

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
7 November 2002 (07.11.2002)

PCT

(10) International Publication Number
WO 02/088312 A2

- (51) International Patent Classification⁷: **C12N**
- (21) International Application Number: PCT/US02/13072
- (22) International Filing Date: 26 April 2002 (26.04.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/286,382 26 April 2001 (26.04.2001) US
- (71) Applicant (for all designated States except US): **PE CORPORATION (NY)** [US/US]; 761 Main Avenue, Norwalk, CT 06859 (US).
- (71) Applicants and
(72) Inventors: **HU, Song** [US/US]; Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **GONG, Fancheng** [US/US]; Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **LADUNGA, Steven, I.** [HU/US]; Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **HIGGINS,**
- (74) Common Representative: **MONTGOMERY, Wayne, W.**; PE Corporation (NY), 761 Main Avenue, Norwalk, CT 06859 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: ISOLATED HUMAN SECRETED PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND USES THEREOF

```

1 ATGCCCCGGC ACATGGGGCT CCTGCTGGTT TGGGTCTGCC TGATTCGCGG
51 TGTGGTCTGGT GGTGTGGTCG GCGCTGTATT TAACGTTCTG GAAGAAGAAA
101 CTATAACTCA TTTCCAGAA GTTACAGATG GGGAGTGTGT CTTTCCATTC
151 CACTATAAAA ATGGAACATA TTATGACTGC ATCAAGTCCA AGGCAAGACA
201 CAAGTGGTGC TCGTTAAACA AGACCTACGA AGGATACTGG AAGTTTTCGA
251 GTGCAGAAGA TTTTGCAAAC TGTGTATTTC CCTTCTGGTA CAGACGCTTG
301 ATCTACTGGG AGTGTACTGA TGATGGGGAA GCATTTGGGA AAAAATGGTG
351 TTCACTGACC AAGAATTTTA ACAAGGACCG AATTGGGAAA TACTGTGAAT
401 GA
(SEQ ID NO:1)

```

FEATURES:

Start Codon: 1
Stop Codon: 400

Homologous proteins:

Top 10 BLAST Hits

	Score	E
gi 3024635 sp P81121 SP1_HORSE SEMINAL PLASMA PROTEIN HSP-1	116	1e-25
gi 3549883 emb CAA07074.1 (AJ006515) seminal plasma protein-1 ...	109	1e-23
gi 6138952 gb AAF04405.1 (AF055981) seminal plasma protein A3 ...	108	2e-23
gi 134452 sp P02784 SFP1_BOVIN SEMINAL PLASMA PROTEIN PDC-109 P...	108	4e-23
gi 134453 sp P04557 SFP3_BOVIN SEMINAL PLASMA PROTEIN BSP-A3 >g...	106	9e-23
gi 2829699 sp P81019 SFP4_BOVIN SEMINAL PLASMA PROTEIN BSP-30 KD	106	1e-22
gi 4335924 gb AAD17519.1 (AF057133) seminal plasma 30K protein...	106	1e-22
gi 1168021 gb AAB35156.1 HSP-1=seminal plasma major protein [h...	105	2e-22
gi 7441785 pir S58424 seminal plasma protein HSP-1 - horse	105	2e-22
gi 11024572 emb CAC14266.1 (AJ278477) epididymal secretory pro...	104	6e-22

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the secreted peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the secreted peptides, and methods of identifying modulators of the secreted peptides.



WO 02/088312 A2



Published:

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

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

ISOLATED HUMAN SECRETED PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND USES THEREOF

FIELD OF THE INVENTION

5 The present invention is in the field of secreted proteins that are related to the seminal plasma protein secreted subfamily, recombinant DNA molecules, and protein production. The present invention specifically provides novel peptides and proteins that effect protein phosphorylation and nucleic acid molecules encoding such peptide and protein molecules, all of which are useful in the development of human therapeutics and diagnostic compositions and
10 methods.

BACKGROUND OF THE INVENTION

Secreted Proteins

Many human proteins serve as pharmaceutically active compounds. Several classes of
15 human proteins that serve as such active compounds include hormones, cytokines, cell growth factors, and cell differentiation factors. Most proteins that can be used as a pharmaceutically active compound fall within the family of secreted proteins. It is, therefore, important in developing new pharmaceutical compounds to identify secreted proteins that can be tested for activity in a variety of animal models. The present invention advances the state of the art by
20 providing many novel human secreted proteins.

Secreted proteins are generally produced within cells at rough endoplasmic reticulum, are then exported to the golgi complex, and then move to secretory vesicles or granules, where they are secreted to the exterior of the cell via exocytosis.

Secreted proteins are particularly useful as diagnostic markers. Many secreted proteins
25 are found, and can easily be measured, in serum. For example, a 'signal sequence trap' technique can often be utilized because many secreted proteins, such as certain secretory breast cancer proteins, contain a molecular signal sequence for cellular export. Additionally, antibodies against particular secreted serum proteins can serve as potential diagnostic agents, such as for diagnosing cancer.

30 Secreted proteins play a critical role in a wide array of important biological processes in humans and have numerous utilities; several illustrative examples are discussed herein. For example, fibroblast secreted proteins participate in extracellular matrix formation. Extracellular

matrix affects growth factor action, cell adhesion, and cell growth. Structural and quantitative characteristics of fibroblast secreted proteins are modified during the course of cellular aging and such aging related modifications may lead to increased inhibition of cell adhesion, inhibited cell stimulation by growth factors, and inhibited cell proliferative ability (Eleftheriou *et al.*, *Mutat Res* 1991 Mar-Nov;256(2-6):127-38).

The secreted form of amyloid beta/A4 protein precursor (APP) functions as a growth and/or differentiation factor. The secreted form of APP can stimulate neurite extension of cultured neuroblastoma cells, presumably through binding to a cell surface receptor and thereby triggering intracellular transduction mechanisms. (Roch *et al.*, *Ann N Y Acad Sci* 1993 Sep 24;695:149-57). Secreted APPs modulate neuronal excitability, counteract effects of glutamate on growth cone behaviors, and increase synaptic complexity. The prominent effects of secreted APPs on synaptogenesis and neuronal survival suggest that secreted APPs play a major role in the process of natural cell death and, furthermore, may play a role in the development of a wide variety of neurological disorders, such as stroke, epilepsy, and Alzheimer's disease (Mattson *et al.*, *Perspect Dev Neurobiol* 1998; 5(4):337-52).

Breast cancer cells secrete a 52K estrogen-regulated protein (see Rochefort *et al.*, *Ann N Y Acad Sci* 1986;464:190-201). This secreted protein is therefore useful in breast cancer diagnosis.

Two secreted proteins released by platelets, platelet factor 4 (PF4) and beta-thromboglobulin (betaTG), are accurate indicators of platelet involvement in hemostasis and thrombosis and assays that measure these secreted proteins are useful for studying the pathogenesis and course of thromboembolic disorders (Kaplan, *Adv Exp Med Biol* 1978;102:105-19).

Vascular endothelial growth factor (VEGF) is another example of a naturally secreted protein. VEGF binds to cell-surface heparan sulfates, is generated by hypoxic endothelial cells, reduces apoptosis, and binds to high-affinity receptors that are up-regulated by hypoxia (Asahara *et al.*, *Semin Interv Cardiol* 1996 Sep;1(3):225-32).

Many critical components of the immune system are secreted proteins, such as antibodies, and many important functions of the immune system are dependent upon the action of secreted proteins. For example, Saxon *et al.*, *Biochem Soc Trans* 1997 May;25(2):383-7, discusses secreted IgE proteins.

For a further review of secreted proteins, see Nilsen-Hamilton *et al.*, *Cell Biol Int Rep* 1982 Sep;6(9):815-36.

Seminal Plasma Protein

Within the ejaculated seminal plasma come spermatozoa, along with prostasomes, seminal plasma proteins and other components. Prostatasomes are secreted membranous organelles from prostate gland, and contains multiple enzyme/component systems to facilitate sperm functions in various ways during male reproduction (Kravets et al., Prostate 43:169-174 (2000)). Their primary function is to enhance sperm capacity such as regulating sperm viability and vitality. Seminal plasma proteins are a group of secreted proteins, which are modified at their side chains through O-linked glycosylation (Calvete et al., FEBS Lett. 350: 203-206. (1995)). The proteins are also implicated in the capacitation or "switching-on" of spermatozoa after their release from male reproductive tract (Fraser, Hum. Reprod. 14 (suppl 1): 38-46 (1999)). Seminal plasma proteins enhance the fertilizing capacity of spermatozoa by interacting with sperm lipids containing phosphorylcholine group for sperm coating and protection as well as glycosaminoglycans present in the female genital tract for sperm residence. For more information, see Calvete, et al., Biochem. J. 310: 615-622 (1995); Calvete et al., FEBS Lett. 407: 201-206 (1997); and Constantine, et al., J. Mol. Biol. 223: 281-298 (1992).

Some problems in male infertility and subfertility result from defects in the physiological mechanisms that need to be activated in spermatozoa following their release from the male reproductive tract. Capacitation encompasses a number of changes that, collectively, confer fertilizing potential on sperm cells. Extrinsic factors modulate capacitation in vitro in ways that could be very relevant to fertilization in vivo, possibly helping to maximize the fertilizing potential of the few cells that reach the site of fertilization. In some men, defects in either of the factors and the systems they modulate could result in defective fertilization. However, by understanding the underlying mechanisms, it may prove possible to develop new diagnostic techniques and new therapeutic treatments to alleviate the infertility. For more information, see Fraser, Hum Reprod 14(suppl 1):38-46 (1999).

Secreted proteins, particularly members of the seminal plasma protein secreted protein subfamily, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown members of this subfamily of secreted proteins. The present invention advances the state of the art by providing previously unidentified human secreted proteins that have homology to members of the seminal plasma protein secreted protein subfamily.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of amino acid sequences of human secreted peptides and proteins that are related to the seminal plasma protein secreted protein subfamily, as well as allelic variants and other mammalian orthologs thereof. These
5 unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate secreted protein activity in cells and tissues that express the secreted protein.

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of a cDNA molecule or transcript sequence that encodes the secreted protein of the present invention. (SEQ ID NO:1) In addition, structure and functional information is provided, such as ATG start, stop and tissue distribution, where
15 available, that allows one to readily determine specific uses of inventions based on this molecular sequence.

FIGURE 2 provides the predicted amino acid sequence of the secreted protein of the present invention. (SEQ ID NO:2) In addition structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence.

FIGURE 3 provides genomic sequences that span the gene encoding the secreted protein of the present invention. (SEQ ID NO:3) In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence. As illustrated in
25 Figure 3, SNPs were identified at 33 different nucleotide positions.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed
30 previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized

within the art as being a secreted protein or part of a secreted protein and are related to the seminal plasma protein secreted protein subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human secreted peptides and proteins that are related to the seminal plasma protein secreted protein subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these secreted peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the secreted protein of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known secreted proteins of the seminal plasma protein secreted protein subfamily and the expression pattern observed. The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known seminal plasma protein family or subfamily of secreted proteins.

Specific Embodiments

Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the secreted protein family of proteins and are related to the seminal plasma protein secreted protein subfamily (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figure 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in Figure 3, will be referred herein as the secreted peptides of the present invention, secreted peptides, or peptides/proteins of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprise the amino acid sequences of the secreted peptides disclosed in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA or Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide 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 peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide 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 secreted peptide 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.

The isolated secreted peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. For example, a nucleic acid molecule encoding the secreted peptide 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. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic

sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequence of such a protein is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

5 The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about
10 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein comprises an amino acid sequence when the amino acid sequence is at
15 least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the
20 secreted peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The secreted peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a secreted peptide operatively linked to a heterologous protein having an amino acid sequence not substantially
25 homologous to the secreted peptide. "Operatively linked" indicates that the secreted peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the secreted peptide.

In some uses, the fusion protein does not affect the activity of the secreted peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for
30 example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant secreted peptide. 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.

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
5 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.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A secreted peptide-encoding
10 nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the secreted peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring
15 recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence
20 information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the secreted peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid
25 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, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length
30 of a reference sequence is aligned for comparison purposes. 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 comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossom 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, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)) (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. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence 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.* (*J. Mol. Biol.* 215:403-10 (1990)). 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.* (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

5 Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the secreted peptides of the present invention as well as being encoded by the same genetic locus as the secreted peptide provided herein.

Allelic variants of a secreted peptide can readily be identified as being a human protein
10 having a high degree (significant) of sequence homology/identity to at least a portion of the secreted peptide as well as being encoded by the same genetic locus as the secreted peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the reference human. As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at
15 least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a secreted peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides information on SNPs that have been found in the gene encoding the
20 protein of the present invention. SNPs were identified at 33 different nucleotide positions. Changes in the amino acid sequence caused by these SNPs is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. Some of these SNPs that are located outside the ORF and in introns may affect gene expression. Positioning of each SNP in an exon, intron, or outside the ORF can
25 readily be determined using the DNA position given for each SNP and the start/stop, exon, and intron genomic coordinates given in Figure 3.

Paralogs of a secreted peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the secreted peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered
30 paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a secreted peptide

encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a secreted peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the secreted peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a secreted peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the secreted peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the secreted peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a secreted peptide by another amino acid of like characteristics. 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 and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

Variant secreted peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind substrate, ability to phosphorylate substrate, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis and can be used to identify critical domains/regions. Functional variants can also contain substitution of similar amino acids that result 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.

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.*, *Science* 244:1081-1085 (1989)), particularly using the results provided in Figure 2. 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 secreted protein activity or in assays such as an *in vitro* proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al. Science* 255:306-312 (1992)).

The present invention further provides fragments of the secreted peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a secreted peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the secreted peptide or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the secreted peptide, e.g., active site or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

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 secreted peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond

formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

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

Accordingly, the secreted peptides of the present invention 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 secreted peptide is fused with another compound, such as a compound to increase the half-life of the secreted peptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature secreted peptide, such as a leader or secretory sequence or a sequence for purification of the mature secreted peptide or a pro-protein sequence.

Protein/Peptide Uses

The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein or ligand (such as, for example, in a secreted protein-effector protein interaction or secreted protein-ligand interaction), the protein can be used to identify the binding partner/ligand so as to develop a system to identify inhibitors of the binding interaction. Any or all of these uses are capable of being developed into reagent grade or kit format for commercialization as commercial products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, secreted proteins isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the secreted protein. A large percentage of pharmaceutical agents are being developed that modulate the activity of secreted proteins, particularly members of the seminal plasma protein subfamily (see Background of the Invention). The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in Figure 1. Such uses can readily be determined using the information provided herein, that which is known in the art, and routine experimentation.

The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to secreted proteins that are related to members of the seminal plasma protein subfamily. Such assays involve any of the known secreted protein functions or activities or properties useful for diagnosis and treatment of

secreted protein-related conditions that are specific for the subfamily of secreted proteins that the one of the present invention belongs to, particularly in cells and tissues that express the secreted protein.

The proteins of the present invention 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 secreted protein, as a biopsy or expanded in cell culture. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the secreted protein.

The polypeptides can be used to identify compounds that modulate secreted protein activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with the secreted protein. Both the secreted proteins of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the secreted protein. These compounds can be further screened against a functional secreted protein to determine the effect of the compound on the secreted protein activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the secreted protein to a desired degree.

Further, the proteins of the present invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the secreted protein and a molecule that normally interacts with the secreted protein, e.g. a substrate or a component of the signal pathway that the secreted protein normally interacts (for example, another secreted protein). Such assays typically include the steps of combining the secreted protein with a candidate compound under conditions that allow the secreted 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 secreted protein and the target.

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.*, *Nature* 354:82-84 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) 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.*, *Cell* 72:767-778 (1993)); 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 fragment of the receptor that competes for substrate binding. Other candidate compounds include mutant secreted proteins or appropriate fragments containing mutations that affect secreted protein function and thus compete for substrate.

Accordingly, a fragment that competes for substrate, for example with a higher affinity, or a
5 fragment that binds substrate but does not allow release, is encompassed by the invention.

Any of the biological or biochemical functions mediated by the secreted 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 or that can be readily
10 identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the secreted protein can be assayed.

Binding and/or activating compounds can also be screened by using chimeric secreted proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of
15 the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a substrate-binding region can be used that interacts with a different substrate than that which is recognized by the native secreted protein. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the
20 specific host cell from which the secreted protein is derived.

The proteins of the present invention are also useful in competition binding assays in methods designed to discover compounds that interact with the secreted protein (e.g. binding partners and/or ligands). Thus, a compound is exposed to a secreted protein polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble
25 secreted protein polypeptide is also added to the mixture. If the test compound interacts with the soluble secreted protein polypeptide, it decreases the amount of complex formed or activity from the secreted protein target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the secreted protein. Thus, the soluble polypeptide that competes with the target secreted protein region is designed to contain peptide sequences
30 corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the secreted 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 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 are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of secreted protein-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 secreted protein-binding protein and a candidate compound are incubated in the secreted 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 secreted protein target molecule, or which are reactive with secreted 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.

Agents that modulate one of the secreted proteins of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

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

In yet another aspect of the invention, the secreted proteins 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
5 WO94/10300), to identify other proteins, which bind to or interact with the secreted protein and are involved in secreted protein activity.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a secreted protein is fused to
10 a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a secreted protein-dependent complex, the DNA-binding and activation domains of the
15 transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the secreted protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a secreted protein-modulating agent, an antisense secreted protein
20 nucleic acid molecule, a secreted protein-specific antibody, or a secreted protein-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The secreted proteins of the present invention are also useful to provide a target for
30 diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. The method involves contacting a biological sample with a

compound capable of interacting with the secreted protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to 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 peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. 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 secreted protein activity in cell-based or cell-free assay, alteration in substrate 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. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. 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 that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides 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. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). 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 effects 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 secreted protein in which one or more of the secreted protein functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other substrate-binding regions that are more or less active in substrate binding, and secreted protein activation. Accordingly, substrate dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Accordingly, methods for treatment include the use of the secreted protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal

antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or F(ab')₂, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, *Antibodies*, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the secreted proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or secreted protein/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention 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/biotin and avidin/biotin; 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.

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention 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. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess 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's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, 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.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. 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 expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins 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 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 protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the secreted peptide to a binding partner such as a substrate. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a secreted peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the secreted peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the 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, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant 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 nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA 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.

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.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide sequences shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, humans genomic sequence (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide 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.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the secreted peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, 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 nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules 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).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the

secreted proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (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 nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. 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.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify gene-modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a

fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene.

Figure 3 provides information on SNPs that have been found in the gene encoding the protein of the present invention. SNPs were identified at 33 different nucleotide positions. Changes in the amino acid sequence caused by these SNPs is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. Some of these SNPs that are located outside the ORF and in introns may affect gene expression. Positioning of each SNP in an exon, intron, or outside the ORF can readily be determined using the DNA position given for each SNP and the start/stop, exon, and intron genomic coordinates given in Figure 3.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. 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. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2. As illustrated in Figure 3, SNPs were identified at 33 different nucleotide positions.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods.

The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in secreted protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA include 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 a secreted protein, such as by measuring a level of a secreted protein-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a secreted protein gene has been mutated.

5 Nucleic acid expression assays are useful for drug screening to identify compounds that modulate secreted protein nucleic acid expression.

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 secreted protein gene, particularly biological and pathological processes that are mediated by the secreted protein in cells and tissues that express
10 it. The method typically includes assaying the ability of the compound to modulate the expression of the secreted protein nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired secreted protein nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the secreted protein nucleic acid or recombinant cells genetically engineered to express specific
15 nucleic acid sequences.

Thus, modulators of secreted protein 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 secreted protein mRNA in the presence of the candidate compound is compared to the level of expression of secreted protein mRNA in the absence of the candidate compound. The
20 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
25 presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate secreted protein nucleic acid expression in cells and tissues that express the secreted protein. Modulation includes
30 both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for secreted protein 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 secreted protein nucleic acid expression in the cells and tissues that express the protein.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the secreted protein 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.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in secreted protein nucleic acid expression, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in secreted protein genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally occurring genetic mutations in the secreted protein 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 secreted protein 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 secreted protein.

Individuals carrying mutations in the secreted protein gene can be detected at the nucleic acid level by a variety of techniques. Figure 3 provides information on SNPs that have been found in the gene encoding the protein of the present invention. SNPs were identified at 33 different nucleotide positions. Changes in the amino acid sequence caused by these SNPs is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. Some of these SNPs that are located outside the ORF and in introns may affect gene expression. Positioning of each SNP in an exon, intron, or outside the ORF can readily be determined using the DNA position given for each SNP and the start/stop, exon, and intron genomic coordinates given in Figure 3. 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 some uses, 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.*, *Science* 241:1077-1080 (1988); and Nakazawa *et al.*, *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res.* 23:675-682 (1995)). 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.

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

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant secreted protein gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

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.*, *Science* 230:1242 (1985)); Cotton *et al.*, *PNAS* 85:4397 (1988); Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and

Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), 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.*, *Nature* 313:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective

5 amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules 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).

10 Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the secreted protein gene in an individual in order to select an appropriate compound or dosage regimen for treatment. Figure 3 provides information on SNPs that have been found in the gene encoding the protein of the present invention. SNPs were identified at 33 different nucleotide positions. Changes in the amino acid sequence caused by these SNPs is indicated in Figure 3 and
15 can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. Some of these SNPs that are located outside the ORF and in introns may affect gene expression. Positioning of each SNP in an exon, intron, or outside the ORF can readily be determined using the DNA position given for each SNP and the start/stop, exon, and intron genomic coordinates given in Figure 3.

20 Thus nucleic acid molecules 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 nucleic acid molecules are thus useful as antisense constructs to control secreted protein
25 gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of secreted protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into secreted protein.

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

translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the secreted protein, such as substrate binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in secreted protein 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 secreted protein to treat the individual.

The invention also encompasses kits for detecting the presence of a secreted protein 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 secreted protein nucleic acid in a biological sample; means for determining the amount of secreted protein nucleic acid in the sample; and means for comparing the amount of secreted protein 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 secreted protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1 and 3).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee *et al.*, PCT application W095/11995 (Chee *et al.*), Lockhart, D. J. *et al.* (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. *et al.* (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown *et al.*, US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential

oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

5 In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may
10 be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be
15 paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler *et al.*) which is incorporated herein in its entirety by
20 reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic
25 instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is
30 produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with

precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large-scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the secreted proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the present invention and or alleles of the secreted protein gene of the present invention. Figure 3 provides information on SNPs that have been found in the gene encoding the protein of the present invention. SNPs were identified at 33 different nucleotide positions. Changes in the amino acid sequence caused by these SNPs is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. Some of these SNPs that are located outside the ORF and in introns may affect gene expression. Positioning of each SNP in an exon, intron, or outside the ORF can readily be determined using the DNA position given for each SNP and the start/stop, exon, and intron genomic coordinates given in Figure 3.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the Human genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified secreted protein gene of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules 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 nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

5 The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

 Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed
10 in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules 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
15 is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

 The regulatory sequence to which the nucleic acid molecules 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*,
20 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
25 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
30 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.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. 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.*, *Molecular Cloning: A Laboratory Manual*, 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

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 nucleic acid molecules 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 nucleic acid molecule 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 peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. 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 peptide 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.*, *Gene* 67:31-40 (1988)), 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.*, *Gene* 69:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

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

The nucleic acid molecules 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.*, *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell* 30:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules 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.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

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

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 nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

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 nucleic acid molecule 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).

5 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
10 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. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989*).

15 Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules 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
20 to the nucleic acid molecule 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.

25 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 nucleic acid molecules 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
30 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 peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as kinases, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with kinases, 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 peptide 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 peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

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

Host cells are also useful for conducting cell-based assays involving the secreted protein or secreted protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native secreted protein is useful for assaying compounds that stimulate or inhibit secreted protein function.

Host cells are also useful for identifying secreted 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 secreted protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native secreted protein.

Genetically engineered host cells can be further 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
5 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 secreted protein and identifying and evaluating modulators of secreted protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

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

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already
15 included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the secreted 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.
20 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
25 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
30 selected systems that 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.* *PNAS* 89:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al.* *Science*

251:1351-1355 (1991). 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
5 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, I. *et al. Nature* 385:810-813 (1997) 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.
10 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 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.

15 Transgenic animals containing recombinant cells that express the peptides 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 effect substrate binding, secreted protein activation, and signal transduction, 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* secreted protein
20 function, including substrate interaction, the effect of specific mutant secreted proteins on secreted protein function and substrate interaction, and the effect of chimeric secreted proteins. It is also possible to assess the effect of null mutations, that is, mutations that substantially or completely eliminate one or more secreted protein functions.

All publications and patents mentioned in the above specification are herein incorporated
25 by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-
30 described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

Claims

That which is claimed is:

1. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown in SEQ ID NO:2;
 - (b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;
 - (c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3; and
 - (d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.
2. An isolated peptide comprising an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown in SEQ ID NO:2;
 - (b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;
 - (c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3; and
 - (d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.
3. An isolated antibody that selectively binds to a peptide of claim 2.

4. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;
- (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;
- (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;
- (d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and
- (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;
- (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;
- (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;
- (d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and
- (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

6. A gene chip comprising a nucleic acid molecule of claim 5.

7. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.

8. A nucleic acid vector comprising a nucleic acid molecule of claim 5.
9. A host cell containing the vector of claim 8.
10. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
11. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
12. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows detection of the presence of the peptide in the sample and then detecting the presence of the peptide.
13. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.
14. A method for identifying a modulator of a peptide of claim 2, said method comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.
15. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.

16. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.

17. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.

18. A method for treating a disease or condition mediated by a human secreted protein, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.

19. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.

20. An isolated human secreted peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence shown in SEQ ID NO:2.

21. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence shown in SEQ ID NO:2.

22. An isolated nucleic acid molecule encoding a human secreted peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 or 3.

23. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 or 3.

```

1  ATGGCCCGGC ACATGGGGCT CCTGCTGGTT TGGGTCTGCC TGATTCGCGG
51 TGTGGTCGGT GGTGTGGTCG GCGCTGTATT TAACGTTCTG GAAGAAGAAA
101 CTATAACTCA TTTTCCAGAA GTTACAGATG GGGAGTGTGT CTTTCCATTC
151 CACTATAAAA ATGGAACATA TTATGACTGC ATCAAGTCCA AGGCAAGACA
201 CAAGTGGTGC TCGTTAAACA AGACCTACGA AGGATACTGG AAGTTTTGCA
251 GTGCAGAAGA TTTTGCAAAC TGTGTATTTT CCTTCTGGTA CAGACGCTTG
301 ATCTACTGGG AGTGTACTGA TGATGGGGAA GCATTTGGGA AAAAATGGTG
351 TTCCTGACC AAGAATTTTA ACAAGGACCG AATTTGGAAT TACTGTGAAT
401 GA
(SEQ ID NO:1)

```

FEATURES:

Start Codon: 1

Stop Codon: 400

Homologous proteins:

Top 10 BLAST Hits

	Score	E
gi 3024635 sp P81121 SP1_HORSE SEMINAL PLASMA PROTEIN HSP-1	116	1e-25
gi 3549883 emb CAA07074.1 (AJ006515) seminal plasma protein-1 ...	109	1e-23
gi 6138952 gb AAF04405.1 (AF055981) seminal plasma protein A3 ...	108	2e-23
gi 134452 sp P02784 SFP1_BOVIN SEMINAL PLASMA PROTEIN PDC-109 P...	108	4e-23
gi 134453 sp P04557 SFP3_BOVIN SEMINAL PLASMA PROTEIN BSP-A3 >g...	106	9e-23
gi 2829699 sp P81019 SFP4_BOVIN SEMINAL PLASMA PROTEIN BSP-30 KD	106	1e-22
gi 4335924 gb AAD17519.1 (AF057133) seminal plasma 30K protein...	106	1e-22
gi 1168021 gb AAB35156.1 HSP-1=seminal plasma major protein [h...	105	2e-22
gi 7441785 pir S58424 seminal plasma protein HSP-1 - horse	105	2e-22
gi 11024572 emb CAC14266.1 (AJ278477) epididymal secretory pro...	104	6e-22

FIGURE 1

1 MARHMGILLLV WVCLIRGVVG GVVGAFFNVL EETITHFPE VTDGECVFPP
 51 HYKNGTYDC IKSARHKWC SLNKTYEGYW KFCSAEDFAN CVFPFWYRRL
 101 IYWECTDDGE AFGKKWCSLT KNFNKDRIWK YCE
 (SEQ ID NO:2)

FEATURES:**Functional domains and key regions:**

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION

N-glycosylation site

Number of matches: 2

1	54-57	NGTY
2	73-76	NKTY

[2] PDOC00006 PS00006 CK2_PHOSPHO_SITE

Casein kinase II phosphorylation site

Number of matches: 3

1	42-45	TDGE
2	56-59	TYTD
3	84-87	SAED

[3] PDOC00008 PS00008 MYRISTYL

N-myristoylation site

Number of matches: 2

1	17-22	GVVGGV
2	24-29	GAVFNV

[4] PDOC00009 PS00009 AMIDATION

Amidation site

112-115	FGKK
---------	------

[5] PDOC00022 PS00023 FIBRONECTIN_2

Type II fibronectin collagen-binding domain

91-132	CVFPFWYRRLIYWECTDDGEAFGKKWCSLTKNFNKDRIWKYC
--------	--

SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	21	0.989	0.37	YES
max. Y	21	0.876	0.34	YES
max. S	10	0.988	0.88	YES
mean S	1-20	0.967	0.48	YES

Most likely cleavage site between pos. 20 and 21: VVG-GV

BLAST Alignment to Top Hits:

>gi|3024635|sp|P81121|SP1_HORSE SEMINAL PLASMA PROTEIN HSP-1

Length = 120

Score = 116 bits (287), Expect = 1e-25

Identities = 47/96 (48%), Positives = 62/96 (63%), Gaps = 1/96 (1%)

Frame = +1

Query: 109 HFPEVT-DGECVFPPHYKNGTYDCIKSKARHKWC SLNKTYEGYWKFC SAEDFAN CVFPF 285

H VT + + CVFPF+Y+ YYDC ++ + ++WCSL TY G WK+C+A D+A C FPF

Sbjct: 24 HSATVTPENKCVFPFNRYRGYRYDCTRTDSFYRWCSLTGTYS GSWKYCAATDYAKCAFPF 83

Query: 286 WYRRLIYWECTDDGEAFGKKWCSLTKNFNKDRIWKYC 396

YR Y CT DG F WCS+T N++ WKYC

Sbjct: 84 VYRGQTYDRCTT DGS LFRISWCSVTPNYDHHGAWKYC 120

(SEQ ID NO:4)

FIGURE 2A

```

>gi|3549883|emb|CAA07074.1| (AJ006515) seminal plasma protein-1
      [Equus caballus]
      Length = 105
Score = 109 bits (271), Expect = 1e-23
Identities = 45/94 (47%), Positives = 60/94 (62%), Gaps = 1/94 (1%)
Frame = +1

Query: 109 HFPEVT-DGECVFPFHYKNGTYDCIKSKARHKWCSLNKTYEGYWKFCSAEDFANCVFPPF 285
          H   VT + +CVFPP+Y+   YYDC ++ + ++WCSL  TY G WK+C+A D+A C FPF
Sbjct: 11  HSATVTPENKCVFPFNRYRGYRYDCTRDTSFYRWCSLTGTYSGSWKYCAATDYAKCAFPF 70

Query: 286 WYRRLIYWECTDDGEAFGKKWCSLTKNFNKDRIWK 390
          YR   Y CT DG F   WCS+T N++   WK
Sbjct: 71  VYRGQTYDRCTTDGSLFRISWCSVTPNYDHHGAWK 105
(SEQ ID NO:5)

>gi|6138952|gb|AAF04405.1| (AF055981) seminal plasma protein A3 [Bos
      taurus]
      Length = 140
Score = 108 bits (268), Expect = 2e-23
Identities = 54/132 (40%), Positives = 74/132 (55%), Gaps = 8/132 (6%)
Frame = +1

Query: 1   MARHMGLLLVWVCL-----IRGVVGGVVGAVFNVL---EEETITHFPEVTDGECVFPFHY 156
          MA +GL L+W +   + V G   + NV+   E++   E D +CVFPP Y
Sbjct: 1   MALRLGLFLIWAGVSMFLQLDPVNGDEQLSEDNVILPKEKKDPASGAETKDNKCVFPFIY 60

Query: 157 KNGTYDCIKSKARHKWCSLNKTYEGYWKFCSAEDFANCVFPFWYRRLIYWECTDDGEAF 336
          N Y+DC   +   WCSL+ Y G WK+C+ D+A CVFPP Y   Y C G F
Sbjct: 61  GNKKYFDCTLHGSFLFLWCSLDADYTGWRKYCTKNDYAKCVFPFIYEGKSYDTCIIIGSTF 120

Query: 337 GKKWCSLTKNFNKDRIWKYC 396
          WCSL+ N+++D +WKYC
Sbjct: 121 MNYWCSLSSNYDEDGVWKYC 140
(SEQ ID NO:6)

>gi|134452|sp|P02784|SFPI_BOVIN SEMINAL PLASMA PROTEIN PDC-109
      PRECURSOR (SEMINAL VESICLE SECRETORY PROTEIN 109)
      (SVSP109) (BSP-A1 AND BSP-A2) >gi|1070648|pir||WTBO
      seminal fluid protein PDC-109 precursor - bovine
      >gi|163735|gb|AAA30766.1| (M22244) seminal vesicle
      secretory protein 109 precursor [Bos taurus]
      >gi|488603|emb|CAA83915.1| (Z33621) SVSP109 [Bos taurus]
      >gi|833775|emb|CAA43021.1| (X60495) bovine seminal
      vesicle secretory protein SVSP109 [Bos taurus]
      >gi|226435|prf||1512339A major seminal protein [Bos
      taurus]
      Length = 134
Score = 108 bits (266), Expect = 4e-23
Identities = 55/132 (41%), Positives = 73/132 (54%), Gaps = 5/132 (3%)
Frame = +1

Query: 1   MARHMGLLLVWVCL-----IRGVVGGVVGAVFNVLEEETITHFPEVTDGECVFPFHYKNG 165
          MA +GL L+W +   + V G   V   ++   PE D ECVFPP Y+N
Sbjct: 1   MALQLGLFLIWAGVSVFLQLDPVNGDQDEGVSTEQDGPaelPE--DEECVFPFVYRNR 58

Query: 166 TYYDCIKSKARHKWCSLNKTYEGYWKFCSAEDFANCVFPFWYRRLIYWECTDDGEAFGKK 345
          ++DC   +   WCSL+ Y G WK+C+ D+A CVFPP Y   Y CT G +
Sbjct: 59  KHFDCTVHGSLFPWCSLDADYVGRWKYCAQRDYAKCVFPFIYGGKKYETCTKIGSMW-MS 117

```

FIGURE 2B

Query: 346 WCSLTKNFNKDRIWKYC 396
 WCSL+ N++KDR WKYC
 Sbjct: 118 WCSLSPNYDKDRAWKYC 134
 (SEQ ID NO:7)

>gi|134453|sp|P04557|SFP3_BOVIN SEMINAL PLASMA PROTEIN BSP-A3
 >gi|89765|pir||A29156 seminal fluid protein A3 - bovine
 Length = 115
 Score = 106 bits (263), Expect = 9e-23
 Identities = 43/93 (46%), Positives = 55/93 (58%)
 Frame = +1

Query: 118 EVTDGECVFPPHYKNGTYDYDCIKSKARHKWCSLNKTYEGYWKFCSAEDFANCVFPFWYRR 297
 E D +CVFPF Y N Y+DC + WCSL+ Y G WK+C+ D+A CVFPF Y
 Sbjct: 23 ETKDNKCVFPFIYGNKKYFDCTLHGSLFLWCSLDADYTGRWKYCTKNDYAKCVFPFIYEG 82

Query: 298 LIYWECTDDGEAFGKKWCSLTKNFNKDRIWKYC 396
 Y C G F WCSL+ N+++D +WKYC
 Sbjct: 83 KSYDTCIKIGSTFMNYWCSLSSNYDEDEGVWKYC 115
 (SEQ ID NO:8)

HMM results:

Model	Description	Score	E-value	N
PF00040	Fibronectin type II domain	88.0	2e-32	2
CE00265	E00265 Seminal_fluid_protein	65.3	3.2e-17	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00040	1/2	46	83 ..	1	42 []	35.0	4.2e-12
CE00265	1/1	31	131 ..	17	117 .]	65.3	3.2e-17
PF00040	2/2	91	132 ..	1	42 []	60.8	5.7e-22

FIGURE 2C

```

1  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
51  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
951 GGGAGGAGGG AAAGAGAGGA GGGACAGGGA GGGGAAGTGG TTGTGGTTAG
1001 TTCCAGGGGC TATAAAATCT TGGACATCAG CTGGTCGACC ACGAGTGTCT
1051 GACAGGAGCA TGGCCCGGCA CATGGGGCTC CTGCTGGTTT GGGTCTGCCT
1101 GATTCCGCGT GTGGTCGGTG GTGTGGTCGG CGCTGTATTT AACGTTCGGT
1151 AAGAAGGTGA GGAGACAGGT GACCCTGGCT GGGACGGAGT CCTGCACATC
1201 GGCTGCTGTA TGTAATCAAA TCCAGTCGC CATGGAGGGC AACAGTCTCC
1251 CTTGCTTTAG GTACCTTAAG AAAGAAAAAG ATACTCTCCC TTCCATGTCA
1301 GAAACACAAT CCATAGGTAA AGATTTTTTA AAAAATAAGA AAGTTCGGCT
1351 AATAAAATAA TGAAAGACTA AGAATGTAAA TTGGAATAAT CTAGTGACAT
1401 GTTACTTGAA AACCATGTGT CTCCTCCCTT CATGCTAGGA GATTTTAAAT
1451 GCCTCCTGCA GAGGTTTTGA AAATGAGACT AATTTTCATA AAGTCGATTT
1501 TATTCCTTTT CCAAAGTCGT GTTGCTGTA ACAATTATGA AGCGAAATAG
1551 AAAAATAAGA AGGGAATTCA CTATGGATGC ATCTGACAAC GTTTGATGAG
1601 GACTAAGCAT AGTTTCAAAA ACAATTTTTT ATGGAATGAC CAAGAAGAA
1651 GATCTTGCAG ATATTCTTTT TAAGTATTCT TCAGGACATC TTTGCATTAA
1701 AAAAATAAAT CACTGCTTGC TATTTTTTTT TAAAGAACAG AAAAGAAAA
1751 ATACTCCCAA TGTAACATG ACCACTGCTT TCTTATCCAC TAGAATTATT
1801 TTATACCTAC TGAAAAATAA CACTGTACTT GGCCACAGGT TTTTTTTTTT
1851 AATCAGACAT CATGTTTGTA GATACGTAAA AAGGAATATA GATAAGTGGA
1901 AAAAGGCAAA TTAAGACGAT CCTAAAGCTT CTCCGCTGTC AATGTCTGGT
1951 CTCCTAATTT ATGGATGTCA TATTTTCCA CATAATACTT GTGCTATTCA
2001 GAGCCTTACT GATGAAGCAT TTCTCTTAAT AATGCATCAG AGAATAAAAA
2051 CTCATGGGCC CCGGGCTCGT TAGCCTGTAT TTTAAATGAT GTGGTGCTAG
2101 TCTACTTTTG ATCTGCAGAA CTGTTGCCTA TTAATATCT CCCTCTCTCT
2151 CCCTTGCAAG CATCTGCCTC CTGGAATTAT TGCTCTTTT TTTTTTTTTT
2201 TTTTTTTTTT TGGAGTCTTG CTCTGTCAAC CAGACTGGAG TGCAGTGGCA
2251 TGATCTCGGC TCACCGCAAC CTCAGCCTCC CAGATTGAG CAATTCTCCT
2301 GCCTCAGCCT CCCGAGTAGC TGGGACTAAA GGCACGTGCC ATCATGCCCC
2351 GCTAAATTTT GTATTTT TAGAGACCAG GTTCCACCAT GTTGGCCAGG
2401 CTGGTCTCGA ACTCTGATC TTTGGTGATT CACCCACCTC AGCCTCCCAA
2451 AGTGCTGGGA TTTATAGGCAG GAGCCACCGC ACACAGCCTA TTGCTCTTTA
2501 AGAGTTTCTT TTGAGCTACT GAGTAGCAAA CTGCAAGTTT TTGATGCTGG
2551 TCGAATTCAA TACAGACAAT AACAGCTGGG TCCCCACAGC AATCCGGCTT
2601 GCAGATAATA AGATGTCGTC ACAAATAAGC GGAAGGTGGG AACTCATAGA
2651 CATCTAGGGC TCAAAGTCTT GCTGGAGGCC ATAGGCTAAG CTTCGTGGGA
2701 ACCCAGAGAT GCTGGAGGGA AATCTGGAAT CCTGAGCTTG CCAGGAGCAG
2751 TCAAGTGGGA CAGAGGAGGT GAAGGGAGTC CCAGTTTTTG ATGAGTGGGG
2801 ATGTATGAGA GCTCTCAGCT ATGGTTCTCT CTCATCCTCA TCAGAGGAGC
2851 ACTACATTTT GGTGTTTTAG AAGCAGTGAC AGACTGGGTG CGGTGGCTCA
2901 CACCTGTAAT CCCCAGCAC TTTGGGAGGCC GAGGCGGGCG GATCACAAGG
2951 TCAGGAGTTC AAGACCAGCC TGGCCAACAT GGTGAAACCC CGTCTCTACT
3001 AAAAATACAA CAATTAGCCG GGCATGGTGG CAGGTGCCCTA TAATCCCAGC
3051 TACTCAGGAG GTGAGGCAG GAGAATCACT TGAACCAAG AGGCGGAGGT
3101 TGCAGTGACC TGAGATCATG CCATTGCACT CCAGCCTGGG CAACAAGAGC
3151 AAGACTCCAT CTCAAAATAA TAATAATAAT AATAATAATA ATAATAATAA
3201 TAATAATAAT AATAAAACGG TGAATTGATA GTTTGCTGG ATGCACAGAT
3251 GCAGAGACTA AATAGCATGA GCATCTCCAC TCATGAGCAT GTGATATGCA
3301 CCCCCACCTC TTCCCACTAG TGCACGATCT AGTGGGAAGA CAGAAGAGGA
3351 AACAGACCAT GTGTATGGCA CATGCTAAAT CAGCTATACG CCTAGCCTGA
3401 AAAAGAAAAA ATAATTGAAT TGGCAAGACT TCACAAAGGA CTGATTGAG
3451 TGATCTTGTT CCAGGGCTAG GAATTTGCCA TGTCACACAT GTAATTACTA
3501 TTCCAACCTG ATGGGTTGCC TAAGGGATTA CATGGGAGTG GGGTCCCTTG
3551 CGGGGAGAAC TGTCCAGGTG TGTAAGACCG GACATGCTTC GTGCACCCAG
3601 AACTCTGGGG TTTCACCTTT TGCCAGTTCA GGCTTTGGTG CGGTAGGACT
3651 GCCCACCACG CGATCGTTTG ATGTCAACAA ACGTTTGTGC TAAAATGTGA

```

FIGURE 3A

```

3701 ATCTTGCATT GCAAGAGCCA TTATTATTTA AAGAGGATGA GGAGGAACCA
3751 TGGCTAAGAG CTCTCATGCA CCCCACACTA CCAAGAACTG AAGGCTCCTC
3801 TCGCCTCCCTC TGTCCCCTC AACTCTGCTC CTGGCAAGCT CAGGAGACCA
3851 GTTTTCCCTC CAGCATCTCC GGGTTCCAAC CAAGCTTAGC CTGTGGCCTC
3901 AAGCAAGACT TTGAGTCCTG GATGTCTGTG AGTCCCACA TCTGAAACAG
3951 GAGGATCGTG ATTCTTGCTG ACTCCCAGGG TTGCTGGAG ATTAATGGG
4001 TTAGCATGCG TGACGCTCAG AACAGAACTG TTATCATTAA GTCATTATTA
4051 CTACTATTAC AGTACCAGAA GTACTATTAC AGTATTAAAG AGGCAGGATA
4101 AGGGTTTGGG TTTAGTAGTC GATACTCACA GAGTGAGTTT AACTACTTTA
4151 AACTTCTCAT ACAAGTGGA TCGTGCAATA TTTGTCCCTC TGTGAATGGT
4201 TTGTTCCACT TAGCATAATG TCCTCCAGGT TCATCCATGT TACTGAAAAC
4251 GGCAGGATTT TCTCTTTTTT TAATTTTTAA AATTTTTCTG ATTTTGTAAAT
4301 TTTTATTTTT ATTTTATTTT ATTTGTTTTT GAGACAAGGG CTCACTCTGT
4351 CACCCAGGCT GGAGTGCAGT GGTGTGATCT CAGCACACTG CAACCTCCGC
4401 CTCCCAGGAC TGAAGTAATT AATCCTCCA CTCAGCCTCC CAGTACGGGA
4451 ACCACAGGAG CACACCACCA TGCCCAGCTA ATTTTTTTGT AATTTTGGTA
4501 GAGATAAGAT TTCGCCAAGT TGCCCAGGCA GGTCTCAAAC TCCTGAGCTC
4551 AAGCGATCTG CCCACCTCCG CCTTCCAAAG TGCTGTGAGC CACCACACCC
4601 AGCCCTCGAC TAAAGTTCTT TGATCAAGGT TGTTTTGACC TAAGGGCAGG
4651 ATCAATAGTA AAGACAAGCT CTGGCCATGC ATAGTTCAAG GTGGGAGGAT
4701 TGCTTGAGCC CAGGAATTCG AGACCAGCCT GGGAAACATA GTGGGACCTA
4751 GTCTCTACAG CAACTTTGA AAATTAGCCA GCTATGGTAG TGTGTGCCTG
4801 TAGTTCCAGC TACTTGGGAG GCTGAGGCAA GAGAATCACT TGAGCCTGGG
4851 AAGTCAAGCC TGCAGTGAGC TATGACAGCG CCACTGTACT CCAACCTGGG
4901 TGACAGAGTG AGATTCTGTC TCCAAAAAAA AAAAAAAA GTAAAGTACTA
4951 AGTACATGTA GTACGTGTTT TTACACAAAG AACAGTAGAT AAAATAGAAA
5001 TTGTAGGGAC ATTCCGAGGA CGGGGTTAAT CAGAAACCAA CACGGCGAAT
5051 TAGCACCCAA GACAGAGGTG TTTTAGCCTC CACAACTGG ATTGTCTAGT
5101 GACACTCAAA TTCTTCGTCT TCATCCTCAC TGTCTGGGGT TGCATTGCAT
5151 AGTGTCTTTC CAGAATTCAT GTTTACCAGA GATCTGTGAA TATGACTGTA
5201 TTTGAATAG GTCCCTGCAG ATGTCGGTAA TTAAGATGTA GTCACAATGG
5251 ATACGGGTTG GCCCTAATCC AATGACTGGT ATCTTTATAA GAAGAGGAAA
5301 CAGACACACA CAGGGAAGAC AAGCACGCAT AGTCACAAGA AAAGATTAAA
5351 GTGATGCTGC CACAAGTCAC AGAATGCCAC AGCTTGCCAG CAGCCACCAA
5401 AAGCAAGGAG AGAGGCATGG AATAGTTTCT GCCTCTGAGC TCCTAAGAAAG
5451 GAACCAACCC TGGCATTTTG GCCTTCTGGC CTCCAGATCT GTGAGACAAT
5501 AAATGTCTAT TGCTCTAGAC CAAACAGTCT ATGGTACTTT GTCTAGATTT
5551 TTTTATTTT TATTTTATTT TATTTTATTT TTAGATAGGG TCTCATTTCTG
5601 TCACCCAGGC TGAGTGCAG TGGCACAATC ACGGCTCACT CAGAGCTTGA
5651 CCTCCCAGGC TCAAGTGATC CTCTTGCTC AGCCTCCCAA GTAGCTGGGA
5701 CTACAGGTGC ACGCCACCAT GCCCAGCTAA TTTTGTGTCT TTTATTTATT
5751 GCAGAGATGG GGTCTTGCTA TGTTGCCAG GCTGGTCTCG AACTCCTGGG
5801 TTCAAGCAGT CTCCTGGCT CAGCCTCCAA AAGTGCTGGG ATTACAAGCG
5851 TAAATCACCA TGCCTGACCA GTAAAATTCT ACAGCACAAA ACCAAGGAAG
5901 AAGCAATCAC AGAGGCCACG CGGCTCTCCT GTCTGGGGGA TGAGCCTCCC
5951 ATGGTTTGGT CCAAGGAC CCAACTCCT CCCATTCT TGTATTGCT
6001 CTTACACAGT GCTCTTCCAG GCCTTCTGAC ATTCTCTAGA TTCGTAAGTT
6051 TTGTCTTCTA TCTTGCTGTG TATTCCATTG AACATAGAAT GGAACAGCAA
6101 GTGCAGCACT TTGGGAAGAC TCCTGACCCA GCTCTTGAAC CCAACAAGA
6151 ATCTTGGCCA ATTTAGGATC GTAGGACTAC CATGTAGCAG GCTACAGCAA
6201 TTCCCCACGA TAGGGAGGCC ACTGCAGCTT TTTTCCCTA GTCTGTTTCA
6251 CCAGCTTCCC CAGACTCCTC CCTCCTCTGG GGTCCGCTAG CTGGGAAGAA
6301 TAAAGTCTGA TAAACTATAT CCTCCTGCGG GCAACTGCAA GAATGCCAAT
6351 AATAGCTGTT GACGATGCTG AAATTCACCT CTCATCCTCT GGAAAAGTGA
6401 TCGGGTTCAG GGGACTTCAT AAGCTCCTGA ACACATTTCA CATGCTTTCA
6451 TTTTATTTAC ATTCACTTAT TCTTTTCCTT TTTTTTTTTT CTAGACAAAG
6501 TCTTGCCCTG TCACTCAGGC TGGAGTGTAG TGGTGTGATC TCAGCTCACT
6551 GCAACCTCCA CGTCCCAGGT TCAAGTGATT CTCCTGCCTC AGCTGCCTCA
6601 GTAGCTGGGA TTATAGGCGC ACACCACCAT GCCCAGCTAA TTTTGTATT
6651 TTTAGTAGAG ACAGGGTTTC ACCATGTTGG CCAGGCTGGT CTCGAACTCC
6701 TGACCTCAGG TGATCTGCCC ACCTCGGCTT CCCATAGTGC TGGGATTACA
6751 GGTGTGAGCC ACCATGCCGG GCCTACTTAC TCTTTTTCTA ATGTGCTTCC
6801 TCTCTTTATA TCTCTGAAGA ACTTTTAACA TTTCTTTCAA GGCAGTTTTT
6851 CTGGTAGCAA ATTCCCTCAA TGTTTGTGTTG TCTGAGAAATC TTTGTCTTTC
6901 TTTCTCCTTC ACTTCTGAAG AATAATTTTG CAGGGTACAG AATTCTAGAT
6951 TGAAGATTTT TTTTCTCTC AACACTGTAA ATGCTTCAAT CTTATCTTCA
7001 TGGTTTCTGG AAAGAAGTTG GACATAATTC TTATCTTTGC TTCTCTGTAG
7051 GTAAGGTATT TTTCCCTCCT TGGCTTCTTT CAAAAGATTT TTCTTTTAAA
7101 AAATTTTATT TTAGATGCAG GAGGTACATG TGCAGATTG TTACATGAGT
7151 ATGTTGCATA AATGGTGAGG TTTGGGATTG AATTGAAGCC ATTACCCAAA
7201 TAATGAACAT AGTGCCAGT AGGCAGTTTT CTTTTCTTTT TTCTTTTTTT
7251 TTTTTTTTTT CAGACAGAGT CTTGTATTGC CCAGACTGGA GTGCAGTGGT
7301 GCGATCTTGG CTCACGTCAA CCTCTGCCTC CCAGGTTCAA GCGATTCTCC
7351 TGCCCTCAGC TCCCAGTAG CTGGGATTAC AGGCGCCAC CACCATGTCC

```

FIGURE 3B

```

7401 GGCTAATTTT TAGTAGAGAC GGGATTTCAT CATGTTGGCC AGGCTGGTCT
7451 TGAACCTCTG ACCTCATGAT CAGTCCACCT CAGCTTCCCA AAGTGCTGGG
7501 ATTACAGGCA TGAGCCACTG CACCCGGCCA TCAATAGGTA GTTCCAAGC
7551 CTGGGTCCCC TCCTTACCTC CCCACTTTTG TAATCCCCAG TGCCTATTGT
7601 TCCCATCTTT GTGTCCCTGT GCACCAATG TTAGTTCCTC ACTTATAAGG
7651 GAGAACATGT GGTATTTGGT TTTCTGTTTC TGCACTAATT TGCTTAGGCT
7701 AATGGCCTCC AGCTGCATCC ATGTTGCTAA AAAGGACATG ATTTTGTCTT
7751 TTGTTATGGC TGCATAAAAC AATTTCGTTT TGGGTTTAAC TGAAAAAATG
7801 AGACTCATTT TTATTTCAAA TAAGGGTATT TTTAAACCTT TATATATTAT
7851 AACAGAAATTT AGCTTGAATT ATGACCAAAG GTGAATTCTT GAGAATGAAA
7901 AGAATTGGAG TGCAGGACTT TCCTGACAAA ATTGTTACGA TTTCCAGTG
7951 CATAGCTTTG TCTTCTTCCA TTCTACACTT TCCTCAAAGG GTCATAAAGA
8001 AAGCCAGTTT TGTATAAGTA TTCTAAGACT AATTCTAGGA CTATATTAAT
8051 TTTTAGGGGC CCTTCTTAAT CCTTTGAATT TAAAAATATA TAATATATAT
8101 ATTTATATTT CATGTATAAT ATACATATTA TATATATTTA TATTTTATAT
8151 ATATATATAT TTTTCTTCTT CAAGCCAGGC ATGGTGGCTT ACACCTGTAA
8201 TCCAGCACTT TTGGGAGGCC GAGGTGGGTG GATCACCTGA GGTGAGGAGT
8251 TTGGGACCAG CCTGACCAAC ATGGCAAAC CCTGTCTCTA CTAATAATAC
8301 AAGATTCAGT CAGTGAATAA TACAGGTGGT GGTGCACACC TGTAGTCCCA
8351 GCTACTTGGG AGGCAGAGGC AGAAGACTCC CTGGAACCCA GGAGGCAGAG
8401 GTTGCACTGA GCCGAGATCG TGCCACTACA CGCCAGCCTG GGTGACAGAG
8451 TGAGACTCCG TGTCAAAAAA AAAAAAATTT TTCTCTTCAA AGCCCTTATG
8501 GAGTCACAAT ATGTGCTCTT TTTAACTTTG TTGTATCACT GTCATTAAAT
8551 TCATACATTC TTGCATCTTT CACAAAGTGA CCGGCTTCAT TTTGTAGTCC
8601 ACTTGCTTTG CAGTTGCATT GTACTTGCAA GGTATTTTTC GAAACATAAC
8651 ATCCCAACAT TCACAACAAA TGATCCACCA GGTGGATGGC AGAATTTCTG
8701 GTGTTAATCA GAGTGGGACG TTTGAGTGGG GACTCATAAAT TTTATAGCAC
8751 CGAGGATGGT GCACTGTGTT TATGATATGG GGATGTTGCA TTCTGATGCT
8801 TACCTGGTGA TACCTGTTCT AAGATCATAG TGCTTAGCAG ACAGGCTATG
8851 TTGGAATCGC TCCTCTGTAT CTCCCCATTC ATGAAATCAG GCCTATCATG
8901 ATCCTCAAAA CCAAAATCCA TTTCTTAATA GTTTCATAT GGAATGCATA
8951 CTGAATGAAC CCAACAAATC CAGGAAGTAG AAACAATTAT AATCCCCATA
9001 TTTCAGAGAA AGACAGCTT AACTCCTGAT GAGTCCATTG ATCTTGGATT
9051 CCCACCTGAA TCTCATGGTT TCTGTAGTCC AGAGTCTACC ATTTAGCCCA
9101 CTCCCTTTTT GCTTAGCCCA GAACCTAGCA CTAGGCTGGG CACATAGTAG
9151 GCATTCAATC CATATTTGTT GTCTACATGA GTGGGTTGAT GGAAGTGGCTG
9201 AATAAATGCA GATCCCTTCA TCGTGAGATC TTCCCTATAT ACTGGAACCA
9251 GTGCACAGTT CAGTGGCATT CAATGCATTG AAATTTGTGCA ATCTATCTCC
9301 AGAGCCTTTT CATTGTCCCA AACTGAAACT CTGCACCCAT TAAATAATAA
9351 CTCCCCATTC CTCCCTCCCT CCAGCTCCTA ACAACCACCA TTCTGCTTTT
9401 TGTCTCTATG TATTTGACTA TTCTAGTTAC CTATATAAAG TAGAATCATA
9451 TAATATTTGT CTTTTTGTGT CTGGCTTATT TTACTTAACC TAGTATTTTC
9501 AAAAGTGCAT CATGTGGCGG CATCTATTCA AATTTTGTTC CTTTTTAATG
9551 CTGAATAATA CTCCATTGTA TGTGTCTACC ATATTTTGTG TATTTACTCA
9601 TTTGCTCAGC GACATGTGGG TTCTTTCCAC CTTTGGGCTA TTATGAATAA
9651 TGCTACCATG AACATGGGTT TACAAATATT TGTTTGAGTC TCTGCTTTCA
9701 ATTATTTTGT ATATATACCC AGAAGTGGAA GGGATCTGGG TTTTAAAAA
9751 CTCAGCTGGT GATTCTAATA TGCAACAAAG CTGTGTAACC ATTGTGCTGG
9801 TCCATTATGA CTATTTCTTA AATAAGATGT GCCTAAGGAA AAACGTTTAA
9851 ATTTCAAGCC ACAATGGCAG CCAGGACAAT TTAGCTGAAA AACAACTTG
9901 TGGAATGAAT AAACACGAGC AAATAGCAAA ATCGTCAGTA ATAAATGAAA
9951 TATACACAAC AAATTCAGAA TTAAATCAT TGCTTTTGCA AAAATCTTTT
10001 CAAAAAATAA AAAATTAAGT CATTGCTCTT TCAAGTGTA TCCACAAATT
10051 TGATGAACCT GTTATGAACA CAGAATATGT CGAGTTATAA TGCTTTAATA
10101 TGAGAGATAT CTGCTAATTA CATTATAGTA CGTGGATGTT CTCTTAACCA
10151 AACATGACAT TTGTAAATTT ACTTGTGGT AAATTTGTGCT TTTCTATCTG
10201 ACATCATCTT AATGGTTATT AAGCATCATC TATTACGGGC TATTTGGGAA
10251 AAAATCAAAAT TTTTCACTCT ATATTTATTC TACAATTCAAT TCTATGTTGT
10301 ATTTCACTCT AAAGTTATGA CTGTTCTTAA AGCCTCAATT AAATGAACTT
10351 TGTAACAGTT ATGTAATATT TAACTAAAAC ATGCTGGTTT TAAATCAAAT
10401 AGCCTTTGCC TAAAACTACT TCATCAGGAA CCATGCAAAG TAACAAAACA
10451 AAAAAATTCAT GTTCTAAGGA GGGGGAGGTA AATATTACAG GTTGAATGGC
10501 TCATGAAGGA ACAATGGTTG AAAAATGCTG CTTAAGAAA AAAAAAAGT
10551 AAAGCTATCA GGCAGTTTCC TTCATTCCAA GAATATTAGG ATATTCAAAC
10601 TGGGGAGGAC TTGACATCCT AATGTTCTCT CTCCATTTTA CAGCAGAGTA
10651 AACTGAGGGC ATGAGGTTGG TAAGTGGTTG AGCCTGTGCC AGGGTTGGG
10701 CCGTTACTTC CTAAGTGAAC AACCCCTCCG GGGGATTCA AATTGAATTG
10751 CCATATGCTT AACTTTTTTA ATTTAATGGA ACCTAACATT TCTGAACCTT
10801 ATCTCTAAGA TGATGGAAC AATGCCCTGCC TCACAAAATT GTTCTGACAA
10851 GTTAAATAAA ATATAGTGCC CAGGATAGAG AAGACGTTCA GAGCATTTAT
10901 TTGCATTTCC TACATACCTC AGAGGTATCA TTCTTATTTT TCCCTCCTAG
10951 AGGAAGAGGT TTCCATTTTG AAAGTGGCAA CTCCAAAGAC ATGAATCTCT
11001 AATAGTAGAA TTAGGTATT CAGTGAGAAG TGGTGACAGA AATAAAAACC
11051 TATAGTGTTT CATATCAGAA ACCTATGATG AAGTGGTTTA TTTTATAGAA

```

FIGURE 3C

```

11101 AAATGTGAGA AAAGTAAATC ATTCCTTTG CCTGCTGAAT TTTTGTGTTT
11151 AAAAAATTAAT TTAATCATTC CTATCCTTTA ATAAATATTG AAAAAACAA
11201 TTAAGCAAA AGGAAAGTAA GAAATCACCC ACAATCACAC TACTAAAAAT
11251 AAGCTCTATT GAACTTTGAT GTCTAACTTT CAAAATAGTA TTATTCTTAC
11301 TAGAATAGAA ATTGGTGAAA TCCTTTTCATC TCITTTTATTT TAAAGTGACA
11351 TAAATATTTG TTCAAAATTG TGTTTTTCAG ATGAATTATC ATCAACTGTG
11401 GGTGAGTTGA TCAAAAAACT TTTATTAGCG TTTCTTTTCA AATTTTGTGA
11451 ATAAAAATGTG AAACCATTTGA GTATTCCCTT TCCTGGTAAG AAATCATACT
11501 TTTCAATAAA GATTACTATT TTATGTACTT ATATATAAAG TAAATTCAT
11551 GTTAAAAAAT TTAAGAAAAA AGAAGAAGAG TCCAGCACT TTGGGAGGCC
11601 GAGGCAGGGG GATCACGAGG TCAGGAGATC GAGACCATCC TGGCTAACAA
11651 GGTGAAACCC CGTCTCTACT AAAAACTAC AAAAAATTAA CTGGGCATGG
11701 TGGCGGGCGC CTGTAGTCCC AGCTACTCCC GAGGCTGAGG CAAGAGAATG
11751 GCGTGAACCT AGGAGGCGGA GCTTGCAGTG AGCCGAGATC ACGCCACTGC
11801 ACTCCAGCCT GAGGTGACAGT CTCAAAAAAA AAAAAAATAA GAAAGAAAGA
11851 AGAAGAGGAG ATAAATCATA ATTTTACTGT TATTTTACTT TTGGGTGATT
11901 AAACATGTCA TTACATGTAA CTTCATAATG TGATTTTATG CAAAAATCAA
11951 GTATTCTGGA AAAGAATGAA ATAGAAAATG CCTTTTGCCA CCTAAAAATG
12001 AGTCTATACAT TTTCTGTAT CTATTTATAT ACTCATCTTT ATTAATGCCT
12051 ACTTGATATT ATAAAGCCAA TTTTATTTTA AAATCCCTC AGGGATGAAT
12101 ATTTTATTAAT TAATTATTAT TTTTAAATCT AATACTAATC ATGCTGAAAT
12151 AAACATTCTT AATCTAAGTG TCATACCAAT TAAACTAGG AAAGCACAGG
12201 GGTGTTACAT TAACACTATT TTAGGTCTA ACATAAGTGA TATTGTAGAA
12251 AGGACAAACT CTATATAAAA TAACAGGCAA AATTATTTTA TGATTGTATA
12301 CCTTAAATCC CTAAGAACAT GAATAAAAT TATTAGTTT AATAAGTAA
12351 TATGGCTATG TGGCTGGATG CAAGATAAAC ATACAACAGT CAATAGTTTT
12401 TCTCTATACT GATAAGAACC AACTGGACAT TGATGGAAT GGAAAAACAT
12451 CTGTTTAAACA TCAGTGACCA AAAAAAGCA CTTTGAATT AAATTTGTGA
12501 GAAATGTGTA GGATTATAT AGAATAAATA AGAAATCAT ATTGAAGAGC
12551 AAAAAAATAC ATTTGAAAA AGAGCAGATG CCATGTTCTT GGATGAAAA
12601 CATAATGCCA TAACATATC ACTTTTCCCT AACTAACAT TTTTATTTAA
12651 TTTTATTTTA CGTAAGAGCC CATTTTAAAG CCTAATAATA GTGAAATAAA
12701 TGTCAGCTCT AAATTTTATT TTTTCAAAAC TATAACTCAT TTTCCAGAA
12751 TTACAGGTAA GTCGATCAGA AAATTTGAT TATTAATGTA TATTTCTTT
12801 TAATTTTGTA AAAGTGAGGA AATCAGTAAT GTTTTCCAG TAAAAAATAA
12851 TCATTTTTC TCCTTATATA AAATGTAACA TAAGTGCTC ATAAACAGAT
12901 AAAGAAACA GAGGTATAGA AGTCACTCTG AACTTTCATC AGCACTAAAT
12951 GTGCTACTC TTTAATTTCT GCCAATCTAA CAGGAAATA TAGATTTTAA
13001 AATTTTATA GAAAAACAT CCAATAATA TTAGAAGTAT GAACAAAGTG
13051 CTGTAGTAGA GAAATATTTT GTCTCATCAG ATATGTAGAC ATACTAAAT
13101 ATTTGTAATAT AAGCAACATG GAATCAGCCT AGAAAAATGG AAGAGGAGGA
13151 GGAGGAGGGA GGAAGGAAGA AGAAGAGGAA GAAGAAGAAG GAGGAGGAGA
13201 AGGAAGAGGA GAGAAGAAAA GAAGAAATAT TTTGTGATGC TTGCTGGCTC
13251 CAGAGCCTGC TCATTTTACC ACCATGCTTC CTACCTCTC TAATAATTC
13301 AAGAACTTT ATAGAAAATA TAGAATGGAT CCCAGGATAT ATTAGAATAA
13351 TCCATGGTTT TAAACATGAT TGTTTAAATG TACAGGTGAT TTATTTTCATC
13401 ATCGATCTGT CATTTACTAA TATGTTTGAA AGAAAAATAA GTTGCAATGC
13451 CCTTTGGCAT CATATACAAA ACTAAATCT AGCTTAATCA AATTTTAAAT
13501 GTAAAAGGAA CAAAATGTT TAATTTAAGA AGAAATGGA AAATATATGT
13551 ATGGCCTAAA GACCAGTAGA TCTTTTAAAG CAAGGCAGAA ATCTTCTGTA
13601 ATACAGGTAA CACATTTTAA AAGTGTGTAA ATGCTTCGTG CTAATGAAAA
13651 TTCAGTAAA GGAGCACAAT TGGTGGGAAA GACTAATTGC CATTTTCTTT
13701 TGACAAAGCA ATATTCATCA ATGTTAAAT TGTATATACA TTGTGTCCTT
13751 GCAACCTGTC TTTGGGATT CTATAGAAAC AAATGCTTCA ATACATAAGG
13801 ACTTAAATAT GTGAAATGTT TACTACAGCT GTATTTACAG TAGGAAAAAT
13851 CTGAATGTGG AAATACTGAA TAAATGTGT TGAATCCATA TCACAGAACAA
13901 TGAATCAGCA TTAGAAAGAA GTTAAGGCCG GACTCAGTGG CTCATGACTG
13951 TAATCCAGC ATTTTGGAA GCCGAGGCAG GAAGATGCTT GAGCTCAGGA
14001 GTTTGAGACC ACCCTGGCCA ACATAGTGAG ATCTCATCTC TACAAAAAAT
14051 TTAATAATTA ACTGGGCATG GTAGCACGCA CTGCACTCC CAGCTACTCC
14101 AGAGGTTGAG GTGGGAGGAT TGCTTGAGCC CAGGAGGATA AGGCTGCAGT
14151 GAACCATGAA GAAGTTAGGG CTACACAATC ACTTTAAGAA GATTTCTGTG
14201 ATGTGTTAAG TCGGAAACTC AAGTATAAAA TATGATTAT TTTGAAGGGT
14251 AGCAAAAAAT TATGTATGTG TGCCTGTATG TGTGTATGCA CCTATAACTA
14301 TATATATATA TATATATATA TATATATCCT GTATTAACCA TACAGCATAT
14351 ATACAAACC ATATAACATA TATAAGTATA TATATAAACC CATAACAAC
14401 AGATATACCC AGGTGACACA CATAACACA TGTATAGTTG CATGGGAGAG
14451 GGAGAGAAAG AGAGATGATG GAAAGAGGTA AAAATAAGAG ATCCAGGGGC
14501 AGGGGGAAG AATTGTTGGG TAGATAAATG CCCACATTTA TGTTCCTACG
14551 CTAGAGAGAG ATTTCTGAAA CTGGGGCAA TGCTAAGCAA AATTTTGAGA
14601 CTGTTTATAA CCTGGAAAAG AAGTAGGGCA GGGCAGGAA TATAAACAA
14651 AAGAGCTTTG TCATTAGAGA GCCAAAGCTC TGTATTGAG TAGCACAATA
14701 ACATTTACTA CAGCCACAGC CAATGACTCT GATTTCAAAG GGCAACAGCG
14751 CCTAGGAAGA CAGTCGCTTG ATAACCTCTC TCCTAACACA CCCAGAAACA

```

FIGURE 3D

```

14801 GAGGAAGAAT TTTGTTTTCT CTAGATGGGG AGTGTGTCTT TCCATTCCAC
14851 TATAAAAATG GAACATATTA TGA CTGCATC AAGTCCAAGG CAAGACACAA
14901 GTGGTGCTCG TTAACAAGA CCTACGAAGG ATACTGGAAG TTTTGCAGTG
14951 CAGAAGGTGA GTGTCCTGTC TCTAAATACC TCAGAGCAGT AATCTTGGTC
15001 TCTGAGGCTG GATCTGAGG CAAAACACAA GAGGGTCTG AAAATAGAAT
15051 CCATAGCCAA TTAAGTTGTG AAAATAAAGG ATGTTTCGATT TCTGACGATA
15101 TAAAAATCAC ATTGCACCAT AGATATGTTA TGAGCATTTT AACCACATTA
15151 TTTTGAAAAG AATCTTTTTA ATAGAACAGC AAAGTGATGT GGGATCTGAA
15201 TCCCCACCA ATATTGTTCC CATGTGCCCT TGGTGAAATA GGGCAGAGCA
15251 GATGATACTG AATAGGAAGC AAAAGTGTGT ATTTGCCCGT GTGTGTGTGA
15301 GCGCACCATG GTGCATACAT GTATGTCTGT GTGAGTGTGC ATGTGTATGT
15351 GTGCACATTT GTTCACGTGC ATGCACATGT GTGCATATGT GTGCATGTGT
15401 GTGCATGGTG TGGTGAGAGT GGGTGCAAAT TAGAAAGGTT TCTTACTCTG
15451 CTTCTTCCTC TTCCTCCTCT GCAGATTTTG CAAACTGTGT ATTTCCCTTC
15501 TGGTACAGAC CTTGATCTA CTGGGAGTGT ACTGATGATG GGAAGCATT
15551 TGGGAAAAAA TGGTGTTTAC TGACCAAGAA TTTTAACAAG GACCGAATTT
15601 GGAAATACTG TGAATGATGG TGAGATTTAC CAGGGGAGGG GGATGAAGAC
15651 GTTTAATCAT TCTCCTTGGA AAGAACCCTAG AGTTTTTGTT GGCAGAGGAG
15701 AACCCTTTC CATTAGAGAG GGTTTTTTTT TGTTTTTTTT TTTTTTCTGG
15751 GATGTGAGTT GAAGGTGATC TAACTATAAA TAACTAGAGG TTTTCGATTC
15801 AGGATTATGG GGTTTTAATA TGGAAATCTG TAGTTTAGCC TCAGCACTTT
15851 AATTGCTAAC TGTGTAAAT TAGTCCCTTT TAAAAATTA GTCTATGGGG
15901 TAGGCGTGGT GCGTTATGCC CATAATCCCA GCGCTTTGGG AGGCCAAAAA
15951 GGAAGGACTG CTTGAGGCCA GGAGTTCAAG ACCAGCCTGG ACCAACATAG
16001 CAAGACCCCG TTTCTACAAA AGAAAAAATT TAAAAATGTA ACTAGGCATG
16051 GTGGCACACA CTTGTAGTCT CAGCTACTCG GGAAGCTGAG GCAGGAGAAT
16101 CACCTGAGTT CAAGAGTTCA AGGTTGCATT GAGTTGTGAT GAAAGAGAGA
16151 CCCACCTCTA AAATAAAATA AATGAAGTAA AATAAGCTTA TGGATTTTGG
16201 GTGATAGTGA TGTGTCAATG TAGATTCATC AATTGCAACA AATGTCCAC
16251 TCTGGTGGAG GATATTGATA ACACAGTGGC TATGTGTACA TGGAAATTTT
16301 TATACCTACT ACTCAACTTT TCTGTGACTT AAAACTGCTC TAAAAATAA
16351 CGTGTATTAA AAATGCATTA CCTTGGCCGG GCACGGTGGC TCACGCCTGT
16401 AATCCAGCA CTTTGGGAGG TCGAGGCGGG TGGATCACGA GGTGAGGAGA
16451 TCGAGACCGT CCTGGCTAAC ATGGTGAAAC CCCGTCTCTT CTAAAAATAC
16501 AAAAAATTAG CCGGCGGTGG TGGCGGGTGC CTGTAGTCCC AGCTACTCAG
16551 GAGGCT
(SEQ ID NO:3)

```

FEATURES:

```

Start: 1060
Exon: 1060-1156
Intron: 1157-12726
Exon: 12727-12756
Intron: 12757-14824
Exon: 14825-14956
Intron: 14957-15474
Exon: 15475-15614
Stop: 15615

```

FIGURE 3E

Allelic Variants (SNPs):

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
1723	-	T	Intron			
1876	G	A	Intron			
4060	C	T	Intron			
4117	A	C	Intron			
4138	T	C	Intron			
5342	A	G	Intron			
5577	A	T	Intron			
6721	A	G	Intron			
7364	C	T	Intron			
7538	G	T	Intron			
7540	A	G	Intron			
7650	G	T	Intron			
7778	A	T	Intron			
8859	G	A	Intron			
9518	C	T	Intron			
9588	G	C	Intron			
9610	G	A	Intron			
10187	T	C	Intron			
10514	A	G	Intron			
10538	A	-	Intron			
10729	C	T	Intron			
10801	A	G	Intron			
10848	C	T	Intron			
11429	C	T	Intron			
11841	G	A	Intron			
11842	-	A	Intron			
13973	T	C	Intron			
13983	A	G	Intron			
14542	G	A	Intron			
15732	G	T	Beyond ORF(3')			
15813	-	A T	Beyond ORF(3')			
16132	-	A	Beyond ORF(3')			
16317	C	T	Beyond ORF(3')			

Context:

DNA

Position

1723 CCTCCCTTCATGCTAGGAGATTTTAAATGCCTCCTGCAGAGGTTTTGAAAATGAGACTAA
TTTTTCATAAAGTCGATTTTATTCCTTTTCCAAAGTCGTGTTGCCTGTAACAATTATGAAG
CGAAATAGAAAAATAAGAAGGGAATTCACATATGGATGCATCTGACACGTTTGATGAGGA
CTAAGCATAGTTTCAAAAACAATTTTATGGAATGACCAAGAAGAATGATCTTGCAGAT
ATTCTTTTAAAGTATTCTTCAGGACATCTTTCATTAAAAAAAATAACACTGCTTGCTA
[-, T]
TTTTTTTTTAAAGAACAGAAAAGAAAATATACTCCCAATGTAACATGACCACTGCTTTCT
TATCCACTAGAATTATTTTATACTTACTGAAAATAACACTGTACTTGGCCACAGGTTTT
TTTTTTTAAATCAGACATCATGTTTGTAGATACGTAAAAAGGAATATAGATAAGTGGAAAA
AGGCAAATTAAGACGATCCTAAAGCTTCTCCGCTGTCAATGTCTGGTCTCCTAATTTATG
GATGTCATATTTTCCACATAATACTTGTGCTATTTCAGAGCCTTACTGATGAAGCATTTT

1876 GATGCATCTGACACGTTTGATGAGGACTAAGCATAGTTTCAAAAACAATTTTTTATGGA
ATGACCAAGAAGAATGATCTTGCAGATATTTCTTTTAAAGTATCTTCAGGACATCTTGC
ATTAAAAAAAATAACACTGCTTGCTATTTTTTTTTTAAAGAACAGAAAAGAAAATATACT
CCCAATGTAACATGACCACTGCTTTCTTATCCACTAGAATTATTTTATACTTACTGAAA
AATAACACTGTACTTGGCCACAGGTTTTTTTTTTAAATCAGACATCATGTTGTAGATAC
[G, A]
TAAAAAGGAATATAGATAAGTGGAAAAAGGCAAATTAAGACGATCCTAAAGCTTCTCCGC
TGTCATGTCTGGTCTCCTAATTTATGGATGTCATATTTTCCACATAATACTTGTGCTA
TTTCAAGCCTTACTGATGAAGCATTTCTCTTAATAATGCATCAGAGAATAAAACATCATG
GGCCCCGGGCTCGTTAGCCTGTATTTTAAATGATGTGGTGTCTACTTTTGATCTGC
AGAACTGTTGCCATTTCATATCTCCCTCTCTCCCTTGCAAGCATCTGCCCTCTGGAC

4060 GCTCTCATGCACCCCCACTCACCAAGAACTGAAGGCTCCTCTCGCCTCCTCTGTCCCACT
CAACTCTGCTCCTGGCAAGCTCAGGAGACCAAGTTTTCCCTCCAGCATCTCCGGTTCCAA
CCAAGCTTAGCCTGTGGCTCAAGCAAGACTTTGAGTCCCTGGATGCTGTGAGTTCCAC
ATCTGAAACAGGAGGATCGTGATCTTGCTGACTCCAGGGTGTCTGGAAGATTAAATGG
GTTAGCATGCGTGACGCTCAGAACAGAACTGTTATCATTAAGTCATTATTACTACTATTA
[C, T]
AGTACCAAAAGTACTATTACAGTATTAAAGAGGCAGGATAAGGGTTTGGATTAGTAGTC

FIGURE 3F

GATACTCACAGAGTGAGTTTAACTACTTTAACTTCTCATACAAGTGAATCGTGCAATA
TTTGTCTCTCTGGAATGGTTTGGTTCACTTAGCATAATGTCCTCCAGGTTTCATCCATGT
TACTGAAACGCGCAGGATTTTCTCTTTTAAATTTTAAAAATTTTCTGATTTTGTAAAT
TTTTATTTTATTTATTTATTTATTTGTTTTTGAGACAAGGGCTCACTCTGTCAACCAGGCT

4117 ACTCAACTCTGCTCCTGGCAAGCTCAGGAGACCAGTTTCCCTCCAGCATCTCCGGGTTCC
CAACCAAGCTTAGCCTGTGGCCTCAAGCAAGACTTTGAGTCCTGGATGTCGTGTGAGTTCC
CACATCTGAAACAGGAGGATCGTGATTCTTGCTGACTCCAGGGTTGCTGGAAAGATTAAA
TGGGTTAGCATGCGTGACGCTCAGAACAGAACTGTTATCATTAAAGTCATTATTACTACTA
TTACAGTACCAAAAGTACTATTACAGTATTAAAGAGGCAGGATAAGGGTTTGGATTAGT
[A, C]
GTCGATACTCACAGAGTGAGTTTAACTACTTTAACTTCTCATACAAGTGAATCGTGCA
ATATTTGTCTCTGTGAATGGTTTGGTTCACTTAGCATAATGTCCTCCAGGTTTCATCCA
TGTTACTGAAACGGCAGGATTTTCTCTTTTAAATTTTAAAAATTTTCTGATTTTGT
AAATTTTATTTTATTTTATTTTATTTGTTTTTGAGACAAGGGCTCACTCTGTCAACCAG
GCTGGAGTGCACTGGTGTGATCTCAGCACACTGCAACCTCCGCTCCCGGGACTGAAGTA

4138 GCTCAGGAGACCAGTTTCCCTCCAGCATCTCCGGGTTCCAACCAAGCTTAGCCTGTGGC
CTCAAGCAAGACTTTGAGTCCTGGATGTCGTGAGTTCCACATCTGAAACAGGAGATC
GTGATTTCTGCTGACTCCAGGGTTGCTGGAAGATTAAATGGGTTAGCATGCGTGACGCT
TGAAACAGAACTGTTATCATTAAAGTCATTATTACTACTATTACAGTACCAAAAGTACTAT
TACAGTATTAAGAGGCAGGATAAGGGTTTGGATTAGTAGTCGATACTCACAGAGTGAG
[T, C]
TTAACTACTTTAACTTCTCATACAAGTGAATCGTGCAATATTTGTCTCTGTGAATG
GTTTGGTTCACTTAGCATAATGTCCTCCAGGTTTCATCCATGTTACTGAAACGGCAGGAT
TTTCTCTTTTAAATTTTAAATTTTCTGATTTTGTAAATTTTATTTTATTTTATTTATTT
TTATTTTATTTTATTTTATTTTATTTGTTTTTGAGACAAGGGCTCACTCTGTCAACCAGGCTGGAGTGCACTGGTGTGAT
CTCAGCACACTGCAACCTCCGCTCCCGGGACTGAAGTAATTAATCTCCAACCTCAGCCT

5342 ACGGCGAATTAGCACCCAAGACAGAGGTGTTTTAGCCTCCACAACTGGATTGCTTAGTG
ACACTCAAATTTCTCGTCTTCATCCTCACTGTCTGGGGTTGCATTGCATAGTGTCCTTCC
AGAATTCATGTTTACCAGAGATCTGTGAATATGACTGTATTTGAAATAGGTCCTTGCAGA
TGTCGGTAATTAGATGTAGTCACAATGGATACGGGTTGGCCCTAATCCAATGACTGGTA
TCTTTATAAGAAGAGGAAACAGACACACAGGGAAGACAAGCACGCATAGTCACAAGAA
[A, G]
AGATTAAAGTGATGCTGCCACAAGTCACAGAATGCCACAGCTTGCCAGCAGCCACCAAAA
GCAGGAGAGAGGCATGGAATAGTTTCTGCCCTGTGAGCTCCTAAGAGGAACCAACCTTG
GCATTTTGGCCTTCTGGCCTCCAGATCTGTGAGACAATAAATGCTATTGCTCTAGACCA
AACAGTCTATGGTACTTTGTCTAGATTTTTTTTATTTTATTTTATTTTATTTTATTTT
AGATAGGGTCTCATTCTGTCAACCAGGCTGGAGTGCACTGGCACAATCACGGCTCACTGC

5577 TGGTATCTTTATAAGAAGAGGAAACAGACACACACAGGGAAGACAAGCACGCATAGTCAC
AAGAAAAGATTAAAGTGATGCTGCCACAAGTCACAGAATGCCACAGCTTGCCAGCAGCCA
CCAAAAGCAAGGAGAGAGGCATGGAATAGTTTCTGCCCTGTGAGCTCCTAAGAGGAACCA
ACCCTGGCATTTTGGCCTTCTGGCCTCCAGATCTGTGAGACAATAAATGTCTATTGCTCT
AGACCAACAGTCTATGGTACTTTGTCTAGATTTTTTTTATTTTATTTTATTTTATTTT
[A, T]
TTTTTAGATAGGGTCTCATTCTGTCAACCAGGCTGGAGTGCACTGGCACAATCACGGCTC
ACTGCAGTCTTGACCTCCAGGCTCAAGTGATCCTCTTGCCCTCAGCTCCCAAGTAGCTG
GGACTACAGGTGCACGCCACCATGCCAGCTAATTTTGTGTCTTTATTTATTTATTTGAGAGA
TGGGGTCTTGCTATGTTGCCAGGCTGGTCTCGAACTCCTGGGTTCAAGCAGTCTCTCTG
GCTCAGCTCCAAAAGTGTGGGATTACAAGCGTAATCACCATGCCTGACCAGTAAAT

6721 AAGCTCCTGAACACATTTACATGCTTTCATTTTATTTACATTCACTTATTCTTTTCCTT
TTTTTTTTTCTAGACAAAGTCTTGCCCTGTCACTCAGGCTGGAGTGAGTGGTGTGATC
TCAGCTCACTGCAACCTCCACGTCCAGGTTCAAGTGATTCTCTGCTCAGCTGCCTCA
GTAGCTGGGATTATAGGCGCACACCACCATGCCAGCTAATTTTGTATTTTATTTAGTAGAG
ACAGGGTTTACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCTGCC
[A, G]
CCTCGGCCCTCCCATAGTGCTGGGATTACAGGTGTGAGCCACCATGCCGGGCTACTTACT
CTTTTCTAATGTGCTTCTCTTTATATCTCTGAAGAATTTTAACTTTCTTTCAAG
GCAGGTTTTCTGTAGCAAAATCCCTCAATGTTTGTGTCTGAGAATCTTTGTCTTTCT
TTCTCTTCACTTCTGAAGAATAATTTTGCAGGTTACAGAATCTAGATTGAAGATTTT
TTTTCTCTCAACACTGTAAATGCTTCTTCTTATCTTCATGGTTTCTGGAAAGAAGTTGG

7364 CCCCCCTCTGGCTTCTTTTCAAAAGATTTTCTTTTAAAAAATTTTATTTTAGATGCAGGAG
GTACATGTGCAGATTTGTACATGAGTATGTTGCATAAATGTTGAGGTTTGGGATTCAAT
TGAAGCCATTACCAAAATAATGAACATAGTGCCAGTAGGCAGTTTCTTTTCTTTTCT
TTTTTTTTTTTTTTTTCAGACAGAGTCTGTATTGCCAGACTGGAGTGCACTGGTGCG
ATCTTGGTCACTGCAACCTCTGCCCTCCAGGTTCAAGCGATTCTCTGCTCAGCTCAGCCTCC
[C, T]
GAGTAGCTGGGATTACAGGCGCCACCACCATGTCCGGCTAATTTTGTAGTAGAGACGGGA
TTTTCATATGTTGGCCAGGCTGGTCTTGAACCTCCTGACCTCATGATCAGTCCACCTCAGC
TTCCCAAAGTGCTGGGATTACAGGCTAGAGCCACTGACCCGGCATCAATAGGTAGTTT

FIGURE 3G

CCAAGCCTGGGTCCCTCCTTACCTCCCACTTTTGTAAATCCCCAGTGCCTATTGTTCCC
ATCTTTGTGTCCCTGTGCACCCAATGTTTAGTTCCCACTTATAAGGGAGAACATGTGGTA

7538 TTTTCTTTTTTTTTTTTTTTTTTTCAGACAGAGTCTTGTATTGCCCAGACTGGAGTGCAGT
GGTGCAGTCTTGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGCCTCA
GCCTCCCGAGTAGCTGGGATTACAGGCGCCCAACCACCATGTCCGGCTAATTTTTAGTAGA
GACGGGATTTTCATCATGTTGGCCAGGCTGGTCTTGAACCTCTGACCTCATGATCAGTCCA
CCTCAGCTTCCCAAAGTGTGGGATTACAGGCATGAGCCACTGCACCCGGCCATCAATAG
[G, T]
TAGTTTCCAAGCCTGGGTCCCTCCTTACCTCCCACTTTTGTAAATCCCCAGTGCCTATT
GTTCCCATCTTTGTGTCCCTGTGCACCCAATGTTTAGTTCCCACTTATAAGGGAGAACAT
GTGGTATTTGGTTTTCTGTTTCTGCACTAATTTGCTTAGGCTAATGGCCTCCAGCTGCAT
CCATGTTGCTAAAAAGGACATGATTTTGTCTTTGTTATGGCTGCATAAAACAATTCGT
TTTGGGTTTAACTGAAAAAATGAGACTCATTTTTATTTCAAATAAGGGTATTTTTAAAC

7540 TTTCTTTTTTTTTTTTTTTTTTTCAGACAGAGTCTTGTATTGCCCAGACTGGAGTGCAGTGG
TGCAGTCTTGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGCCTCAGC
CTCCCGAGTAGCTGGGATTACAGGCGCCCAACCACCATGTCCGGCTAATTTTTAGTAGAGA
CGGGATTTTCATCATGTTGGCCAGGCTGGTCTTGAACCTCTGACCTCATGATCAGTCCACC
TCAGCTTCCCAAAGTGTGGGATTACAGGCATGAGCCACTGCACCCGGCCATCAATAGGT
[A, G]
GTTTCCAAGCCTGGGTCCCTCCTTACCTCCCACTTTTGTAAATCCCCAGTGCCTATTGT
TCCCATCTTTGTGTCCCTGTGCACCCAATGTTTAGTTCCCACTTATAAGGGAGAACATGT
GGTATTTGGTTTTCTGTTTCTGCACTAATTTGCTTAGGCTAATGGCCTCCAGCTGCATCC
ATGTTGCTAAAAAGGACATGATTTTGTCTTTGTTATGGCTGCATAAAACAATTCGT
TGGGTTTAACTGAAAAAATGAGACTCATTTTTATTTCAAATAAGGGTATTTTTAAACCT

7650 CTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCGCCCAACCACCATGTCCGGCTAATTT
TTAGTAGAGACGGGATTTCATCATGTTGGCCAGGCTGGTCTTGAACCTCTGACCTCATGA
TCAGTCCACCTCAGCTTCCCAAAGTGTGGGATTACAGGCATGAGCCACTGCACCCGGCC
ATCAATAGGTAGTTTCCAAAGCCTGGGTCCCTCCTTACCTCCCACTTTTGTAAATCCCCA
GTGCCTATTGTTCCCATCTTTGTGTCCCTGTGCACCCAATGTTTAGTTCCCACTTATAAG
[G, T]
GAGAACATGTGGTATTTGGTTTTCTGTTTCTGCACTAATTTGCTTAGGCTAATGGCCTCC
AGCTGCATCCATGTTGCTAAAAAGGACATGATTTTGTCTTTGTTATGGCTGCATAAAAC
AATTCGTTTTGGGTTTAACTGAAAAAATGAGACTCATTTTTATTTCAAATAAGGGTATT
TTTAAACCTTTATATATATAACAGAATTTAGCTTGAATTATGACCAAAGGTGAATTCCT
GAGAATGAAAAGATTGGGATGCAGGACTTTCCTGACAAAATTGTTACGATTTTCCAGTG

7778 CCTCAGCTTCCCAAAGTGTGGGATTACAGGCATGAGCCACTGCACCCGGCCATCAATAG
GTAGTTTCCAAGCCTGGGTCCCTCCTTACCTCCCACTTTTGTAAATCCCCAGTGCCTAT
TGTTCCCATCTTTGTGTCCCTGTGCACCCAATGTTTAGTTCCCACTTATAAGGGAGAAC
TGTGGTATTTGGTTTTCTGTTTCTGCACTAATTTGCTTAGGCTAATGGCCTCCAGCTGCA
TCCATGTTGCTAAAAAGGACATGATTTTGTCTTTGTTATGGCTGCATAAAACAATTCG
[A, T]
TTTGGGTTTAACTGAAAAAATGAGACTCATTTTTATTTCAAATAAGGGTATTTTTAAAC
CTTATATATTATAACAGAATTTAGCTTGAATTATGACCAAAGGTGAATTCCTGAGAATGA
AAAGAAATGGGATGCAGGACTTTCCTGACAAAATTGTTACGATTTTCCAGTGCATAGCTT
TGCTCTCTCCATTCTACACTTTCCTCAAAGGGTCATAAGAAAGCCAGTTTGTATAAG
TATCTAAGACTAATCTAGGACTATATTAATTTTTAGGGGCCCTTCTTAATCCTTTGAA

8859 TCTTGCATCTTTTCAAAAGTGACCGGCTTCATTTTGTAGTCCACTTGCTTTGCAGTTGCA
TTGTACTTGCAAGGTATTTTTTGAACATAACATCCCAACATTCACAACAATGATCCAC
CAGGTGGATGGCAGAATTTCTGGTGTAAATCAGAGTGGGACGTTTGAGTGGGGACTCATA
ATTTTATAGCACCGAGGATGGTGCACCTGTGTTTATGATATGGGGATGTTGCATTCTGATG
CTTACCTGGTGATACCTGTTCTAAGATCATAGTGCTTAGCAGACAGGCTATGTTGGAATC
[G, A]
CTCGTCTGTATCTCCCCATTATGAAATCAGGCCTATCATGATCCTCAAACCAAAATCC
ATTTCTTAATAGTTTCACTATGGAATGCATACTGAATGAACCAACAATCCAGGAAGTA
GAAACAATTATAATCCCATATTTTCAAGAGAAAGACAAGCTTAACTCCTGATGAGTCCATT
GATCTTGGATTCCCACTGAATCTCATGGTTTTCTGTAGTCCAGAGTCTACCATTTAGCC
ACTCCCTTTTTGGGTAGCCAGAACCTAGCACTAGGCTGGGCACATAGTAGGCATTCAAT

9518 TCATCGTGAGATCTTCCCTATATACTGGAACCAAGTGCACAGTTTCAAGTGGCATTTCAATGCA
TTCAAATTTGTCAATCTATCTCCAGAGCCTTTTCATTGTGCCAAACTGAAACTCTGCACC
CATTAATAATAACTCCCATTTCTCCCTCCCTCCAGCTCCTAACAACCACTTCTGCT
TTTTGCTCTATGTATTGACTATTCTAGTTACCTCATATAAGTAGAATCATATAATATT
TGCTTTTTTGTGTCTGGCTTATTTTACTTAACCTAGTATTTTCAAAGTGCATCCATGTGG
[C, T]
GGCATCTATTCAAATTTTGTTCCTTTTTAATGCTGAATAATACTCCATTGTATGTGTCTA
CCATATTTTGTATTATTTACTCATTTGCTCACGGACATGTGGGTCTTTCCACCTTTTGGC
TATTATGAATAATGCTACCATGAACATGGGTTTACAAATATTTGTTTGAGTCTCTGCTTT
CAATTATTTTGTATATATACCAAGAGTGGAGGGATCTGGGTTTTTAAAAACTCAGCTG
GTGATTCTAATATGCAACAAAGCTTGTGAACCATTTGCTGGTCCATTATGACTATTTCT

FIGURE 3H

9588 GCAATCTATCTCCAGAGCCTTTTCATTGTCCCAAACCTGAACTCTGCACCCATTAAATAA
TAACTCCCCATTCCCTCCCTCCCTCCAGCTCCTAACACCACCATTCTGCTTTTGTCTCT
ATGTATTTGACTATTTCTAGTTACCTCATATAAGTAGAATCATATAATATTTGTCTTTTG
TGTCTGGCTTATTTTACTTAACCTAGTATTTTCAAAGTGCATCCATGTGGCGGCATCTAT
TCAAATTTTGTCTCTTTTAATGCTGAATAATACTCCATTGTATGTGTCTACCATATTTT
[G, C]
TTTATTACTCATTTGCTCACGGACATGTGGGTTCTTTCCACCTTTTGGCTATTATGAAT
AATGCTACCATGAACATGGGTTTACAAATATTTGTTGAGTCTCTGCTTTCAATTATTTT
GTATATATACCCAGAAGTGAAGGGATCTGGGTTTTTAAAAACTCAGCTGGTGATTTCTAA
TATGCAACAAAGCTTGTGAACCATTGTGCTGGTCCATTATGACTATTTCTTAAATAAGAT
GTGCCTAAGGAAAAACGTTTTAATTTCAAGCCACAATGGCAGCCAGGACAATTTAGCTGA

9610 TCATTGTCCCAAACCTGAACTCTGCACCCATTAAATAATAACTCCCCATTCCCTCCCTCCC
TCCAGCTCCTAACACCACCATTCTGCTTTTGTCTCTATGTATTTGACTATTCTAGTTA
CCTCATATAAGTAGAATCATATAATATTTGTCTTTTGTGTCTGGCTTATTTTACTTAAC
CTAGTATTTTCAAAGTGCATCCATGTGGCGGCATCTATTCAAATTTTGTCTCTTTTAAAT
GCTGAATAATACTCCATTGTATGTGTCTACCATATTTTGTTFATTACTCATTGCTCAC
[G, A]
GACATGTGGGTTCTTTCCACCTTTTGGCTATTATGAATAATGCTACCATGAACATGGGTT
TACAAATATTTGTTGAGTCTCTGCTTTCAATTATTTGTATATATACCCAGAAGTGGAA
GGGATCTGGGTTTTTAAAAACTCAGCTGGTGATTTCTAATATGCAACAAAGCTTGTGAACC
ATTGTGCTGGTCCATTATGACTATTTCTTAAATAAGATGTGCCTAAGGAAAAACGTTTTA
ATTTCAAGCCACAATGGCAGCCAGGACAATTTAGCTGAAAAACAACTTGTGGAATGAAT

10187 GAAAAACAAACTTGTGGAATGAATAAACACGAGCAAAATAGCAAAATCGTCAGTAATAAAT
GAAATATACACAACAAATTGAGAAATTAATAATCATTGCTTTTGCAAAAATCTTTTCAAAAA
ATAAAAAATTAAGTCATTGCTCTTTCAAGTGAATCCACAAATTTGATGAAGTATG
AACACAGAATATGTGAGTTATAATGCTTTAATATGAGAGATATCTGCTAATTACATTAT
AGTACGTGGATGTTCTCTTAACCAACATGACATTTGTAAATTTACTTGTGGTAAATTG
[T, C]
GCTTTTCTATCTGACATCATCTTAATGGTTATTAAGCATCATCTATTACGGGCTATTGG
GAAAAATCAAAATTTTTCACCTATATTTATCTACAATTCATTCTATGTTGTATTTCAT
CCTAAAGTTATGACTGTTCTTAAAGCCTCAATTAATGAACCTTTGTAACAGTTATGTAAT
ATTTAACTTAAACATGCTGGTTTAAATCAATAGCCTTTGCCTAAATCACTTCATCAG
GAACCATGCAAGTAACAAAACAAAAAATTCAGTTCTAAGCAGGGGGAGGTAATATTA

10514 GGTATTAAGCATCATCTATTACGGGCTATTTGGGAAAAAATCAAATTTTCACTCTATA
TTTATTTACAAATTCATCTATGTTGTATTTTCATCCTAAAGTTATGACTGTTCTTAAAGC
CTCAATTAATGAACCTTTGTAACAGTTATGTAATATTTAACTAAAACATGCTGGTTTAA
ATCAAAATAGCCTTTGCCTAAATCACTTCATCAGGAACCATGCAAGTAACAAAACAAAA
AATTCAGTTCTAAGGAGGGGGAGGTAAATATTACAGTTGAATGGCTCATGAAGGAACA
[A, G]
TGGTTGAAAAATGCTGCTTTAAGAAAAAAGTAAAGCTATCAGGCAGTTTCCTTCA
TTCCAAGAAATATTAGGATATTCAAACTGGGGAGGACTTGACATCCTAATGTTCTCTCTCC
ATTTTACAGCAGATAAAGTAACTGAGGGCAGATGGTTGGTAAGTGGTTGAGCCTGTGCCAGGG
TTTGGGCGGTTACTTCTACTGAACCAACCCCTCCGGGGGATTCAAATTGAATTGCCAT
ATGCTTAACCTTTTAAATTTAATGAACTTAACATTTCTGAACCTTTATCTCTAAGATGAT

10538 GGCTATTTGGGAAAAAATCAAATTTTCACTCTATATTTATCTACAATTCATTCTATGT
TGATATTTTCATCCTAAAGTTATGACTGTTCTTAAAGCCTCAATTAATGAACCTTTGTAACA
GTTATGTAATATTTAACTAAAACATGCTGGTTTAAATCAAATAGCCTTTGCCTAAATC
ACTTCATCAGGAACCATGCAAGTAACAAAACAAAAAATTCAGTTCTAAGGAGGGGGAG
GTAATATTTACAGTTGAATGGCTCATGAAGGAACAATGGTTGAAAAATGCTGCTTTAAG
[A, -]
AAAAAAGTAAAGCTATCAGGCAGTTTCCTTCATTCCAAGAAATATTAGGATATTCAA
ACTGGGGAGGACTTGACATCCTAATGTTCTCTCTCCATTTTACAGCAGAGTAAACTGAGG
GCAGATGGTTGGTAAGTGGTTGAGCCTGTGCCAGGGTTTGGGCCGTTACTTCTACTGAA
CCAACCCCTCCGGGGGATTCAAATTGAATTGCCATATGCTTAACTTTTTAAATTTAATG
GAACCTAACATTTCTGAACCTTATCTCTAAGATGATGGAACAATGCCTGCCTCACAAAA

10729 AACCATGCAAGTAACAAAACAAAAAATTCAGTTCTAAGGAGGGGGAGGTAATATTAC
AGGTTGAATGGCTCATGAAGGAACAATGGTTGAAAAATGCTGCTTTAAGAAAAAAGGAA
GTAAAGCTATCAGGCAGTTTCCTTCATTCCAAGAATATTAGGATATTCAAACTGGGGAGG
ACTTGACATCCTAATGTTCTCTCTCCATTTTACAGCAGAGTAAACTGAGGGCAGATGGTT
GGTAAGTGGTTGAGCCTGTGCCAGGGTTTGGGCCGTTACTTCTACTGAACCAACCCCTC
[C, T]
GGGGGATTCAAATTGAATTGCCATATGCTTAACCTTTTTTAAATTTAATGGAACCTAACAT
TTCTGAACCTTTATCTCTAAGATGATGGAACAATGCCTGCCTCACAAATTTGTTCTGACA
AGTTAAATAAATATAGTGCCAGGATAGAGAAGACGTTCAAGCATTTTATTTGCATTTT
CTACATACCTCAGAGGTATCATCTTATTTTCCCTCCTAGAGGAAGAGGTTCCATTTT
GAAAGTGCAACTCCAAGACATGAATCTCTAATAGTAGAATTCAGGTATTCAGTGAGAA

10801 TCATGAAGGAACAATGGTTGAAAAATGCTGCTTTAAGAAAAAAGTAAAGCTATCA

FIGURE 3I

- GGCAGTTTCCTTCATTCCAAGAATATTAGGATATTCAAAC TGGGGAGGACTTGACATCCT
AATGTTCTCTCTCCATTTTACAGCAGAGTAAACTGAGGGCAGATGGTTGGTAAGTGGTTG
AGCCTGTGCCAGGGTTTGGGCCGTTACTTCTACTGAACCAACCCCTCCGGGGGGATTCA
AATTGAATTGCCATATGCTTAAC TTTTAAATTAATGGAACCTAACATTTCTGAACTTT
[A, G]
TCTCTAAGATGATGGAACAATGCCTGCCTCACAAAATTGTTCTGACAAGTTAAATAAAA
TATAGTGGCCAGGATAGAGAAGACGTTAGAGCATTTATTTGCATTTCTTACATACCTCA
GAGGTATCATTTCTATTTTCCCTCCTAGAGGAAGAGGTTTCCATTTGAAAAGTGGCAAC
TCCAAAGACATGAATCTCTAATAGTAGAATTCAGGTATTAGTGAGAAAGTGGTGACAGAA
ATAAAACCTATAGTGTCCATATCAGAAACCTATGATGAAGTGGTTATTTTATAGAAA
- 10848 AGTAAAGCTATCAGGCAGTTTCTTCATTCCAAGAATATTAGGATATTCAAAC TGGGGAG
GACTTGACATCCTAATGTTCTCTCCATTTTACAGCAGAGTAAACTGAGGGCAGATGGT
TGGTAAGTGGTTGAGCCTGTGCCAGGGTTTGGGCCGTTACTTCTACTGAACCAACCCCT
CCGGGGGGATTCAAAATGAATTGCCATATGCTTAAC TTTTAAATTAATGGAACCTAAC
ATTTCTGAAC TTTATCTCTAAGATGATGGAACAATGCCTGCCTCACAAAATTGTTCTGA
[C, T]
AAGTTAAATAAAATATAGTGGCCAGGATAGAGAAGACGTTAGAGCATTTATTTGCATTT
CCTACATACCTCAGAGGTATCATTTCTATTTTCCCTCCTAGAGGAAGAGGTTTCCATTT
TGAAAGTGGCAACTCCAAAGACATGAATCTCTAATAGTAGAATTCAGGTATTAGTGAGA
AGTGGTGACAGAAATAAAACCTATAGTGTCCATATCAGAAACCTATGATGAAGTGGTT
TATTTTATAGAAAAATGTGAGAAAAGTAAATCATTTCCCTTGCCTGCTGAATTTTTTGT
- 11429 TGCCTGCTGAATTTTTTGT TTTTAAAAATAATTTTAAATCATTCCTATCCTTTAATAATATT
TGAAAAATACAAATTAAGCAAAAGGAAAGTAAAGAAATCACCACAAATCACACTACTAAAA
ATAAGCTCTATTGAAC TTTGATGTCTAACTTTCAAAATAGTATTATTTCTACTAGAATAG
AAATTGGTGAAATCCTTTTATCTCTTTTATTTTAAAGTGACATAAATATTGTTCAAAT
TGTGTTTTTTCAGATGAATTATCATCAACTGTGGGTGAGTTGATCAAAAAAC TTTTATTAG
[C, T]
GTTTCTTTTCAAATTTTGTCAATAAAATGTGAAACCATTGAGTATTCCTTTCTCGGTAA
GAAATCATACTTTTTCATTAAAGATTACTATTTTATGTACTTATATATAAGTAAATTTCA
TGT TAAAAATGTAAAGAAAAAGAAAGAGTCCAGCAC TTTGGGAGGCCGAGGCAGGG
GGATCAGGAGGTCAGGAGATCGAGACCATCCTGGCTAACAAAGGTGAAACCCCGTCTCTAC
TAAAAAACTACAAAAAATTAAC TGGGCATGGTGGCGGGCGCTGTAGTCCAGCTACTCC
- 11841 TAAATTTTCATGTTAAAAATGTTAAGAAAAAAGAAAGAGTCCAGCAC TTTGGGAGGCC
GAGGCAGGGGGATCACGAGGTCAGGAGATCGAGACCATCCTGGCTAACAAAGGTGAAACCC
CGTCTCTACTAAAAAACTACAAAAAATTAAC TGGGCATGGTGGCGGGCGCTGTAGTCCC
AGCTACTCCCGAGGCTGAGGCAAGAGAATGGCGTGAACCTAGGAGGCGGAGCTTGCAGTG
AGCCGAGATCACGCCACTGCAC TCCAGCCTGGGTGACAGTCTCAAAAAAAAAAAAAAAAAA
[G, A]
AAAGAAAGAAGAAGAGGAGATAAATCATAATTTTACTGTTATTTTACTTTTGGGTGATTA
AACATGTCAATTACATGTAAC TTCATAATGTGATTTTGTAGCAAAAATCAAGTATTCTGGAA
AAGAATGAAATGAAAAATGCCTTTTGCACCTAAAAATGAAGTACAAATTTTCTGTATC
TATTTATATACTCATCTTTATTAATGCCTACTTGATATTATAAGCCAATTTTCATTTTAA
AATTCCTCAGGGATGAATATTTTAAATTAATTATTTATTTTAAATCTAATACTAATCA
- 11842 AAATTTTCATGTTAAAAATGTTAAGAAAAAAGAAAGAGTCCAGCAC TTTGGGAGGCCG
AGGCAGGGGGATCACGAGGTCAGGAGATCGAGACCATCCTGGCTAACAAAGGTGAAACCC
GTCTCTACTAAAAAACTACAAAAAATTAAC TGGGCATGGTGGCGGGCGCTGTAGTCCCA
GCTACTCCCGAGGCTGAGGCAAGAGAATGGCGTGAACCTAGGAGGCGGAGCTTGCAGTGA
GCCGAGATCACGCCACTGCAC TCCAGCCTGGGTGACAGTCTCAAAAAAAAAAAAAAAAAAAG
[-, A]
AAGAAAGAAGAAGAGGAGATAAATCATAATTTTACTGTTATTTTACTTTTGGGTGATTAA
ACATGTCAATTACATGTAAC TTCATAATGTGATTTTGTAGCAAAAATCAAGTATTCTGGAAA
AGAATGAAATAGAAAAATGCCTTTTGCACCTAAAAATGAAGTACAAATTTTCTGTATCT
ATTTATATACTCATCTTTATTAATGCCTACTTGATATTATAAGCCAATTTTCATTTTAAA
ATTCCCTCAGGGATGAATATTTTAAATTAATTATTTATTTTAAATCTAATACTAATCAT
- 13973 GTGGGAAGACTAATTGCCATTTTCTTTGACAAAGCAATATTCATCAATGTTAAAAATG
TATATACATTGTTGCTTGTCAACCTTGCTTTGGGGATTCTATAGAAACAAATGCTTCAAT
ACATAAGGACTTAAATATGTGAAATGTTTACTACAGCTGTATTTACAGTAGGAAAAATCT
GAATGTGGAAATCTGAATAAATGTTGTTGAATCCATATCACAGAACATGAATCAGCAT
AGAAAGAAGTTAAGCCGAGCTCAGTGGCTCATGACTGTAATCCAGCATTTTGGGAAGC
[T, C]
GAGGCAGGAAGATGCTTGAGCTCAGGAGTTTGAGACCACCTGGCCAACATAGTGAGATC
TCATCTCTACAAAAAATTTAAAAATTAAC TGGGCATGGTAGCACGCACCTGCAGTCCAG
CTACTCCAGAGGTTGAGGTGGGAGGATTGCTTGAGCCAGGAGGATAAGGCTGCAGTGAA
CCATGAAGAAGTTAGGGCTACACAATCACTTTAAGAAGATTCTGTGATGTGTTAAGTCTG
GAAACTCAAGTATAAAATATGATTCATTTGAAGGTAGCAAAAAATATGTATGTGTGC
- 13983 CTAATTGCCATTTTCTTTGACAAAGCAATATTCATCAATGTTAAAAATGTATATACATT
GTGTCTTGTCAACCTTGCTTTGGGGATTCTATAGAAACAAATGCTTCAATACATAAGGAC
TTAAATATGTGAAATGTTTACTACAGCTGTATTTACAGTAGGAAAAATCTGAATGTGGAA

FIGURE 3J

ATACTGAATAAATTGTGTTGAATCCATATCACAGAACATGAATCAGCATTAGAAAGAAGT
TAAGGCCGGACTCAGTGGCTCATGACTGTAATCCCAGCATTTTGGAGCCGAGGCAGGA
[A, G]
GATGCTTGAGCTCAGGAGTTTGAGACCACCCTGGCCAACATAGTGAGATCTCATCTCTAC
AAAAAATTAAAAATTAACTGGGCATGGTAGCACGCACCTGCAGTCCAGCTACTCCAGA
GGTTGAGGTGGGAGGATTGCTTGAGCCAGGAGGATAAGGCTGCAGTGAACCATGAAGAA
GTTAGGGCTACACAATCATTAAAGAAGATTCTGTGATGTGTTAAGTCGGAACTCAAG
TATAAAATATGATTCATTTGAAGGTAGCAAAAAATTATGTATGTGTGCCTGTATGTGT
14542 TTGAAGGGTAGCAAAAAATTATGTATGTGTGCCTGTATGTGTGTATGCACCTATAACTAT
ATATATATATATATATATATATATATCTGTATTAACCATACAGCATATATACAAAACCA
TATAACATATATAAGTATATATATAAACCATACAACACAGATATACCCAGGTGACACAC
ATACACACATGTATAGTTGCATGGGAGAGGGAGAGAAAGAGAGATGTGGAAGAGGTAA
AAATAAGAGATCCAGGGGCAGGGGGGAAGAATTGTTGGGTAGATAAATGCCACATTTAT
[G, A]
TTTCTACGCTTAGGAGAGATTCTGAAACTGGGGCAAATGCTAAGCAAAATTTTGAGACT
GTTTATTAACCTGGAAGAAGTAGGGCAGGGGCAGGAATATAAAACAAAAGAGCTTTGTCTC
ATTAGAGAGCCAAAGCTCTGCTATTGAGTAGCACAATAACATTTACTACAGCCACAGCCA
ATGACTCTGATTTCAAAGGGCAACAGCGCTTAGGAAGACAGTCGCTTGATAACCTCTCTC
CTAACACACCCAGAAACAGAGGAAGAATTTTGTTCCTCTAGATGGGAGTGTGTCTTTC
15732 AGAAAGGTTTCTTACTCTGCTTCTCCTCCTCCTCCTGCAGATTTTGCAAACTGTGTA
TTTCCCTTCTGGGTACAGACGCTTGATCTACTGGGAGTGACTGATGATGGGGAAGCATT
GGGAAAAAATGGTGTTCAGTACCAAGAATTTTAACAAGGACCGAATTTGGAATACTGT
GAATGATGGTGAGATTTACCAGGGGAGGGGATGAAGACGTTAATCATTTCTCCTTGGA
AGAACCTAGAGTTTGTGTTGGCAGAGGAGAACCTTTCTCATTAGAGAGGGTTTTTTTTT
[G, T]
TTTTTTTTTTTTTCTGGGATGTGAGTTGAAGGTGATCTAATAAATAACTAGAGGTT
TTCGATTGAGGATTATGGGGTTTAAATATGGAATCTGTAGTTTAGCCTCAGCACTTTAA
TTGCTAAGTGTGTAAAAATTAGTCCCTTTAAAAAATTAGTCTATGGGGTAGGCGTGGTGG
CTTATGCCCATATCCAGCGCTTTGGGAGGCCAAAAGGAAGGACTGCTTGAGGCCAGG
AGTTCAAGACCAGCCTGACCAACATAGCAAGACCCGTTTCTACAAAAGAAAAAATTTA
15813 TTGATCTACTGGGAGTGACTGATGATGGGGAAGCATTGGGAAAAAATGGTGTTCAGTG
ACCAAGAATTTTAAACAAGGACCGAATTTGGAATACTGTGAATGATGGTGAGATTTACCA
GGGGAGGGGGATGAAGACGTTTAAATCATTTCTCCTTGGAAGAACCTAGAGTTTGTGTTGG
CAGAGGAGAACCCTTCTCATTAGAGAGGGTTTTTTTTTGTTTTTTTTTTCTGGGA
TGTGAGTTGAAGTGATCTAATATAAATAACTAGAGGTTTTCGATTGAGGATTATGGGG
[-, A, T]
TTTAATATGGAATCTGTAGTTTAGCCTCAGCACTTTAATTGCTAAGTGTGTAAATAG
TCCCTTTTAAAAAATTAGTCTATGGGGTAGGCGTGGTGGCTTATGCCATAATCCAGCG
CTTTGGGAGGCCAAAAAGGAAGGACTGCTTGAGGCCAGGAGTTCAAGACCAGCCTGGACC
AACATAGCAAGACCCCGTTTCTACAAAAGAAAAAATTAAAAATGTAAGTGGCATGGTG
GCACACACTTGTAGTCTCAGCTACTCGGGAAGCTGAGGCAGGAGAAATCACCTGAGTTCAA
16132 AGTTTAGCCTCAGCACTTTAATTGCTAAGTGTGTAAATAGTCCCTTTTAAAAAATTAAAG
TCTATGGGGTAGGCGTGGTGGCTTATGCCATAATCCAGCGCTTTGGGAGGCCAAAAAG
GAAGGACTGCTTGAGGCCAGGAGTTCAAGACCAGCCTGGACCAACATAGCAAGACCCGT
TTCTACAAAAGAAAAAATTAAAAATGTAAGTGGCATGGTGGCACACACTTGTAGTCTC
AGCTACTCGGGAAGCTGAGGCAGGAGAAATCACCTGAGTTCAAGAGTTCAAGGTTGCATTG
[-, A]
GTTGTGATGAAAGAGAGACCCACCTCTAAAAATAAAATAAATGAAGTAAATAAGCTTATG
GATTTTGGGTGATAGTGATGTGCAATGTAGATTCAATCAATTGCAACAAATGTTCCACTC
TGGTGGAGGATATTGATAACACAGTGGCTATGTGTACATGGAATTTTATACCTACTAC
TCAACTTTTCTGTGACTTAAACTGCTCTAAAAATAACGTGTATTAAAAATGCATTACC
TTGGCCGGGCACGGTGGCTCACGCCTGTAATCCAGCACTTTGGGAGGTCGAGGCGGGTG
16317 CAAAAGAAAAAATTAAAAATGTAAGTGGCATGGTGGCACACACTTGTAGTCTCAGCTA
CTCGGGAAGCTGAGGCAGGAGAAATCACCTGAGTTCAAGAGTTCAAGGTTGCATTGAGTTG
TGATGAAAGAGAGACCCACCTCTAAAAATAAAATAAATGAAGTAAATAAGCTTATGGATT
TTGGGTGATAGTGATGTGCAATGTAGATTCAATCAATTGCAACAAATGTTCCACTCTGGT
GGAGGATATTGATAACACAGTGGCTATGTGTACATGGAATTTTATACCTACTACTCAA
[C, T]
TTTTCTGTGACTTAAACTGCTCTAAAAATAACGTGTATTAAAAATGCATTACCTTGGC
CGGGCACGGTGGCTCACGCCTGTAATCCAGCACTTTGGGAGGTCGAGGCGGGTGGATCA
CGAGGTGAGGAGATCGAGACCGTCTGGCTAACATGGTGAACCCCGTCTCTCTAAAAA
TACAAAAAATTAGCCGGCGTGGTGGCGGGTGCCTGTAGTCCAGCTACTCAGGAGGCT

FIGURE 3K

SEQUENCE LISTING

<110> PE CORPORATION (NY)

<120> ISOLATED HUMAN SECRETED PROTEINS,
NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
USES THEREOF

<130> CL001230PCT

<140> TO BE ASSIGNED

<141> 2002-04-05

<150> 60/286,382

<151> 2001-04-26

<160> 8

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 402

<212> DNA

<213> Homo sapiens

<400> 1

```

atggcccggc acatggggct cctgctgggt tgggtctgcc tgattcgcgg tgggtcggt 60
ggtgtgggtcg gcgtgtatt taacgttctg gaagaagaaa ctataactca tttccagaa 120
gttacagatg gggagtgtgt ctttccattc cactataaaa atggaacata ttatgactgc 180
atcaagtcca aggcaagaca caagtgggtgc tcgttaaaca agacctacga aggatactgg 240
aagttttgca gtgcagaaga ttttgcaaac tgtgtatttc ctttctggta cagacgcttg 300
atctactggg agtgtactga tgatggggaa gcatttggga aaaaatgggtg ttcactgacc 360
aagaatttta acaaggaccg aatttggaaa tactgtgaat ga 402

```

<210> 2

<211> 133

<212> PRT

<213> Homo sapiens

<400> 2

```

Met Ala Arg His Met Gly Leu Leu Leu Val Trp Val Cys Leu Ile Arg
 1          5          10          15
Gly Val Val Gly Val Val Gly Ala Val Phe Asn Val Leu Glu Glu
 20          25          30
Glu Thr Ile Thr His Phe Pro Glu Val Thr Asp Gly Glu Cys Val Phe
 35          40          45
Pro Phe His Tyr Lys Asn Gly Thr Tyr Tyr Asp Cys Ile Lys Ser Lys
 50          55          60
Ala Arg His Lys Trp Cys Ser Leu Asn Lys Thr Tyr Glu Gly Tyr Trp
 65          70          75          80
Lys Phe Cys Ser Ala Glu Asp Phe Ala Asn Cys Val Phe Pro Phe Trp
 85          90          95
Tyr Arg Arg Leu Ile Tyr Trp Glu Cys Thr Asp Asp Gly Glu Ala Phe
100          105          110
Gly Lys Lys Trp Cys Ser Leu Thr Lys Asn Phe Asn Lys Asp Arg Ile
115          120          125
Trp Lys Tyr Cys Glu
130

```

<210> 3

<211> 16556

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)...(16556)
 <223> n = A,T,C or G

<400> 3

```

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 960
aaagacagga gggacaggga ggggaagtgg ttgtggttag ttccaggggc tataaaatct 1020
tggaacatcag ctggctcgacc acgagtgtct gacaggagca tggcccgcca catggggctc 1080
ctgtctggtt gggctctgcct gattctgcgt gtggctcggt gtgtggtcgg cgctgtattt 1140
aacgttcttg aagaaggtga ggagacaggt gaccctggct gggacggagt cctgcacatc 1200
ggctgtctga tgaatcaaaa tcccagtcgc catggagggc aacagctctc cttgctttag 1260
gtaccttaag aaagaaaaag atactctccc ttccatgtca gaaacacaat ccataggtaa 1320
agatttttta aaaaataaga aagtctcggt aataaaataa tgaaagacta agaattgtaa 1380
ttggaatac ctagtacat gttacttgaa aaccatgtgt ctctccctt catgctagga 1440
gatttttaaat gcctcctgca gaggttttga aaatgagact aattttcata agtcgattt 1500
tattcctttt ccaaagtcgt gttgcctgta acaattatga agcgaatag aaaaataaga 1560
aggaattca ctatggatgc atctgacaac gtttgatgag gactaagcat agtttcaaaa 1620
acaattttt atggaatgac caagaagaat gatcttgca atatttctt taagtattct 1680
tcaggacatc ttgcattaa aaaaaataa cactgcttgc tattttttt taagaacag 1740
aaaagaaaat atactccaa tgtaactatg accactgctt tcttatccac tagaattatt 1800
ttatacttac tgaaaaataa cactgtactt ggccacaggt ttttttttta aatcagacat 1860
catgtttgta gatacgtaaa aaggaatata gataagtga aaaaggcaa ttaagacgat 1920
cctaaagctt ctccgctgtc aatgtctggt ctccataatt atggatgtca tatttttcca 1980
cataatactt gtgtatttca gagccttact gatgaagcat ttctcttaat aatgcatcag 2040
agaataaaaa ctcatgggc ccgggctcgt tagcctgtat tttaaatgat gtggtgctag 2100
tctacttttg atctgcagaa ctgttgcta ttactatct cctctctct ccttgcag 2160
catctgcctc ctggacttat tgtctctttt ttttttttt ttttttttta tggagtcttg 2220
ctctgtcacc cagactggag tgcagtggca tgatctcggc tcaccgcaac ctccagctcc 2280
cagattcgag caattctcct gcctcagcct ccgagtagc tgggactaaa ggcacgtgcc 2340
atcatgccca gctaattttt gtatttttag tagagaccag gttccaccat gttggccagg 2400
ctggtctcga actcctgac ttgtgtgatt caccacctc agcctccaa agtgctggga 2460
ttataggcag gagccaccgc acacagccta ttgtctctta agagtctt ttgagctact 2520
gagtagcaaa ctgcaagttt ttgatgctgg tcgaattcaa tacagacaat aacagctggg 2580
tcccacagc aatccggctt gcagataata agatgtcgtc acaataagc ggaaggtggg 2640
aactcataga catctagggc tcaaagtcct gctggaggcc ataggctaag cttcgttggg 2700
accagagat gctggaggga aatctggact cctgagcttg ccaggagcag tcaagtggga 2760
cagaggaggt gaaggagtc ccagtttttg atgagtggg atgtatgaga gctctcagct 2820
atggttcctc ctcatcctca tcagaggagc actacatttt ggtgttttag aagcagtgc 2880
agactgggtg cgtggctca cacctgtaat cccagcact ttgggaggcc gaggcggcg 2940
gatcacaagg tcaggagttc aagaccagcc tggccaacat ggtgaaacc cgtctctact 3000
aaaaatacaa caattagccg ggcattggtg caggtgccta taatccagc tactcaggag 3060
gctgaggcag gagaatcact tgaaccaag aggcggaggt tgcagtgacc tgagatcatg 3120
ccattgcact ccagcctggg caacaagagc aagactccat ctcaaaataa taataataat 3180
aataataata ataataata taataataat aataaaacgg tgacttgata ggtttgctgg 3240
atgcacagat gcagagacta aatagcatga gcatctccac tcatgagcat gtgatatgca 3300
ccccacctc ttccactag tgcacgatct agtgggaaga cagaagagga aacagaccat 3360
gtgtatggca catgctaaat cagctatacg cctagcctga aaaagaaaa ataattgaat 3420
tggcaagact tcacaaagga cttgattgag tgatcttgtt ccagggctag gaatttgcca 3480
tgtccaacat gtaattacta ttccaacttg atgggttgcc taagggatta catgggagtg 3540
gggtcccttg cggggagaac tgtccaggtg tgtagaccg gacatgcttc gtgccaccag 3600
aactctgggg ttttcacttt tgccagttca ggctttggtg cggtaggact gccaccaac 3660
cgatcgtttg atgtcaacaa acgtttgtgc taaaatgtga atcttgcat gcaagagcca 3720
ttattattta aagaggatga ggaggaacca tggctaagag ctctcatgca cccctactca 3780
ccaagaactg aaggctcctc tcgctcctc tgtccactc aactctgctc ctggcaagct 3840

```

```

caggagacca gttttccctc cagcatctcc gggttccaac caagcttagc ctgtggcctc 3900
aagcaagact ttgagtcctg gatgtctgtg agttcccaca tctgaaacag gaggatcgtg 3960
attcttgctg actcccaggg ttgctggaag attaaatggg ttagcatgcg tgacgctcag 4020
aacagaactg ttatcattaa gtcattatta ctactattac agtaccaaaa gtactattac 4080
agtattaaag aggcaggata aggggttgga tttagtagtc gatactcaca gagtgaagttt 4140
aactacttta aacttctcat acaagtggaa tctgtgaata tttgtccttc tgtgaatggt 4200
ttggttcact tagcataatg tcctccaggt tcattccatg tactgaaaac ggcaggattt 4260
tctctttttt taatttttaa aatttttctg attttgtaat ttttattttt attttatttt 4320
atltgttttt gagacaaggg ctactctgt caccagggt ggagtgcagt ggtgtgatct 4380
cagcacactg caacctccgc ctcccgggac tgaagtaatt aatcctccaa ctgagctcc 4440
cagtacggga accacaggag cacaccacca tgcccagcta atttttttgt aattttggta 4500
gagataagat ttcgccaaagt tgcccaggca ggtctcaaac tcctgagctc aagcgatctg 4560
cccactccg ccttccaaag tgctgtgagc caccacacc agcctgcct aaaggttctt 4620
ttggttccat tgttttgacc taagggcagg atccaatagta aagacaagct ctggccattg 4680
atagttcaag gtgggaggat tgcttgagcc caggaattcg agaccagcct gggaaacata 4740
gtgggacctg gtctctacag caacttttga aaattagcca gctatggtag tgtgtgcctg 4800
tagttccagc tacttgggag gctgaggcaa gagaatcact tgagcctggg aagtcaagcc 4860
tgcaagtgcg tatgacagcg ccactgtact ccaacctggg tgacagagtg agattctgtc 4920
tccaaaaaaa aaaaaaaaaa gtaagtacta agtacatgta gtacgtgttc ttacacaaag 4980
aacagtagat aaaatagaaa ttgtagggac attccgagga cggggttaat cagaaaaccaa 5040
cacggcgaat tagcacccaa gacagagggtg ttttagcctc cacaaactgg attgtctagt 5100
gacactcaaa ttcttctgtc tcactctcac tgtctggggg tgcatgtcat agtgtccttc 5160
cagaattcat gtttaccaga gatctgtgaa tatgactgta tttgaaatag gtccttgtag 5220
atgtcggtaa ttaagatgta gtcacaatgg atacgggttg gccctaatacc aatgactggt 5280
atctttataa gaagaggaaa cagacacaca cagggaagac aagcacgcag agtcacaaga 5340
aaagattaaa gtgatgctgc cacaagtcac agaattgccac agcttgccag cagccaccaa 5400
aagcaaggag agaggcatgg aatagtttct gcctctgagc tcctaagaag gaaccaaccc 5460
tggtattttg gccttctggc ctccagatct gtgagacaat aaatgtctat tgctctagac 5520
caaacagtct atggtacttt gtctagattt tttttatttt tattttattt tattttattt 5580
ttagataggg tctcattctg tcacccaggc tggagtgcag tggcacatc acggctcact 5640
gcagtcttga cctcccaggc tcaagtgtac ctcttgctc agcctccaa gtagctggga 5700
ctacaggtgc acgccacat gccagctaa tttttgtgtc tttatttatt gcagagatgg 5760
ggtcttgcta tgttgcccag gctggtctcg aactcctggg ttcaagcagt cctcctggct 5820
cagcctccaa aagtgtggg attacaagcg taaatcacca tgcctgacca gtaaaattct 5880
acagacaaaa accaaggaag aagcaatcac agaggccacg cggctctcct gtctggggga 5940
tgagcctccc atggttttgt acccaaggac cccaactcct cccattcat tgttatttgt 6000
cttcaccagt gctctccag gccttctgac attctctaga ttctgaagt ttgtcttcta 6060
tcttgctgtg tattccattg aacatagaat ggaacagcaa gtgcagcact ttgggaagac 6120
tcctgaccca gctctgaac ccaacaaga atcttgcca agttaggatc gtaggactac 6180
catgtagcag gctacagcaa ttcccacga tagggaggcc actgcagctt tttttcccta 6240
gtctgttoga ccagcttccc cagactcctc cctcctctgg ggetccgtag cttggaagaa 6300
taaagtctga taaactatat cctcctgcgg gcaactgcaa gaatgccaat atatgtgtt 6360
gacgatgctg aaattcactg ctcatcctct ggaaaagtga tcgggttcag gggacttcat 6420
aagctcctga acacatttca catgctttta tttttttac attcacttat tcttttctt 6480
tttttttttt ctagacaaaag tcttgccctg tcaactcaggc tggagtgtag tgggtgtgatc 6540
tcagctcact gcaacctcca cgtcccagggt tcaagtgtat ctctgcctc agctgcctca 6600
gtagctggga ttataggcgc acaccacat gccagctaa tttttgtatt tttagttag 6660
acagggtttc accatgttg ccaggtggt ctcgaactcc tgacctcagg tgatctgccc 6720
acctcggcct cccatagtgcc tgggattaca ggtgtgagcc accatgccg gcctacttac 6780
tctttttcta atgtgcttcc tctctttata tctctgaaga acttttaaca tttcttcaa 6840
ggcaggtttt ctggtagcaa attccctcaa tgtttgtttg tctgagaatc tttgtctttc 6900
tttctccttc acttctgaag aataattttg cagggtacag aattctagat tgaagatttt 6960
ttttctctc aacactgtaa atgtctcatt cttatcttca tggtttctgg aaagaagttg 7020
gacataatc ttatctttgc ttctctgtag gtaaggatt tttccccctc tggcttcttt 7080
caaaagattt ttctttttaa aaattttatt ttagatgcag gaggtacatg tgcagatttg 7140
ttacatgagt atgttgcata aatggtgagg tttgggattc aattgaagcc attacccaaa 7200
taatgaacat agtgccaggt aggcagtttt ctttttcttt ttcttttttt ttttttttt 7260
cagacagagt cttgtattgc ccagactgga gtgcagtggg gcgatcttgg ctactgcaa 7320
cctctgcctc ccaggttcaa gcgattctcc tgccctagcc tcccagtag ctgggattac 7380
aggcgccac caccatgtcc ggctaatttt tagtagagac gggatttcat catgttggcc 7440
aggctggct tgaactcctg acctcatgat cagtccactc cagcttccca aagtgtcggg 7500
attacaggca tgagcactg caccggcca tcaataggta gtttccaagc ctgggtcccc 7560
tccttacctc cccacttttg taatccccag tgccatttgt tccatcttt gtgtccctgt 7620
gcacccaatg tttagttccc acttataagg gagaacatgt ggtatttggg tttctgtttc 7680
tgcactaatt tgcttaggct aatggcctcc agctgcatcc atgttgctaa aaaggacatg 7740
atlttgttct ttgttatggc tgcaataaac aatttcgttt tgggtttaac tgaaaaaatg 7800
agactcattt ttatttcaaa taagggtatt tttaaaacct tatatattat aacagaattt 7860
agcttgaatt atgacaaaag gtgaattctt gagaatgaaa agaattggga tgcaggactt 7920

```

```

tcctgacaaa attgttacga ttttccagtg catagctttg tcttcttcca ttctacactt 7980
tcctcaaagg gtcataaaga aagccagttt tgtataagta ttctaagact aattctagga 8040
ctatattaat ttttaggggc ccttcttaat cctttgaatt taaaaatata taatatatat 8100
atttatattt tatgtataat atacatatta tatatattta tattttatat atatatatat 8160
tttttttctt caagccaggc atggttggtt acacctgtaa tcccagcact ttgggaggcc 8220
gaggtgggtg gatcacctga ggtcaggagt ttgggaccag cctgaccaac atggcaaac 8280
cctgtctcta ctaaaaatac aagaattagc caagtaaaaa tacagtggtt ggtgcacacc 8340
tgtagtccca gctacttggg aggcagaggc agaagactcc ctggaacca ggagcgagag 8400
gttgagtgta gccgagatcg tgccactaca cgccagcctg ggtgacagag tgagactccg 8460
tgtcaaaaaa aaaaaaaaaa ttctcttcaa agcccttatg gagtcacaat atgtgctctt 8520
tttaactttg ttgtatcact gtcattaaat tcatacattc ttgcatcttt cacaagtgta 8580
ccggttcat tttgtagtcc acttgctttg cagttgcatt gtacttgcaa ggtatttttt 8640
gaacataaac atcccaacat tcacaacaaa tgatccacca ggtggatggc agaatttctg 8700
gtgttaatac agtggggacg tttagtggtg gactcataat tttatagcac cgagtaggtt 8760
gcactgtgtt tatgatattg ggatgttgca ttctgatgct tacctggtga tacctgttct 8820
aagatcatag tgcttagcag acaggctatg ttggaatcgc tcgtctgtat ctccccattc 8880
atgaaatcag gcctatcatg atcctcaaaa ccaaaatcca tttcctaata gtttactat 8940
ggaatgcata ctgaatgaac ccaacaaatc caggaagtag aaacaattat aatccccata 9000
tttcagagaa agacaagctt aactcctgat gagtccattg atcttggtt cccacctgaa 9060
tctcatgggt tctgtagtcc agagtctacc atttcagcca ctcccttttt ggtagccca 9120
gaacctagca ctaggctggg cacatagtag gcattcaatc catatttgtt gtctacatga 9180
gtgggttgat ggactggctg aataaatgca gatcccttca tcgtgagatc ttccctatat 9240
actggaacca gtgcacagtt cagtggcatt caatgcattc aaattgtgca atctatctcc 9300
agagcctttt cattgtccca aactgaaact ctgcacccat taaataataa ctccccattc 9360
ctccctccct ccagctccta acaaccacca ttctgctttt tgtctctatg tatttgacta 9420
ttctatggtt ctcatataag tagaatcata taatatttgt ctttttgtgt ctggcttatt 9480
ttacttaacc tagtattttc aaagtgcac catgtggcgg catctattca aattttgttc 9540
ctttttaatg ctgaataata ctccattgta tgtgtctacc atattttgtt tatttactca 9600
tttgctcagc gacatgtggg ttctttccac cttttggcta ttatgaataa tgctaccatg 9660
aacatgggtt tacaaatatt tgtttgagtc tctgctttca attattttgt atatatacc 9720
agaagtggaa gggatctggg tttttaaaaa ctgagctggt gattctaata tgcaacaaag 9780
cttgtgaacc attgtgctgg tccattatga ctatttctta aataagatgt gcctaaggaa 9840
aaacgtttta atttcaagcc acaatggcag ccaggacaat ttagctgaaa aacaaacttg 9900
tggaatgaat aaacacgagc aaatagcaaa atcgtcagta ataaatgaaa tatacacaac 9960
aaattcagaa ttaaaatcat tgcttttgca aaaatctttt caaaaaataa aaaattaagt
10020
cattgctctt tcaagtgtaa tccacaaatt tgatgaacta gttatgaaca cagaatatgt
10080
cgagttataa tgctttaata tgagagatat ctgctaatta cattatagta cgtggatgtt
10140
ctcttaacca aacatgacat ttgtaaattt acttgttggt aaattgtgct tttctatctg
10200
acatcatctt aatggttatt aagcatcatc tattacgggc tatttgggaa aaaatcaaat
10260
ttttcactct atattttatt tacaattcat tctatgttgt atttcacct aaagtatga
10320
ctgttcttaa agcctcaatt aaatgaactt tgtaacagtt atgtaatat taactaaaac
10380
atgctgggtt taaatcaaat agcctttgcc taaaatcact tcatcaggaa ccatgcaaag
10440
taacaaaaca aaaaattcaa gttctaagga gggggaggta aatattacag gttgaatggc
10500
tcatgaagga acaatgggtt aaaaatgctg ctttaagaaa aaaaaaaagt aaagctatca
10560
ggcagtttcc ttcattccaa gaatattagg atattcaaac tggggaggac ttgacatctt
10620
aatgttctct ctccatttta cagcagagta aactgagggc agatggttgg taagtgggtg
10680
agcctgtgcc agggtttggg ccgttacttc ctactgaacc aaccctccg gggggattca
10740
aattgaattg ccatatgctt aactttttta atttaatgga acctaacatt tctgaacttt
10800
atctctaaga tgatggaaac aatgcctgcc tcacaaaatt gttctgacaa gttaaataaa
10860
atatagtgcc caggatagag aagacgttca gagcatttat ttgcatttcc tacatacctc
10920
agaggtatca ttcttatttt tccctcctag aggaagaggt ttccattttg aaagtggcaa
10980

```

ctccaaagac atgaatctct aatagtagaa ttcaggtatt cagtgagaag tggtagacaga
 11040
 aataaaaacc tatagtgttc catatcagaa acctatgatg aagtgggttta ttttatagaa
 11100
 aaatgtgaga aaagtaaadc atttcctttg cctgctgaat tttttgtttt aaaaaataatt
 11160
 ttaatcattc ctatccttta ataataattg aaaaatacaa ttaaagcaaa aggaaagtaa
 11220
 gaaatcacc cacaatcacac tactaaaaat aagctctatt gaactttgat gtctaacttt
 11280
 caaaatagta ttattctttac tagaatagaa attggtgaaa tcctttcatc tcttttattt
 11340
 taaagtgaca taaatatttg ttcaaaattg tgtttttcag atgaattatc atcaactgtg
 11400
 ggtgagttga tcaaaaaact tttattagcg tttcttttca aatttttgca ataaaatgtg
 11460
 aaaccattga gtattccctt tcctggtgaa aaatcatact tttcattaaa gattactatt
 11520
 ttatgtactt atatataaag taaatttcat gttaaaaatg ttaagaaaaa agaagaagag
 11580
 tcccagcact ttgggaggcc gaggcagggg gatcacgagg tcaggagatc gagaccatcc
 11640
 tggctaacaa ggtgaaaccc cgtctctact aaaaaactac aaaaaattaa ctgggcatgg
 11700
 tggcgggccc ctgtagtccc agctactccc gaggctgagg caagagaatg gcgtgaacct
 11760
 aggaggcgga gcttgcagtg agccgagatc acgccactgc actccagcct gggtagacgt
 11820
 ctcaaaaaaa aaaaaaaaaa gaaagaaaga agaagaggag ataaatcata attttactgt
 11880
 tattttactt ttgggtgatt aaacatgtca ttacatgtaa cttcataatg tgatttttag
 11940
 caaaaatcaa gtattctgga aaagaatgaa atagaaaatg ctttttgcca cctaaaaatg
 12000
 aagtacaaat ttttctgtat ctatttataat actcatcttt attaatgcct acttgatatt
 12060
 ataaagccaa tttcatttta aaattccctc agggatgaat attttttaat taattattat
 12120
 tttttaatct aataactaatc atgctgaaat aaacattctt aatctaagtg tcataccaat
 12180
 taaaactagg aaagcacagg ggtgttacat taacactatt ttaggttcta acataagtga
 12240
 tattgtagaa aggacaaact ctatataaaa taacaggcaa aattatttta tgattgtata
 12300
 ccttaaatcc ctaagaacat gaataaaaat tattagtttt aataaagtaa tatggctatg
 12360
 tggctggatg caagataaac atacaacagt caatagtttt tctctatact gataagaacc
 12420
 aactggacat tgatggaaat ggaaaacatt ctgtttaaca tcagtgaacca aaaaaagca
 12480
 ctttggattt aaatttgtga gaaatgtgta ggatttatat agaataaata agaaaatcat
 12540
 attgaagagc aaaaaaatc attttgaaaa agagcagatg ccatgttctt ggatggaaaa
 12600
 cataatgcca taacaatatc actttttcct aaactaacat ttttatttaa ttttatttta
 12660
 cgtaagagcc catttttaaag cctaataata gtgaaataaa tgtcagctct aaattttatt
 12720
 tttcagaaac tataactcat tttccagaag ttacaggtaa gtcgatcaga aaactttgat
 12780
 tattaatgta tattttcttt taattttgta aaagtgagga aatcagtaat gttttccag
 12840
 taaaaaaaaa tcattttttc tccttatata aaatgtaaca taagtgcctc ataaacagat
 12900
 aaagaaaaca gaggtataga agtcactctg aactttcatc agcactaaat gtgtctactc
 12960
 ttttaattct gccaatctaa caggaaataa tagattttta aatttatata gaaaaaacat
 13020

ccaataataa ttagaagtat gaacaaagtg ctgtagtaga gaaatatttt gtctcatcag
 13080
 atatgtagac atactaaaat attgtaatat aagcaacatg gaatcagcct agaaaaatgg
 13140
 aagaggagga ggaggaggga ggaaggaaga agaagaggaa gaagaagaag gaggaggaga
 13200
 aggaagagga gagaagaaaa gaagaaatat tttgtgatgc ttgctggctc cagagcctgc
 13260
 tcattttacc accatgcttc cttacctctc taataatttc aagaaacttt atagaaaata
 13320
 tagaatggat cccaggatat attagaataa tccatggttt taaacatgat tgtttaaatg
 13380
 tacaggatgat ttatttcac atcgatctgt cattactaaa tatgtttgaa agaaaaataa
 13440
 gttgcattgc cctttggcat catatacaaa actaaattct agcttaatca aattttaaat
 13500
 gtaaaaggaa caaaattgtt taatttaaga agaaaaatgga aaatatatgt atggcctaaa
 13560
 gaccagtaga tcttttaaag caaggcagaa atcttctgta atacaggtaa cacattttta
 13620
 aagtgtgtaa atgcttcgtg ctaatgaaaa ttcaggtaaa ggagcacaat tgggtgggaa
 13680
 gactaattgc cattttcctt tgacaaagca atattcatca atgttaaaat tgtatataca
 13740
 ttgtgtcctt gcaaccttgc tttggggatt ctatagaaac aaatgcttca atacataagg
 13800
 acttaaatat gtgaaatgtt tactacagct gtatttacag taggaaaaat ctgaatgtgg
 13860
 aaatactgaa taaatttgtt tgaatccata tcacagaaca tgaatcagca ttagaaagaa
 13920
 gttaaggccg gactcagtgg ctcatgactg taatcccagc atttttggaa gccgaggcag
 13980
 gaagatgctt gagctcagga gtttgagacc accctggcca acatagtgag atctcatctc
 14040
 tacaaaaaat ttaaaaatta actgggcatg gtagcacgca cctgcagtcc cagctactcc
 14100
 agagggttag gtgggaggat tgcttgagcc caggaggata aggctgcagt gaaccatgaa
 14160
 gaagttaggg ctacacaatc actttaagaa gatttctgtg atgtgttaag tcggaaactc
 14220
 aagtataaaa tatgattcat tttgaagggt agcaaaaaat tatgtatgtg tgctgtatg
 14280
 tgtgtatgca cctataacta tatatatata tatatatata tatatatcct gtattaacca
 14340
 tacagcatat atacaaaacc atataacata tataagtata tatataaacc catacaacac
 14400
 agatataccc aggtgacaca catacacaca tgtatagttg catgggagag ggagagaaag
 14460
 agagatgatg gaaagaggta aaaataagag atccaggggc aggggggaag aattgttggg
 14520
 tagataaatg cccacattta tgtttctacg cttaggagag atttctgaaa ctggggcaaa
 14580
 tgctaagcaa aattttgaga ctgtttataa cctggaaaag aagtagggca ggggcaggaa
 14640
 tataaaacaa aagagctttg tcattagaga gccaaagctc tgctattgag tagcacaata
 14700
 acatttacta cagccacagc caatgactct gatttcaaag ggcaacagcg cctaggaaga
 14760
 cagtgcgttg ataacctctc tctaacaca cccagaaaca gaggaagaat tttgttttct
 14820
 ctagatgggg agtgtgtcct tccattccac tataaaaaatg gaacatatta tgactgcac
 14880
 aagtccaagg caagacacaa gtggtgctcg ttaacaaga cctacgaagg atactggaag
 14940
 ttttgcatg cagaaggatg gtgtcctgtc tctaaatacc tcagagcagt aatcttggtc
 15000
 tctgaggctg ggatctgagg caaaacacaa gaggggttctg aaaatagaat ccatagccaa
 15060

ttaagtgttg aaaataaagg atgttcgatt tctgacgata taaaaatcac attgcaccat
 15120
 agatatgtta tgagcatttc aaccacatta ttttgaaaag aatcttttta atagaacagc
 15180
 aaagtgatgt gggatctgaa tccccacca atattgttcc catgtgcctt tggtgaaata
 15240
 gggcagagca gatgatactg aataggaagc aaaagtgtgt atttgcgctg gtgtgtgtga
 15300
 gcgcaccagt gtgcatacat gtatgtctgt gtgagtgtgc atgtgtatgt gtgcacattt
 15360
 gttcacgtgc atgcacatgt gtgcataatgt gtgcatgtgt gtgcatgggt tgggtggagg
 15420
 ggggtgcaaat tagaaagggt tcttactctg cttcttcctc ttcctcctct gcagattttg
 15480
 caaactgtgt atttccttc tggtacagac gcttgatcta ctgggagtgt actgatgatg
 15540
 gggaagcatt tgggaaaaaa tgggtgttcac tgaccaagaa ttttaacaag gaccgaattt
 15600
 ggaaatactg tgaatgatgg tgagattttac caggggaggg ggatgaagac gtttaatcat
 15660
 tctccttgga aagaacctag agtttttgtt ggcagaggag aaccctttct cattagagag
 15720
 ggtttttttt tgtttttttt ttttttctgg gatgtgagtt gaaggtgatc taactataaa
 15780
 taactagagg ttttcgattc aggattatgg ggttttaata tggaaatctg tagtttagcc
 15840
 tcagcacttt aattgctaac tgtgtaaaat tagtcccttt taaaaattaa gtctatgggg
 15900
 taggcgtggg ggcttatgcc cataatccca gcgctttggg aggccaaaaa ggaaggactg
 15960
 cttgaggcca ggagttcaag accagcctgg accaacatag caagaccccg tttctacaaa
 16020
 agaaaaaatt taaaaatgta actaggcatg gtggcacaca cttgtagtct cagctactcg
 16080
 ggaagctgag gcaggagaat cacctgagtt caagagttca aggttgcatg gagttgtgat
 16140
 gaaagagaga cccacctcta aaataaaata aatgaagtaa aataagctta tggattttgg
 16200
 gtgatagtga tgtgtcaatg tagattcatc aattgcaaca aatgttcac tctggtggag
 16260
 gatattgata acacagtggc tatgtgtaca tggaaatttt tatacctact actcaacttt
 16320
 tctgtgactt aaaactgctc taaaaaataa cgtgtattaa aaatgcatta ccttggccgg
 16380
 gcacggtggc tcacgcctgt aatcccagca ctttgggagg tcgaggcggg tggatcacga
 16440
 ggtcaggaga tcgagaccgt cctggctaac atggtgaaac cccgtctctt ctaaaaatac
 16500
 aaaaaattag ccgggcgtgg tggcgggtgc ctgtagtccc agctactcag gaggct
 16556

<210> 4

<211> 97

<212> PRT

<213> Equus caballus

<400> 4

His	Ser	Ala	Thr	Val	Thr	Pro	Glu	Asn	Lys	Cys	Val	Phe	Pro	Phe	Asn
1				5					10					15	
Tyr	Arg	Gly	Tyr	Arg	Tyr	Tyr	Asp	Cys	Thr	Arg	Thr	Asp	Ser	Phe	Tyr
		20						25					30		
Arg	Trp	Cys	Ser	Leu	Thr	Gly	Thr	Tyr	Ser	Gly	Ser	Trp	Lys	Tyr	Cys
		35					40					45			
Ala	Ala	Thr	Asp	Tyr	Ala	Lys	Cys	Ala	Phe	Pro	Phe	Val	Tyr	Arg	Gly
		50				55					60				
Gln	Thr	Tyr	Asp	Arg	Cys	Thr	Thr	Asp	Gly	Ser	Leu	Phe	Arg	Ile	Ser
65					70				75					80	
Trp	Cys	Ser	Val	Thr	Pro	Asn	Tyr	Asp	His	His	Gly	Ala	Trp	Lys	Tyr

85 90 95

Cys

<210> 5
 <211> 95
 <212> PRT
 <213> Equus caballus

<400> 5
 His Ser Ala Thr Val Thr Pro Glu Asn Lys Cys Val Phe Pro Phe Asn
 1 5 10 15
 Tyr Arg Gly Tyr Arg Tyr Tyr Asp Cys Thr Arg Thr Asp Ser Phe Tyr
 20 25 30
 Arg Trp Cys Ser Leu Thr Gly Thr Tyr Ser Gly Ser Trp Lys Tyr Cys
 35 40 45
 Ala Ala Thr Asp Tyr Ala Lys Cys Ala Phe Pro Phe Val Tyr Arg Gly
 50 55 60
 Gln Thr Tyr Asp Arg Cys Thr Thr Asp Gly Ser Leu Phe Arg Ile Ser
 65 70 75 80
 Trp Cys Ser Val Thr Pro Asn Tyr Asp His His Gly Ala Trp Lys
 85 90 95

<210> 6
 <211> 140
 <212> PRT
 <213> Bos taurus

<400> 6
 Met Ala Leu Arg Leu Gly Leu Phe Leu Ile Trp Ala Gly Val Ser Met
 1 5 10 15
 Phe Leu Gln Leu Asp Pro Val Asn Gly Asp Glu Gln Leu Ser Glu Asp
 20 25 30
 Asn Val Ile Leu Pro Lys Glu Lys Lys Asp Pro Ala Ser Gly Ala Glu
 35 40 45
 Thr Lys Asp Asn Lys Cys Val Phe Pro Phe Ile Tyr Gly Asn Lys Lys
 50 55 60
 Tyr Phe Asp Cys Thr Leu His Gly Ser Leu Phe Leu Trp Cys Ser Leu
 65 70 75 80
 Asp Ala Asp Tyr Thr Gly Arg Trp Lys Tyr Cys Thr Lys Asn Asp Tyr
 85 90 95
 Ala Lys Cys Val Phe Pro Phe Ile Tyr Glu Gly Lys Ser Tyr Asp Thr
 100 105 110
 Cys Ile Ile Ile Gly Ser Thr Phe Met Asn Tyr Trp Cys Ser Leu Ser
 115 120 125
 Ser Asn Tyr Asp Glu Asp Gly Val Trp Lys Tyr Cys
 130 135 140

<210> 7
 <211> 134
 <212> PRT
 <213> Bos taurus

<400> 7
 Met Ala Leu Gln Leu Gly Leu Phe Leu Ile Trp Ala Gly Val Ser Val
 1 5 10 15
 Phe Leu Gln Leu Asp Pro Val Asn Gly Asp Gln Asp Glu Gly Val Ser
 20 25 30
 Thr Glu Pro Thr Gln Asp Gly Pro Ala Glu Leu Pro Glu Asp Glu Glu
 35 40 45
 Cys Val Phe Pro Phe Val Tyr Arg Asn Arg Lys His Phe Asp Cys Thr
 50 55 60
 Val His Gly Ser Leu Phe Pro Trp Cys Ser Leu Asp Ala Asp Tyr Val
 65 70 75 80
 Gly Arg Trp Lys Tyr Cys Ala Gln Arg Asp Tyr Ala Lys Cys Val Phe

```

      85      90      95
Pro Phe Ile Tyr Gly Gly Lys Lys Tyr Glu Thr Cys Thr Lys Ile Gly
      100      105      110
Ser Met Trp Met Ser Trp Cys Ser Leu Ser Pro Asn Tyr Asp Lys Asp
      115      120      125
Arg Ala Trp Lys Tyr Cys
      130

```

<210> 8

<211> 93

<212> PRT

<213> Bos taurus

<400> 8

```

Glu Thr Lys Asp Asn Lys Cys Val Phe Pro Phe Ile Tyr Gly Asn Lys
 1      5      10      15
Lys Tyr Phe Asp Cys Thr Leu His Gly Ser Leu Phe Leu Trp Cys Ser
      20      25      30
Leu Asp Ala Asp Tyr Thr Gly Arg Trp Lys Tyr Cys Thr Lys Asn Asp
      35      40      45
Tyr Ala Lys Cys Val Phe Pro Phe Ile Tyr Glu Gly Lys Ser Tyr Asp
      50      55      60
Thr Cys Ile Lys Ile Gly Ser Thr Phe Met Asn Tyr Trp Cys Ser Leu
      65      70      75      80
Ser Ser Asn Tyr Asp Glu Asp Gly Val Trp Lys Tyr Cys
      85      90

```