gln Gly Ala Tyr Arg Glu Ile Ala Gly Arg Val Thr Ile Thr Asp 205 210 215

ttc aac ggt gct ggc cag gag gat ggg aca cgc ttc cac agra cag ggc 843
Phe Asn Ser Val Pro Glu Glu Asp Gly Thr Arg Phe His Arg Gin Ala 220 225 230

agc aag tgt gac agc cac gcg acc cac ctg gca ggt tgt gtc aag ggc 891
Ser Lys Cys Asp Ser His Gly Thr His Leu Ala Gly Val Val Ser Gly 235 240 245
cgg gat gct ggt gtg gcc aag ggc acc agc ctg cac agc ctg cgt gtg
Arg Asp Ala Gly Val Ala Lys Gly Thr Ser Leu His Ser Leu Arg Val
250 255 260 265

tc aac tgt cca ggg aag ggc aca gtc agc ggc acc ctc ata ggc ctg
Leu Asn Cys Glu Gly Lys Gly Thr Val Ser Gly Thr Leu Ile Gly Leu
270 275 280

gag ttt att cgg aag gct cag cta atc cag ccc tcg ggg cca ctc gtg
Glu Phe Ile Arg Lys Ser Gln Leu Ile Gln Pro Ser Gly Pro Leu Val
285 290 295

gtt ctg ctg ccc ctg gcc ggt ggg tat agc cgc atc ctc aac gct gcc
Val Leu Leu Pro Leu Ala Gly Gly Tyr Ser Arg Ile Leu Asn Ala Ala
300 305 310

tgc cag cac ctg gcc agg act ggg gtg gtg ctg gtt gca gca gct ggg
Cys Gln His Leu Ala Arg Thr Gly Val Val Leu Val Ala Ala Gly
315 320 325

aac ttg cgg gag gac gcc tgc ctc tac ttc cca gct tct gct cca gag
Asn Phe Arg Asp Ala Cys Leu Tyr Ser Pro Ala Ser Ala Pro Glu
330 335 340 345

gtc atc atc gtc ggg gcc acg aat gcc cag cac cag cca gtt acc ttg
Val Ile Thr Val Gly Ala Thr Asn Ala Glu Asp Gln Pro Val Thr Leu
350 355 360

ggg act tgt ggg act aat ttt gga cgc tgt gtt gat ctc tgt ttt gcc ccc
Gly Thr Leu Gly Thr Asn Phe Gly Arg Cys Val Asp Leu Phe Ala Pro
365 370 375

ggg aag gag atc atc gga ggc tcc cgg ctc aag gac aca tgc ttc atg
Gly Lys Asp Ile Ile Gly Ala Ser Ser Asp Cys Ser Thr Cys Phe Met
380 385 390

tca cag cgt cgg acc tca cag gct gct gcc cac gtc gcc ggc att gtc
Ser Gln Ser Gly Thr Ser Gln Ala Ala Ala His Val Gly Val Ala
395 400 405

gct cgg atg ctc agc cgg gag ccc aca ctt acc ctc cgg gag cgg cgg
Ala Arg Met Leu Ser Arg Glu Pro Thr Leu Thr Leu Ala Glu Leu Arg
410 415 420 425

cag agg ctc atc cac ttc tct acc aac gac gtc atc atc aac atg gcc tgg
Gln Arg Leu Ile His Phe Ser Thr Lys Asp Val Ile Asn Met Ala Trp
430 435 440

tcc ctt gag gsc cag cag gtt ctg ccc ccc aac ctg gtg ccc aca ctc
cag Gln Phe Glu Asp Gln Gln Val Leu Thr Pro Asn Leu Val Ala Thr Leu
445 450 455

ccc ccc agg acc cat gag aca ggc ggg cag ctc tgt agg acg tgt
Pro Pro Ser Thr His Glu Gly Lys Gly Leu Cys Arg Thr Val
460 465 470

tgg tgc gcc cac tcc ggg ccc act cgg aca gct aca gct aca gcc ccc
gln Trp Ser Ala His Ser Gly Pro Thr Arg Thr Arg Ala Thr Ala Arg
475 480 485

tgt gcc cca gaa gag gag cag ctc tgt ccc aac ctg ctc aac gcg
Cys Ala Pro Glu Glu Glu Leu Leu Ser Ser Phe Ser Arg Ser
490 495 500 505

ggg agg cgt cgt gat tgt att gag gcc ata gga ggc gag cag cag gtc
Gly Arg Arg Arg Gly Asp Trp Ile Glu Ala Ile Gly Gly Gln Gln Val
1707
tgc aag gcc ctc aat gca ttt ggg ggt gag gtg tct tat gcc gtc gcg
Cys Lys Ala Leu Asn Ala Phe Gly Gly Glu Val Tyr Ala Val Ala
525 530 535
aga tgc tgc ctg gtt ccc cat gcc aac tgc agc aac aac acc cct
Arg Cys Cys Leu Val Pro His Ala Asn Cys Ser Ile His Asn Thr Pro
1803
540 545 550
gca gcc aga gct gcc ctg gag acc cat gtc cac gcc cac cag aag gac
Ala Arg Ala Gly Leu Glu Thr His Val His Cys His Glu Lys Asp
1851
555 560 565
cat gtt tct cca gcc tgc agc ttc cat tgg gaa gtg gaa gac ctt agt
His Val Leu Thr Gly Cys Ser Phe His Trp Glu Val Glu Asp Leu Ser
1899
570 575 580 585
gtc cgg agg cag cct gcc ctt ggc cag ctt cgc cag cgg cag cag tgc
Val Arg Glu Arg Glu Arg Leu Arg Ser Arg Arg Glu Gly Glu Cys
1947
590 595 600
gtt gcc cac cag gcc gtc agt gtg tac tgc tgc tgc gtt agt ccc caa
Val Gly His Glu Ala Ala Ser Tyr Ala Ser Cys His Ala Ser Pro
1995
605 610 615
ggg ctc gaa tgg aac atg gaa gag cat ggc atc tca ggt cct tca gag
Gly Leu Glu Cys Lys Ile Lys Glu His Gly Ile Ser Gly Ser Glu
2043
620 625 630
cag gtc gct gtt gcc tgt gaa gca gga tgt acc tgt acct gga tgt aat
Gln Val Ala Val Ala Cys Glu Ala Gly Trp Thr Leu Thr Gly Cys Asn
2091
635 640 645
gtg ctc cct ggg cca tcc ctc act ctc gta gcc tac agc gtc gac aac
Val Leu Pro Gly Ala Ser Leu Thr Leu Gly Ala Tyr Ser Val Asp Asn
2139
650 655 660 665
ctg tgt gca cag gtc cat gac act gcc aga gca gag aag aag acc agt
Leu Cys Val Ala Val Arg Val His Asp Thr Ala Arg Ala Asp Arg Thr Ser
2187
670 675 680
gga gaa gcc cca gta gct gcc gcc atc gcc tgc gcg aac ggg cct tca
Gly Glu Ala Thr Val Ala Ala Ala Ile Cys Cys Arg Ser Arg Pro Ser
2235
695 690 695
gca aag gcc tcc tgg gtt cag tga
Ala Lys Ala Ser Trp Val Glu
2259
700
<table>
<thead>
<tr>
<th>Ser</th>
<th>Lys</th>
<th>Glu</th>
<th>Ala</th>
<th>Trp</th>
<th>Arg</th>
<th>Leu</th>
<th>Pro</th>
<th>Gly</th>
<th>Thr</th>
<th>Tyr</th>
<th>Ile</th>
<th>Val</th>
<th>Val</th>
<th>Leu</th>
<th>Met</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Glu</td>
<td>Thr</td>
<td>Gin</td>
<td>Arg</td>
<td>Leu</td>
<td>Gin</td>
<td>Ile</td>
<td>Glu</td>
<td>Gin</td>
<td>Thr</td>
<td>Ala</td>
<td>His</td>
<td>Arg</td>
<td>Leu</td>
<td>Gin</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>110</td>
</tr>
<tr>
<td>Thr</td>
<td>Arg</td>
<td>Ala</td>
<td>Ala</td>
<td>Arg</td>
<td>Gly</td>
<td>Tyr</td>
<td>Val</td>
<td>Ile</td>
<td>Lys</td>
<td>Val</td>
<td>Leu</td>
<td>His</td>
<td>Ile</td>
<td>Phe</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>120</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>125</td>
</tr>
<tr>
<td>Tyr</td>
<td>Asp</td>
<td>Leu</td>
<td>Phe</td>
<td>Pro</td>
<td>Gly</td>
<td>Phe</td>
<td>Leu</td>
<td>Val</td>
<td>Lys</td>
<td>Met</td>
<td>Ser</td>
<td>Ser</td>
<td>Ser</td>
<td>Asp</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>135</td>
</tr>
<tr>
<td>Gly</td>
<td>Leu</td>
<td>Ala</td>
<td>Leu</td>
<td>Lys</td>
<td>Leu</td>
<td>Pro</td>
<td>His</td>
<td>Val</td>
<td>Glu</td>
<td>Tyr</td>
<td>Ile</td>
<td>Glu</td>
<td>Asp</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>145</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>140</td>
</tr>
<tr>
<td>Phe</td>
<td>Val</td>
<td>Phe</td>
<td>Ala</td>
<td>Gin</td>
<td>Ser</td>
<td>Ile</td>
<td>Pro</td>
<td>Trp</td>
<td>Asn</td>
<td>Leu</td>
<td>Arg</td>
<td>Gin</td>
<td>Arg</td>
<td>Ile</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>165</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>170</td>
</tr>
<tr>
<td>Ala</td>
<td>Trp</td>
<td>His</td>
<td>Gin</td>
<td>Thr</td>
<td>Glu</td>
<td>Asp</td>
<td>Arg</td>
<td>Ser</td>
<td>Pro</td>
<td>Asp</td>
<td>Gly</td>
<td>Ser</td>
<td>Ser</td>
<td>Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>180</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>185</td>
</tr>
<tr>
<td>Val</td>
<td>Glu</td>
<td>Val</td>
<td>Tyr</td>
<td>Leu</td>
<td>Leu</td>
<td>Gly</td>
<td>Thr</td>
<td>Ser</td>
<td>Ile</td>
<td>Gin</td>
<td>Gin</td>
<td>Gly</td>
<td>Ala</td>
<td>Tyr</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>190</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>195</td>
</tr>
<tr>
<td>Ile</td>
<td>Ala</td>
<td>Gly</td>
<td>Arg</td>
<td>Val</td>
<td>Thr</td>
<td>Ile</td>
<td>Thr</td>
<td>Asp</td>
<td>Phe</td>
<td>Asn</td>
<td>Ser</td>
<td>Val</td>
<td>Pro</td>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>210</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>215</td>
</tr>
<tr>
<td>Asp</td>
<td>Gly</td>
<td>Thr</td>
<td>Arg</td>
<td>Phe</td>
<td>His</td>
<td>Arg</td>
<td>Gin</td>
<td>Ala</td>
<td>Ser</td>
<td>Lys</td>
<td>Cys</td>
<td>Asp</td>
<td>Ser</td>
<td>His</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>225</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>230</td>
</tr>
<tr>
<td>Thr</td>
<td>His</td>
<td>Leu</td>
<td>Ala</td>
<td>Gly</td>
<td>Val</td>
<td>Ser</td>
<td>Gly</td>
<td>Arg</td>
<td>Asp</td>
<td>Ala</td>
<td>Gin</td>
<td>Gly</td>
<td>Val</td>
<td>Ala</td>
<td>Lys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>245</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>250</td>
</tr>
<tr>
<td>Gly</td>
<td>Thr</td>
<td>Ser</td>
<td>Leu</td>
<td>His</td>
<td>Ser</td>
<td>Leu</td>
<td>Arg</td>
<td>Val</td>
<td>Leu</td>
<td>Asn</td>
<td>Cys</td>
<td>Gin</td>
<td>Gly</td>
<td>Lys</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>260</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>265</td>
</tr>
<tr>
<td>Thr</td>
<td>Val</td>
<td>Ser</td>
<td>Gly</td>
<td>Thr</td>
<td>Leu</td>
<td>Ile</td>
<td>Gly</td>
<td>Leu</td>
<td>Glu</td>
<td>Phe</td>
<td>Ile</td>
<td>Arg</td>
<td>Lys</td>
<td>Ser</td>
<td>Gin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>280</td>
</tr>
<tr>
<td>Leu</td>
<td>Ile</td>
<td>Gin</td>
<td>Pro</td>
<td>Ser</td>
<td>Gly</td>
<td>Pro</td>
<td>Leu</td>
<td>Val</td>
<td>Leu</td>
<td>Leu</td>
<td>Pro</td>
<td>Leu</td>
<td>Ala</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>290</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>295</td>
</tr>
<tr>
<td>Gly</td>
<td>Tyr</td>
<td>Ser</td>
<td>Arg</td>
<td>Ile</td>
<td>Leu</td>
<td>Asn</td>
<td>Ala</td>
<td>Ala</td>
<td>Cys</td>
<td>Gin</td>
<td>His</td>
<td>Leu</td>
<td>Ala</td>
<td>Arg</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>310</td>
</tr>
<tr>
<td>Gly</td>
<td>Val</td>
<td>Val</td>
<td>Leu</td>
<td>Val</td>
<td>Ala</td>
<td>Ala</td>
<td>Ala</td>
<td>Gin</td>
<td>Asn</td>
<td>Phe</td>
<td>Arg</td>
<td>Asp</td>
<td>Ala</td>
<td>Cys</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>325</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>330</td>
</tr>
<tr>
<td>Gly</td>
<td>Tyr</td>
<td>Ser</td>
<td>Pro</td>
<td>Ala</td>
<td>Ser</td>
<td>Ala</td>
<td>Pro</td>
<td>Glu</td>
<td>Val</td>
<td>Ile</td>
<td>Thr</td>
<td>Val</td>
<td>Gly</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>340</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>345</td>
</tr>
<tr>
<td>Asn</td>
<td>Ala</td>
<td>Gin</td>
<td>Asp</td>
<td>Gin</td>
<td>Pro</td>
<td>Val</td>
<td>Thr</td>
<td>Leu</td>
<td>Gly</td>
<td>Thr</td>
<td>Leu</td>
<td>Gly</td>
<td>Thr</td>
<td>Asn</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>355</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>360</td>
</tr>
<tr>
<td>Gly</td>
<td>Arg</td>
<td>Cys</td>
<td>Val</td>
<td>Asp</td>
<td>Leu</td>
<td>Phe</td>
<td>Ala</td>
<td>Pro</td>
<td>Gly</td>
<td>Lys</td>
<td>Asp</td>
<td>Ile</td>
<td>Ile</td>
<td>Gly</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>370</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>375</td>
</tr>
<tr>
<td>Ser</td>
<td>Ser</td>
<td>Asp</td>
<td>Cys</td>
<td>Ser</td>
<td>Thr</td>
<td>Cys</td>
<td>Phe</td>
<td>Met</td>
<td>Ser</td>
<td>Gin</td>
<td>Ser</td>
<td>Gly</td>
<td>Thr</td>
<td>Ser</td>
<td>Gin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>385</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>390</td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
<td>Asl</td>
<td>Gin</td>
<td>Ala</td>
<td>Gin</td>
<td>Ala</td>
<td>Arg</td>
<td>Met</td>
<td>Leu</td>
<td>Ser</td>
<td>Arg</td>
<td>Glu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>405</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>410</td>
</tr>
<tr>
<td>Pro</td>
<td>Thr</td>
<td>Leu</td>
<td>Thr</td>
<td>Leu</td>
<td>Ala</td>
<td>Glu</td>
<td>Leu</td>
<td>Arg</td>
<td>Gin</td>
<td>Arg</td>
<td>Leu</td>
<td>Ile</td>
<td>His</td>
<td>Phe</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>420</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>425</td>
</tr>
<tr>
<td>Thr</td>
<td>Lys</td>
<td>Asp</td>
<td>Val</td>
<td>Ile</td>
<td>Asn</td>
<td>Met</td>
<td>Ala</td>
<td>Trp</td>
<td>Phe</td>
<td>Pro</td>
<td>Glu</td>
<td>Asp</td>
<td>Gin</td>
<td>Gin</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>435</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>440</td>
</tr>
<tr>
<td>Leu</td>
<td>Thr</td>
<td>Pro</td>
<td>Asn</td>
<td>Leu</td>
<td>Val</td>
<td>Ala</td>
<td>Thr</td>
<td>Leu</td>
<td>Pro</td>
<td>Ser</td>
<td>Thr</td>
<td>His</td>
<td>Glu</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>455</td>
</tr>
<tr>
<td>Gly</td>
<td>Gly</td>
<td>Gln</td>
<td>Leu</td>
<td>Leu</td>
<td>Cys</td>
<td>Arg</td>
<td>Thr</td>
<td>Val</td>
<td>Trp</td>
<td>Ser</td>
<td>Ala</td>
<td>Arg</td>
<td>His</td>
<td>Ser</td>
<td>Gin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>465</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>470</td>
</tr>
<tr>
<td>Thr</td>
<td>Arg</td>
<td>Thr</td>
<td>Ala</td>
<td>Thr</td>
<td>Ala</td>
<td>Arg</td>
<td>Cys</td>
<td>Ala</td>
<td>Pro</td>
<td>Glu</td>
<td>Leu</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>485</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>490</td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>Cys</td>
<td>Ser</td>
<td>Ser</td>
<td>Phe</td>
<td>Ser</td>
<td>Arg</td>
<td>Ser</td>
<td>Gly</td>
<td>Arg</td>
<td>Arg</td>
<td>Asp</td>
<td>Arg</td>
<td>Gly</td>
<td>Asp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>505</td>
</tr>
<tr>
<td>Ile</td>
<td>Glu</td>
<td>Ala</td>
<td>Ile</td>
<td>Gly</td>
<td>Gly</td>
<td>Gln</td>
<td>Gin</td>
<td>Val</td>
<td>Cys</td>
<td>Lys</td>
<td>Ala</td>
<td>Leu</td>
<td>Asn</td>
<td>Ala</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>515</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>520</td>
</tr>
<tr>
<td>Gly</td>
<td>Glu</td>
<td>Gly</td>
<td>Val</td>
<td>Tyr</td>
<td>Ala</td>
<td>Val</td>
<td>Ala</td>
<td>Arg</td>
<td>Cys</td>
<td>Cys</td>
<td>Leu</td>
<td>Pro</td>
<td>His</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>530</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>535</td>
</tr>
<tr>
<td>Ala</td>
<td>Asn</td>
<td>Cys</td>
<td>Ser</td>
<td>Ile</td>
<td>His</td>
<td>Asn</td>
<td>Thr</td>
<td>Pro</td>
<td>Ala</td>
<td>Ala</td>
<td>Ala</td>
<td>Arg</td>
<td>Ala</td>
<td>Gly</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>545</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>550</td>
</tr>
<tr>
<td>Thr</td>
<td>His</td>
<td>Val</td>
<td>His</td>
<td>Cys</td>
<td>His</td>
<td>Gin</td>
<td>Lys</td>
<td>Asp</td>
<td>His</td>
<td>Val</td>
<td>Leu</td>
<td>Thr</td>
<td>Gly</td>
<td>Cys</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>565</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>570</td>
</tr>
<tr>
<td>Phe</td>
<td>His</td>
<td>Trp</td>
<td>Glu</td>
<td>Val</td>
<td>Glu</td>
<td>Asp</td>
<td>Leu</td>
<td>Ser</td>
<td>Val</td>
<td>Arg</td>
<td>Gin</td>
<td>Pro</td>
<td>Ala</td>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>585</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>590</td>
</tr>
<tr>
<td>Arg</td>
<td>Ser</td>
<td>Arg</td>
<td>Gin</td>
<td>Pro</td>
<td>Gly</td>
<td>Cys</td>
<td>Val</td>
<td>Gly</td>
<td>His</td>
<td>Gin</td>
<td>Ala</td>
<td>Ala</td>
<td>Ser</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>595</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>600</td>
</tr>
<tr>
<td>Val</td>
<td>Tyr</td>
<td>Ala</td>
<td>Ser</td>
<td>Cys</td>
<td>His</td>
<td>Ala</td>
<td>Pro</td>
<td>Gly</td>
<td>Leu</td>
<td>Glu</td>
<td>Cys</td>
<td>Lys</td>
<td>Ile</td>
<td>Lys</td>
<td></td>
</tr>
</tbody>
</table>
<210> 7
<211> 3372
<212> DNA
<213> Rattus norvegicus

<220>
<221> CDS
<222> (40)...(2277)

<400> 7
ggcgtccagc aacccacaccc taaaaggttt ttcatacttt atg tgg agc cgc agt 54
Met Trp Thr Arg Ser 1

ctg cca ctg ggc tcc cgt tct tct gac cga gac ctg agg act gag 102
Leu Pro Leu Gly Ser Arg Ser Leu Ser Asp Arg Asp Leu Arg Thr Glu 10

cca gtg ctt ggc tcc cct agg gac atc aca ggc tgc agt ccc aga gcc 150
Pro Val Leu Gly Ser Pro Arg Asp Ile Thr Ala Cys Ser Pro Arg Ala 25

cag tgc ccc ggc ttc acg tcc ttc cgc agg cct cgt gca cct ctc ctc 198
Gln Cys Pro Ala Phe Thr Ser Phe Pro Arg Pro Arg Ala Pro Leu Leu 40

gcc ccg atg ggc atc gcc tgc tct aca tgg ttt ccg tgg ctc ctg tgg 246
Ala Pro Met Gly Ile Arg Cys Ser Thr Trp Leu Arg Trp Pro Leu Ser 55

ccg cag ctg ctg tgg tgt cta ctg tgc ccc aca ggc tcc cgt gcc 294
Pro Gln Leu Leu Leu Leu Leu Leu Cys Pro Thr Gly Ser Arg Ala 70

cag gcc gag gac gga gag tcc gaa gag ctg atg ctc gcc ctc ccg tcc 342
Gln Asp Glu Asp Gly Asp Tyr Glu Glu Leu Met Leu Ala Pro Ser 90

cag gag gat agc ctg gtt gat gag gcc tca cac gtg gcc acc gcc acc 390
Gln Glu Asp Ser Leu Val Asp Glu Ala Ser His Val Ala Thr Ala Thr 105

ttc cgc cgt tgc tcc aag gag gcc tgg agg ctg cca gga acc tac cgc 438
Phe Arg Arg Cys Ser Lys Glu Ala Trp Arg Leu Pro Gly Thr Tyr Val 120

gtg ctg ctg ctg gag gag gcc cag cgg ctc cag gtt gaa cca act gcc 486
Val Val Leu Met Glu Glu Thr Gin Arg Leu Gin Val Glu Gin Thr Ala 135

cat cgc ctg cag acc tgg gcc gcc cgc cgg cgg tat gtc atc aag gtt 534
His Arg Leu Gin Thr Ala Ala Arg Gly Tyr Val Ile Lys Val 150

cag tgc ctc ttc tat gac ttc ttc cct gcc ttc tgg atg agc 582

582
Leu His Val Phe Tyr Asp Leu Phe Pro Gly Phe Leu Val Lys Met Ser
170 175 180
agt gac ctg ttg gcc ctg ccg ccg aag ttg ccc cat gtg gag tac atc
Ser Asp Leu Leu Gly Leu Ala Leu Iys Leu Pro His Val Glu Tyr Ile
185 190 195
gag gaa gac tca tta gtc ttc gcc cag agc atc cca tgg aac ctg gag
Glu Glu Asp Ser Leu Val Phe Ala Gln Ser Ile Pro Trp Asn Leu Glu
200 205 210
cgg att atc cca cgq tgg cag cag cag gaa gat agc tcc cct gac
Arg Ile Ile Pro Ala Trp Gln Glu Thr Glu Asp Ser Ser Pro Asp
215 220 225

gga agt agc cag gtg gag gtg tat ctc tta gat acc agc atc cag agt
Gly Ser Ser Gin Val Glu Val Tyr Leu Asp Ser Thr Ser Ile Gin Ser
230 235 240 245
ggc cac cgq gag atc ggg gag aga gtt acc atc act gac tcc aac agt
Gly His Arg Glu Ile Glu Gly Arg Val Thr Ile Thr Asp Phe Asn Ser
250 255 260

gtt cct gag gag ggg aca cgt ttc cac aga cag gcg agc aag tgt
Val Pro Glu Asp Gly Thr Arg Phe His Arg Gln Ala Ser Lys Cys
265 270 275

gac agc cat ggc acc cac cta gca ggt gtg gtc agc ggc cgq gat gct
Asp Ser His Gly Thr His Leu Ala Gly Val Val Ser Gly Arg Asp Ala
280 285 290

gtt tgt gcc aag ggc acc agt ctg cac agt ctg tgt ctc aac tgt
Gly Val Ala Lys Gly Thr Ser Leu His Ser Leu Arg Val Leu Asn Cys
295 300 305

caa ggg aag ggc aca gtc agc ggc acc ctc ata ggc ctg gag ttt att
Gln Gly Lys Gly Thr Val Ser Val Gly Thr Leu Ile Gly Leu Phe Ile
310 315 320 325
cgg aag agc cag cta atc cat cct cgg ggg cca ctc gtt tgt ctc tgt
Arg Lys Ser Gin Ser Ile Gin Pro Ser Gly Pro Leu Val Val Leu
330 335 340
ccc ctg ggc ggt ggg tat agc cgq atc ctt aac act gcc tgc cag cgq
Pro Leu Ala Gly Gly Tyr Ser Arg Ile Leu Asn Thr Ala Cys Gin Arg
345 350 355

cgg gcc acc act ggg gta gtg ctc gca gtt gca gct ggg aat ttc cga
Leu Ala Arg Thr Gly Val Val Leu Val Ala Ala Gly Asn Phe Arg
360 365 370

gat gat gcc tgc ctc tac tcc cca gcc tct gct cca gag gtc att aca
Asp Asp Ala Cys Leu Tyr Ser Leu Pro Ser Ala Pro Glu Val Ile Thr
375 380 385

gtt ggg gcc act aat gcc cag gac ccq gtc acc ctg ggg act ttt
Val Gly Ala Thr Asn Ala Gin Asp Gln Pro Val Thr Leu Gly Thr Leu
390 395 400 405
ggg aca aac ttt gga cgq gtt gtg gat ctc ttt gcc ccc ggg aag gac
Gly Thr Asn Phe Gly Arg Cys Val Asp Leu Phe Ala Pro Gly Lys Asp
410 415 420

cgg acc gcc tgg gcc atc gtt gcc aat ggt ctc atc ggc atc tca cag aag
Ile Ile Gly Ala Ser Ser Asp Cys Ser Thr Cys Tyr Met Ser Gin Ser
425 430 435
15
<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Protein Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGG</td>
<td>Arginine</td>
<td>1398</td>
</tr>
<tr>
<td>ACT</td>
<td>Threonine</td>
<td>1446</td>
</tr>
<tr>
<td>TCT</td>
<td>Serine</td>
<td>1494</td>
</tr>
<tr>
<td>CAC</td>
<td>Histidine</td>
<td>1542</td>
</tr>
<tr>
<td>GCC</td>
<td>Glutamine</td>
<td>1590</td>
</tr>
<tr>
<td>CCA</td>
<td>Proline</td>
<td>1638</td>
</tr>
<tr>
<td>GAA</td>
<td>Glutamic Acid</td>
<td>1686</td>
</tr>
<tr>
<td>CGG</td>
<td>Arginine</td>
<td>1734</td>
</tr>
<tr>
<td>CCA</td>
<td>Proline</td>
<td>1782</td>
</tr>
<tr>
<td>GCT</td>
<td>Alanine</td>
<td>1830</td>
</tr>
<tr>
<td>GCC</td>
<td>Glutamine</td>
<td>1878</td>
</tr>
<tr>
<td>CCA</td>
<td>Proline</td>
<td>1926</td>
</tr>
<tr>
<td>CAG</td>
<td>Histidine</td>
<td>1974</td>
</tr>
<tr>
<td>GAA</td>
<td>Glutamic Acid</td>
<td>2022</td>
</tr>
<tr>
<td>TGC</td>
<td>Cysteine</td>
<td>2070</td>
</tr>
<tr>
<td>GGC</td>
<td>Glycine</td>
<td>2118</td>
</tr>
<tr>
<td>GAA</td>
<td>Glutamic Acid</td>
<td>2166</td>
</tr>
</tbody>
</table>

This is a partial listing of the amino acid sequences found in the protein.
gca cga atc cgt gat gct gtt aga ggc gac agg acc aag gag gaa gcc
Ala Arg Ile Arg Asp Ala Gly Arg Ala Asp Arg Thr Ser Glu Glu Ala
710 715 720 725

acg gta gct gct gcc atc tgc tgc cgg gcc cgg cct tgg gca aag gcc
Thr Val Ala Ala Ala Ile Cys Cys Arg Ser Arg Pro Ser Ala Lys Ala
730 735 740

tcc tgt gtc cac cag tgcagcctc aggcagcgtc tgtctcttgt gcctggagcga
Ser Trp Val His Gln
745

2214

2262

2317

2377

2437

2497

2557

2617

2677

2737

2797

2857

2917

2977

3037

3097

3157

3217

3277

3337

3397

3457

3517

3577

3637

3732

<210> 8
<211> 746
<212> PRT
<213> Rattus norvegicus

<400> 8
Met Thr Thr Arg Ser Leu Pro Leu Gly Ser Arg Ser Arg Ser Leu Ser Asp Arg
1 5 10 15
Asp Leu Arg Thr Glu Pro Val Leu Gly Ser Pro Arg Asp Ile Thr Ala
20 25 30
Cys Ser Pro Arg Ala Gin Cys Pro Ala Phe Thr Ser Phe Pro Arg Pro
35 40 45
Arg Ala Pro Leu Leu Ala Pro Met Gly Ile Arg Cys Ser Thr Trp Leu
50 55 60
Arg Thr Pro Leu Ser Pro Gin Leu Leu Leu Leu Leu Cys Pro
65 70 75 80
Thr Gly Ser Ala Gin Asp Gin Asp Gly Asp Tyr Glu Glu Leu Met
85 90 95
Leu Ala Leu Pro Ser Gin Glu Asp Ser Leu Val Asp Glu Ala Ser His
100 105 110
Val Ala Thr Ala Thr Phe Arg Arg Cys Ser Lys Glu Ala Trp Arg Leu
115 120 125
Pro Gly Thr Tyr Val Val Val Met Glu Glu Thr Gin Arg Leu Gin
130 135 140
Val Glu Gin Thr Ala His Arg Leu Gin Thr Thr Ala Arg Arg Gly
145 150 155 160
Tyr Val Ile Lys Val Leu His Val Phe Tyr Asp Leu Phe Pro Gly Phe
165 170 175
Leu Val Lys Met Ser Ser Asp Leu Leu Gly Leu Ala Leu Lys Leu Pro
180 185 190
His Val Gly Tyr Ile Glu Gin Asp Ser Leu Val Phe Ala Gin Ser Ile
195 200 205
Pro Thr Asn Leu Glu Arg Ile Ile Pro Ala Trp Gin Gin Thr Glu Glu
210 215 220
Asp Ser Ser Pro Asp Gly Ser Ser Gin Val Glu Val Tyr Leu Leu Asp

17
<table>
<thead>
<tr>
<th></th>
<th>225</th>
<th>230</th>
<th>235</th>
<th>240</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thr Ser Ile Gln Ser Gly His Arg Glu Ile Glu Gly Arg Val Thr Ile</td>
<td>245</td>
<td>250</td>
<td>255</td>
<td></td>
</tr>
<tr>
<td>Thr Asp Phe Asn Ser Val Pro Glu Glu Asp Gly Thr Arg Phe His Arg</td>
<td>260</td>
<td>265</td>
<td>270</td>
<td></td>
</tr>
<tr>
<td>Gln Ala Ser Lys Cys Asp Ser His Gly Thr His Leu Ala Gly Val Val</td>
<td>275</td>
<td>280</td>
<td>285</td>
<td></td>
</tr>
<tr>
<td>Ser Gly Arg Asp Ala Gly Val Ala Lys Gly Thr Ser Leu His Ser Leu</td>
<td>290</td>
<td>295</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Arg Val Leu Asn Cys Gln Gly Lys Gly Thr Val Ser Gly Thr Leu Ile</td>
<td>305</td>
<td>310</td>
<td>315</td>
<td>320</td>
</tr>
<tr>
<td>Gly Leu Glu Phe Ile Arg Lys Ser Gln Leu Ile Glu Gln Pro Ser Gly Pro</td>
<td>325</td>
<td>330</td>
<td>335</td>
<td></td>
</tr>
<tr>
<td>Leu Val Val Leu Leu Pro Leu Ala Gly Gly Tyr Ser Arg Ile Leu Asn</td>
<td>340</td>
<td>345</td>
<td>350</td>
<td></td>
</tr>
<tr>
<td>Thr Ala Cys Gln Arg Leu Ala Thr Gly Val Leu Val Ala Ala</td>
<td>355</td>
<td>360</td>
<td>365</td>
<td></td>
</tr>
<tr>
<td>Ala Gly Asn Phe Arg Asp Ala Cys Leu Tyr Ser Pro Ala Ser Ala</td>
<td>370</td>
<td>375</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>Pro Glu Val Ile Thr Val Gly Ala Thr Asn Ala Gln Asp Gln Pro Val</td>
<td>385</td>
<td>390</td>
<td>395</td>
<td>400</td>
</tr>
<tr>
<td>Thr Leu Gly Thr Leu Gly Thr Asn Phe Gly Arg Cys Val Asp Leu Phe</td>
<td>405</td>
<td>410</td>
<td>415</td>
<td></td>
</tr>
<tr>
<td>Ala Pro Gly Lys Asp Ile Ile Gly Ala Ser Ser Asp Cys Ser Thr Cys</td>
<td>420</td>
<td>425</td>
<td>430</td>
<td></td>
</tr>
<tr>
<td>Tyr Met Ser Gin Ser Gly Thr Ser Gin Ala Ala Ala His Val Ala Gly</td>
<td>435</td>
<td>440</td>
<td>445</td>
<td></td>
</tr>
<tr>
<td>Ile Val Ala Met Met Leu Asn Arg Asp Pro Ala Leu Thr Leu Ala Glu</td>
<td>450</td>
<td>455</td>
<td>460</td>
<td></td>
</tr>
<tr>
<td>Leu Arg Gln Arg Leu Ile Leu Phe Ser Thr Lys Asp Val Ile Asn Met</td>
<td>465</td>
<td>470</td>
<td>475</td>
<td>480</td>
</tr>
<tr>
<td>Ala Trp Phe Pro Glu Asp Gin Arg Val Leu Thr Pro Asn Arg Val Ala</td>
<td>485</td>
<td>490</td>
<td>495</td>
<td></td>
</tr>
<tr>
<td>Thr Leu Pro Pro Ser Thr Gin Glu Thr Gly Gly Gin Leu Leu Cys Arg</td>
<td>500</td>
<td>505</td>
<td>510</td>
<td></td>
</tr>
<tr>
<td>Thr Val Trp Ser Ala His Ser Gly Pro Thr Arg Thr Ala Thr Ala Thr</td>
<td>515</td>
<td>520</td>
<td>525</td>
<td></td>
</tr>
<tr>
<td>Ala Arg Cys Ala Pro Glu Glu Leu Leu Ser Cys Ser Ser Phe Ser</td>
<td>530</td>
<td>535</td>
<td>540</td>
<td></td>
</tr>
<tr>
<td>Arg Ser Gly Arg Arg Arg Gly Asp Arg Ile Glu Ala Ile Gly Gly Gln</td>
<td>545</td>
<td>550</td>
<td>555</td>
<td>560</td>
</tr>
<tr>
<td>Gln Val Cys Lys Ala Leu Asn Ala Phe Gly Gly Glu Gly Val Tyr Ala</td>
<td>570</td>
<td>575</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val Ala Arg Cys Cys Leu Leu Pro Arg Val Asn Cys Ser Ile His Asn</td>
<td>580</td>
<td>585</td>
<td>590</td>
<td></td>
</tr>
<tr>
<td>Thr Pro Ala Ala Arg Ala Gly Pro Gin Thr Pro Val His Cys His Gin</td>
<td>595</td>
<td>600</td>
<td>605</td>
<td></td>
</tr>
<tr>
<td>Lys Asp His Val Leu Thr Gly Cys Ser Phe His Trp Glu Glu Asn</td>
<td>610</td>
<td>615</td>
<td>620</td>
<td></td>
</tr>
<tr>
<td>Leu Arg Ala Gln Gin Gin Pro Leu Leu Arg Ser Arg His Gin Pro Gly</td>
<td>625</td>
<td>630</td>
<td>635</td>
<td>640</td>
</tr>
<tr>
<td>Gln Cys Val Gly His Gin Glu Val Asn Ala Ser Cys His</td>
<td>645</td>
<td>650</td>
<td>655</td>
<td></td>
</tr>
<tr>
<td>Ala Pro Gly Leu Glu Cys Lys Ile Lys Glu His Gly Ile Ala Gly Pro</td>
<td>660</td>
<td>665</td>
<td>670</td>
<td></td>
</tr>
<tr>
<td>Ala Glu Gin Val Thr Val Ala Cys Glu Ala Gly Trp Thr Leu Thr Gly</td>
<td>675</td>
<td>680</td>
<td>685</td>
<td></td>
</tr>
<tr>
<td>Cys Asn Val Leu Pro Gly Ala Ser Leu Pro Leu Gly Ala Tyr Ser Val</td>
<td>690</td>
<td>695</td>
<td>700</td>
<td></td>
</tr>
<tr>
<td>Asp Asn Val Cys Val Ala Arg Ile Arg Asp Ala Gly Arg Ala Asp Arg</td>
<td>705</td>
<td>710</td>
<td>715</td>
<td>720</td>
</tr>
<tr>
<td>Thr Ser Glu Ala Thr Val Ala Ala Ala Ile Cys Cys Arg Ser Arg</td>
<td>725</td>
<td>730</td>
<td>735</td>
<td></td>
</tr>
<tr>
<td>Pro Ser Ala Lys Ala Ser Trp Val His Gln</td>
<td>740</td>
<td>745</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>