

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 February 2002 (28.02.2002)

PCT

(10) International Publication Number
WO 02/16591 A2

(51) International Patent Classification⁷: **C12N 15/00**

(21) International Application Number: PCT/US01/26212

(22) International Filing Date: 21 August 2001 (21.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/226,504 21 August 2000 (21.08.2000) US
60/250,932 30 November 2000 (30.11.2000) US

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

(72) Inventors; and

(75) Inventors/Applicants (for US only): **CURTIS, Rory, A., J.** [GB/US]; 31 Constitution Drive, Southborough, MA 01772 (US). **CHUN, Miyoung** [US/US]; 88 Country Club Lane, Belmont, MA 02478 (US).

(74) Agents: **MANDRAGOURAS, Amy, E.** et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (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, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

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

Published:

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

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

(54) Title: 49937, 49931, AND 49933, NOVEL HUMAN TRANSPORTER FAMILY MEMBERS AND USES THEREOF

(57) Abstract: The invention provides isolated nucleic acid molecules, designated HEAT nucleic acid molecules, which encode novel transporter family members. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing HEAT nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a HEAT gene has been introduced or disrupted. The invention still further provides isolated HEAT proteins, fusion proteins, antigenic peptides and anti-HEAT antibodies. Diagnostic and therapeutic methods utilizing compositions of the invention are also provided.



WO 02/16591 A2

**49937, 49931, AND 49933, NOVEL HUMAN TRANSPORTER FAMILY
MEMBERS AND USES THEREOF**

Related Applications

- 5 This application claims the benefit of U.S. Provisional Application No. 60/226,504, filed on August 21, 2000, and U.S. Provisional Application No. 60/250,932, filed on November 30, 2000, incorporated herein in their entirety by this reference.

Background of the Invention

- 10 The E1-E2 ATPase family is a large superfamily of cation transport enzymes that contains at least 80 members found in diverse organisms such as bacteria, archaea, and eukaryotes (Palmgren, M.G. and Axelsen, K.B. (1998) *Biochim. Biophys. Acta* 1365:37-45). These enzymes are involved in ATP hydrolysis-
- 15 dependent transmembrane movement of a variety of inorganic cations (*e.g.*, H⁺, Na⁺, K⁺, Ca²⁺, Cu²⁺, Cd⁺, and Mg²⁺ ions) across a concentration gradient, whereby the enzyme converts the free energy of ATP hydrolysis into electrochemical ion gradients. E1-E2 ATPases are also known as "P-type" ATPases, referring to the existence of a covalent high-energy phosphoryl-enzyme intermediate in the chemical
- 20 reaction pathway of these transporters. The superfamily contains four major groups: Ca²⁺ transporting ATPases; Na⁺/K⁺- and gastric H⁺/K⁺ transporting ATPases; plasma membrane H⁺ transporting ATPases of plants, fungi, and lower eukaryotes; and all bacterial P-type ATPases (Kuhlbrandt *et al.* (1998) *Curr. Opin. Struct. Biol.* 8:510-516).
- 25 E1-E2 ATPases are phosphorylated at a highly conserved DKTG sequence. Phosphorylation at this site is thought to control the enzyme's substrate affinity. Most E1-E2 ATPases contain ten alpha-helical transmembrane domains, although additional domains may be present. A majority of known gated-pore translocators contain twelve alpha-helices, including Na⁺/H⁺ antiporters (West (1997) *Biochim. Biophys. Acta* 1331:213-234).
- 30 Members of the E1-E2 ATPase superfamily are able to generate electrochemical ion gradients which enable a variety of processes in the cell such as absorption, secretion, transmembrane signaling, nerve impulse transmission, excitation/contraction coupling, and growth and differentiation (Scarborough (1999) *Curr. Opin. Cell Biol.* 11:517-522). These molecules are thus critical to normal cell
- 35 function and well-being of the organism.

Summary of the Invention

The present invention is based, at least in part, on the discovery of novel calcium transporter family members, referred to interchangeably herein as "P-type ATPase", "E1-E2 ATPase", "human E1-E2 ATPase", or "HEAT" nucleic acid and protein molecules (e.g., HEAT-1, HEAT-2 and HEAT-3). The HEAT nucleic acid and protein molecules of the present invention are useful as modulating agents in regulating a variety of cellular processes, e.g., tone regulation in vascular smooth muscle cells, cellular growth and/or proliferation, and/or angiogenesis. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding HEAT proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of HEAT-encoding nucleic acids.

In one embodiment, the invention features an isolated nucleic acid molecule that includes the nucleotide sequence set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:10. In another embodiment, the invention features an isolated nucleic acid molecule that encodes a polypeptide including the amino acid sequence set forth in SEQ ID NO:2, 6, or 9. In another embodiment, the invention features an isolated nucleic acid molecule that includes the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number _____, _____, or _____.

In still other embodiments, the invention features isolated nucleic acid molecules including nucleotide sequences that are substantially identical (e.g., 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9% identical) to the nucleotide sequence set forth as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, OR SEQ ID NO:10. The invention further features isolated nucleic acid molecules including at least 30 contiguous nucleotides of the nucleotide sequence set forth as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, OR SEQ ID NO:10. In another embodiment, the invention features isolated nucleic acid molecules which encode a polypeptide including an amino acid sequence that is substantially identical (e.g., 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9% identical) to the amino acid sequence set forth as SEQ ID NO:2, 6, or 9. Also featured are nucleic acid molecules which encode allelic variants of the polypeptide having the amino acid sequence set forth as SEQ ID NO:2, 6, or 9. In addition to isolated nucleic acid molecules encoding full-length polypeptides, the present invention also features nucleic acid molecules which encode fragments, for example, biologically active or

antigenic fragments, of the full-length polypeptides of the present invention (*e.g.*, fragments including at least 10 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2, 6, or 9). In still other embodiments, the invention features nucleic acid molecules that are complementary to, antisense to, or hybridize under stringent conditions to the isolated nucleic acid molecules described herein.

In a related aspect, the invention provides vectors including the isolated nucleic acid molecules described herein (*e.g.*, HEAT-encoding nucleic acid molecules). Such vectors can optionally include nucleotide sequences encoding heterologous polypeptides. Also featured are host cells including such vectors (*e.g.*, host cells including vectors suitable for producing HEAT nucleic acid molecules and polypeptides).

In another aspect, the invention features isolated HEAT polypeptides and/or biologically active or antigenic fragments thereof. Exemplary embodiments feature a polypeptide including the amino acid sequence set forth as SEQ ID NO:2, 6, or 9, a polypeptide including an amino acid sequence at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9% identical to the amino acid sequence set forth as SEQ ID NO:2, 6, or 9, a polypeptide encoded by a nucleic acid molecule including a nucleotide sequence at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9% identical to the nucleotide sequence set forth as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, OR SEQ ID NO:10. Also featured are fragments of the full-length polypeptides described herein (*e.g.*, fragments including at least 10 contiguous amino acid residues of the sequence set forth as SEQ ID NO:2, 6, or 9) as well as allelic variants of the polypeptide having the amino acid sequence set forth as SEQ ID NO:2, 6, or 9.

The HEAT polypeptides and/or biologically active or antigenic fragments thereof, are useful, for example, as reagents or targets in assays applicable to treatment and/or diagnosis of cardiovascular disorders. In one embodiment, a HEAT polypeptide or fragment thereof has a HEAT activity. In another embodiment, a HEAT polypeptide or fragment thereof has at least one or more of the following domains or motifs: a transmembrane domain, an E1-E2 ATPase domain, an E1-E2 ATPases phosphorylation site, an N-terminal large extramembrane domain, a C-terminal large extramembrane domain, a P-type ATPase sequence 1 motif, a P-type ATPase sequence 2 motif, and/or a P-type ATPase sequence 3 motif and, optionally, has a HEAT activity. In a related aspect, the invention features antibodies (*e.g.*, antibodies which specifically bind to any one of the polypeptides, as described herein)

as well as fusion polypeptides including all or a fragment of a polypeptide described herein.

The present invention further features methods for detecting HEAT polypeptides and/or HEAT nucleic acid molecules, such methods featuring, for example, a probe, primer or antibody described herein. Also featured are kits for the detection of HEAT polypeptides and/or HEAT nucleic acid molecules. In a related aspect, the invention features methods for identifying compounds which bind to and/or modulate the activity of a HEAT polypeptide or HEAT nucleic acid molecule described herein. Also featured are methods for modulating a HEAT activity.

In other embodiments, the invention provides methods for identifying a subject having a cardiovascular disorder, or at risk for developing a cardiovascular disorder; methods for identifying a compound capable of treating a cardiovascular disorder characterized by aberrant HEAT nucleic acid expression or HEAT polypeptide activity; and methods for treating a subject having a cardiovascular disorder characterized by aberrant HEAT polypeptide activity or aberrant HEAT nucleic acid expression.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figures 1A-1D depict the nucleotide sequence of the human HEAT-1 cDNA and the corresponding amino acid sequence. The nucleotide sequence corresponds to nucleic acids 1 to 4055 of SEQ ID NO:1. The amino acid sequence corresponds to amino acids 1 to 1180 of SEQ ID NO:2. The coding region without the 5' and 3' untranslated regions of the human HEAT-1 gene is shown in SEQ ID NO:3.

Figure 2 depicts a structural, hydrophobicity, and antigenicity analysis of the human HEAT-1 polypeptide. The locations of the 12 transmembrane domains, as well as the E1-E2 ATPase domain, are indicated.

Figure 3 depicts the results of a search in the HMM database in PFAM which resulted in the identification of an "E1-E2 ATPase domain" in the human HEAT-1 polypeptide (SEQ ID NO:2).

Figures 4A-4B depict a Clustal W (1.74) multiple sequence alignment of the human HEAT-1 amino acid sequence (Fbh49937; SEQ ID NO:2) and the amino acid sequence of a *C. elegans* cation-transporting ATPase (YH2Melegans; SEQ ID NO:4; GenBank Accession No. Q27533). Amino acid identities are indicated by stars.

Figure 5 depicts the expression levels of human HEAT-1 mRNA in various human cell types and tissues, as determined by Taqman analysis. Column: (1) normal artery; (2) normal vein; (3) aortic smooth muscle cells (early); (4) coronary

smooth muscle cells; (5) umbilical vein endothelial cells (static); (6) umbilical vein endothelial cells (shear); (7) normal heart; (8) heart (congestive heart failure); (9) kidney; (10) skeletal muscle; (11) normal adipose tissue; (12) pancreas; (13) primary osteoblasts; (14) differentiated osteoclasts; (15) normal skin; (16) normal spinal cord; 5 (17) normal brain cortex; (18) normal brain hypothalamus; (19) nerve; (20) dorsal root ganglion; (21) glial cells (astrocytes); (22) glioblastoma; (23) normal breast; (24) breast tumor; (25) normal ovary; (26) ovarian tumor; (27) normal prostate; (28) prostate tumor; (29) epithelial cells (prostate); (30) normal colon; (31) colon tumor; (32) normal lung; (33) lung tumor; (34) lung (chronic obstructive pulmonary disease); 10 (35) colon (inflammatory bowel disease); (36) normal liver; (37) liver fibrosis; (38) dermal cells (fibroblasts); (39) normal spleen; (40) normal tonsil; (41) lymph node; (42) small intestine; (43) skin (decubitus); (44) synovium; (45) bone marrow mononuclear cells; (46) activated peripheral blood mononuclear cells.

Figure 6 depicts the expression levels of human HEAT-1 mRNA in various 15 vascular rich organs, as determined by Taqman analysis. Column: (1) confluent microvascular endothelial cells; (2) aortic smooth muscle cells; (3) fetal heart; (4) normal heart atrium; (5) normal heart atrium; (6) normal heart ventricle; (7) normal heart ventricle; (8) normal heart ventricle; (9) normal heart ventricle; (10) normal heart ventricle; (11) diseased heart ventricle; (12) diseased heart ventricle; (13) 20 diseased heart ventricle; (14) normal kidney; (15) normal kidney; (16) normal kidney; (17) normal kidney; (18) normal kidney; (19) hypertensive kidney; (20) hypertensive kidney; (21) hypertensive kidney; (22) hypertensive kidney; (23) skeletal muscle; (24) skeletal muscle; (25) liver; (26) liver; (27) normal fetal adrenal gland; (28) Wilms tumor; (29) Wilms tumor; (30) normal spinal cord; (31) diseased cartilage.

25 *Figure 7* depicts the expression levels of human HEAT-1 mRNA in various human and monkey vessels, as determined by Taqman analysis. Column: (1) human aortic smooth muscle cells; (2) human microvascular endothelial cells; (3) human adipose tissue; (4) human normal carotid artery; (5) human normal carotid artery; (6) human normal muscular artery; (7) human diseased iliac artery; (8) human diseased 30 tibial artery; (9) human diseased aorta; (10) human normal saphenous vein; (11) human normal saphenous vein; (12) human normal saphenous vein; (13) human normal saphenous vein; (14) human diseased saphenous vein; (15) human normal vein; (16) human normal vein; (17) human normal vein; (18) monkey normal coronary artery; (19) monkey normal coronary artery; (20) monkey normal coronary 35 artery; (21) monkey normal coronary artery; (22) monkey normal vein; (23) no transcriptase control.

Figure 8 depicts the expression levels of human HEAT-1 mRNA in various human coronary vascular cell types, as well as other cell types, as determined by Taqman analysis. Column: (1) aortic smooth muscle cells; (2) aortic smooth muscle cells; (3) aortic smooth muscle cells; (4) coronary smooth muscle cells; (5) coronary smooth muscle cells; (6) coronary smooth muscle cells; (7) coronary smooth muscle cells; (8) macrophages; (9) macrophages treated with IFN γ ; (10) macrophages treated with CD40; (11) macrophages treated with LPS; (12) umbilical vein endothelial cells; (13) microvascular endothelial cells; (14) aortic endothelial cells; (15) aortic endothelial cells; (16) cortex renal epithelium; (17) renal proximal tubule epithelium; (18) mesangial cells; (19) skeletal muscle; (20) skeletal muscle; (21) lung fibroblasts.

Figure 9 depicts the expression levels of human HEAT-1 mRNA in various human endothelial cell paradigms, as determined by Taqman analysis. Column: (1) umbilical vein endothelial cells (static); (2) umbilical vein endothelial cells (laminar shear stress); (3) umbilical vein endothelial cells (static); (4) umbilical vein endothelial cells (laminar shear stress); (5) umbilical vein endothelial cells (proliferating); (6) umbilical vein endothelial cells (confluent); (7) umbilical vein endothelial cells (without growth factor treatment); (8) umbilical vein endothelial cells (treated with IL-1); (9) cardiac microvascular endothelial cells (proliferating); (10) cardiac microvascular endothelial cells (confluent); (11) cardiac microvascular endothelial cells (proliferating); (12) cardiac microvascular endothelial cells (confluent); (13) lung microvascular endothelial cells (proliferating); (14) lung microvascular endothelial cells (confluent); (15) lung microvascular endothelial cells (without growth factor treatment); (16) lung microvascular endothelial cells (proliferating); (17) lung microvascular endothelial cells (confluent); (18) aortic cells (control 4h); (19) aortic cells (TNF treated 4h); (20) aortic cells (control 14h); (21) aortic cells (TNF treated 14h); (22) 293 cells; (23) lung microvascular endothelial cells (Matrigel 5h); (24) lung microvascular endothelial cells (Matrigel 25h); (25) lung microvascular endothelial cells (proliferating); (26) lung microvascular endothelial cells (without growth factor treatment).

Figure 10 depicts the expression levels of human HEAT-1 mRNA in cells subjected to various laminar shear stress treatments, as determined by Taqman analysis. Column: (1) static (control); (2) laminar shear stress (LSS); (3) LSS + 1h up; (4) LSS + 1h down; (5) static (control); (6) LSS; (7) LSS + 6h up; (8) static (control); (9) LSS; (10) LSS + 6h down.

Figures 11A-11E depict the nucleotide sequence of the human HEAT-2 cDNA and the corresponding amino acid sequence. The nucleotide sequence corresponds to nucleic acids 1 to 7249 of SEQ ID NO:5. The amino acid sequence corresponds to

amino acids 1 to 1256 of SEQ ID NO:6. The coding region without the 5' or 3' untranslated regions of the human HEAT-2 gene is shown in SEQ ID NO:7.

Figure 12 depicts the results of a search in the HMM database which resulted in the identification of an "E1-E2 ATPase domain" in the human HEAT-2

5 polypeptide (SEQ ID NO:6).

Figure 13 depicts a structural, hydrophobicity, and antigenicity analysis of human HEAT-2. The locations of the 12 transmembrane domains, as well as the E1-E2 ATPase domain, are indicated.

10 *Figures 14A-14B* depict a Clustal W (1.74) multiple sequence alignment of the human HEAT-2 amino acid sequence (Fbh49931FL; SEQ ID NO:6) and the amino acid sequence of a *C. elegans* cation-transporting ATPase (YH2Melegans; SEQ ID NO:4; GenBank Accession No. Q27533). Amino acid identities are indicated by stars. The twelve transmembrane domains, as well as the phosphorylation site, are indicated by boxes.

15 *Figure 15* depicts the expression levels of human HEAT-2 mRNA in various human cell types and tissues, as determined by Taqman analysis. Column: (1) normal aorta; (2) normal fetal heart; (3) normal heart; (4) heart (congestive heart failure); (5) normal vein; (6) aortic smooth muscle cells; (7) normal spinal cord; (8) brain (normal cortex); (9) brain (hypothalamus); (10) glial cells (astrocytes); (11) 20 brain (glioblastoma); (12) normal breast; (13) breast tumor (infiltrating ductal carcinoma); (14) normal ovary; (15) ovarian tumor; (16) pancreas; (17) normal prostate; (18) prostate tumor; (19) normal colon; (20) colon tumor; (21) colon (inflammatory bowel disease); (22) normal kidney; (23) normal liver; (24) fibrotic liver; (25) normal fetal liver; (26) normal lung; (27) lung tumor; (28) lung (chronic obstructive pulmonary disease); (29) normal spleen; (30) normal tonsil; (31) normal lymph node; (32) normal thymus; (33) epithelial cells (from prostate); (34) aortic endothelial cells; (35) skeletal muscle; (36) dermal fibroblasts; (37) normal skin; (38) 25 normal adipose tissue; (39) primary osteoblasts; (40) undifferentiated osteoblasts; (41) differentiated osteoblasts; (42) osteoclasts; (43) aortic smooth muscle cells (early); (44) aortic smooth muscle cells (late); (45) human umbilical vein endothelial cells (shear); (46) human umbilical vein endothelial cells (static).

30 *Figure 16* depicts the expression levels of human HEAT-2 mRNA in various vascular rich organs, as determined by Taqman analysis. Column: (1) normal human heart; (2) normal human heart; (3) normal human heart; (4) normal human heart; (5) 35 normal human heart; (6) normal human heart; (7) normal human heart; (8) normal human heart; (9) diseased human heart; (10) diseased human right ventricle; (11) diseased human left ventricle; (12) normal monkey heart; (13) normal monkey heart; (14) normal monkey heart; (15) normal human kidney; (16) normal human kidney;

(17) normal human kidney; (18) normal human kidney; (19) normal human kidney; (20) human hypertensive kidney; (21) human hypertensive kidney; (22) human hypertensive kidney; (23) human hypertensive kidney; (24) human hypertensive kidney; (25) human liver; (26) human liver; (27) human liver; (28) human skeletal muscle; (29) human skeletal muscle; (30) human skeletal muscle.

Figure 17 depicts the expression levels of human HEAT-2 mRNA in various human and monkey vessels, as determined by Taqman analysis. Column: (1) human adipose tissue; (2) human normal artery; (3) human normal artery; (4) human carotid artery; (5) human carotid artery; (6) human normal artery; (7) human diseased artery; (8) human diseased artery; (9) human diseased artery; (10) human normal vein; (11) human normal vein; (12) human vein; (13) human vein; (14) human normal vein; (15) human varicose vein; (16) confluent human microvascular endothelial cells; (17) human aortic smooth muscle cells; (18) monkey aorta; (19) monkey aorta; (20) monkey aorta; (21) monkey artery; (22) monkey artery; (23) monkey renal artery; (24) monkey renal artery; (25) monkey renal artery; (26) monkey renal artery; (27) monkey renal artery; (28) monkey coronary artery; (29) monkey coronary artery; (30) monkey coronary artery; (31) monkey coronary artery; (32) monkey coronary artery; (33) monkey coronary artery; (34) monkey coronary artery.

Figure 18 depicts the expression levels of human HEAT-2 mRNA in various cell types and tissues, as determined by transcriptional profiling analysis. Column: (1) human aortic smooth muscle cells; (2) human coronary artery smooth muscle cells; (3) human umbilical vein endothelial cells; (4) human microvascular endothelial cells (lung); (5) monkey aorta; (6) monkey vein; (7) monkey heart; (8) monkey liver.

Figure 19 depicts the expression levels of human HEAT-2 mRNA in various human coronary vascular cell types, as well as other cell types, as determined by Taqman analysis. Column: (1) aortic smooth muscle cells; (2) aortic smooth muscle cells; (3) aortic smooth muscle cells; (4) aortic smooth muscle cells; (5) coronary smooth muscle cells; (6) coronary smooth muscle cells; (7) coronary smooth muscle cells; (8) coronary smooth muscle cells; (9) macrophages; (10) macrophages treated with IFN γ ; (11) macrophages treated with CD40; (12) macrophages treated with LPS; (13) umbilical vein endothelial cells; (14) microvascular endothelial cells; (15) aortic endothelial cells; (16) coronary artery endothelial cells; (17) coronary artery endothelial cells; (18) cortex renal epithelium; (19) renal proximal tubule epithelium; (20) mesangial cells; (21) skeletal muscle; (22) skeletal muscle; (23) lung fibroblasts.

Figure 20 depicts the expression levels of human HEAT-2 mRNA in various human endothelial cell paradigms of shear stress, as determined by Taqman analysis. Column: (1) umbilical vein endothelial cells (static); (2) umbilical vein endothelial

cells (shear regulated); (3) umbilical vein endothelial cells (proliferating); (4) umbilical vein endothelial cells (confluent); (5) umbilical vein endothelial cells (without growth factor treatment); (6) umbilical vein endothelial cells (Interleukin-1 stimulated); (7) microvascular endothelial cells (proliferating); (8) microvascular endothelial cells (confluent); (9) microvascular endothelial cells (proliferating); (10) microvascular endothelial cells (confluent); (11) microvascular endothelial cells (proliferating); (12) microvascular endothelial cells (confluent); (13) microvascular endothelial cells (without growth factor treatment); (14) coronary microvascular endothelial cells (proliferating); (15) coronary microvascular endothelial cells (confluent); (16) microvascular endothelial cells (5% serum plus growth factors); (17) microvascular endothelial cells (5% serum without growth factors); (18) microvascular endothelial cells (hEGF treated); (19) microvascular endothelial cells (VEGF treated); (20) microvascular endothelial cells (bFGF treated); (21) microvascular endothelial cells (IGF treated); (22) 293 cells; (23) umbilical vein endothelial cells (static 25h); (24) umbilical vein endothelial cells (laminar shear stress); (25) umbilical vein endothelial cells (laminar shear stress + 1h up); (26) umbilical vein endothelial cells (laminar shear stress + 1h down); (27) umbilical vein endothelial cells (static 30h); (28) umbilical vein endothelial cells (laminar shear stress); (29) umbilical vein endothelial cells (laminar shear stress + 6h up); (30) umbilical vein endothelial cells (static 30h); (31) umbilical vein endothelial cells (laminar shear stress); (32) umbilical vein endothelial cells (laminar shear stress + 6h down).

Figure 21 depicts the expression level of human HEAT-2 mRNA human microvascular endothelial cells under conditions of tube formation (growth on Matrigel). Column: (1) Matrigel (5h); (2) Matrigel (25h); (3) proliferating; (4) confluent.

Figures 22A-22D depict the nucleotide sequence of the human HEAT-3 cDNA and the corresponding amino acid sequence. The nucleotide sequence corresponds to nucleic acids 1 to 3919 of SEQ ID NO:8. The amino acid sequence corresponds to amino acids 1 to 1204 of SEQ ID NO:9. The coding region without the 5' or 3' untranslated regions of the human HEAT-3 gene is shown in SEQ ID NO:10.

Figure 23 depicts the results of a search in the HMM database which resulted in the identification of an "E1-E2 ATPase domain" in the human HEAT-3 polypeptide (SEQ ID NO:9).

Figure 24 depicts a structural, hydrophobicity, and antigenicity analysis of human HEAT-3. The locations of the 12 transmembrane domains, as well as the E1-E2 ATPase domain, are indicated.

Figures 25A-25B depict a Clustal W (1.74) multiple sequence alignment of the human HEAT-3 amino acid sequence (Fbh49933FL1; SEQ ID NO:8) and the amino acid sequence of a *C. elegans* cation-transporting ATPase (YE56elegans; SEQ ID NO:12; GenBank Accession No. P90747). Amino acid identities are indicated by stars. The twelve transmembrane domains, as well as the phosphorylation site, are indicated by boxes.

Figure 26 depicts the expression levels of human HEAT-3 mRNA in various human cell types and tissues, as determined by Taqman analysis. Column: (1) normal aorta; (2) normal fetal heart; (3) normal heart; (4) heart (congestive heart failure); (5) normal vein; (6) normal spinal cord; (7) normal brain cortex; (8) normal brain hypothalamus; (9) glial cells (astrocytes); (10) glioblastoma (brain); (11) normal breast; (12) breast tumor (infiltrating ductal carcinoma); (13) normal ovary; (14) ovarian tumor; (15) pancreas; (16) normal prostate; (17) prostate tumor; (18) normal colon; (19) colon tumor; (20) colon (inflammatory bowel disease); (21) normal kidney; (22) normal liver; (23) liver fibrosis; (24) normal fetal liver; (25) normal lung; (26) lung tumor; (27) lung (chronic obstructive pulmonary disease); (28) normal spleen; (29) normal tonsil; (30) normal lymph node; (31) normal thymus; (32) epithelial cells (prostate); (33) endothelial cells (aortic); (34) normal skeletal muscle; (35) fibroblasts (dermal); (36) normal skin; (37) normal adipose tissue; (38) primary osteoblasts; (39) undifferentiated osteoblasts; (40) differentiated osteoblasts; (41) osteoclasts; (42) aortic smooth muscle cells (early); (43) aortic smooth muscle cells (late); (44) umbilical vein endothelial cells (laminar shear stress); (45) umbilical vein endothelial cells (static); (46) undifferentiated osteoclasts.

Figure 27 depicts the expression levels of human HEAT-3 mRNA in various vascular rich organs, as determined by Taqman analysis. Column: (1) normal heart; (2) normal heart; (3) normal heart; (4) normal heart; (5) normal heart; (6) normal heart; (7) normal heart; (8) normal heart; (9) diseased heart; (10) diseased right ventricle; (11) normal fetal heart; (12) normal kidney; (13) normal kidney; (14) normal kidney; (15) normal kidney; (16) normal kidney; (17) hypertensive kidney; (18) hypertensive kidney; (19) hypertensive kidney; (20) hypertensive kidney; (21) hypertensive kidney; (22) skeletal muscle; (23) skeletal muscle; (24) skeletal muscle; (25) liver; (26) liver; (27) normal monkey heart; (28) normal monkey heart; (29) normal monkey heart; (30) normal monkey heart; (31) smooth muscle cells (SMC); (32) confluent human microvascular endothelial cells (HMVECs); (33) M human umbilical vein endothelial cells (HUVECs); (34) human umbilical vein endothelial cells (HUVECs) - vehicle; (35) M human amniotic endothelial cells (HAECs); (36) human amniotic endothelial cells (HAECs) - vehicle.

Figure 28 depicts the expression levels of human HEAT-3 mRNA in various human vessels, as determined by Taqman analysis. Column: (1) aortic smooth muscle cells; (2) microvascular endothelial cells; (3) adipose tissue; (4) normal artery; (5) normal artery; (6) normal artery; (7) diseased artery; (8) diseased artery; (9) diseased aorta; (10) normal vein; (11) normal vein; (12) normal vein; (13) normal vein; (14) diseased vein; (15) normal vein; (16) normal vein; (17) normal vein.

Figure 29 depicts the expression levels of human HEAT-3 mRNA in various human coronary vascular cell types, as well as other cell types, as determined by Taqman analysis. Column: (1) aortic smooth muscle cells; (2) aortic smooth muscle cells; (3) coronary smooth muscle cells; (4) coronary smooth muscle cells; (5) coronary smooth muscle cells; (6) coronary smooth muscle cells; (7) macrophages; (8) macrophages treated with IFN γ ; (9) macrophages treated with CD40; (10) macrophages treated with LPS; (11) microvascular endothelial cells; (12) aortic endothelial cells; (13) coronary artery endothelial cells; (14) cortex renal epithelium; (15) renal proximal tubule epithelium; (16) mesangial cells; (17) skeletal muscle.

Figure 30 depicts an alignment of a region important in calcium binding from HEAT-1, HEAT-2, HEAT-3 with similar sequences from a number of E1-E2 ATPases of various substrate specificities from a number of different organisms. This region includes the sixth transmembrane domain from each of HEAT-1, HEAT-2, and HEAT-3, as well as a number of amino acid residues adjacent to the sixth transmembrane domain. Amino acid residues determined to be important for calcium binding by mutagenesis of a SERCA calcium-transporting E1-E2 ATPase are indicated ("SERCA mutagenesis"). Amino acid residues in this region that are critical for calcium binding are indicated in bold. Substrate specificities are as follows: Type V (calcium), Ca²⁺ (calcium), Cu²⁺ (copper), Na⁺/K⁺ (sodium/potassium), and PL (phospholipid).

Figure 31 depicts the expression levels of human HEAT-3 mRNA in various human vessels, as determined by Taqman analysis. Column: (1) LC smooth muscle cells; (2) LC smooth muscle cells; (3) aortic smooth muscle cells; (4) human microvascular endothelial cells; (5) normal human carotid artery; (6) normal human carotid artery; (7) normal human muscular artery; (8) human diseased iliac artery; (9) human diseased tibial artery; (10) human diseased aorta; (11) human normal saphenous vein; (12) human normal saphenous vein; (13) human normal saphenous vein; (14) human normal saphenous vein; (15) human diseased saphenous vein; (16) human normal vein; (17) human normal saphenous vein.

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of novel calcium transporter family members, referred to interchangeably herein as "P-type ATPase", "E1-E2 ATPase", "human E1-E2 ATPase", or "HEAT" nucleic acid and protein molecules (*e.g.*, HEAT-1, HEAT-2 and HEAT-3). These novel molecules are members of the E1-E2 ATPase superfamily and are highly expressed in human vessels, endothelial cells, and vascular smooth muscle cells, *e.g.*, coronary vascular smooth muscle cells.

The E1-E2 ATPases are involved in ATP hydrolysis-dependent transmembrane movement of inorganic cations (*e.g.*, Ca^{2+} ions) across a concentration gradient. E1-E2 ATPases are phosphorylated at a highly conserved DKTG sequence. Phosphorylation at this site is thought to control the enzyme's substrate affinity. Most E1-E2 ATPases contain ten alpha-helical transmembrane domains, although additional domains may be present. Members of the E1-E2 ATPase superfamily are able to generate electrochemical ion gradients which enable a variety of processes in the cell such as absorption, secretion, transmembrane signaling, nerve impulse transmission, excitation/contraction coupling, and growth and differentiation.

As indicated in the Examples presented herein, the HEAT molecules of the present invention, *e.g.*, HEAT-2, are up-regulated during shear, proliferation, and tube formation of endothelial cells and, thus, are believed to be involved in angiogenesis. Calcium ions are involved in the regulation of many cellular activities. In vascular smooth muscle cells, transient increases in intracellular calcium levels mediate contraction. Thus, maintenance of a low steady-state level of calcium is critical to maintaining proper cell function. Additionally, since the main determinant of the contraction-relaxation cycle of smooth muscle is calcium, calcium concentration is an important factor in the regulation of vascular tone. The normal concentration of calcium in the cell is in the submicromolar range, while the concentration in the extracellular compartment is in the millimolar range. In order to maintain intracellular calcium concentration in the submicromolar range, several mechanisms are operative in most cells. In smooth muscle cells, these regulatory mechanisms include calcium extrusion via Ca^{2+} -transporting E1-E2 ATPases at the plasma membrane and at the sarcoplasmic/endoplasmic reticulum.

Thus, as the HEAT molecules of the present invention are Ca^{2+} -transporting E1-E2 ATPases, and are highly expressed in vessels, endothelial cells, and vascular smooth muscle cells, these molecules are believed to be involved in vasotone regulation of vascular smooth muscle cells, *e.g.*, coronary vascular smooth muscle cells. For example, activation of a HEAT molecule of the invention, *e.g.*, HEAT-3, may result in decreased cytosolic calcium concentrations, thus reducing vascular tone.

Inhibition of a HEAT molecule of the invention, *e.g.*, HEAT-3, may result in decreased intracellular calcium store, which may subsequently lower the calcium release by vasopressor stimulation, thereby reducing vascular smooth muscle tone.

Accordingly, the HEAT molecules of the present invention provide novel
5 diagnostic targets and therapeutic agents for cardiovascular disorders. As used herein, the term “cardiovascular disorder” includes a disorder, disease or condition which affects the cardiovascular system, *e.g.*, the heart or blood vessels. Cardiovascular disorders can detrimentally affect cellular functions such as calcium transport and inter- or intra-cellular communication; and tissue functions such as angiogenesis,
10 vascular smooth muscle tone, vascular function, and cardiac function. Examples of cardiovascular disorders include cardiovascular disorders include hypertension, arteriosclerosis, ischemia reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism, tachycardia, bradycardia, pressure overload, aortic bending, coronary
15 artery ligation, vascular heart disease, atrial fibrillation, Jervell syndrome, Lange-Nielsen syndrome, long-QT syndrome, congestive heart failure, sinus node dysfunction, angina, heart failure, atrial fibrillation, atrial flutter, dilated cardiomyopathy, idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, arrhythmia, atherosclerosis, transplant atherosclerosis,
20 varicose veins, migraine headaches, cluster headaches, vascular disease, diabetic vascular disease, pulmonary vascular disease, peripheral vascular disease, renovascular hypertension, intravascular tumor, pulmonary vasculitis, vascular tone disorders in pregnancy, pulmonary capillaritis, peripheral arterial disease, idiopathic hypereosinophilic syndrome, aortic aneurysm, respiratory disease, vasospasm,
25 systemic sclerosis, preeclampsia, graft vessel disease, cardiac allograft vasculopathy, vascular ischemic injury, familial amyloidotic polyneuropathy, acute atherosclerosis, cardiovascular disease, Kawasaki disease, ischemic syndromes, chronic heart failure, and fibrosis.

The HEAT molecules of the present invention further provide novel
30 diagnostic targets and therapeutic agents for cellular proliferation, growth, or differentiation disorders. Cellular proliferation, growth, or differentiation disorders include those disorders that affect cell proliferation, growth, or differentiation processes. As used herein, a “cellular proliferation, growth, or differentiation
35 process” is a process by which a cell increases in number, size or content, or by which a cell develops a specialized set of characteristics which differ from that of other cells. The HEAT molecules of the present invention are upregulated in various endothelial cell paradigms of shear, proliferation, and tube formation (see Figures 5, 9, 10, 15, 20, and 21), indicating that the HEAT molecules of the present invention are involved in

cellular growth and proliferation. Thus, the HEAT molecules of the present invention may play a role in disorders characterized by aberrantly regulated growth, proliferation, or differentiation. Such disorders include cancer, *e.g.*, carcinoma, sarcoma, or leukemia; tumor angiogenesis and metastasis; skeletal dysplasia; hepatic disorders; myelodysplastic syndromes; and hematopoietic and/or myeloproliferative disorders.

Other disorders related to angiogenesis and which may, therefore, be treated using the molecules described herein, include diabetic retinopathy, neovascularization (*e.g.*, intraocular neovascularization), psoriasis, endometriosis, Grave's disease, ischemic disease, chronic inflammatory diseases, macular degeneration, neovascular glaucoma, retinal fibroplasia, uveitis, eye diseases associated with choroidal neovascularization and iris neovascularization, hereditary hemorrhagic telangiectasia, fibrodysplasia ossificans progressiva, idiopathic pulmonary fibrosis, autosomal dominant polycystic kidney disease, synovitis, familial exudative vitreoretinopathy (FEVR), Alagille syndrome, Knobloch syndrome, disseminated lymphangiomatosis, toxic epidermal necrolysis, Von Hippel Lindau disease (VHL), microbial-related dysplastic and neoplastic angiomatous proliferative processes (*e.g.*, verruga peruana (VP)), Proteus syndrome (PS), Castleman's disease, and Klippel-Trenaunay-Weber syndrome.

Additional disorders that may be treated using the molecules of the present invention include disorders affecting tissues in which HEAT protein is expressed (*e.g.*, vessels, endothelial cells, and vascular smooth muscle cells).

The term "family" when referring to the protein and nucleic acid molecules of the present invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well as other distinct proteins of human origin or alternatively, can contain homologues of non-human origin, *e.g.*, rat or mouse proteins. Members of a family can also have common functional characteristics.

For example, the family of HEAT proteins of the present invention comprises at least one "transmembrane domain," preferably at least 2, 3, or 4 transmembrane domains, more preferably 5, 6, 7, 8, or 9 transmembrane domains, even more preferably 10 or 11 transmembrane domains, and most preferably, 12 transmembrane domains. As used herein, the term "transmembrane domain" includes an amino acid sequence of about 15 amino acid residues in length which spans the plasma membrane. More preferably, a transmembrane domain includes about at least 20, 25,

30, 35, 40, or 45 amino acid residues and spans the plasma membrane.

Transmembrane domains are rich in hydrophobic residues, and typically have an alpha-helical structure. In a preferred embodiment, at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids of a transmembrane domain are hydrophobic, *e.g.*, leucines, isoleucines, tyrosines, or tryptophans. Transmembrane domains are described in, for example, Zagotta, W.N. *et al.* (1996) *Annu. Rev. Neurosci.* 19:235-263, the contents of which are incorporated herein by reference. Amino acid residues 8-25, 47-65, 231-253, 256-276, 428-448, 464-484, 936-954, 963-987, 994-1015, 1049-1065, 1079-1102, and 1118-1134 of the human HEAT-1 protein (SEQ ID NO:2) are predicted to comprise transmembrane domains (see Figure 2). Amino acid residues 29-50, 211-227, 234-253, 294-317, 410-434, 449-469, 941-960, 968-985, 1000-1020, 1076-1092, 1105-1129, 1144-1160 of the human HEAT-2 protein (SEQ ID NO:6) are predicted to comprise transmembrane domains (see Figures 12 and 13). Amino acid residues 65-89, 99-116, 242-258, 265-281, 445-464, 493-509, 990-1007, 1015-1031, 1049-1073, 1103-1119, 1134-1151, 1171-1187 of the human HEAT-3 protein (SEQ ID NO:9) are also predicted to comprise transmembrane domains (see Figures 24 and 25).

In another embodiment, members of the HEAT family of proteins include at least one "E1-E2 ATPase domain" in the protein or corresponding nucleic acid molecule. As used herein, the term "E1-E2 ATPase" domain includes a protein domain having at least about 70-110 amino acid residues and a bit score of at least 30 when compared against an E1-E2 ATPase Hidden Markov Model (HMM), *e.g.*, PFAM Accession Number PF00122. Preferably, an E1-E2 ATPase domain includes a protein having an amino acid sequence of about 80-100, or more preferably about 87, 89, or 90 amino acid residues, and a bit score of at least 35, 40, 50, or more preferably, 37.0, 51.4, or 53.4. To identify the presence of an E1-E2 ATPase domain in a HEAT protein, and make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein is searched against a database of known protein motifs and/or domains (*e.g.*, the HMM database). The E1-E2 ATPase domain (HMM) has been assigned the PFAM Accession number PF00122 (see the PFAM website, available online through Washington University in Saint Louis). A search was performed against the HMM database resulting in the identification of an E1-E2 ATPase domain in the amino acid sequence of human HEAT-1 at about residues 299-387 of SEQ ID NO:2. The results of the search are set forth in Figure 3. A search was also performed against the HMM database resulting in the identification of an E1-E2 ATPase domain in the amino acid sequence of human HEAT-2 at about residues 278-365 of SEQ ID NO:6. The results of the search are set forth in Figure 12. A search was further performed against the HMM database

resulting in the identification of an E1-E2 ATPase domain in the amino acid sequence of human HEAT-3 at about residues 302-392 of SEQ ID NO:9. The results of the search are set forth in Figure 23.

A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28:405-420, and a detailed description of HMMs can be found, for example, in Gribskov *et al.* (1990) *Methods Enzymol.* 183:146-159; Gribskov *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.* (1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.* (1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference.

- 10 Preferably an E1-E2 ATPase domain is at least about 70-110 amino acid residues and has an "E1-E2 ATPase activity", for example, the ability to interact with a HEAT substrate or target molecule (*e.g.*, ATP or a cation such as Ca^{2+}); to transport a HEAT substrate or target molecule (*e.g.*, a cation such as Ca^{2+}) from one side of a biological membrane to the other; to adopt an E1 conformation or an E2
- 15 conformation; to convert a HEAT substrate or target molecule to a product (*e.g.*, to hydrolyze ATP); to interact with a second non-HEAT protein; to modulate intra- or inter-cellular signaling and/or gene transcription (*e.g.*, either directly or indirectly); to modulate vascular smooth muscle tone; to modulate cellular growth and/or proliferation; and/or to modulate angiogenesis. Accordingly, identifying the presence
- 20 of an "E1-E2 ATPase domain" can include isolating a fragment of a HEAT molecule (*e.g.*, a HEAT polypeptide) and assaying for the ability of the fragment to exhibit one of the aforementioned E1-E2 ATPase domain activities.

- In another embodiment, a HEAT molecule of the present invention may also be identified based on its ability to adopt an E1 conformation or an E2 conformation.
- 25 As used herein, an "E1 conformation" of a HEAT protein includes a 3-dimensional conformation of a HEAT protein which does not exhibit HEAT activity (*e.g.*, the ability to transport Ca^{2+}), as defined herein. An E1 conformation of a HEAT protein usually occurs when the HEAT protein is unphosphorylated. As used herein, an "E2 conformation" of a HEAT protein includes a 3-dimensional conformation of a HEAT
- 30 protein which exhibits HEAT activity (*e.g.*, the ability to transport Ca^{2+}), as defined herein. An E2 conformation of a HEAT protein usually occurs when the HEAT protein is phosphorylated.

- In another embodiment, a HEAT protein of the present invention is identified based on the presence of an "E1-E2 ATPases phosphorylation site" in the protein or
- 35 corresponding nucleic acid molecule. An E1-E2 ATPases phosphorylation site functions in accepting a phosphate moiety and has the following consensus sequence: D-K-T-G-T-[LIVM]-[TI] (SEQ ID NO:11), wherein D is phosphorylated. The use of amino acids in brackets indicates that the amino acid at the indicated position may be

any one of the amino acids within the brackets, *e.g.*, [TI] indicates any of one of either T (threonine) or I (isoleucine). The E1-E2 ATPases phosphorylation site has been assigned ProSite Accession Number PS00154. To identify the presence of an E1-E2 ATPases phosphorylation site in a HEAT protein, and to make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein may be searched against a database of known protein domains (*e.g.*, the ProSite database) using the default parameters (available online through the Swiss Institute for Bioinformatics). A search was performed against the ProSite database resulting in the identification of an E1-E2 ATPases phosphorylation site in the amino acid sequence of human HEAT-1 (SEQ ID NO:2) at about residues 513-519. A similar search resulted in the identification of an E1-E2 phosphorylation site in the amino acid sequence of human HEAT-2 (SEQ ID NO:6) at about residues 498-504 (see Figures 14A-14B) and in the amino acid sequence of human HEAT-3 (SEQ ID NO:9) at about residues 533-539 (see Figures 25A-25B).

Preferably an E1-E2 ATPases phosphorylation site has a “phosphorylation site activity,” for example, the ability to be phosphorylated; to be dephosphorylated; to regulate the E1-E2 conformational change of the HEAT protein in which it is contained; to regulate transport of Ca^{2+} across a biological membrane by the HEAT protein in which it is contained; and/or to regulate the activity (as defined herein) of the HEAT protein in which it is contained. Accordingly, identifying the presence of an “E1-E2 ATPases phosphorylation site” can include isolating a fragment of a HEAT molecule (*e.g.*, a HEAT polypeptide) and assaying for the ability of the fragment to exhibit one of the aforementioned phosphorylation site activities.

The family of HEAT proteins of the present invention also comprises at least one “large extramembrane domain” in the protein or corresponding nucleic acid molecule. As used herein, a “large extramembrane domain” includes a domain having greater than 20 amino acid residues that is found between transmembrane domains, preferably on the cytoplasmic side of the plasma membrane, and does not span or traverse the plasma membrane. A large extramembrane domain preferably includes at least one, two, three, four or more motifs or consensus sequences characteristic of P-type or E1-E2 ATPases, *i.e.*, includes one, two, three, four, or more “P-type ATPase consensus sequences or motifs”. As used herein, the phrase “P-type ATPase consensus sequences or motifs” includes any consensus sequence or motif known in the art to be characteristic of P-type ATPases, including, but not limited to, the P-type ATPase sequence 1 motif (as defined herein), the P-type ATPase sequence 2 motif (as defined herein), the P-type ATPase sequence 3 motif (as defined herein), and the E1-E2 ATPases phosphorylation site (as defined herein).

In one embodiment, the family of HEAT proteins of the present invention comprises at least one "N-terminal" large extramembrane domain in the protein or corresponding nucleic acid molecule. As used herein, an "N-terminal" large extramembrane domain is found in the N-terminal 1/3rd of the protein, preferably between the fourth and fifth transmembrane domains of a HEAT protein, and includes about 50-270, 50-250, 60-230, 70-210, 80-190, 90-170, or preferably, 92, 151, or 163 amino acid residues. In a preferred embodiment, an N-terminal large extramembrane domain includes at least one P-type ATPase sequence 1 motif (as described herein). An N-terminal large extramembrane domain was identified in the amino acid sequence of human HEAT-1 at about residues 277-427 of SEQ ID NO:2. An N-terminal large extramembrane domain was also identified in the amino acid sequence of human HEAT-2 at about residues 318-409 of SEQ ID NO:6 and in the amino acid sequence of human HEAT-3 at about residues 282-444 of SEQ ID NO:9.

The family of HEAT proteins of the present invention also comprises at least one "C-terminal" large extramembrane domain in the protein or corresponding nucleic acid molecule. As used herein, a "C-terminal" large extramembrane domain is found in the C-terminal 2/3rds of the protein, preferably between the sixth and seventh transmembrane domains of a HEAT protein and includes about 340-590, 360-570, 380-550, 400-530, 420-510, 440-490, or preferably, 451, 471, or 480 amino acid residues. In a preferred embodiment, a C-terminal large extramembrane domain includes at least one or more of the following motifs: a P-type ATPase sequence 2 motif (as described herein), a P-type ATPase sequence 3 motif (as defined herein), and/or an E1-E2 ATPases phosphorylation site (as defined herein). A C-terminal large extramembrane domain was identified in the amino acid sequence of human HEAT-1 at about residues 485-935 of SEQ ID NO:2, in the amino acid sequence of human HEAT-2 at about residues 470-940 of SEQ ID NO:6, and in the amino acid sequence of human HEAT-3 at about residues 510-989 of SEQ ID NO:9.

In another embodiment, a HEAT protein of the present invention includes at least one "P-type ATPase sequence 1 motif" in the protein or corresponding nucleic acid molecule. As used herein, a "P-type ATPase sequence 1 motif" is a conserved sequence motif diagnostic for P-type ATPases (Tang, X. *et al.* (1996) *Science* 272:1495-1497; Fagan, M.J. and Saier, M.H. (1994) *J. Mol. Evol.* 38:57). A P-type ATPase sequence 1 motif is involved in the coupling of ATP hydrolysis with transport (*e.g.*, transport of Ca²⁺). The consensus sequence for a P-type ATPase sequence 1 motif is [DNS]-[QENR]-[SA]-[LIVSAN]-[LIV]-[TSN]-G-E-[SN] (SEQ ID NO:13). The use of amino acids in brackets indicates that the amino acid at the indicated position may be any one of the amino acids within the brackets, *e.g.*, [SA] indicates any of one of either S (serine) or A (alanine). In a preferred embodiment, a

P-type ATPase sequence 1 motif is contained within an N-terminal large extramembrane domain. In another preferred embodiment, a P-type ATPase sequence 1 motif in the HEAT proteins of the present invention has at least 1, 2, 3, 4, 5, 6, 7, 8 or more amino acid residues which match the consensus sequence for a P-type ATPase sequence 1 motif. A P-type ATPase sequence 1 motif was identified in the amino acid sequence of human HEAT-1 at about residues 341-349 of SEQ ID NO:2, in the amino acid sequence of human HEAT-2 at about residues 318-326 of SEQ ID NO:6, and in the amino acid sequence of human HEAT-3 at about residues 348-356 of SEQ ID NO:9.

In another embodiment, a HEAT protein of the present invention includes at least one "P-type ATPase sequence 2 motif" in the protein or corresponding nucleic acid molecule. As used herein, a "P-type ATPase sequence 2 motif" is a conserved sequence motif diagnostic for P-type ATPases (Tang, X. *et al.* (1996) *Science* 272:1495-1497; Fagan, M.J. and Saier, M.H. (1994) *J. Mol. Evol.* 38:57). Preferably, a P-type ATPase sequence 2 motif overlaps with and/or includes an E1-E2 ATPases phosphorylation site (as defined herein). The consensus sequence for a P-type ATPase sequence 2 motif is [LIV]-[CAML]-[STFL]-D-K-T-G-T-[LI]-T (SEQ ID NO:14). The use of amino acids in brackets indicates that the amino acid at the indicated position may be any one of the amino acids within the brackets, *e.g.*, [LI] indicates any of one of either L (leucine) or I (isoleucine). In a preferred embodiment, a P-type ATPase sequence 2 motif is contained within a C-terminal large extramembrane domain. In another preferred embodiment, a P-type ATPase sequence 2 motif in the HEAT proteins of the present invention has at least 1, 2, 3, 4, 5, 6, 7, 8, 9 or more amino acid residues which match the consensus sequence for a P-type ATPase sequence 2 motif. A P-type ATPase sequence 2 motif was identified in the amino acid sequence of human HEAT-1 at about residues 510-519 of SEQ ID NO:2, in the amino acid sequence of human HEAT-2 at about residues 495-504 of SEQ ID NO:6, and in the amino acid sequence of human HEAT-3 at about residues 530-539 of SEQ ID NO:9.

In yet another embodiment, a HEAT protein of the present invention includes at least one "P-type ATPase sequence 3 motif" in the protein or corresponding nucleic acid molecule. As used herein, a "P-type ATPase sequence 3 motif" is a conserved sequence motif diagnostic for P-type ATPases (Tang, X. *et al.* (1996) *Science* 272:1495-1497; Fagan, M.J. and Saier, M.H. (1994) *J. Mol. Evol.* 38:57). A P-type ATPase sequence 3 motif is involved in ATP binding. The consensus sequence for a P-type ATPase sequence 3 motif is [TIV]-G-D-G-X-N-D-[ASG]-P-[ASV]-L (SEQ ID NO:15). X indicates that the amino acid at the indicated position may be any amino acid (*i.e.*, is not conserved). The use of amino acids in brackets indicates that the

amino acid at the indicated position may be any one of the amino acids within the brackets, *e.g.*, [TIV] indicates any of one of either T (threonine), I (isoleucine), or V (valine). In a preferred embodiment, a P-type ATPase sequence 3 motif is contained within a C-terminal large extramembrane domain. In another preferred embodiment,

5 a P-type ATPase sequence 3 motif in the HEAT proteins of the present invention has at least 1, 2, 3, 4, 5, 6, 7, 8 or more amino acid residues (including the amino acid at the position indicated by "X") which match the consensus sequence for a P-type ATPase sequence 3 motif. A P-type ATPase sequence 3 motif was identified in the amino acid sequence of human HEAT-1 at about residues 876-886 of SEQ ID NO:2,

10 in the amino acid sequence of human HEAT-2 at about residues 881-891 of SEQ ID NO:6, and in the amino acid sequence of human HEAT-3 at about residues 862-872 of SEQ ID NO:9.

Isolated HEAT proteins of the present invention have an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2, 6, or 9, or are

15 encoded by a nucleotide sequence sufficiently homologous to SEQ ID NO:1, 3, 5, 7, 8, or 10. As used herein, the term "sufficiently homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence

20 such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%,

25 99.8%, 99.9% or more homology or identity across the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as sufficiently homologous. Furthermore, amino acid or nucleotide sequences which share at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%,

30 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more homology or identity and share a common functional activity are defined herein as sufficiently homologous.

In a preferred embodiment, a HEAT protein includes at least one or more of the following domains or motifs: a transmembrane domain, an E1-E2 ATPase domain, an E1-E2 ATPases phosphorylation site, an N-terminal large extramembrane

35 domain, a C-terminal large extramembrane domain, a P-type ATPase sequence 1 motif, a P-type ATPase sequence 2 motif, and/or a P-type ATPase sequence 3 motif, and has an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%,

99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more homologous or identical to the amino acid sequence of SEQ ID NO:2, 6, or 9, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____. In yet another preferred embodiment, a HEAT

5 protein includes at least one or more of the following domains or motifs: a transmembrane domain, an E1-E2 ATPase domain, an E1-E2 ATPases phosphorylation site, an N-terminal large extramembrane domain, a C-terminal large extramembrane domain, a P-type ATPase sequence 1 motif, a P-type ATPase sequence 2 motif, and/or a P-type ATPase sequence 3 motif, and is encoded by a

10 nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10. In another preferred embodiment, a HEAT protein includes at least one or more of the following domains or motifs: a transmembrane domain, an E1-E2 ATPase domain, an E1-E2 ATPases phosphorylation site, an N-terminal large extramembrane domain, a C-terminal large

15 extramembrane domain, a P-type ATPase sequence 1 motif, a P-type ATPase sequence 2 motif, and/or a P-type ATPase sequence 3 motif, and has a HEAT activity.

As used interchangeably herein, a "HEAT activity", "biological activity of

20 HEAT" or "functional activity of HEAT", includes an activity exerted or mediated by a HEAT protein, polypeptide or nucleic acid molecule on a HEAT responsive cell or on a HEAT substrate, as determined *in vivo* or *in vitro*, according to standard techniques. In one embodiment, a HEAT activity is a direct activity, such as an association with a HEAT target molecule. As used herein, a "target molecule" or

25 "binding partner" is a molecule with which a HEAT protein binds or interacts in nature, such that HEAT-mediated function is achieved. A HEAT target molecule can be a non-HEAT molecule or a HEAT protein or polypeptide of the present invention. In an exemplary embodiment, a HEAT target molecule is a HEAT substrate (*e.g.*, a Ca^{2+} ion; ATP; or a non-HEAT protein). A HEAT activity can also be an indirect

30 activity, such as a cellular signaling activity mediated by interaction of the HEAT protein with a HEAT substrate (*e.g.*, regulation of vascular smooth muscle tone, cellular growth and/or proliferation, and/or angiogenesis).

In a preferred embodiment, a HEAT activity is at least one of the following activities: (i) interaction with a HEAT substrate or target molecule (*e.g.*, a Ca^{2+} ion;

35 ATP; or a non-HEAT protein); (ii) transport of a HEAT substrate or target molecule (*e.g.*, a Ca^{2+} ion) from one side of a biological membrane to the other; (iii) the ability to be phosphorylated or dephosphorylated; (iv) adoption of an E1 conformation or an E2 conformation; (v) conversion of a HEAT substrate or target molecule to a product

- (*e.g.*, hydrolysis of ATP to ADP and free phosphate); (vi) interaction with a second non-HEAT protein; (vii) modulation of intra- or inter-cellular signaling and/or gene transcription (*e.g.*, either directly or indirectly); (viii) modulation of vascular smooth muscle tone; (ix) modulation of cellular growth and/or proliferation; and/or (x) modulation of angiogenesis.

The nucleotide sequence of the isolated human HEAT-1 cDNA and the predicted amino acid sequence encoded by the HEAT-1 cDNA are shown in Figures 1A-1D and in SEQ ID NOs:1 and 2, respectively. A plasmid containing the human HEAT-1 cDNA was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Number _____. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit were made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

The human HEAT-1 gene, which is approximately 4055 nucleotides in length, encodes a protein having a molecular weight of approximately 129.8 kD and which is approximately 1180 amino acid residues in length.

The nucleotide sequence of the isolated human HEAT-2 cDNA and the predicted amino acid sequence encoded by the HEAT-2 cDNA are shown in Figures 11A-11E and in SEQ ID NOs:5 and 6, respectively. A plasmid containing the human HEAT-2 cDNA was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Number _____. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit were made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

The human HEAT-2 gene, which is approximately 7249 nucleotides in length, encodes a protein having a molecular weight of approximately 138.2 kD and which is approximately 1256 amino acid residues in length.

The nucleotide sequence of the isolated human HEAT-3 cDNA and the predicted amino acid sequence encoded by the HEAT-3 cDNA are shown in Figures 22A-22D and in SEQ ID NOs:8 and 9, respectively. A plasmid containing the human HEAT-3 cDNA was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Number _____. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms

for the Purposes of Patent Procedure. This deposit were made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

5 The human HEAT-3 gene, which is approximately 3919 nucleotides in length, encodes a protein having a molecular weight of approximately 132.5 kD and which is approximately 1204 amino acid residues in length.

Various aspects of the invention are described in further detail in the following subsections:

10

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode HEAT proteins or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify HEAT-encoding
15 nucleic acid molecules (*e.g.*, HEAT mRNA) and fragments for use as PCR primers for the amplification or mutation of HEAT nucleic acid molecules. As used herein, the term “nucleic acid molecule” is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-
20 stranded or double-stranded, but preferably is double-stranded DNA.

The term “isolated nucleic acid molecule” includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term “isolated” includes nucleic acid molecules which are separated from the
25 chromosome with which the genomic DNA is naturally associated. 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. For example, in various embodiments, the isolated HEAT nucleic acid molecule can contain less than about 5
30 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an “isolated” nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or
35 other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession

- Number _____, _____, or _____, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as
- 5 Accession Number _____, _____, or _____, as hybridization probes, HEAT nucleic acid molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, J. *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).
- 10 Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____ can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide
- 15 sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____.

- A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified
- 20 can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to HEAT nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

- In one embodiment, an isolated nucleic acid molecule of the invention
- 25 comprises the nucleotide sequence shown in SEQ ID NO:1, 3, 5, 7, 8, or 10. This cDNA may comprise sequences encoding the human HEAT-1 protein (*e.g.*, the “coding region”, from nucleotides 210-3749), as well as 5’ untranslated sequences (nucleotides 1-209) and 3’ untranslated sequences (nucleotides 3750-4055) of SEQ ID NO:1. The cDNA may also comprise sequences encoding human the HEAT-2
- 30 protein (*e.g.*, the “coding region”, from nucleotides 225-3992), as well as 5’ untranslated sequences (nucleotides 1-224) and 3’ untranslated sequences (nucleotides 3993-7249) of SEQ ID NO:5. The cDNA may also comprise sequences encoding the human HEAT-3 protein (*e.g.*, the “coding region”, from nucleotides 68-3679), as well as 5’ untranslated sequences (nucleotides 1-67) and 3’ untranslated
- 35 sequences (nucleotides 3680-3919) of SEQ ID NO:8. Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:1 (*e.g.*, nucleotides 210-3749, corresponding to SEQ ID NO:3), SEQ ID NO:5 (*e.g.*, nucleotides 225-3992, corresponding to SEQ ID NO:7), or SEQ ID NO:8 (*e.g.*, nucleotides 68-3679,

corresponding to SEQ ID NO:10). Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention comprises SEQ ID NO:3 and nucleotides 1-209 of SEQ ID NO:1. In another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:3 and nucleotides 3750-4055 of SEQ ID NO:1. In yet another embodiment, an isolated nucleic acid molecule of the invention comprises SEQ ID NO:7 and nucleotides 1-224 of SEQ ID NO:5. In another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:7 and nucleotides 3993-7249 of SEQ ID NO:5. In still another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:10 and nucleotides 1-67 of SEQ ID NO:8. In another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:10 and nucleotides 3680-3919 of SEQ ID NO:8. In yet another embodiment, the nucleic acid molecule consists of the nucleotide sequence set forth as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, OR SEQ ID NO:10. In still another embodiment, the nucleic acid molecule can comprise the coding region of SEQ ID NO:1 (*e.g.*, nucleotides 210-3749, corresponding to SEQ ID NO:3), as well as a stop codon (*e.g.*, nucleotides 3750-3752 of SEQ ID NO:1). In another embodiment, the nucleic acid molecule can comprise the coding region of SEQ ID NO:5 (*e.g.*, nucleotides 225-3992, corresponding to SEQ ID NO:7), as well as a stop codon (*e.g.*, nucleotides 3993-3995 of SEQ ID NO:5). In another embodiment, the nucleic acid molecule can comprise the coding region of SEQ ID NO:8 (*e.g.*, nucleotides 68-3679, corresponding to SEQ ID NO:10), as well as a stop codon (*e.g.*, nucleotides 3680-3682 of SEQ ID NO:5). In yet other embodiments, the nucleic acid molecule can comprise nucleotides 1-28 of SEQ ID NO:1, nucleotides 4016-4055 of SEQ ID NO:1, nucleotides 1-175 of SEQ ID NO:5, nucleotides 410-1225 of SEQ ID NO:5, or nucleotides 224-462 of SEQ ID NO:8.

In still another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide

sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, thereby forming a stable duplex.

In still another embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more identical to the nucleotide sequence shown in SEQ ID NO:1, 3, 5, 7, 8, or 10 (*e.g.*, to the entire length of the nucleotide sequence), or to the nucleotide sequence (*e.g.*, the entire length of the nucleotide sequence) of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, or a portion or complement of any of these nucleotide sequences. In one embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least (or no greater than) 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 1994, 2000, 2050, 2073, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 2950, 3000, 3050, 3100, 3150, 3200, 3250, 3300, 3350, 3400, 3441, 3450, 3500, 3550, 3600, 3650, 3700, 3750, 3800, 3841, 3850, 3900, 3950, 4000, 4050, 4100, 4150, 4200, 4250, 4300, 4350, 4400, 4450, 4500, 4550, 4600, 4650, 4700, 4750, 4800, 4850, 4900, 4950, 5000, 5050, 5100, 5150, 5200, 5250, 5300, 5350, 5400, 5450, 5500, 5550, 5600, 5650, 5700, 5750, 5800, 5850, 5900, 5950, 6000, 6050, 6100, 6150, 6200, 6250, 6300, 6350, 6400, 6450, 6500, 6550, 6600, 6650, 6700, 6750, 6800, 6850, 6900, 6950, 7000, 7050, 7100, 7150, 7200 or more nucleotides in length and hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, for example, a fragment which can be used as a probe or primer or a fragment encoding a portion of a HEAT protein, *e.g.*, a biologically active portion of a HEAT protein. The nucleotide sequence determined from the cloning of the HEAT gene allows for the generation of probes and primers designed for use in identifying and/or cloning other HEAT family members, as well as HEAT homologues from other species. The probe/primer (*e.g.*, oligonucleotide) typically comprises substantially purified oligonucleotide. The oligonucleotide

typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, of an anti-sense sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____.

Exemplary probes or primers are at least (or no greater than) 12 or 15, 20 or 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75 or more nucleotides in length and/or comprise consecutive nucleotides of an isolated nucleic acid molecule described herein. Also included within the scope of the present invention are probes or primers comprising contiguous or consecutive nucleotides of an isolated nucleic acid molecule described herein, but for the difference of 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 bases within the probe or primer sequence. Probes based on the HEAT nucleotide sequences can be used to detect (*e.g.*, specifically detect) transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. In another embodiment a set of primers is provided, *e.g.*, primers suitable for use in a PCR, which can be used to amplify a selected region of a HEAT sequence, *e.g.*, a domain, region, site or other sequence described herein. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differ by no greater than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 bases when compared to a sequence disclosed herein or to the sequence of a naturally occurring variant. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a HEAT protein, such as by measuring a level of a HEAT-encoding nucleic acid in a sample of cells from a subject, *e.g.*, detecting HEAT mRNA levels or determining whether a genomic HEAT gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of a HEAT protein" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, which encodes a polypeptide having a HEAT biological activity (the biological activities of the HEAT proteins are described herein), expressing the encoded portion of the HEAT

protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the HEAT protein. In an exemplary embodiment, the nucleic acid molecule is at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 1994, 2000, 2050, 2073, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 2950, 3000, 3050, 3100, 3150, 3200, 3250, 3300, 3350, 3400, 3441, 3450, 3500, 3550, 3600, 3650, 3700, 3750, 3800, 3841, 3850, 3900, 3950, 4000, 4050, 4100, 4150, 4200, 4250, 4300, 4350, 4400, 4450, 4500, 4550, 4600, 4650, 4700, 4750, 4800, 4850, 4900, 4950, 5000, 5050, 5100, 5150, 5200, 5250, 5300, 5350, 5400, 5450, 5500, 5550, 5600, 5650, 5700, 5750, 5800, 5850, 5900, 5950, 6000, 6050, 6100, 6150, 6200, 6250, 6300, 6350, 6400, 6450, 6500, 6550, 6600, 6650, 6700, 6750, 6800, 6850, 6900, 6950, 7000, 7050, 7100, 7150, 7200 or more nucleotides in length and encodes a protein having a HEAT activity (as described herein).

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, due to degeneracy of the genetic code and thus encode the same HEAT proteins as those encoded by the nucleotide sequence shown in SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence which differs by at least 1, but no greater than 5, 10, 20, 50 or 100 amino acid residues from the amino acid sequence shown in SEQ ID NO:2, 6, or 9, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number _____, _____, or _____. In yet another embodiment, the nucleic acid molecule encodes the amino acid sequence of human HEAT. If an alignment is needed for this comparison, the sequences should be aligned for maximum homology.

Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologues (different locus), and orthologues (different organism) or can be non naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. 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 (as compared in the encoded product).

Allelic variants result, for example, from DNA sequence polymorphisms within a population (*e.g.*, the human population) that lead to changes in the amino acid sequences of the HEAT proteins. Such genetic polymorphism in the HEAT genes may exist among individuals within a population due to natural allelic
5 variation. As used herein, the terms “gene” and “recombinant gene” refer to nucleic acid molecules which include an open reading frame encoding a HEAT protein, preferably a mammalian HEAT protein, and can further include non-coding regulatory sequences, and introns.

Accordingly, in one embodiment, the invention features isolated nucleic acid
10 molecules which encode a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 6, or 9, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, wherein the nucleic acid molecule hybridizes to a complement of a nucleic acid molecule comprising SEQ ID NO:1, 3,
15 5, 7, 8, or 10, for example, under stringent hybridization conditions.

Allelic variants of HEAT, *e.g.*, human HEAT-1, HEAT-2, or HEAT-3, include both functional and non-functional HEAT proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the HEAT protein that maintain the ability to, *e.g.*, bind or interact with a HEAT substrate or target
20 molecule, transport a HEAT substrate or target molecule across a biological membrane, hydrolyze ATP, be phosphorylated or dephosphorylated, adopt an E1 conformation or an E2 conformation, and/or modulate cellular signaling, vascular smooth muscle tone, cellular growth and/or proliferation, and/or angiogenesis. Functional allelic variants will typically contain only a conservative substitution of
25 one or more amino acids of SEQ ID NO:2, 6, or 9, or a substitution, deletion or insertion of non-critical residues in non-critical regions of the protein.

Non-functional allelic variants are naturally occurring amino acid sequence variants of the HEAT protein, *e.g.*, human HEAT-1, HEAT-2, or HEAT-3, that do not have the ability to, *e.g.*, bind or interact with a HEAT substrate or target molecule,
30 transport a HEAT substrate or target molecule across a biological membrane, hydrolyze ATP, be phosphorylated or dephosphorylated, adopt an E1 conformation or an E2 conformation, and/or modulate cellular signaling, vascular smooth muscle tone, cellular growth and/or proliferation, and/or angiogenesis. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion,
35 or premature truncation of the amino acid sequence of SEQ ID NO:2, 6, or 9, or a substitution, insertion, or deletion in critical residues or critical regions of the protein.

The present invention further provides non-human orthologues (*e.g.*, non-human orthologues of the human HEAT-1, HEAT-2, or HEAT-3 proteins). Orthologues of the human HEAT proteins are proteins that are isolated from non-human organisms and possess the same HEAT substrate or target molecule binding mechanisms, Ca^{2+} transporting activity, ATPase activity, and/or modulation of cellular signaling mechanisms of the human HEAT proteins. Orthologues of the human HEAT proteins can readily be identified as comprising an amino acid sequence that is substantially homologous to SEQ ID NO:2, 6, or 9.

Moreover, nucleic acid molecules encoding other HEAT family members and, thus, which have a nucleotide sequence which differs from the HEAT sequences of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____ are intended to be within the scope of the invention. For example, another HEAT cDNA can be identified based on the nucleotide sequence of human HEAT-1, HEAT-2, or HEAT-3. Moreover, nucleic acid molecules encoding HEAT proteins from different species, and which, thus, have a nucleotide sequence which differs from the HEAT sequences of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____ are intended to be within the scope of the invention. For example, a mouse or monkey HEAT cDNA can be identified based on the nucleotide sequence of a human HEAT-1, HEAT-2, or HEAT-3.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the HEAT cDNAs of the invention can be isolated based on their homology to the HEAT nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Nucleic acid molecules corresponding to natural allelic variants and homologues of the HEAT cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the HEAT gene.

Orthologues, homologues, and allelic variants can be identified using methods known in the art (*e.g.*, by hybridization to an isolated nucleic acid molecule of the present invention, for example, under stringent hybridization conditions). In one embodiment, an isolated nucleic acid molecule of the invention is at least 15, 20, 25, 30 or more nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____. In other embodiment, the nucleic acid is at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650,

700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 1994, 2000, 2050, 2073, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 2950, 3000, 3050, 3100, 3150, 3200, 3250, 3300, 3350, 3400, 3441, 3450, 3500, 3550, 3600, 3650, 3700, 3750, 3800, 3841, 3850, 3900, 3950, 4000, 4050, 4100, 4150, 4200, 4250, 4300, 4350, 4400, 4450, 4500, 4550, 4600, 4650, 4700, 4750, 4800, 4850, 4900, 4950, 5000, 5050, 5100, 5150, 5200, 5250, 5300, 5350, 5400, 5450, 5500, 5550, 5600, 5650, 5700, 5750, 5800, 5850, 5900, 5950, 6000, 6050, 6100, 6150, 6200, 6250, 6300, 6350, 6400, 6450, 6500, 6550, 6600, 6650, 6700, 6750, 6800, 6850, 6900, 6950, 7000, 7050, 7100, 7150, 7200 or more nucleotides in length.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences that are significantly identical or homologous to each other remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% identical to each other 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*, Ausubel *et al.*, eds., John Wiley & Sons, Inc. (1995), sections 2, 4, and 6. Additional stringent conditions can be found in *Molecular Cloning: A Laboratory Manual*, Sambrook *et al.*, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), chapters 7, 9, and 11. A preferred, non-limiting example of stringent hybridization conditions includes hybridization in 4X sodium chloride/sodium citrate (SSC), at about 65-70°C (or alternatively hybridization in 4X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 1X SSC, at about 65-70°C. A preferred, non-limiting example of highly stringent hybridization conditions includes hybridization in 1X SSC, at about 65-70°C (or alternatively hybridization in 1X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 0.3X SSC, at about 65-70°C. A preferred, non-limiting example of reduced stringency hybridization conditions includes hybridization in 4X SSC, at about 50-60°C (or alternatively hybridization in 6X SSC plus 50% formamide at about 40-45°C) followed by one or more washes in 2X SSC, at about 50-60°C. Ranges intermediate to the above-recited values, *e.g.*, at 65-70°C or at 42-50°C are also intended to be encompassed by the present invention. SSPE (1xSSPE is 0.15M NaCl, 10mM NaH₂PO₄, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1X SSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes each after hybridization is complete. The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-

10°C less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, $T_m(^{\circ}\text{C}) = 2(\# \text{ of A} + \text{T bases}) + 4(\# \text{ of G} + \text{C bases})$. For hybrids between 18 and 49 base pairs in length, $T_m(^{\circ}\text{C}) = 81.5 + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\% \text{G} + \text{C}) - (600/\text{N})$, where

5 N is the number of bases in the hybrid, and $[\text{Na}^+]$ is the concentration of sodium ions in the hybridization buffer ($[\text{Na}^+]$ for 1X SSC = 0.165 M). It will also be recognized by the skilled practitioner that additional reagents may be added to hybridization and/or wash buffers to decrease non-specific hybridization of nucleic acid molecules to membranes, for example, nitrocellulose or nylon membranes, including but not

10 limited to blocking agents (e.g., BSA or salmon or herring sperm carrier DNA), detergents (e.g., SDS), chelating agents (e.g., EDTA), Ficoll, PVP and the like. When using nylon membranes, in particular, an additional preferred, non-limiting example of stringent hybridization conditions is hybridization in 0.25-0.5M NaH_2PO_4 , 7% SDS at about 65°C, followed by one or more washes at 0.02M NaH_2PO_4 , 1% SDS at 65°C

15 (see e.g., Church and Gilbert (1984) *Proc. Natl. Acad. Sci. USA* 81:1991-1995), or alternatively 0.2X SSC, 1% SDS.

Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a

20 “naturally-occurring” nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In addition to naturally-occurring allelic variants of the HEAT sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:1, 3, 5, 7, 8,

25 or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, thereby leading to changes in the amino acid sequence of the encoded HEAT proteins, without altering the functional ability of the HEAT proteins. For example, nucleotide substitutions leading to amino acid substitutions at “non-essential” amino acid residues can be

30 made in the sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____. A “non-essential” amino acid residue is a residue that can be altered from the wild-type sequence of HEAT-2 or HEAT-3 (e.g., the sequence of SEQ ID NO:2, 6, or 9) without altering the biological activity, whereas an “essential”

35 amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the HEAT proteins of the present invention, e.g., those present in an E1-E2 ATPase domain or an E1-E2 ATPases phosphorylation site, are predicted to be particularly unamenable to alteration. Furthermore, additional

amino acid residues that are conserved between the HEAT proteins of the present invention and other members of the E1-E2 ATPase family are not likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding HEAT proteins that contain changes in amino acid residues that are not essential for activity. Such HEAT proteins differ in amino acid sequence from SEQ ID NO:2, 6, or 9, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more homologous to SEQ ID NO:2, 6, or 9, *e.g.*, to the entire length of SEQ ID NO:2, 6, or 9.

An isolated nucleic acid molecule encoding a HEAT protein homologous to the protein of SEQ ID NO:2, 6, or 9 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____ by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A “conservative amino acid substitution” is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine, tryptophan), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a HEAT protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a HEAT coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for HEAT biological activity to identify mutants

that retain activity. Following mutagenesis of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

- 5 In a preferred embodiment, a mutant HEAT protein can be assayed for the ability to (i) interact with a HEAT substrate or target molecule (*e.g.*, a Ca^{2+} ion; ATP; or a non-HEAT protein;); (ii) transport of HEAT substrate or target molecule (*e.g.*, a Ca^{2+} ion) from one side of a biological membrane to the other; (iii) be phosphorylated or dephosphorylated; (iv) adopt an E1 conformation or an E2 conformation; (v)
- 10 convert a HEAT substrate or target molecule to a product (*e.g.*, hydrolyze ATP to ADP and free phosphate); (vi) interact with a second non-HEAT protein; (vii) modulate intra- or inter-cellular signaling and/or gene transcription (*e.g.*, either directly or indirectly); (viii) modulate vascular smooth muscle tone; (ix) modulate cellular growth and/or proliferation; and/or (x) modulate angiogenesis.

- 15 In addition to the nucleic acid molecules encoding HEAT proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. In an exemplary embodiment, the invention provides an isolated nucleic acid molecule which is antisense to a HEAT nucleic acid molecule (*e.g.*, is antisense to the coding strand of a HEAT nucleic acid molecule). An
- 20 “antisense” nucleic acid comprises a nucleotide sequence which is complementary to a “sense” nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire HEAT coding strand, or
- 25 to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to “coding region sequences” of the coding strand of a nucleotide sequence encoding HEAT-1, HEAT-2, or HEAT-3. The term “coding region sequences” refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the coding region sequences of human HEAT-1, HEAT-2,
- 30 or HEAT-3, corresponding to SEQ ID NO:3, 7, or 10, respectively). In another embodiment, the antisense nucleic acid molecule is antisense to a “noncoding region” of the coding strand of a nucleotide sequence encoding HEAT-1, HEAT-2, or HEAT-3. The term “noncoding region” refers to 5’ (*e.g.*, nucleotides 1-209 of SEQ ID NO:1 (HEAT-1), nucleotides 1-224 of SEQ ID NO:5 (HEAT-2), or nucleotides 1-67 of
- 35 SEQ ID NO:8 (HEAT-3)) and/or 3’ sequences (*e.g.*, nucleotides 3750-4055 of SEQ ID NO:1 (HEAT-1), nucleotides 3993-7249 of SEQ ID NO:5 (HEAT-2) or nucleotides 3680-3919 of SEQ ID NO:8 (HEAT-3)) which flank the coding region

sequences that are not translated into amino acids (also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding HEAT-1, HEAT-2, and HEAT-3 disclosed herein (*e.g.*, SEQ ID NOs:3, 7, and 10, respectively), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to coding region sequences of HEAT-1, HEAT-2, or HEAT-3 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the HEAT-1, HEAT-2, or HEAT-3 mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a HEAT protein to thereby inhibit

expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-*o*-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haseloff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave HEAT mRNA transcripts to thereby inhibit translation of HEAT mRNA. A ribozyme having specificity for a HEAT-encoding nucleic acid can be designed based upon the nucleotide sequence of a HEAT cDNA disclosed herein (*i.e.*, SEQ ID NO:1, 3, 5, 7, 8, or 10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a HEAT-encoding mRNA. See, *e.g.*, Cech *et al.*, U.S. Patent No. 4,987,071; and Cech *et al.*, U.S. Patent No. 5,116,742. Alternatively, HEAT mRNA can be used to select a catalytic RNA having a specific ribonuclease

activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, HEAT gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the HEAT gene (*e.g.*, the HEAT-1 promoter and/or enhancers; the HEAT-2 promoter and/or enhancers; or the HEAT-3 promoter and/or enhancers) to form triple helical structures that prevent transcription of the HEAT gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioessays* 14(12):807-15.

10 In yet another embodiment, the HEAT nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup, B. and Nielsen, P.E. (1996) *Bioorg. Med.* 15 *Chem.* 4(1):5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup and Nielsen (1996) *supra* and Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:14670-675.

PNAs of HEAT nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of HEAT nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene (*e.g.*, by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes (*e.g.*, S1 nucleases (Hyrup and Nielsen (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup and Nielsen (1996) *supra*; Perry-O'Keefe *et al.* (1996) *supra*).

In another embodiment, PNAs of HEAT can be modified (*e.g.*, to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of HEAT nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA

- portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup and Nielsen (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in
- 5 Hyrup and Nielsen (1996) *supra* and Finn, P.J. *et al.* (1996) *Nucleic Acids Res.* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acids Res.*
- 10 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn, P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5:1119-11124).
- 15 In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO 88/09810) or the blood-brain barrier (see, *e.g.*, PCT
- 20 Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, *e.g.*, Krol *et al.* (1988) *Biotechniques* 6:958-976) or intercalating agents (See, *e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered
- 25 cleavage agent).

II. Isolated HEAT Proteins and Anti-HEAT Antibodies

- One aspect of the invention pertains to isolated or recombinant HEAT proteins and polypeptides, and biologically active portions thereof, as well as polypeptide
- 30 fragments suitable for use as immunogens to raise anti-HEAT antibodies. In one embodiment, native HEAT proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, HEAT proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a HEAT protein or polypeptide can be
- 35 synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the HEAT protein is derived, or substantially free from

chemical precursors or other chemicals when chemically synthesized. The language “substantially free of cellular material” includes preparations of HEAT protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language “substantially
5 free of cellular material” includes preparations of HEAT protein having less than about 30% (by dry weight) of non-HEAT protein (also referred to herein as a “contaminating protein”), more preferably less than about 20% of non-HEAT protein, still more preferably less than about 10% of non-HEAT protein, and most preferably less than about 5% non-HEAT protein. When the HEAT protein or biologically
10 active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language “substantially free of chemical precursors or other chemicals”
15 includes preparations of HEAT protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language “substantially free of chemical precursors or other chemicals” includes preparations of HEAT protein having less than about 30% (by dry weight) of chemical precursors or non-HEAT chemicals, more
20 preferably less than about 20% chemical precursors or non-HEAT chemicals, still more preferably less than about 10% chemical precursors or non-HEAT chemicals, and most preferably less than about 5% chemical precursors or non-HEAT chemicals.

As used herein, a “biologically active portion” of a HEAT protein includes a fragment of a HEAT protein which participates in an interaction between a HEAT
25 molecule and a non-HEAT molecule (*e.g.*, a HEAT substrate such as Ca^{2+} , ATP or a non-HEAT protein). Biologically active portions of a HEAT protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the HEAT amino acid sequences, *e.g.*, the amino acid sequences shown in SEQ ID NO:2, 6, or 9, which include sufficient amino acid residues to exhibit at least one activity of
30 a HEAT protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the HEAT protein, *e.g.*, the ability to interact with a HEAT substrate or target molecule (*e.g.*, a Ca^{2+} ion; ATP; or a non-HEAT protein;); the ability to transport a HEAT substrate or target molecule (*e.g.*, a Ca^{2+} ion) from one side of a biological membrane to the other; the ability to be phosphorylated or
35 dephosphorylated; the ability to adopt an E1 conformation or an E2 conformation; the ability to convert a HEAT substrate or target molecule to a product (*e.g.*, the ability to hydrolyze ATP to ADP and free phosphate); the ability to interact with a second non-HEAT protein; the ability to modulate intra- or inter-cellular signaling and/or gene

transcription (*e.g.*, either directly or indirectly); the ability to modulate vascular smooth muscle tone; the ability to modulate cellular growth and/or proliferation; and/or the ability to modulate angiogenesis. A biologically active portion of a HEAT protein can be a polypeptide which is, for example, 10, 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 691, 700, 750, 800, 850, 900, 950, 1000, 1100, 1050, 1060, 1072, 1100, 1150, 1200, 1250 or more amino acids in length. Biologically active portions of a HEAT protein can be used as targets for developing agents which modulate a HEAT mediated activity, *e.g.*, any of the aforementioned HEAT activities.

10 In one embodiment, a biologically active portion of a HEAT protein comprises at least at least one or more of the following domains or motifs: a transmembrane domain, an E1-E2 ATPase domain, an E1-E2 ATPases phosphorylation site, an N-terminal large extramembrane domain, a C-terminal large extramembrane domain, a P-type ATPase sequence 1 motif, a P-type ATPase
15 sequence 2 motif, and/or a P-type ATPase sequence 3 motif. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native HEAT protein.

Another aspect of the invention features fragments of the protein having the
20 amino acid sequence of SEQ ID NO:2, 6, or 9, for example, for use as immunogens. In one embodiment, a fragment comprises at least 8 amino acids (*e.g.*, contiguous or consecutive amino acids) of the amino acid sequence of SEQ ID NO:2, 6, or 9, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number _____, _____, or _____. In another embodiment, a
25 fragment comprises at least 8, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more amino acids (*e.g.*, contiguous or consecutive amino acids) of the amino acid sequence of SEQ ID NO:2, 6, or 9, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number _____, _____, or _____.

In a preferred embodiment, a HEAT protein has an amino acid sequence
30 shown in SEQ ID NO:2, 6, or 9. In other embodiments, the HEAT protein is substantially identical to SEQ ID NO:2, 6, or 9, and retains the functional activity of the protein of SEQ ID NO:2, 6, or 9, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. In another embodiment, the HEAT protein is a protein which comprises an amino acid
35 sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more identical to SEQ ID NO:2, 6, or 9.

In another embodiment, the invention features a HEAT protein which is encoded by a nucleic acid molecule consisting of a nucleotide sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more identical to a nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or a complement thereof. This invention further features a HEAT protein which is encoded by a nucleic acid molecule consisting of a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or a complement thereof.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (*e.g.*, when aligning a second sequence to the HEAT-1 amino acid sequence of SEQ ID NO:2 having 1180 amino acid residues, at least 354, preferably at least 472, more preferably at least 590, even more preferably at least 708, and even more preferably at least 826, 944, or 1062 amino acid residues are aligned; when aligning a second sequence to the HEAT-2 amino acid sequence of SEQ ID NO:6 having 1256 amino acid residues, at least 377, preferably at least 502, more preferably at least 628, even more preferably at least 755, and even more preferably at least 879, 1005 or 1130 amino acid residues are aligned; when aligning a second sequence to the HEAT-3 amino acid sequence of SEQ ID NO:9 having 1204 amino acid residues, at least 361, preferably at least 482, more preferably at least 602, even more preferably at least 722, and even more preferably at least 844, 963 or 1084 amino acid residues are aligned). 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 between two sequences can be accomplished using a mathematical algorithm. 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 online through the Genetics Computer Group), using either a Blossum 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 (available at online through the Genetics Computer Group), 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. A preferred, non-limiting example of parameters to be used in conjunction with the GAP program include a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of Meyers and Miller (*Comput. Appl. Biosci.* 4:11-17 (1988)) which has been incorporated into the ALIGN program (version 2.0 or version 2.0U), 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 public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to HEAT 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 HEAT protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See the website for the National Center for Biotechnology Information.

The invention also provides HEAT chimeric or fusion proteins. As used herein, a HEAT "chimeric protein" or "fusion protein" comprises a HEAT polypeptide operatively linked to a non-HEAT polypeptide. A "HEAT polypeptide" refers to a polypeptide having an amino acid sequence corresponding to HEAT-1, HEAT-2, or HEAT-3, for example, whereas a "non-HEAT polypeptide" refers to a

polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the HEAT protein, *e.g.*, a protein which is different from the HEAT protein and which is derived from the same or a different organism.

Within a HEAT fusion protein the HEAT polypeptide can correspond to all or a portion of a HEAT protein. In a preferred embodiment, a HEAT fusion protein comprises at least one biologically active portion of a HEAT protein. In another preferred embodiment, a HEAT fusion protein comprises at least two biologically active portions of a HEAT protein. Within the fusion protein, the term “operatively linked” is intended to indicate that the HEAT polypeptide and the non-HEAT polypeptide are fused in-frame to each other. The non-HEAT polypeptide can be fused to the N-terminus or C-terminus of the HEAT polypeptide.

For example, in one embodiment, the fusion protein is a GST-HEAT fusion protein in which the HEAT sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant HEAT. In another embodiment, the fusion protein is a HEAT protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of HEAT can be increased through use of a heterologous signal sequence.

In another embodiment, a HEAT “chimeric protein” or “fusion protein” comprises a HEAT protein of the present invention wherein one or more domains or motifs in the HEAT protein are replaced with the corresponding domains or motifs from another HEAT protein or another E1-E1 or P-type ATPase. Such HEAT chimeric proteins are useful, for example, for determining how individual motifs or domains contribute to or influence the activity of a HEAT protein. Such HEAT chimeric proteins are also useful, for example, in creating HEAT proteins with certain activities of one HEAT protein and certain activities of a different HEAT protein, or with certain activities of a HEAT protein and certain activities of a different E1-E2 or P-type ATPase.

The HEAT fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The HEAT fusion proteins can be used to affect the bioavailability of a HEAT substrate. Use of HEAT fusion proteins may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a HEAT protein; (ii) mis-regulation of the HEAT gene; and (iii) aberrant post-translational modification of a HEAT protein.

Moreover, the HEAT-fusion proteins of the invention can be used as immunogens to produce anti-HEAT antibodies in a subject, to purify HEAT substrates, and in screening assays to identify molecules which inhibit or enhance the interaction of HEAT with a HEAT substrate.

5 Preferably, a HEAT chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini,
10 filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two
15 consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.*, John Wiley & Sons:1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A HEAT-encoding nucleic acid can be cloned into such an expression
20 vector such that the fusion moiety is linked in-frame to the HEAT protein.

The present invention also pertains to variants of the HEAT proteins which function as either HEAT agonists (mimetics) or as HEAT antagonists. Variants of the HEAT proteins can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of a HEAT protein. An agonist of the HEAT proteins can retain
25 substantially the same, or a subset, of the biological activities of the naturally occurring form of a HEAT protein. An antagonist of a HEAT protein can inhibit one or more of the activities of the naturally occurring form of the HEAT protein by, for example, competitively modulating a HEAT-mediated activity of a HEAT protein. Thus, specific biological effects can be elicited by treatment with a variant of limited
30 function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the HEAT protein.

In one embodiment, variants of a HEAT protein which function as either
35 HEAT agonists (mimetics) or as HEAT antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of a HEAT protein for HEAT protein agonist or antagonist activity. In one embodiment, a variegated library of HEAT variants is generated by combinatorial mutagenesis at the nucleic acid level

and is encoded by a variegated gene library. A variegated library of HEAT variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential HEAT sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of HEAT sequences therein. There are a variety of methods which can be used to produce libraries of potential HEAT variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector.

10 Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential HEAT sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acids Res.* 11:477.

15 In addition, libraries of fragments of a HEAT protein coding sequence can be used to generate a variegated population of HEAT fragments for screening and subsequent selection of variants of a HEAT protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a HEAT coding sequence with a nuclease under conditions wherein

20 nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be

25 derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the HEAT protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are

30 adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of HEAT proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the

35 combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the

screening assays to identify HEAT variants (Arkin and Youvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delagrave *et al.* (1993) *Protein Eng.* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated HEAT library. For example, a library of expression vectors can be transfected into a cell line which ordinarily responds to HEAT in a particular HEAT substrate-dependent manner. The transfected cells are then contacted with HEAT and the effect of the expression of the mutant on signaling by the HEAT substrate can be detected by measuring *e.g.*, Ca^{2+} transport (*e.g.*, by measuring Ca^{2+} levels inside the cell or its various cellular compartments, or in the extracellular medium), hydrolysis of ATP, phosphorylation or dephosphorylation of the HEAT protein, and/or gene transcription. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the HEAT substrate, or which score for increased or decreased levels of Ca^{2+} transport or ATP hydrolysis, and the individual clones further characterized.

An isolated HEAT protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind HEAT using standard techniques for polyclonal and monoclonal antibody preparation. A full-length HEAT protein can be used or, alternatively, the invention provides antigenic peptide fragments of HEAT for use as immunogens. The antigenic peptide of HEAT comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2, 6, or 9 and encompasses an epitope of HEAT such that an antibody raised against the peptide forms a specific immune complex with HEAT. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of HEAT that are located on the surface of the protein, *e.g.*, hydrophilic regions, as well as regions with high antigenicity (see, for example, Figures 2, 13, and 24).

A HEAT immunogen typically is used to prepare antibodies by immunizing a suitable subject (*e.g.*, rabbit, goat, mouse, or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed HEAT protein or a chemically-synthesized HEAT polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic HEAT preparation induces a polyclonal anti-HEAT antibody response.

Accordingly, another aspect of the invention pertains to anti-HEAT antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as HEAT. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind HEAT. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of HEAT. A monoclonal antibody composition thus typically displays a single binding affinity for a particular HEAT protein with which it immunoreacts.

Polyclonal anti-HEAT antibodies can be prepared as described above by immunizing a suitable subject with a HEAT immunogen. The anti-HEAT antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized HEAT. If desired, the antibody molecules directed against HEAT can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the anti-HEAT antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497 (see also Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem.* 255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol. Today* 4:72), the EBV-hybridoma technique (Cole *et al.* (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally Kenneth, R.H., in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); Lerner, E.A. (1981) *Yale J. Biol. Med.*, 54:387-402; Gefter, M.L. *et al.* (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a HEAT immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds HEAT.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-HEAT monoclonal antibody (see, *e.g.*, Galfre, G. *et al.* (1977) *Nature* 266:55052; Gefter *et al.* (1977) *supra*; Lerner (1981) *supra*; Kenneth (1980) *supra*). Moreover, the
5 ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (*e.g.*, a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an
10 immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, *e.g.*, the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from
15 ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are
20 detected by screening the hybridoma culture supernatants for antibodies that bind HEAT, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-HEAT antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display
25 library) with HEAT to thereby isolate immunoglobulin library members that bind HEAT. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in
30 generating and screening antibody display library can be found in, for example, Ladner *et al.*, U.S. Patent No. 5,223,409; Kang *et al.*, PCT International Publication No. WO 92/18619; Dower *et al.*, PCT International Publication No. WO 91/17271; Winter *et al.*, PCT International Publication No. WO 92/20791; Markland *et al.*, PCT International Publication No. WO 92/15679; Breitling *et al.*, PCT International
35 Publication No. WO 93/01288; McCafferty *et al.*, PCT International Publication No. WO 92/01047; Garrard *et al.*, PCT International Publication No. WO 92/09690; Ladner *et al.*, PCT International Publication No. WO 90/02809; Fuchs *et al.* (1991) *Biotechnology (NY)* 9:1369-1372; Hay *et al.* (1992) *Hum. Antibodies Hybridomas*

- 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J.* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clackson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrard *et al.* (1991) *Biotechnology (NY)* 9:1373-1377; Hoogenboom *et al.* (1991) *Nucleic Acids Res.* 19:4133-4137; Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty *et al.* (1990) *Nature* 348:552-554.

Additionally, recombinant anti-HEAT antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.*, International Application No. PCT/US86/02269; Akira *et al.*, European Patent Application No. 184,187; Taniguchi, M., European Patent Application No. 171,496; Morrison *et al.*, European Patent Application 173,494; Neuberger *et al.*, PCT International Publication No. WO 86/01533; Cabilly *et al.*, U.S. Patent No. 4,816,567; Cabilly *et al.*, European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Cancer Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison, S.L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *Biotechniques* 4:214; Winter, U.S. Patent No. 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyen *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

An anti-HEAT antibody (*e.g.*, monoclonal antibody) can be used to isolate HEAT by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-HEAT antibody can facilitate the purification of natural HEAT from cells and of recombinantly produced HEAT expressed in host cells. Moreover, an anti-HEAT antibody can be used to detect HEAT protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the HEAT protein. Anti-HEAT antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable

prosthetic group complexes include streptavidin/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
5 bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, for example recombinant
10 expression vectors, containing a HEAT nucleic acid molecule or vectors containing a nucleic acid molecule which encodes a HEAT protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA
15 segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome
20 of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present
25 specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

30 The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a
35 recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the

host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel (1990) *Methods Enzymol.* 185:3-7. Regulatory sequences include those which direct constitutive
5 expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like.
10 The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, HEAT proteins, mutant forms of HEAT proteins, fusion proteins, and the like).

Accordingly, an exemplary embodiment provides a method for producing a
15 protein, preferably a HEAT protein, by culturing in a suitable medium a host cell of the invention (*e.g.*, a mammalian host cell such as a non-human mammalian cell) containing a recombinant expression vector, such that the protein is produced.

The recombinant expression vectors of the invention can be designed for expression of HEAT proteins in prokaryotic or eukaryotic cells. For example, HEAT
20 proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel (1990) *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

25 Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of
30 recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to
35 purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5

(Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in HEAT activity assays (*e.g.*, direct assays or competitive assays described in detail below), or to generate antibodies
5 specific for HEAT proteins, for example. In a preferred embodiment, a HEAT fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells, which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (*e.g.*, six (6) weeks).

10 Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.* (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.* (1990) *Methods Enzymol.* 185:60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid *trp-lac* fusion promoter. Target
15 gene expression from the pET 11d vector relies on transcription from a T7 *gn10-lac* fusion promoter mediated by a coexpressed viral RNA polymerase (T7 *gn1*). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 *gn1* gene under the transcriptional control of the *lacUV 5* promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to
20 express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S. (1990) *Methods Enzymol.* 185:119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.* (1992) *Nucleic Acids Res.*
25 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the HEAT expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari *et al.* (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz (1982)
30 *Cell* 30:933-943), pJRY88 (Schultz *et al.* (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (Invitrogen Corp., San Diego, CA).

Alternatively, HEAT proteins can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of
35 proteins in cultured insect cells (*e.g.*, Sf9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J. *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to HEAT mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue-specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced

under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*

- “Antisense RNA as a molecular tool for genetic analysis”, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a HEAT nucleic acid molecule of the invention is introduced, *e.g.*, a HEAT nucleic acid molecule within a vector (*e.g.*, a recombinant expression vector) or a HEAT nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms “host cell” and “recombinant host cell” are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a HEAT protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms “transformation” and “transfection” are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be 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), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a HEAT protein or can be introduced on a separate vector. Cells stably transfected with the introduced

nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a HEAT protein. Accordingly, the invention further provides methods for producing a HEAT protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of the invention (into which a recombinant expression vector encoding a HEAT protein has been introduced) in a suitable medium such that a HEAT protein is produced. In another embodiment, the method further comprises isolating a HEAT protein from the medium or the host cell.

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which HEAT coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous HEAT sequences have been introduced into their genome or homologous recombinant animals in which endogenous HEAT sequences have been altered. Such animals are useful for studying the function and/or activity of a HEAT protein and for identifying and/or evaluating modulators of HEAT activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous HEAT gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a HEAT-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection or retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The HEAT cDNA sequence of SEQ ID NO:1, 5, or 8 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of a human HEAT gene, such as a rat or mouse HEAT gene, can be used as a transgene. Alternatively, a HEAT gene

homologue, such as another HEAT family member, can be isolated based on hybridization to the HEAT cDNA sequences of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____ (described further in subsection I above) and used as a transgene.

- 5 Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a HEAT transgene to direct expression of a HEAT 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. 10 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified 15 based upon the presence of a HEAT transgene in its genome and/or expression of HEAT mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a HEAT protein can further be bred to other transgenic animals carrying other transgenes.

- 20 To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a HEAT gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the HEAT gene. The HEAT gene can be a human gene (*e.g.*, the cDNA of SEQ ID NO:3, 7, or 10), but more preferably, is a non-human homologue of a human HEAT gene (*e.g.*, a 25 cDNA isolated by stringent hybridization with the nucleotide sequence of SEQ ID NO:1, 5, or 8). For example, a mouse HEAT gene can be used to construct a homologous recombination nucleic acid molecule, *e.g.*, a vector, suitable for altering an endogenous HEAT gene in the mouse genome. In a preferred embodiment, the homologous recombination nucleic acid molecule is designed such that, upon 30 homologous recombination, the endogenous HEAT gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the homologous recombination nucleic acid molecule can be designed such that, upon homologous recombination, the endogenous HEAT gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory 35 region can be altered to thereby alter the expression of the endogenous HEAT protein). In the homologous recombination nucleic acid molecule, the altered portion of the HEAT gene is flanked at its 5' and 3' ends by additional nucleic acid sequence of the HEAT gene to allow for homologous recombination to occur between the

exogenous HEAT gene carried by the homologous recombination nucleic acid molecule and an endogenous HEAT gene in a cell, *e.g.*, an embryonic stem cell. The additional flanking HEAT nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the homologous recombination nucleic acid molecule (see, *e.g.*, Thomas, K.R. and Capecchi, M.R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The homologous recombination nucleic acid molecule is introduced into a cell, *e.g.*, an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced HEAT gene has homologously recombined with the endogenous HEAT gene are selected (see *e.g.*, Li, E. *et al.* (1992) *Cell* 69:915). The selected cells can then be injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (see *e.g.*, Bradley, A., in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, E.J. ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination nucleic acid molecules, *e.g.*, vectors, or homologous recombinant animals are described further in Bradley, A. (1991) *Curr. Opin. Biotechnol.* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec *et al.*; WO 91/01140 by Smithies *et al.*; WO 92/0968 by Zijlstra *et al.*; and WO 93/04169 by Berns *et al.*

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

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmot, I. *et al.* (1997) *Nature* 385:810-813 and PCT International Publication Nos. WO 97/07668 and WO

97/07669. In brief, a cell, *e.g.*, a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The
5 reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

10 IV. Pharmaceutical Compositions

The HEAT nucleic acid molecules, proteins, fragments thereof, anti-HEAT antibodies, and HEAT modulators (also referred to herein as “active compounds”) of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule,
15 protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language “pharmaceutically acceptable carrier” is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active
20 substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible
25 with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline
30 solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be
35 adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

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

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

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and

the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant
5 such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

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

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

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

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

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit
35 form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the

invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

- 5 Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the
- 10 ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.
- 15 The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound
- 20 used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more
- 25 accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

- As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20
- 30 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present.
- 35 Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (*i.e.*, including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention.

Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (*e.g.*, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is

understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug
 5 combination, and the degree of expression or activity to be modulated.

In certain embodiments of the invention, a modulator of HEAT activity is administered in combination with other agents (*e.g.*, a small molecule), or in conjunction with another, complementary treatment regime. For example, in one embodiment, a modulator of HEAT activity is used to treat a cardiovascular disorder
 10 (*e.g.*, atherosclerosis, hypertension, and/or vascular disease). Accordingly, modulation of HEAT activity may be used in conjunction with, for example, another agent used to treat the disorder. For example, non-limiting examples of agents used to treat cardiovascular disorders include niacin; clofibrate; bile acid binding resins (*e.g.*, cholestipol and cholestyramine); neomycin; statins (*e.g.*, atorvastatin,
 15 fluvastatin, lovastatin, pravastatin, and simvastatin); diuretics (*e.g.*, loop diuretics such as furosemide, bumetanide, torsemide, and ethacrynic acid; thiazide diuretics such as hydrochlorothiazide, chlorothiazide, methyclothiazide, and bendroflumethiazide; osmotic diuretics such as mannitol, glycerine, and isosorbide; potassium-sparing diuretics such as spironolactone; and sodium channel blockers such
 20 as amiloride and triamterene); angiotensin converting enzyme inhibitors (*e.g.*, enalapril, capoten, and lisinopril); angiotensin II receptor blockers (*e.g.*, losartan, valsartan, and candasartan); aldosterone antagonists (*e.g.*, spironolactone); alpha 2 adrenergic agonists (*e.g.*, methyldopa and clonidine); alpha 1 adrenergic blockers (*e.g.*, prazosin and terazosin); beta blockers (*e.g.*, propranolol, metoprolol, and
 25 atenolol); combination alpha/beta blockers (*e.g.*, carvedilol); nitrates (*e.g.*, nitroglycerin and isosorbide dinitrate); calcium channel blockers (*e.g.*, verapamil, nicardipine, amlodipine, diltiazem, and nifedipine); digoxin; folic acid; sodium channel blockers (*e.g.*, quinidine, lidocaine, and procainamide); vasodilators (*e.g.*, minoxidil and hydralazine); thrombolytic agents (*e.g.*, streptokinase and urokinase);
 30 tissue plasminogen activators (*e.g.*, alteplase, reteplase, and tenecteplase); antiplatelet agents (*e.g.*, aspirin, clopidogrel, and dipyridimole); and anticoagulants (*e.g.*, heparin, enoxaparin, dalteparin, ardeparin, and warfarin).

Further, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A
 35 cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-

dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon *et al.* "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy" in *Monoclonal Antibodies And Cancer Therapy*, Reisfeld *et al.* (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom *et al.* "Antibodies For Drug Delivery" in *Controlled Drug Delivery (2nd Ed.)*, Robinson *et al.* (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review" in *Monoclonal Antibodies '84: Biological And Clinical Applications*, Pinchera *et al.* (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy" in *Monoclonal Antibodies For Cancer Detection And Therapy*, Baldwin *et al.* (eds.), pp. 303-16 (Academic Press 1985); and Thorpe *et al.* "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates" *Immunol. Rev.* 62:119-58 (1982). Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent No.

5,328,470) or by stereotactic injection (see *e.g.*, Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where
5 the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

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

10

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, protein fragments, antibodies, peptides, peptidomimetics, and small molecules described herein can be used in one or more of the following methods: a) screening assays; b) predictive
15 medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (*e.g.*, therapeutic and prophylactic). As described herein, a HEAT protein of the invention has one or more of the following activities: (i) interaction with a HEAT substrate or target molecule (*e.g.*, a Ca^{2+} ion; ATP; or a non-HEAT protein;); (ii) transport of a HEAT substrate or target
20 molecule (*e.g.*, a Ca^{2+} ion) from one side of a biological membrane to the other; (iii) ability to be phosphorylated or dephosphorylated; (iv) adoption of an E1 conformation or an E2 conformation; (v) conversion of a HEAT substrate or target molecule to a product (*e.g.*, hydrolysis of ATP to ADP and free phosphate); (vi) interaction with a second non-HEAT protein; (vii) modulation of intra- or inter-
25 cellular signaling and/or gene transcription (*e.g.*, either directly or indirectly); (viii) modulation of vascular smooth muscle tone; (ix) modulation of cellular growth and/or proliferation; and/or (x) modulation of angiogenesis.

The isolated nucleic acid molecules of the invention can be used, for example, to express HEAT protein (*e.g.*, via a recombinant expression vector in a host cell in
30 gene therapy applications), to detect HEAT mRNA (*e.g.*, in a biological sample) or a genetic alteration in a HEAT gene, and to modulate HEAT activity, as described further below. The HEAT proteins can be used to treat disorders characterized by insufficient or excessive HEAT protein activity or HEAT nucleic acid expression, for example, cardiovascular disorders.

35 In addition, the HEAT proteins can be used to screen for naturally occurring HEAT substrates, to screen for drugs or compounds which modulate HEAT activity, as well as to treat disorders characterized by insufficient or excessive production of HEAT protein or production of HEAT protein forms which have decreased, aberrant

or unwanted activity compared to HEAT wild type protein (*e.g.*, a cardiovascular disorder).

Moreover, the anti-HEAT antibodies of the invention can be used to detect and isolate HEAT proteins, regulate the bioavailability of HEAT proteins, and
5 modulate HEAT activity.

A. Screening Assays:

The invention provides a method (also referred to herein as a “screening assay”) for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*,
10 peptides, peptidomimetics, small molecules or other drugs) which bind to HEAT proteins, have a stimulatory or inhibitory effect on, for example, HEAT expression or HEAT activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a HEAT substrate.

In one embodiment, the invention provides assays for screening candidate or
15 test compounds which are substrates of a HEAT protein or polypeptide or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a HEAT protein or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous
20 approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the ‘one-bead one-compound’ library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the
25 other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:45).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example, in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:6909; Erb
et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med.*
30 *Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips
35 (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP ‘409), plasmids (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990)

Science 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a HEAT protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate HEAT activity is determined. Determining the ability of the test compound to modulate HEAT activity can be accomplished by monitoring, for example: (i) interaction with a HEAT substrate or target molecule (*e.g.*, a Ca^{2+} ion; ATP; or a non-HEAT protein); (ii) transport of a HEAT substrate or target molecule (*e.g.*, a Ca^{2+} ion) from one side of a biological membrane to the other; (iii) ability to be phosphorylated or dephosphorylated; (iv) adoption of an E1 conformation or an E2 conformation; (v) conversion of a HEAT substrate or target molecule to a product (*e.g.*, hydrolysis of ATP to ADP and free phosphate); (vi) interaction with a second non-HEAT protein; (vii) modulation of intra- or inter-cellular signaling and/or gene transcription (*e.g.*, either directly or indirectly); (viii) modulation of vascular smooth muscle tone; (ix) modulation of cellular growth and/or proliferation; and/or (x) modulation of angiogenesis.

Determining the ability of the test compound to modulate angiogenesis can be accomplished by testing the compound in a chicken chorioallantoic membrane (CAM) assay. In a CAM assay, the ability of test compounds to modulate bFGF induced angiogenesis from the CAM, can be determined. The CAM assay is performed essentially as described in Liekens, S. *et al.* (1997) *Oncology Res.* 9:173-181, the contents of which are incorporated herein by reference, and may be performed with the modifications described below. Briefly, fresh fertilized chicken eggs are incubated for 3 days at 37°C. On the third day, the shell is cracked and the egg is placed into a tissue culture plate and incubated at 38°C. For the assay, bFGF and the compound to be tested are attached on a matrix of collagen on a nylon mesh. The mesh is then used to cover the chorioallantoic membrane and the eggs are incubated at 37°C. If angiogenesis occurs, new capillaries form and grow through the mesh within 24 hours. The ability of the test compounds (at various concentrations) to modulate the bFGF-induced angiogenesis can then be determined.

The ability of a HEAT protein to be phosphorylated (*e.g.*, be autophosphorylated) can be determined by, for example, an *in vitro* kinase assay. Briefly, a HEAT protein, *e.g.*, an immunoprecipitated HEAT protein from a cell line expressing such a protein, can be incubated with radioactive ATP, *e.g.*, [γ - ^{32}P]-ATP, in a buffer containing MgCl_2 and MnCl_2 , *e.g.*, 10 mM MgCl_2 and 5 mM MnCl_2 . Following the incubation, the immunoprecipitated HEAT protein can be separated by SDS-polyacrylamide gel electrophoresis under reducing conditions, transferred to a

membrane, *e.g.*, a PVDF membrane, and autoradiographed. The appearance of detectable bands on the autoradiograph indicates that the HEAT protein has been phosphorylated. Phosphoaminoacid analysis of the phosphorylated HEAT protein can also be performed in order to determine which residues on the HEAT protein are phosphorylated (*e.g.*, to test whether the D amino acid residue in the E1-E2 ATPases phosphorylation site has been phosphorylated). Briefly, the radiophosphorylated protein band can be excised from the SDS gel and subjected to partial acid hydrolysis. The products can then be separated by one-dimensional electrophoresis and analyzed on, for example, a phosphorimager and compared to ninhydrin-stained phosphoaminoacid standards.

The ability of the test compound to modulate HEAT binding to a substrate or to bind to another HEAT protein or subunit can also be determined. Determining the ability of the test compound to modulate HEAT binding to a substrate can be accomplished, for example, by coupling the HEAT substrate with a radioisotope or enzymatic label such that binding of the HEAT substrate to HEAT can be determined by detecting the labeled HEAT substrate in a complex. Alternatively, HEAT could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate HEAT binding to a HEAT substrate in a complex. Determining the ability of the test compound to bind HEAT can be accomplished, for example, by coupling the compound with a radioisotope or enzymatic label such that binding of the compound to HEAT can be determined by detecting the labeled HEAT compound in a complex. For example, compounds (*e.g.*, HEAT substrates) can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a compound (*e.g.*, a HEAT substrate) to interact with HEAT without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a compound with HEAT without the labeling of either the compound or the HEAT. McConnell, H.M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (*e.g.*, Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and HEAT.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a HEAT target molecule (*e.g.*, a HEAT substrate) with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the HEAT target molecule. Determining the ability of the test compound to modulate the activity of a HEAT target molecule can be accomplished, for example, by determining the ability of a HEAT protein to bind to or interact with the HEAT target molecule, by determining the cellular location of the target molecule, or determining whether the target molecule (*e.g.*, ATP) has been hydrolyzed.

10 In one embodiment, the ability of a test compound to modulate the activity of a HEAT protein or a HEAT target molecule can be determined by measuring microvessel contraction (as described in, for example, Example 4 and in Bischoff, A. *et al.* (2000) *Br. J. Pharmacol.* 130:1871-1877 and in Volpe and Cosentino (2000) *J. Cardiovasc. Pharmacol.* 35 (4 Suppl 2):S45-48). In another embodiment, the ability of a test compound to modulate the activity of a HEAT protein or a HEAT target molecule can be determined by measuring intracellular Ca^{2+} concentration (as described in, for example, Example 5 and in Bischoff, A. *et al.* (2000) *supra*). In still another embodiment, the ability of a test compound to modulate the activity of a HEAT protein or a HEAT target molecule can be determined by measuring calcium transport by the HEAT protein (as described in, for example, Example 6 and in Maruyama, K. and MacLennan, D.H. (1988) *Proc. Natl. Acad. Sci. USA* 85:3314-3318).

Determining the ability of the HEAT protein, or a biologically active fragment thereof, to bind to or interact with a HEAT target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the HEAT protein to bind to or interact with a HEAT target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting the cellular location of the target molecule, detecting a metabolite of the target molecule (*e.g.*, detecting the byproducts of ATP hydrolysis), detecting catalytic/enzymatic activity of the target molecule upon an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a target-regulated cellular response (*e.g.*, a change in vascular smooth muscle tone).

In yet another embodiment, an assay of the present invention is a cell-free assay in which a HEAT protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the HEAT

protein or biologically active portion thereof is determined. Preferred biologically active portions of the HEAT proteins to be used in assays of the present invention include fragments which participate in interactions with non-HEAT molecules, *e.g.*, fragments with high surface probability scores (see, for example, Figures 2, 13, and 24). Binding of the test compound to the HEAT protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the HEAT protein or biologically active portion thereof with a known compound which binds HEAT to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a HEAT protein, wherein determining the ability of the test compound to interact with a HEAT protein comprises determining the ability of the test compound to preferentially bind to HEAT or biologically active portion thereof as compared to the known compound.

In another embodiment, the assay is a cell-free assay in which a HEAT protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the HEAT protein or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of a HEAT protein can be accomplished, for example, by determining the ability of the HEAT protein to bind to a HEAT target molecule by one of the methods described above for determining direct binding. Determining the ability of the HEAT protein to bind to a HEAT target molecule can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of a HEAT protein can be accomplished by determining the ability of the HEAT protein to further modulate the activity of a downstream effector of a HEAT target molecule. For example, the activity of the effector molecule on an appropriate target can be determined or the binding of the effector to an appropriate target can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting a HEAT protein or biologically active portion thereof with a known compound which binds the HEAT protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the

HEAT protein, wherein determining the ability of the test compound to interact with the HEAT protein comprises determining the ability of the HEAT protein to preferentially bind to or modulate the activity of a HEAT target molecule.

The cell-free assays of the present invention are amenable to use of both
5 soluble and/or membrane-bound forms of isolated proteins (*e.g.*, HEAT proteins or biologically active portions thereof). In the case of cell-free assays in which a membrane-bound form of an isolated protein is used it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the isolated protein is maintained in solution. Examples of such solubilizing agents include non-ionic
10 detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton[®] X-100, Triton[®] X-114, Thesit[®], Isotridecypoly(ethylene glycol ether)_n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or
15 N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either HEAT or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test
20 compound to a HEAT protein, or interaction of a HEAT protein with a substrate or target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or
25 both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/HEAT fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized micrometer plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or HEAT
30 protein, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from
35 the matrix, and the level of HEAT binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a HEAT protein or a HEAT substrate or target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated HEAT protein, substrates, or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with HEAT protein or target molecules but which do not interfere with binding of the HEAT protein to its target molecule can be derivatized to the wells of the plate, and unbound target or HEAT protein trapped in the wells by antibody conjugation. 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 HEAT protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the HEAT protein or target molecule.

In another embodiment, modulators of HEAT expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of HEAT mRNA or protein in the cell is determined. The level of expression of HEAT mRNA or protein in the presence of the candidate compound is compared to the level of expression of HEAT mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of HEAT expression based on this comparison. For example, when expression of HEAT mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of HEAT mRNA or protein expression. Alternatively, when expression of HEAT mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of HEAT mRNA or protein expression. The level of HEAT mRNA or protein expression in the cells can be determined by methods described herein for detecting HEAT mRNA or protein.

In yet another aspect of the invention, the HEAT 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 WO94/10300) to identify other proteins which bind to or interact with HEAT ("HEAT-binding proteins" or "HEAT-bp") and are involved in HEAT activity. Such HEAT-binding proteins are also likely to be involved in the propagation of signals by the HEAT proteins or HEAT targets as, for

example, downstream elements of a HEAT-mediated signaling pathway.

Alternatively, such HEAT-binding proteins may be HEAT inhibitors.

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 HEAT protein is fused to 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 HEAT-dependent complex, the DNA-binding and activation domains of the 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 HEAT protein.

In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell-free assay, and the ability of the agent to modulate the activity of a HEAT protein can be confirmed *in vivo*, *e.g.*, in an animal such as an animal model for vascular disease, hypertension, angiogenesis, and/or cellular transformation or tumorigenesis.

For example, the following animal models may be used in the methods of the invention: the mouse model of cardiomyopathy in the HIV-1 transgenic mouse treated with zidovudine (Lewis, W. *et al.* (2000) *Lab Invest.* 80:187-97); the pressure-overloaded guinea pig model with cardiac hypertrophy and failure (Ahmmed, G.U. *et al.* (2000) *Circ. Res.* 86:558-70); the hypertensive transgenic mouse model that lacks fat and has lipoatrophic diabetes (Reitmann, M.L. *et al.* (1999) *Ann. N.Y. Acad. Sci.* 192:289-96; Moitra J. *et al.* (1998) *Genes Dev.* 12:3168-81); various mouse models of hypertension (Cvetokovic, B. and Sigmund, C.D. (2000) *Kidney Int.* 57:863-74; Merrill, D.C. *et al.* (1997) *Proc. Assoc. Am. Physicians* 109:533-46); various rat models for hypertension (Rapp, J.P. (2000) *Physiol. Rev.* 80:135-72; Yamori, Y. (1999) *Clin. Exp. Pharmacol. Physiol.* 26:568-72; Nara, Y. *et al.* (1999) *Clin. Exp. Pharmacol. Physiol.* 17:481-7; Boulanger, C.M. (1999) *J. Mol. Cell. Cardiol.* 31:39-49; Wookey, P.J. *et al.* (1998) *Miner Electrolyte Metab.* 24:389-99; Aitman, T.J. (1998) *Pathol. Biol. (Paris)* 46:693-4; Engler, S. *et al.* (1998) *Regul. Pept.* 77:3-8; Zolk, O. *et al.* (1998) *Cardiovasc. Res.* 39:242-56; Pinto, Y.M. *et al.* (1998)

Cardiovasc. Res. 39:77-88; Yagil, Y. and Yagil, C. (1998) *Kidney Int.* 53:1493-500; Packer, C.S. (1994) *Proc. Soc. Exp. Biol. Med.* 207:148-74; Griffith, S.L. *et al.* (1994) *J. Appl. Physiol.* 77:406-14); animal models of lower extremity chronic venous disease (Dalsing, M.C. *et al.* (1998) *Ann. Vasc. Surg.* 12:487-94); the pecten oculi of
 5 the chicken, a model system for studying vascular differentiation and barrier maturation (Wolburg, H. *et al.* (1999) *Int. Rev. Cytol.* 187:111-59); a VEGF transgenic animal model for atherosclerosis and angiogenesis (Sueishi, K. *et al.* (1997) *Ann. N.Y. Acad. Sci.* 811:311-324); the chick embryo chorioallantoic membrane, a model for *in vivo* research on angiogenesis (Ribatti, D. *et al.* (1996) *Int.*
 10 *J. Dev. Biol.* 40:1189-97); various rat models of angiogenesis (Fan, T.P. *et al.* (1992) *E. X. S.* 61:308-14; Norrby, K. (1992) *E. X. S.* 61:282-6); the porcine model for coronary artery spasm (Shimokawa, H. (2000) *Jpn. Circ. J.* 64:1-12; Kuga, T. *et al.* (2000) *J. Cardiovasc. Pharmacol.* 35:822-8; Kandabashi, T. *et al.* (2000) *Circulation* 101:1319-23); various animal models of neointima formation in atherosclerosis-prone
 15 arteries (De Meyer, G.R. and Bult, H. (1997) *Vasc. Med.* 2:179-89); the hamster cheek pouch model for vascular smooth muscle function and microcirculation research (Svensjo, E. (1990) *Eur. Respir. J. Suppl.* 12:595s-600s); and the squid axon model for studying plasma membrane mechanism for calcium regulation (DiPolo, R. and Beauge, L. (1987) *Hypertension* 10:115-9). There are also numerous animal
 20 models for tumorigenesis, including: mice that develop spontaneous tumors, either naturally, or as a result of addition of an exogenous, tumor causing transgene, a knock-out of an endogenous gene, or injection of exogenous tumor cells, and mice that develop tumors as a result of infection with oncogene-containing viruses.

This invention further pertains to novel agents identified by the above-
 25 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 (*e.g.*, an animal model such as any of those described above). For example, an agent identified as described herein (*e.g.*, a HEAT modulating agent, an antisense HEAT nucleic acid molecule, a HEAT-specific antibody, or a HEAT binding partner) can be
 30 used in an animal 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 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.

35 In another aspect, cell-based systems, as described herein, may be used to identify compounds which may act to ameliorate cardiovascular disease symptoms. For example, such cell systems may be exposed to a compound, suspected of exhibiting an ability to ameliorate cardiovascular disease symptoms, at a sufficient

concentration and for a time sufficient to elicit such an amelioration of cardiovascular disease symptoms in the exposed cells. After exposure, the cells are examined to determine whether one or more of the cardiovascular disease cellular phenotypes has been altered to resemble a more normal or more wild type, non-cardiovascular disease phenotype. Cellular phenotypes that are associated with cardiovascular disease states include aberrant vascular tone, angiogenesis, or tube formation.

In addition, animal-based cardiovascular disease systems, such as those described herein, may be used to identify compounds capable of ameliorating cardiovascular disease symptoms. Such animal models may be used as test substrates for the identification of drugs, pharmaceuticals, therapies, and interventions which may be effective in treating cardiovascular disease. For example, animal models may be exposed to a compound, suspected of exhibiting an ability to ameliorate cardiovascular disease symptoms, at a sufficient concentration and for a time sufficient to elicit such an amelioration of cardiovascular disease symptoms in the exposed animals. The response of the animals to the exposure may be monitored by assessing the reversal of disorders associated with cardiovascular disease, for example, by measuring blood pressure, by measuring the degree of vascularization of a tumor or eye, or by measuring the size of atherosclerotic plaques before and after treatment. In addition, the animals may be monitored by assessing the reversal of disorders associated with cardiovascular disease, for example, reduction in hypertension, reduction in atherosclerosis, or reduction in tumor burden, tumor size, and invasive and/or metastatic potential before and after treatment.

With regard to intervention, any treatments which reverse any aspect of cardiovascular disease symptoms should be considered as candidates for human cardiovascular disease therapeutic intervention. Dosages of test agents may be determined by deriving dose-response curves.

Additionally, gene expression patterns may be utilized to assess the ability of a compound to ameliorate cardiovascular disease symptoms. For example, the expression pattern of one or more genes may form part of a "gene expression profile" or "transcriptional profile" which may be then be used in such an assessment. "Gene expression profile" or "transcriptional profile", as used herein, includes the pattern of mRNA expression obtained for a given tissue or cell type under a given set of conditions. Such conditions may include, but are not limited to, the presence of hypertension, the presence of atherosclerotic plaques, or the presence of a vascularized tumor, *e.g.*, a colon or lung tumor, including any of the control or experimental conditions described herein. Other conditions may include, for example, tube formation or shear stress. Gene expression profiles may be generated, for example, by utilizing a differential display procedure, Northern analysis and/or

RT-PCR. In one embodiment, HEAT gene sequences may be used as probes and/or PCR primers for the generation and corroboration of such gene expression profiles.

Gene expression profiles may be characterized for known states, either cardiovascular disease or normal, within the cell- and/or animal-based model systems.

- 5 Subsequently, these known gene expression profiles may be compared to ascertain the effect a test compound has to modify such gene expression profiles, and to cause the profile to more closely resemble that of a more desirable profile.

- For example, administration of a compound may cause the gene expression profile of a cardiovascular disease model system to more closely resemble the control
10 system. Administration of a compound may, alternatively, cause the gene expression profile of a control system to begin to mimic a cardiovascular disease state. Such a compound may, for example, be used in further characterizing the compound of interest, or may be used in the generation of additional animal models.

15 B. Detection Assays

- Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their
20 respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

- 25 Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the HEAT nucleotide sequences, described herein, can be used to map the location of the HEAT genes on a chromosome. The mapping of the HEAT sequences to
30 chromosomes is an important first step in correlating these sequences with genes associated with disease.

- Briefly, HEAT genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the HEAT nucleotide sequences. Computer analysis of the HEAT sequences can be used to predict primers that do not
35 span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the HEAT sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes (D'Eustachio, P. *et al.* (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the HEAT nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a HEAT sequence to its chromosome include *in situ* hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome-specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to

noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

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

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

2. Tissue Typing

The HEAT sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the HEAT nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The HEAT nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1, 5, or 8 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3, 7, or 10 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

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

3. Use of Partial HEAT Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.*, another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by

restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1, 5, or 8 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the

5 HEAT nucleotide sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of SEQ ID NO:1, 5, or 8 having a length of at least 20 bases, preferably at least 30 bases.

The HEAT nucleotide sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used

10 in, for example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, vascular smooth muscle tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such HEAT probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, HEAT primers or probes can be used

15 to screen tissue culture for contamination (*i.e.*, screen for the presence of a mixture of different types of cells in a culture).

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in

20 which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining HEAT protein and/or nucleic acid expression as well as HEAT activity, in the context of a biological sample (*e.g.*, blood, serum, cells, or tissue) to thereby

25 determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant or unwanted HEAT expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with HEAT protein, nucleic acid expression, or activity. For example, mutations in a

30 HEAT gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with HEAT protein, nucleic acid expression or activity.

Another aspect of the invention pertains to monitoring the influence of agents

35 (*e.g.*, drugs, compounds) on the expression or activity of HEAT in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of HEAT protein, polypeptide or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or
5 an agent capable of detecting HEAT protein, polypeptide or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes HEAT protein such that the presence of HEAT protein or nucleic acid is detected in the biological sample. In another aspect, the present invention provides a method for detecting the presence of HEAT activity in a biological sample by contacting the biological sample with an agent capable of
10 detecting an indicator of HEAT activity such that the presence of HEAT activity is detected in the biological sample. A preferred agent for detecting HEAT mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to HEAT mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length HEAT nucleic acid, such as the nucleic acid of SEQ ID NO:1, 3, 5, 7, 8, or 10, or the DNA
15 insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to HEAT mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

20 A preferred agent for detecting HEAT protein is an antibody capable of binding to HEAT protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody
25 by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin.

30 The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect HEAT mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of HEAT mRNA include Northern
35 hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of HEAT protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of HEAT genomic DNA include Southern hybridizations. Furthermore, *in vivo*

techniques for detection of a HEAT protein include introducing into a subject a labeled anti-HEAT antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

5 The present invention also provides diagnostic assays for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding a HEAT protein; (ii) aberrant expression of a gene encoding a HEAT protein; (iii) mis-regulation of the gene; and (iii) aberrant post-translational modification of a HEAT protein, wherein a wild-type form of the
10 gene encodes a protein with a HEAT activity. "Misexpression or aberrant expression", as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes, but is not limited to, expression at non-wild type levels (*e.g.*, over or under expression); a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed (*e.g.*, increased
15 or decreased expression (as compared with wild type) at a predetermined developmental period or stage); a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological
20 activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene (*e.g.*, a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus).

25 In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

 In another embodiment, the methods further involve obtaining a control
30 biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting HEAT protein, mRNA, or genomic DNA, such that the presence of HEAT protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of HEAT protein, mRNA or genomic DNA in the control sample with the presence of HEAT protein, mRNA or genomic
35 DNA in the test sample.

 The invention also encompasses kits for detecting the presence of HEAT in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting HEAT protein or mRNA in a biological sample; means for

determining the amount of HEAT in the sample; and means for comparing the amount of HEAT 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 HEAT protein or nucleic acid.

5

2. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant or unwanted HEAT expression or activity. As used herein, the term
10 “aberrant” includes a HEAT expression or activity which deviates from the wild type HEAT expression or activity. Aberrant expression or activity includes increased or decreased expression or activity, as well as expression or activity which does not follow the wild type developmental pattern of expression or the subcellular pattern of expression. For example, aberrant HEAT expression or activity is intended to include
15 the cases in which a mutation in the HEAT gene causes the HEAT gene to be under-expressed or over-expressed and situations in which such mutations result in a non-functional HEAT protein or a protein which does not function in a wild-type fashion, *e.g.*, a protein which does not interact with or transport a HEAT substrate, or one which interacts with or transports a non-HEAT substrate. As used herein, the term
20 “unwanted” includes an unwanted phenomenon involved in a biological response such as deregulated cation transport. For example, the term unwanted includes a HEAT expression or activity which is undesirable in a subject.

The assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a
25 disorder associated with a misregulation in HEAT protein activity or nucleic acid expression, such as a vascular disorder, an angiogenesis disorder, or a cell growth or proliferation disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disorder associated with a misregulation in HEAT protein activity or nucleic acid expression, such as a vascular disorder, an
30 angiogenesis disorder, or a cell growth or proliferation disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant or unwanted HEAT expression or activity in which a test sample is obtained from a subject and HEAT protein or nucleic acid (*e.g.*, mRNA or genomic DNA) is detected, wherein the presence of HEAT protein or nucleic acid is diagnostic for a
35 subject having or at risk of developing a disease or disorder associated with aberrant or unwanted HEAT expression or activity. As used herein, a “test sample” refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted HEAT expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a drug or toxin sensitivity disorder or a cell proliferation and/or differentiation disorder. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant or unwanted HEAT expression or activity in which a test sample is obtained and HEAT protein or nucleic acid expression or activity is detected (*e.g.*, wherein the abundance of HEAT protein or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant or unwanted HEAT expression or activity).

The methods of the invention can also be used to detect genetic alterations in a HEAT gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in HEAT protein activity or nucleic acid expression, such as a vascular disorder, an angiogenesis disorder, or a cell growth or proliferation disorder. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a HEAT-protein, or the mis-expression of the HEAT gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a HEAT gene; 2) an addition of one or more nucleotides to a HEAT gene; 3) a substitution of one or more nucleotides of a HEAT gene; 4) a chromosomal rearrangement of a HEAT gene; 5) an alteration in the level of a messenger RNA transcript of a HEAT gene; 6) aberrant modification of a HEAT gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a HEAT gene; 8) a non-wild type level of a HEAT-protein; 9) allelic loss of a HEAT gene, and 10) inappropriate post-translational modification of a HEAT-protein. As described herein, there are a large number of assays known in the art which can be used for detecting alterations in a HEAT gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, *e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, *e.g.*, Landegran *et al.* (1988) *Science* 241:1077-

1080; and Nakazawa *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the HEAT-gene (see Abravaya *et al.* (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (*e.g.*,
5 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 HEAT gene under conditions such that hybridization and amplification of the HEAT 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
10 sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

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

In an alternative embodiment, mutations in a HEAT gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel
25 electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, 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.

30 In other embodiments, genetic mutations in HEAT can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Hum. Mutat.* 7:244-255; Kozal, M.J. *et al.* (1996) *Nat. Med.* 2:753-759). For example, genetic mutations in HEAT can be identified in two dimensional arrays
35 containing light-generated DNA probes as described in Cronin *et al.* (1996) *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows

the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the HEAT gene and detect mutations by comparing the sequence of the sample HEAT with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of 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.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the HEAT gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type HEAT sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in HEAT cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine

DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a HEAT sequence, *e.g.*, a wild-type HEAT sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in HEAT genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci. USA* 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control HEAT nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet.* 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys. Chem.* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or

a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention.

- 5 Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

- The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a HEAT gene.

- Furthermore, any cell type or tissue in which HEAT is expressed (*e.g.*, vessels, endothelial cells, and/or vascular smooth muscle cells) may be utilized in the prognostic assays described herein.

25

3. Monitoring of Effects During Clinical Trials

- Monitoring the influence of agents (*e.g.*, drugs) on the expression or activity of a HEAT protein (*e.g.*, the modulation of cellular signaling, Ca^{2+} transport, gene expression, vascular smooth muscle tone regulation, angiogenesis, and/or cell growth or proliferation mechanisms) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase HEAT gene expression, protein levels, or upregulate HEAT activity, can be monitored in clinical trials of subjects exhibiting decreased HEAT gene expression, protein levels, or downregulated HEAT activity.
- 35 Alternatively, the effectiveness of an agent determined by a screening assay to decrease HEAT gene expression, protein levels, or downregulate HEAT activity, can be monitored in clinical trials of subjects exhibiting increased HEAT gene expression, protein levels, or upregulated HEAT activity. In such clinical trials, the expression or

activity of a HEAT gene, and preferably, other genes that have been implicated in, for example, a cardiovascular disorder can be used as a “read out” or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes, including HEAT, that are
5 modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates HEAT activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cardiovascular disorders (*e.g.*, disorders characterized by deregulated cellular signaling, Ca^{2+} transport, gene expression, vascular smooth muscle tone regulation,
10 angiogenesis, and/or cell growth or proliferation mechanisms), for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of HEAT and other genes implicated in the cardiovascular disorder, respectively. The levels of gene expression (*e.g.*, a gene expression pattern) can be quantified by northern blot analysis or RT-PCR, as described herein, or alternatively
15 by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of HEAT or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

20 In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) including the steps of (i) obtaining a pre-administration sample from a subject prior to administration of
25 the agent; (ii) detecting the level of expression of a HEAT protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the HEAT protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the HEAT protein,
30 mRNA, or genomic DNA in the pre-administration sample with the HEAT protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of HEAT to higher levels than detected, *i.e.*, to increase the effectiveness of
35 the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of HEAT to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent. According to such an embodiment, HEAT

expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

D. Methods of Treatment:

5 The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) or having a cardiovascular disorder, *e.g.*, a disorder associated with aberrant or unwanted HEAT expression or activity (*e.g.*, hypertension or atherosclerosis). As used herein, “treatment” of a subject includes the application or administration of a therapeutic agent to a subject, or
10 application or administration of a therapeutic agent to a cell or tissue from a subject, who has a diseases or disorder, has a symptom of a disease or disorder, or is at risk of (or susceptible to) a disease or disorder, with the purpose of curing, healing, alleviating, relieving, altering, remedying, ameliorating, improving, or affecting the disease or disorder, the symptom of the disease or disorder, or the risk of (or
15 susceptibility to) the disease or disorder. As used herein, a “therapeutic agent” includes, but is not limited to, small molecules, peptides, polypeptides, antibodies, ribozymes, and antisense oligonucleotides.

With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained
20 from the field of pharmacogenomics. “Pharmacogenomics”, as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient’s genes determine his or her response to a drug (*e.g.*, a patient’s “drug response phenotype”, or “drug response
25 genotype”). Thus, another aspect of the invention provides methods for tailoring an individual’s prophylactic or therapeutic treatment with either the HEAT molecules of the present invention or HEAT modulators according to that individual’s drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the
30 treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a
35 disease or condition associated with an aberrant or unwanted HEAT expression or activity, by administering to the subject a HEAT or an agent which modulates HEAT expression or at least one HEAT activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted HEAT expression or activity can be

identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the HEAT aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on
5 the type of HEAT aberrancy, for example, a HEAT molecule, HEAT agonist or HEAT antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

10 Another aspect of the invention pertains to methods of modulating HEAT expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell capable of expressing HEAT with an agent that modulates one or more of the activities of HEAT protein activity associated with the cell, such that HEAT activity
15 in the cell is modulated. An agent that modulates HEAT protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a HEAT protein (*e.g.*, a HEAT substrate), a HEAT antibody, a HEAT agonist or antagonist, a peptidomimetic of a HEAT agonist or antagonist, or other small molecule. In one embodiment, the agent stimulates one or more HEAT
20 activities. Examples of such stimulatory agents include active HEAT protein and a nucleic acid molecule encoding HEAT that has been introduced into the cell. In another embodiment, the agent inhibits one or more HEAT activities. Examples of such inhibitory agents include antisense HEAT nucleic acid molecules, anti-HEAT antibodies, and HEAT inhibitors. These modulatory methods can be performed *in*
25 *vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a HEAT protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an
30 agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) HEAT expression or activity. In another embodiment, the method involves administering a HEAT protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted HEAT expression or activity.

35 Stimulation of HEAT activity is desirable *in situations* in which HEAT is abnormally downregulated and/or in which increased HEAT activity is likely to have a beneficial effect. For example, stimulation of HEAT activity is desirable *in situations* in which a HEAT is downregulated and/or in which increased HEAT

activity is likely to have a beneficial effect. Likewise, inhibition of HEAT activity is desirable *in situations* in which HEAT is abnormally upregulated and/or in which decreased HEAT activity is likely to have a beneficial effect.

5 3. Pharmacogenomics

The HEAT molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on HEAT activity (*e.g.*, HEAT gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically)
10 cardiovascular disorders (*e.g.*, disorders characterized by aberrant modulation of cellular signaling, Ca²⁺ transport, gene expression, vascular smooth muscle tone, angiogenesis, and/or cell growth or proliferation mechanisms) associated with aberrant or unwanted HEAT activity. In conjunction with such treatment, pharmacogenomics (*i.e.*, the study of the relationship between an individual's
15 genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining
20 whether to administer a HEAT molecule or HEAT modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a HEAT molecule or HEAT modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected
25 persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering
30 the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate transporter deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics,
35 nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related

markers (*e.g.*, a “bi-allelic” gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants). Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a “SNP” is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the “candidate gene approach” can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug’s target is known (*e.g.*, a HEAT protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-transporter 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular

basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the “gene expression profiling”, can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (*e.g.*, a HEAT molecule or HEAT modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a HEAT molecule or HEAT modulator, such as a modulator identified by one of the exemplary screening assays described herein.

15

4. Use of HEAT Molecules as Surrogate Markers

The HEAT molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the HEAT molecules of the invention may be detected, and may be correlated with one or more biological states *in vivo*. For example, the HEAT molecules of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states.

As used herein, a “surrogate marker” is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (*e.g.*, with the presence or absence of a tumor). The presence or quantity of such markers is independent of the causation of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (*e.g.*, early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (*e.g.*, an assessment of cardiovascular disease may be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection may be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of

the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35:258-264; and James (1994) *AIDS Treatment News Archive* 209.

The HEAT molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug may be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug may be sufficient to activate multiple rounds of marker (*e.g.*, a HEAT marker) transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-HEAT antibodies may be employed in an immune-based detection system for a HEAT protein marker, or HEAT-specific radiolabeled probes may be used to detect a HEAT mRNA marker. Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.*, U.S. Patent No. 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90:229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3:S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3:S16-S20.

The HEAT molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, *e.g.*, McLeod *et al.* (1999) *Eur. J. Cancer* 35(12):1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or

which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA, or protein (*e.g.*, HEAT protein or RNA) for specific tumor markers in a subject, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in HEAT DNA may correlate HEAT drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

10 E. Electronic Apparatus Readable Media and Arrays

Electronic apparatus readable media comprising HEAT sequence information is also provided. As used herein, "HEAT sequence information" refers to any nucleotide and/or amino acid sequence information particular to the HEAT molecules of the present invention, including but not limited to full-length nucleotide and/or amino acid sequences, partial nucleotide and/or amino acid sequences, polymorphic sequences including single nucleotide polymorphisms (SNPs), epitope sequences, and the like. Moreover, information "related to" said HEAT sequence information includes detection of the presence or absence of a sequence (*e.g.*, detection of expression of a sequence, fragment, polymorphism, etc.), determination of the level of a sequence (*e.g.*, detection of a level of expression, for example, a quantitative detection), detection of a reactivity to a sequence (*e.g.*, detection of protein expression and/or levels, for example, using a sequence-specific antibody), and the like. As used herein, "electronic apparatus readable media" refers to any suitable medium for storing, holding, or containing data or information that can be read and accessed directly by an electronic apparatus. Such media can include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact discs; electronic storage media such as RAM, ROM, EPROM, EEPROM and the like; and general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having recorded thereon HEAT sequence information of the present invention.

As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatuses; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as a personal digital assistants (PDAs), cellular phone, pager and the like; and local and distributed processing systems.

As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the HEAT sequence information.

- 5 A variety of software programs and formats can be used to store the sequence information on the electronic apparatus readable medium. For example, the sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, represented in the form of an ASCII file, or stored in a database application, such as
- 10 DB2, Sybase, Oracle, or the like, as well as in other forms. Any number of dataprocessor structuring formats (*e.g.*, text file or database) may be employed in order to obtain or create a medium having recorded thereon the HEAT sequence information.

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

- 20 The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has a cardiovascular disease or disorder or a pre-disposition to a cardiovascular disease or disorder, wherein the method comprises the steps of determining HEAT sequence information associated with the subject and based on the HEAT sequence information, determining whether
- 25 the subject has a cardiovascular disease or disorder or a pre-disposition to a cardiovascular disease or disorder, and/or recommending a particular treatment for the disease, disorder, or pre-disease condition.

- The present invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a cardiovascular disease or
- 30 disorder or a pre-disposition to a disease associated with HEAT wherein the method comprises the steps of determining HEAT sequence information associated with the subject, and based on the HEAT sequence information, determining whether the subject has a cardiovascular disease or disorder or a pre-disposition to a cardiovascular disease or disorder, and/or recommending a particular treatment for the
- 35 disease, disorder or pre-disease condition. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the subject.

The present invention also provides in a network, a method for determining whether a subject has a cardiovascular disease or disorder or a pre-disposition to a cardiovascular disease or disorder associated with HEAT, said method comprising the steps of receiving HEAT sequence information from the subject and/or information
5 related thereto, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to HEAT and/or a cardiovascular disease or disorder, and based on one or more of the phenotypic information, the HEAT information (*e.g.*, sequence information and/or information related thereto), and the acquired information, determining whether the subject has
10 a cardiovascular disease or disorder or a pre-disposition to a cardiovascular disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

The present invention also provides a business method for determining whether a subject has a cardiovascular disease or disorder or a pre-disposition to a
15 cardiovascular disease or disorder, said method comprising the steps of receiving information related to HEAT (*e.g.*, sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to HEAT and/or related to a cardiovascular disease or disorder, and based on one or more of the phenotypic information, the
20 HEAT information, and the acquired information, determining whether the subject has a cardiovascular disease or disorder or a pre-disposition to a cardiovascular disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

The invention also includes an array comprising a HEAT sequence of the
25 present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression, one of which can be HEAT. This allows a profile to be developed showing a battery of genes specifically
30 expressed in one or more tissues.

In addition to such qualitative determination, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that
35 tissue. This is useful, for example, in ascertaining the relationship of gene expression between or among tissues. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined.

Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of a cardiovascular disease or disorder, progression of cardiovascular disease or disorder, and processes, such a cellular transformation associated with the cardiovascular disease or disorder.

The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (*e.g.*, ascertaining the effect of HEAT expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (*e.g.*, including HEAT) that could serve as a molecular target for diagnosis or therapeutic intervention.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application, as well as the sequence listing and the figures, are incorporated herein by reference.

EXAMPLES

30

EXAMPLE 1: IDENTIFICATION AND CHARACTERIZATION OF HUMAN HEAT cDNA

In this example, the identification and characterization of the genes encoding human HEAT-1 (clone 49937), human HEAT-2 (clone 49931), and human HEAT-3 (clone 49933) is described.

Isolation of the human HEAT cDNAs

The invention is based, at least in part, on the discovery of genes encoding novel members of the E1-E2 ATPase family. The entire sequence of human clones Fbh49937, Fbh49931, and Fbh49933 were determined and found to contain open
5 reading frames termed human "HEAT-1," human "HEAT-2," and human "HEAT-3," respectively.

The nucleotide sequence encoding the human HEAT-1 is shown in Figures 1A-1D and is set forth as SEQ ID NO:1. The protein encoded by this nucleic acid molecule comprises about 1180 amino acids and has the amino acid sequence shown
10 in Figures 1A-1D and set forth as SEQ ID NO:2. The coding region (open reading frame) of SEQ ID NO:1 is set forth as SEQ ID NO:3. Clone Fbh49937, comprising the coding region of human HEAT-1, was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession No. _____.

15 The nucleotide sequence encoding the human HEAT-2 is shown in Figures 11A-11E and is set forth as SEQ ID NO:5. The protein encoded by this nucleic acid molecule comprises about 1256 amino acids and has the amino acid sequence shown in Figures 11A-11E and set forth as SEQ ID NO:6. The coding region (open reading frame) of SEQ ID NO:1 is set forth as SEQ ID NO:7. Clone Fbh49931, comprising
20 the coding region of human HEAT-2, was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession No. _____.

The nucleotide sequence encoding the human HEAT-3 is shown in Figures 22A-22D and is set forth as SEQ ID NO:8. The protein encoded by this nucleic acid
25 molecule comprises about 1204 amino acids and has the amino acid sequence shown in Figures 22A-22D and set forth as SEQ ID NO:9. The coding region (open reading frame) of SEQ ID NO:8 is set forth as SEQ ID NO:10. Clone Fbh49933, comprising the coding region of human HEAT-3, was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on
30 _____, and assigned Accession No. _____.

Analysis of the human HEAT Molecules

The amino acid sequences of human HEAT-1, HEAT-2, and HEAT-3 were analyzed using the program PSORT (available online; see Nakai, K. and Kanehisa,
35 M. (1992) *Genomics* 14:897-911) to predict the localization of the proteins within the cell. This program assesses the presence of different targeting and localization amino acid sequences within the query sequence. The results of the analyses show that human HEAT-1 may be localized to the endoplasmic reticulum, mitochondria,

secretory vesicles, or vacuoles. The results of these analyses further show that human HEAT-2 may be localized to the endoplasmic reticulum or the mitochondria and that human HEAT-3 may be localized to endoplasmic reticulum, the mitochondria, or vacuoles.

- 5 Analyses of the amino acid sequence of human HEAT-1, HEAT-2, and HEAT-3 were performed using MEMSAT. These analyses resulted in the identification of twelve possible transmembrane domains in the amino acid sequence of human HEAT-2 at residues 29-50, 211-227, 234-253, 294-317, 410-434, 449-469, 941-960, 968-985, 1000-1020, 1076-1092, 1105-1129, and 1144-1160 of SEQ ID
- 10 NO:6 (Figures 13 and 14). These analyses further resulted in the identification of twelve possible transmembrane domains in the amino acid sequence of human HEAT-3 at residues 65-89, 99-116, 242-258, 265-281, 445-464, 493-509, 990-1007, 1015-1031, 1049-1073, 1049-1073, 1103-1119, 1134-1151, and 1171-1187 of SEQ ID NO:9 (Figures 24 and 25). The analysis of human HEAT-1 predicted twelve
- 15 possible transmembrane domains in the amino acid sequence of human HEAT-1 (SEQ ID NO:2) at about residues 8-25, 47-65, 256-276, 428-448, 464-484, 900-920, 936-954, 963-987, 994-1015, 1049-1065, 1079-1102, and 1118-1134. The potential transmembrane domain at about residues 900-920 has a notably low score of only 0.4 by MEMSAT analysis. Further analysis of the amino acid sequence of SEQ ID NO:2
- 20 (*e.g.*, alignment with, for example, a known *C. elegans* E1-E2 ATPase cation transporter) resulted in the identification of a twelfth transmembrane domain at about amino acid residues 231-253 of SEQ ID NO:2. Accordingly, the human HEAT-1 protein of SEQ ID NO:2 is predicted to have at least twelve transmembrane domains, for example, at about residues 8-25, 47-65, 231-253, 256-276, 428-448, 464-484,
- 25 936-954, 963-987, 994-1015, 1049-1065, 1079-1102, and 1118-1134.

- Searches of the amino acid sequences of human HEAT-1, HEAT-2, and HEAT-3 were also performed against the HMM database (Figures 3, 12, and 23, respectively). These searches resulted in the identification of an "E1-E2 ATPase" domain in the amino acid sequence of HEAT-1 at about residues 299-387 (score =
- 30 51.4) of SEQ ID NO:2 (Figure 3). These searches also resulted in the identification of an "E1-E2 ATPase" domain in the amino acid sequence of human HEAT-2 at about residues 278-365 (score = 53.4) of SEQ ID NO:6. These searches further resulted in the identification of an "E1-E2 ATPase" domain in the amino acid sequence of human HEAT-3 at about residues 302-392 (score = 37.0) of SEQ ID
- 35 NO:9.

Searches of the amino acid sequence of human HEAT-1 were performed against the Prosite database. These searches resulted in the identification of an "E1-E2 ATPases phosphorylation site" at about residues 513-519 of SEQ ID NO:2. These

searches also resulted in the identification in the amino acid sequence of human HEAT-1 of a number of potential N-glycosylation sites, cAMP- and cGMP-dependent protein kinase phosphorylation sites, protein kinase C phosphorylation sites, casein kinase II phosphorylation sites, and N-myristoylation sites.

5 Searches of the amino acid sequence of human HEAT-2 were also performed against the Prosite database. These searches resulted in the identification of an "E1-E2 ATPases phosphorylation site" at about residues 498-504 of SEQ ID NO:6 (Figures 14A-14B). These searches also resulted in the identification in the amino acid sequence of human HEAT-2 of a number of potential N-glycosylation sites, cAMP- and cGMP-dependent protein kinase phosphorylation sites, protein kinase C phosphorylation sites, casein kinase II phosphorylation sites, tyrosine phosphorylation sites, and N-myristoylation sites.

10 Searches of the amino acid sequence of human HEAT-3 were further performed against the Prosite database. These searches resulted in the identification of an "E1-E2 ATPases phosphorylation site" at about residues 533-539 of SEQ ID NO:9 (Figures 25A-25B). These searches also resulted in the identification in the amino acid sequence of human HEAT-3 of a number of potential N-glycosylation sites, cAMP- and cGMP-dependent protein kinase phosphorylation sites, protein kinase C phosphorylation sites, casein kinase II phosphorylation sites, and N-myristoylation sites.

20 The amino acid sequence of human HEAT-2 was used as a database query using the BLASTP program. This search established that human HEAT-2 has the highest homology to a putative yeast Ca^{2+} -transporting ATPase (high score = 798, probability = $2.9\text{e-}87$).

25

Tissue Expression Analysis of HEAT-1, HEAT-2, and HeAT-3 mRNA Using Transcriptional Profiling and Taqman Analysis

This example describes the tissue distribution of human HEAT-2 mRNA, as determined using transcriptional profiling analysis and the TaqMan™ procedure. For transcriptional profiling analysis, an array of several thousand cDNA clones are spotted onto a nylon membrane and probed with a complex probe prepared by radiolabeling cDNA made from mRNA from, for example, normal tissue, and another, separate probe made from mRNA from another tissue, for example, diseased tissue. Expression levels of each gene in the first (*e.g.*, normal) and the second (*e.g.*, diseased) tissue are then compared. Transcriptional profiling thus allows assessment of the expression level of several thousand genes in an mRNA sample at the same time.

The Taqman™ procedure is a quantitative, reverse transcription PCR-based approach for detecting mRNA. The RT-PCR reaction exploits the 5' nuclease activity of AmpliTaq Gold™ DNA Polymerase to cleave a TaqMan™ probe during PCR. Briefly, cDNA is generated from the samples of interest and used as the starting material for PCR amplification. In addition to the 5' and 3' gene-specific primers, a gene-specific oligonucleotide probe (complementary to the region being amplified) is included in the reaction (*i.e.*, the Taqman™ probe). The TaqMan™ probe includes the oligonucleotide with a fluorescent reporter dye covalently linked to the 5' end of the probe (such as FAM (6-carboxyfluorescein), TET (6-carboxy-4,7,2',7'-tetrachlorofluorescein), JOE (6-carboxy-4,5-dichloro-2,7-dimethoxyfluorescein), or VIC) and a quencher dye (TAMRA (6-carboxy-N,N,N',N'-tetramethylrhodamine) at the 3' end of the probe.

During the PCR reaction, cleavage of the probe separates the reporter dye and the quencher dye, resulting in increased fluorescence of the reporter. Accumulation of PCR products is detected directly by monitoring the increase in fluorescence of the reporter dye. When the probe is intact, the proximity of the reporter dye to the quencher dye results in suppression of the reporter fluorescence. During PCR, if the target of interest is present, the probe specifically anneals between the forward and reverse primer sites. The 5'-3' nucleolytic activity of the AmpliTaq™ Gold DNA Polymerase cleaves the probe between the reporter and the quencher only if the probe hybridizes to the target. The probe fragments are then displaced from the target, and polymerization of the strand continues. The 3' end of the probe is blocked to prevent extension of the probe during PCR. This process occurs in every cycle and does not interfere with the exponential accumulation of product. RNA was prepared using the trizol method and treated with DNase to remove contaminating genomic DNA. cDNA was synthesized using standard techniques. Mock cDNA synthesis in the absence of reverse transcriptase resulted in samples with no detectable PCR amplification of the control gene confirms efficient removal of genomic DNA contamination.

30

Endothelial Cell Paradigms

To induce tube formation, human microvascular endothelial cells isolated from the lung (HMVECs) were plated on Matrigel to induce capillary-like tube formation. At 5 hours, the cells were actively forming tubes, and RNA was harvested. Additional RNA samples were prepared from cells 25 hours after plating on Matrigel when tube formation was complete, and from actively proliferating and confluent HMVECs grown on plastic.

Cells were also treated with laminar shear stress (LSS) of 7 dyn/cm² for 24-30 hours, LSS plus one or six additional hours of 12 dyn/cm² ("1h up" or "6h up"), or LSS plus one or six additional hours of 2 dyn/cm² ("1h down" or "6h down").

5 *HEAT-1*

The expression levels of human HEAT-1 mRNA in various human and monkey cell types and tissues was first determined using the Taqman procedure. As shown in Figure 5, HEAT-1 is highly expressed in coronary artery vascular smooth muscle cells, prostate epithelial cells, pancreas, and brain (including cortex, hypothalamus, dorsal root ganglion cells, and glial cells/astrocytes).

The expression levels of human HEAT-1 mRNA in various human vascular rich organs was then determined using the Taqman procedure. As shown in Figure 6, HEAT-1 is highly expressed in Wilms' tumor, normal spinal cord, and microvascular endothelial cells.

15 In another experiment, the expression levels of human HEAT-1 mRNA in various human and monkey vessels was determined using the Taqman procedure. As shown in Figure 7, HEAT-1 is highly expressed in vessels such as arteries and veins.

The expression levels of human HEAT-1 mRNA in various human coronary vascular cell types was also determined using the Taqman procedure. As shown in Figure 8, HEAT-1 is highly expressed in coronary and vascular smooth muscle cells, as compared to other cell types.

The expression levels of human HEAT-1 mRNA in various human endothelial cell paradigms was determined using the Taqman procedure. As shown in Figures 5, 9, and 10, human HEAT-1 is upregulated during shear stress of endothelial cells. As shown in Figure 9, human HEAT-1 is upregulated during proliferation and tube formation of endothelial cells. These data strongly link human HEAT-1 to a role in angiogenesis.

HEAT-2

30 The expression levels of human HEAT-2 mRNA in various human and monkey cell types and tissues was first determined using transcriptional profiling. As shown in Figure 18, HEAT-2 is highly expressed in coronary artery vascular smooth muscle cells, as compared to other tissues such as aortic vascular smooth muscle cells, umbilical vein endothelial cells, microvascular endothelial cells, heart, liver, aorta, and vein. The expression levels of human HEAT-2 mRNA in various human cell types and tissues was confirmed in a second experiment using the Taqman procedure (see Figure 15).

The expression levels of human HEAT-2 mRNA in various human vascular rich organs was then determined using the Taqman procedure. As shown in Figure 16, HEAT-2 is highly expressed in the heart.

5 In another experiment, the expression levels of human HEAT-2 mRNA in various human and monkey vessels was determined using the Taqman procedure and *in situ* hybridization. As shown in Figure 17, HEAT-2 is highly expressed in vessels such as arteries and veins.

The expression levels of human HEAT-2 mRNA in various human coronary vascular cell types was also determined using the Taqman procedure. As shown in 10 Figure 19, HEAT-2 is highly expressed in coronary vascular smooth muscle cells, as compared to other cell types.

The expression levels of human HEAT-2 mRNA in various human endothelial cell paradigms was determined using the Taqman procedure. As shown in Figures 15 and 20, HEAT-2 is upregulated during shear and proliferation of endothelial cells. 15 These data strongly link HEAT-2 to a role in angiogenesis.

As shown in Figure 21, human HEAT-2 is upregulated during tube formation of endothelial cells. The expression level of human HEAT-2 is 5-fold higher in the 5-hour Matrigel sample than in any other sample, indicating that expression is significantly induced during the process of capillary-like tube formation. There is 20 also significantly higher expression in proliferating HMVECs than in confluent HMVECs grown on plastic. These results indicate a pro-angiogenic function for human HEAT-2.

Human HEAT-2 mRNA expression was also detected by *in situ* hybridization analysis in human endothelial cells and myocytes in the heart and in endothelial cells 25 and inflammatory cells in ApoE knockout mouse diseased aortic roots.

HEAT-3

The expression levels of human HEAT-3 mRNA in various human and monkey cell types and tissues was first determined using the Taqman procedure. As 30 shown in Figure 26, HEAT-3 is highly expressed in coronary artery vascular smooth muscle cells, prostate epithelial cells, pancreas, and brain (including cortex, hypothalamus, and glial cells/astrocytes).

The expression levels of human HEAT-3 mRNA in various human vascular rich organs was then determined using the Taqman procedure. As shown in Figure 35 27, HEAT-3 is expressed in the heart, kidney, and skeletal muscle.

In another experiment, the expression levels of human HEAT-3 mRNA in various vessels was determined using the Taqman procedure. As shown in Figures 28 and 31, HEAT-3 is highly expressed in vessels such as arteries and veins.

The expression levels of human HEAT-3 mRNA in various human coronary vascular cell types was also determined using the Taqman procedure. As shown in Figure 29, HEAT-3 is highly expressed in coronary and aortic vascular smooth muscle cells, as well as in renal proximal tubule epithelium, as compared to other cell types.

Tissue Distribution of HEAT mRNA Using *in situ* Analysis

This example describes the tissue distribution of human HEAT mRNA, as was determined using *in situ* hybridization analysis. For *in situ* analysis, various tissues, e.g., vascular tissue or heart tissue, were first frozen on dry ice. Ten-micrometer-thick sections of the tissues were postfixed with 4% formaldehyde in DEPC-treated 1X phosphate-buffered saline at room temperature for 10 minutes before being rinsed twice in DEPC 1X phosphate-buffered saline and once in 0.1 M triethanolamine-HCl (pH 8.0). Following incubation in 0.25% acetic anhydride-0.1 M triethanolamine-HCl for 10 minutes, sections were rinsed in DEPC 2X SSC (1X SSC is 0.15 M NaCl plus 0.015 M sodium citrate). Tissue was then dehydrated through a series of ethanol washes, incubated in 100% chloroform for 5 minutes, and then rinsed in 100% ethanol for 1 minute and 95% ethanol for 1 minute and allowed to air dry.

Hybridizations were performed with ^{35}S -radiolabeled (5×10^7 cpm/ml) cRNA probes. Probes were incubated in the presence of a solution containing 600 mM NaCl, 10 mM Tris (pH 7.5), 1 mM EDTA, 0.01% sheared salmon sperm DNA, 0.01% yeast tRNA, 0.05% yeast total RNA type X1, 1X Denhardt's solution, 50% formamide, 10% dextran sulfate, 100 mM dithiothreitol, 0.1% sodium dodecyl sulfate (SDS), and 0.1% sodium thiosulfate for 18 hours at 55°C.

After hybridization, slides were washed with 2X SSC. Sections were then sequentially incubated at 37°C in TNE (a solution containing 10 mM Tris-HCl (pH 7.6), 500 mM NaCl, and 1 mM EDTA), for 10 minutes, in TNE with 10 µg of RNase A per ml for 30 minutes, and finally in TNE for 10 minutes. Slides were then rinsed with 2X SSC at room temperature, washed with 2X SSC at 50°C for 1 hour, washed with 0.2X SSC at 55°C for 1 hour, and 0.2X SSC at 60°C for 1 hour. Sections were then dehydrated rapidly through serial ethanol-0.3 M sodium acetate concentrations before being air dried and exposed to Kodak Biomax MR scientific imaging film for 24 hours and subsequently dipped in NB-2 photoemulsion and exposed at 4°C for 7 days before being developed and counter stained.

Using *in situ* hybridization analysis, HEAT-2 mRNA was found to be expressed in human endothelial-cells and myocytes in the heart and in endothelial cells and inflammatory cells in ApoE knockout mice diseased aortic roots.

**EXAMPLE 2: EXPRESSION OF RECOMBINANT HEAT PROTEIN IN
BACTERIAL CELLS**

In this example, human HEAT is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated
5 and characterized. Specifically, human HEAT is fused to GST and this fusion polypeptide is expressed in *E. coli*, e.g., strain PEB199. Expression of the GST-HEAT fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel
10 electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

**EXAMPLE 3: EXPRESSION OF RECOMBINANT HEAT PROTEIN IN
COS CELLS**

15 To express the human HEAT gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire HEAT protein and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is
20 cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

To construct the plasmid, the HEAT DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by
25 approximately twenty nucleotides of the HEAT coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the HEAT coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the
30 vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the HEAT gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5 α , SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates,
35 and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the HEAT-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J. *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the HEAT polypeptide is detected by radiolabeling (³⁵S-methionine or ³⁵S-cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with ³⁵S-methionine (or ³⁵S-cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the HEAT coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the HEAT polypeptide is detected by radiolabeling and immunoprecipitation using a HEAT specific monoclonal antibody.

EXAMPLE 4: ASSESMENT OF MICROVESSEL CONTRACTION

This example describes the assessment of microvessel contraction using rat microvessels, as described in, for example, Bischoff, A. *et al.* (2000) *Br. J. Pharmacol.* 130:1871-1877. Microvessels (*e.g.*, mesenteric or renal microvessels such as interlobar arteries) are prepared from rats (*e.g.*, adult Wistar rats) as described in Chen *et al.* (1996) *Naunyn-Schmiedeberg's Arch. Pharmacol.* 353:314-323 and Chen *et al.* (1997) *J. Auton. Pharmacol.* 17:137-146. Rats are killed by either decapitation or an overdose of thiobutabarbital. The vessels are mounted on 40 µm diameter stainless steel wires in a myograph chamber for isometric recording of tension development. The vessels are then bathed in Krebs-Henseleit buffer of the following composition: 119 mM NaCl, 25 mM NaHCO₃, 4.7 mM KCl, 1.18 mM KH₂PO₄, 1.17 mM MgSO₄, 2.5 mM CaCl₂, 0.026 mM EDTA, and 5.5 mM D-glucose. The buffer temperature is maintained at 37°C, and the chamber is gassed with 5% CO₂/95% O₂ to maintain a pH of 7.4. Additionally, 5 µM cocaine and 1 µM (±)-propranolol may be added to block neuronal catecholamine uptake and β-adrenoceptor activation by high noradrenaline concentration. Following equilibration, the vessels are challenged several times with 125 mM KCl and 10 µM

noradrenaline. The vessels are then treated with 100 μ M carbachol; vessels with a relaxation response of at least 50% indicate a functionally intact epithelium.

EXAMPLE 5: ASSESMENT OF INTRACELLULAR FREE CALCIUM
CONCENTRATIONS IN CULTURED RAT AORTIC SMOOTH MUSCLE
CELLS

This example describes the assessment of intracellular free calcium concentrations in cultured rat aortic smooth muscle cells, as described in, for example, Bischoff, A. *et al.* (2000) *Br. J. Pharmacol.* 130:1871-1877. Vascular smooth muscle cells are prepared from rat thoracic aorta according to Rosskoph *et al.* (1995) *Cell Physiol. Biochem.* 5:276-285). Briefly, freshly prepared aortae are incubated for 30 minutes at room temperature with 125 U/ml collagenase I in Hank's balanced salt solution (HBSS) of the following composition: 118 mM NaCl, 5 mM KCl, 1 mM CaCl_2 , 1 mM MgCl_2 , 5 mM D-glucose, and 15 mM HEPES pH 7.4. Thereafter, remaining connective tissue and endothelium are removed, the aortae are cut into small pieces and incubated for 4-6 hours at 37°C in DMEM/F12 medium with 100 U/ml penicillin, 100 μ g/ml streptomycin, and 250 ng/ml amphotericin B. Treatment with collagenase (125 U/ml) and elastase (0.5 mg/ml) in HBSS without Ca^{2+} and Mg^{2+} follow for 2 hours at 37°C. The reaction is stopped by addition of DMEM/F12 medium containing 20% fetal calf serum and penicillin, streptomycin, and amphotericin B, and the cells are plated onto 60-mm cell culture plates. The cells are used between passage 3 and 6. The Ca^{2+} concentration measurements are performed as described in Meyer zu Heringdorf *et al.* (1996) *Naunyn-Schmiedeberg's Arch. Pharmacol.* 354:397-403. Briefly, the cells are loaded with 1 μ M fura2/AM for 1 hour at room temperature in HBSS, washed with HBSS, and used for fluorescence measurements within the next hour. Ca^{2+} concentrations are measured in a continuously stirred cell suspension at room temperature in a Hitachi F2000 spectrofluorometer as described in Meyer zu Heringdorf *et al.* (1996) *supra*.

EXAMPLE 6: CALCIUM TRANSPORT ASSAY

This example describes the assessment of calcium transport by HEAT molecules in cultured COS-1 cells, as described in, for example, Maruyama, K. and MacLennan, D.H. (1988) *Proc. Natl. Acad. Sci. USA* 85:3314-3318.

Cell Culture and DNA Transfection

COS-1 or HEK-293 cells are maintained in Dulbecco's modified Eagle's medium (DMEM) with 0.1 mM α -MEM nonessential amino acids, 4 mM L-glutamine, 100 units of penicillin per ml, 100 μ g of streptomycin per ml, and 10%

fetal calf serum under 5% CO₂/95% air at 37°C. Transfection of HEAT-containing DNA is carried out by the DEAE dextran-chloroquine shock method (Sompayrac, L.M. and Danna, K.J. (1981) *Proc. Natl. Acad. Sci. USA* 78:7575-7578; Gorman, C. (1985) in *DNA Cloning: A Practical Approach*, ed. Gover, D.M. (IRL, Washington, DC), Vol. 2, pp. 143-190) with 25 µg of cesium chloride gradient-purified DNA and 1.5 mg of DEAE dextran per 10 cm Petri dish. Cells are then incubated for 3 hours at 37°C in 6 ml of DMEM containing 300 µg of chloroquine, washed, and cultured in DMEM for 48 or 72 hours. Control cells are treated in the same way with vector DNA or with no added DNA.

10

Isolation of Microsomal Fraction

For isolation of a microsomal fraction (Resh, M.D. and Erikson, R.L. (1985) *J. Cell Biol.* 100:409-417; Yamada, S. and Ikemoto, N. (1980) *J. Biol. Chem.* 255:3108-3119), cells from five 10 cm Petri dishes are washed twice with 5 ml of a solution of 0.137 M NaCl/2.7 mM KCl/8 mM Na₂HPO₄/1.5 mM KH₂PO₄ (PBS), harvested in a solution of 5 mM EDTA in PBS and washed with 5 ml of PBS. The cells are swollen at 0°C for 10 minutes in 2 ml of a hypotonic solution of 10 mM Tris-HCl, pH 7.5/0.5 mM MgCl₂, and then phenylmethylsulfonyl fluoride and Trasylol are added to 0.1 mM and 100 units/ml, respectively. The cells are homogenized with 30 strokes in a glass Dounce homogenizer, and the homogenate is diluted with an equal volume of a solution of 0.5 M sucrose/6 mM 2-mercaptoethanol, 40 µM CaCl₂/300 mM KCl/10 mM Tris-HCl, pH 7.5. The suspension is centrifuged at 10,000 × g for 20 minutes to pellet nuclei and mitochondria. The supernatant is brought to a concentration of 0.6 M KCl by the addition of 0.9 ml of a 2.5 M solution. The suspension is centrifuged at 100,000 × g for 60 minutes to sediment the microsomal fraction. The pellet is suspended in a solution containing 0.25 M sucrose, 0.15 M KCl, 3 mM 2-mercaptoethanol, 20 µM CaCl₂, 10 mM Tris-HCl (pH 7.5), and centrifuged again at 100,000 × g for 60 minutes. The final pellet, containing approximately 100 µg of protein, is suspended in the same solution at a protein concentration of 1 mg/ml.

30

Ca²⁺ Transport Assay

Ca²⁺ transport activity is assayed in a reaction mixture containing 20 µM Mops-KOH (pH 6.8), 100 mM KCl, 5 mM CaCl₂, 5 mM ATP, 0.45 mM CaCl₂ (containing ⁴⁵Ca at a specific activity of 10⁶ cpm/µmol), 0.5 mM EGTA, and 5 mM potassium oxalate. The uptake reaction is initiated by the addition of 10 µg of microsomal protein to 1 ml of reaction mixture at room temperature. At different time points, 0.15 ml samples are filtered through a 0.3 µm Millipore filter and washed

35

with 10 ml of 0.15 M KCl. Radioactivity on the filter is measured by liquid scintillation counting. For the measurement of Ca^{2+} ion dependency, free Ca^{2+} concentration is calculated by the computer program of Fabiato and Fabiato ((1979) *J. Physiol. (London)* 75:463-505). For the measurement of ATP dependency, an ATP
5 regenerating system consisting of 2.5 mM phosphoenolpyruvate and 50 μg of pyruvate kinase per ml is used.

Measurement of Phosphorylated HEAT Intermediate

Microsomal protein (5 μg) is added to 0.1 ml of a solution of 20 mM Mops,
10 pH 6.8/100 mM KCl/5 mM MgCl_2 /0.5 mM EGTA in the presence or absence of 0.5 mM CaCl_2 . The reaction, at ice temperature, is started by the addition of 5 μM ATP (10⁶ cpm/nmol) and stopped after 5 seconds by the addition of 0.6 ml of a mixture of 5% trichloroacetic acid and 5 mM potassium phosphate. Incorporation of ³²P is
15 determined either by collecting the protein on a filter for scintillation counting or by separating the protein in acidic NaDodSO₄/polyacrylamide gels for autoradiography Sarkadi, B. et al. (1986) *J. Biol. Chem.* 261:9552-9557).

EXAMPLE 7: ANALYSIS OF HEAT-3 ACTIVITY

The full-length HEAT-3 was inserted into the multiple cloning site in the
20 pCDNA3 vector. The DNA for the clone was amplified and transfected into HEK-293 cells using calcium phosphate precipitation. After 72 hours, the cells were harvested, microsomal fractions isolated, and ⁴⁵Ca-uptake measured as a function of calcium concentration using a filter assay.

Two different HEAT-3 fusion proteins were generated. One HEAT-3 fusion
25 protein was created by inserting the 3X Flag epitope at the 3' end of the HEAT-3 gene. Another HEAT-3 fusion protein was created by inserting the green fluorescent protein (GFP) at the 3' end of the HEAT-3 gene. Fluorescence of this protein could be observed with the naked eye and by confocal microscopy. Measurement of expression using Western blotting with an anti-GFP antibody showed that HEAT-3 is
30 well expressed in the microsomal fraction.

Confocal microscopy showed that the expression pattern of HEAT-3 is similar to SERCA1, indicating that HEAT-3 is targeted to and localized in the endoplasmic reticulum of HEK-293 cells.

Ca^{2+} uptake experiments (as described in Example 6) were performed using
35 both the Flag and GFP fusion proteins. An increase in Ca^{2+} uptake of HEAT-3 was shown over GFP vector alone. Five independent experiments were performed to confirm the increase of calcium uptake with HEAT-3 as compared to vector alone. The Flag fusion protein showed an increase of calcium uptake as compared to Flag

vector alone. In these experiments, the Vmax for HEAT-3 was lower (about 20X) than the Vmax for SERCA1 under the same conditions. The KCa for HEAT-3 was about 6.00 pCa units, as compared with about 6.38 pCa units for SERCA1. In the presence of ATP, there is 2-3 fold more calcium uptake compared to uptake in the
5 absence of ATP, indicating that the calcium uptake by HEAT-3 is ATP dependent..

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the
10 invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed:

1. An isolated nucleic acid molecule selected from the group consisting of:
 - (a) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:1, 5, or 8; and
 - (b) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:3, 7, or 10.
2. An isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2, 6, or 9.
3. An isolated nucleic acid molecule comprising the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number _____, _____, or _____.
4. An isolated nucleic acid molecule which encodes a naturally-occurring allelic variant of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2, 6, or 9.
5. An isolated nucleic acid molecule selected from the group consisting of:
 - (a) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to the nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or a complement thereof;
 - (b) a nucleic acid molecule comprising a fragment of at least 30 nucleotides of a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10, or a complement thereof;
 - (c) a nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 60% identical to the amino acid sequence of SEQ ID NO:2, 6, or 9; and
 - (d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 6, or 9, wherein the fragment comprises at least 10 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2, 6, or 9.
6. An isolated nucleic acid molecule which hybridizes to a complement of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 under stringent conditions.

7. An isolated nucleic acid molecule comprising a nucleotide sequence which is complementary to the nucleotide sequence of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.
- 5
8. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5, and a nucleotide sequence encoding a heterologous polypeptide.
- 10
9. A vector comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.
10. The vector of claim 9, which is an expression vector.
- 15
11. A host cell transfected with the expression vector of claim 10.
12. A method of producing a polypeptide comprising culturing the host cell of claim 11 in an appropriate culture medium to, thereby, produce the polypeptide.
- 20
13. An isolated polypeptide selected from the group consisting of:
- a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 6, or 9, wherein the fragment comprises at least 10 contiguous amino acids of SEQ ID NO:2, 6, or 9;
- 25
- b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 6, or 9, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to complement of a nucleic acid molecule consisting of SEQ ID NO:1, 3, 5, 7, 8, or 10 under stringent conditions;
- 30
- c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 8, or 10; and
- d) a polypeptide comprising an amino acid sequence which is at least 60% identical to the amino acid sequence of SEQ ID NO:2, 6, or 9.
- 35
14. The isolated polypeptide of claim 13 comprising the amino acid sequence of SEQ ID NO:2, 6, or 9.

15. The polypeptide of claim 13, further comprising heterologous amino acid sequences.

16. An antibody which selectively binds to a polypeptide of claim 13.

5

17. A method for detecting the presence of a polypeptide of claim 13 in a sample comprising:

a) contacting the sample with a compound which selectively binds to the polypeptide; and

10 b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of claim 13 in the sample.

18. The method of claim 17, wherein the compound which binds to the polypeptide is an antibody.

15

19. A kit comprising a compound which selectively binds to a polypeptide of claim 13 and instructions for use.

20. A method for detecting the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in a sample comprising:

20 a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and
b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample to thereby detect the presence of a nucleic acid
25 molecule of any one of claims 1, 2, 3, 4, or 5 in the sample.

21. The method of claim 20, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

30 22. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 and instructions for use.

23. A method for identifying a compound which binds to a polypeptide of claim 13 comprising:

35 a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound; and

b) determining whether the polypeptide binds to the test compound.

24. The method of claim 23, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

- a) detection of binding by direct detection of test compound/polypeptide binding;
- b) detection of binding using a competition binding assay; and
- c) detection of binding using an assay for HEAT activity.

25. A method for modulating the activity of a polypeptide of claim 13 comprising contacting the polypeptide or a cell expressing the polypeptide with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

26. A method for identifying a compound which modulates the activity of a polypeptide of claim 13 comprising:

- a) contacting a polypeptide of claim 13 with a test compound; and
- b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

20

27. A method of identifying a subject having a cardiovascular disorder, or at risk for developing a cardiovascular disorder comprising:

- a) contacting a sample obtained from said subject comprising nucleic acid molecules with a hybridization probe comprising at least 25 contiguous nucleotides of SEQ ID NO:1, 5, or 8; and
- b) detecting the presence of a nucleic acid molecule in said sample that hybridizes to said probe, thereby identifying a subject having a cardiovascular disorder.

28. A method of identifying a subject having a cardiovascular disorder, or at risk for developing a cardiovascular disorder comprising:

- a) contacting a sample obtained from said subject comprising nucleic acid molecules with a first and a second amplification primer, said first primer comprising at least 25 contiguous nucleotides of SEQ ID NO:1, 5, or 8 and said second primer comprising at least 25 contiguous nucleotides from the complement of SEQ ID NO:1, 5, or 8;
- b) incubating said sample under conditions that allow nucleic acid amplification; and

c) detecting the presence of a nucleic acid molecule in said sample that is amplified, thereby identifying a subject having a cardiovascular disorder, or at risk for developing a cardiovascular disorder.

5 29. A method of identifying a subject having a cardiovascular disorder, or at risk for developing a cardiovascular comprising:

a) contacting a sample obtained from said subject comprising polypeptides with a HEAT binding substance; and

10 b) detecting the presence of a polypeptide in said sample that binds to said HEAT binding substance, thereby identifying a subject having a cardiovascular disorder, or at risk for developing a cardiovascular disorder.

30. A method for identifying a compound capable of treating a cardiovascular disorder characterized by aberrant HEAT nucleic acid expression or
15 HEAT polypeptide activity comprising assaying the ability of the compound to modulate HEAT nucleic acid expression or HEAT polypeptide activity, thereby identifying a compound capable of treating a cardiovascular disorder characterized by aberrant HEAT nucleic acid expression or HEAT polypeptide activity.

20 31. A method for treating a subject having a cardiovascular disorder characterized by aberrant HEAT polypeptide activity or aberrant HEAT nucleic acid expression comprising administering to the subject a HEAT modulator, thereby treating said subject having a cardiovascular disorder.

TACTATAGGGAGTCGACCCACGCGTCCGGCCGCGCCGAGCCTGGTGGCCCAGGTGCCCCGCCCGCGTCAGCCCTGCTCC
AGCCCCGCGCTAGCCCAGCGCCCTCGCCCCGGGCCGTCCGGACCGCGCCCCCGCCCAGGGCCTTGCGCACGCCGGGGC

CCAGGCCGAGGGCCGAGCGCCGGGGCCGGCGATGAGCGCGAGGAGCCGGC M S A D S S P 7
ATG AGC GCA GAC AGC AGC CCT 21

L V G S T P T G Y G T L T I G T S I D P 27
CTC GTG GGC AGC ACG CCC ACC GGT TAT GGG ACC CTG ACG ATA GGG ACA TCA ATA GAT CCC 81

L S S S V S S V R L S G Y C G S P W R V 47
CTC AGC TCC TCA GTT TCA TCC GTG AGG CTC AGC GGC TAC TGT GGC AGT CCA TGG AGG GTC 141

I G Y H V V V W M M A G I P L L L F R W 67
ATC GGC TAT CAC GTC GTG GTC TGG ATG ATG GCT GGG ATC CCT TTG CTG CTC TTC CGT TGG 201

K P L W G V R L R L R P C N L A H A E T 87
AAG CCC CTG TGG GGC GTG CGG CTG CGG CTC CGG CCC TGC AAC CTG GCC CAC GCC GAA ACA 261

L V I E I R D K E D S S W Q L F T V Q V 107
CTC GTT ATC GAA ATA AGA GAC AAA GAG GAT AGT TCC TGG CAG CTC TTC ACT GTC CAG GTG 321

Q T E A I G E G S L E P S P Q S Q A E D 127
CAG ACT GAG GCC ATC GGC GAG GGC AGC CTG GAG CCG TCC CCA CAG TCC CAG GCA GAG GAT 381

G R S Q A A V G A V P E G A W K D T A Q 147
GGC CGG AGC CAG GCG GCA GTT GGG GCG GTA CCA GAG GGT GCC TGG AAG GAT ACG GCC CAG 441

L H K S E E A V S V G Q K R V L R Y Y L 167
CTC CAC AAG AGC GAG GAG GCG GTG AGT GTC GGA CAG AAG CGG GTG CTG CGG TAT TAC CTC 501

F Q G Q R Y I W I E T Q Q A F Y Q V S L 187
TTC CAG GGC CAG CGC TAT ATC TGG ATC GAG ACC CAG CAA GCC TTC TAC CAG GTC AGC CTC 561

L D H G R S C D D V H R S R H G L S L Q 207
CTG GAC CAT GGC CGC TCT TGT GAC GAC GTC CAC CGC TCC CGC CAT GGC CTC AGC CTC CAG 621

D Q M V R K A I Y G P N V I S I P V K S 227
GAC CAA ATG GTG AGG AAG GCC ATT TAC GGC CCC AAC GTG ATC AGC ATA CCG GTC AAG TCC 681

Y P Q L L V D E A L N P Y Y G F Q A F S 247
TAC CCC CAG CTG CTG GTG GAC GAG GCA CTG AAC CCC TAC TAT GGG TTC CAG GCC TTC AGC 741

I A L W L A D H Y Y W Y A L C I F L I S 267
ATC GCG CTG TGG CTG GCT GAC CAC TAC TAC TGG TAC GCC CTG TGC ATC TTC CTC ATT TCC 801

S I S I C L S L Y K T R K Q S Q T L R D 287
TCC ATC TCC ATC TGC CTG TCG CTG TAC AAG ACC AGA AAG CAA AGC CAG ACT CTA AGG GAC 861

M V K L S M R V C V C R P G G E E E W V 307
ATG GTC AAG TTG TCC ATG CGG GTG TGC GTG TGC CGG CCA GGG GGA GAG GAA GAG TGG GTG 921

D S S E L V P G D C L V L P Q E G G L M 327
GAC TCC AGT GAG CTA GTG CCC GGA GAC TGC CTG GTG CTG CCC CAG GAG GGT GGG CTG ATG 981

Fig. 1A

P	C	D	A	A	L	V	A	G	E	C	M	V	N	E	S	S	L	T	G	347
CCC	TGT	GAT	GCC	GCC	CTG	GTG	GCC	GGC	GAG	TGC	ATG	GTG	AAT	GAG	AGC	TCT	CTG	ACA	GGA	1041
E	S	I	P	V	L	K	T	A	L	P	E	G	L	G	P	Y	C	A	E	367
GAG	AGC	ATT	CCA	GTG	CTG	AAG	ACG	GCA	CTG	CCG	GAG	GGG	CTG	GGG	CCC	TAC	TGT	GCA	GAG	1101
T	H	R	R	H	T	L	F	C	G	T	L	I	L	Q	A	R	A	Y	V	387
ACA	CAC	CGG	CGG	CAC	ACA	CTC	TTC	TGC	GGG	ACC	CTC	ATC	TTG	CAG	GCC	CGG	GCC	TAT	GTG	1161
G	P	H	V	L	A	V	V	T	R	T	G	F	C	T	A	K	G	G	L	407
GGA	CCG	CAC	GTC	CTG	GCA	GTG	GTG	ACC	CGC	ACA	GGG	TTC	TGC	ACG	GCA	AAA	GGG	GGC	CTG	1221
V	S	S	I	L	H	P	R	P	I	N	F	K	F	Y	K	H	S	M	K	427
GTG	AGC	TCC	ATC	TTG	CAC	CCC	CGG	CCC	ATC	AAC	TTC	AAG	TTC	TAT	AAA	CAC	AGC	ATG	AAG	1281
F	V	A	A	L	S	V	L	A	L	L	G	T	I	Y	S	I	F	I	L	447
TTT	GTG	GCT	GCC	CTC	TCT	GTC	CTG	GCT	CTC	CTC	GGC	ACC	ATC	TAC	AGC	ATC	TTC	ATC	CTC	1341
Y	R	N	R	V	P	L	N	E	I	V	I	R	A	L	D	L	V	T	V	467
TAC	CGA	AAC	CGG	GTG	CCT	CTG	AAT	GAG	ATT	GTA	ATC	CGG	GCT	CTC	GAC	CTG	GTG	ACC	GTG	1401
V	V	P	P	A	L	P	A	A	M	T	V	C	T	L	Y	A	Q	S	R	487
GTG	GTG	CCA	CCT	GCC	CTG	CCT	GCT	GCC	ATG	ACT	GTG	TGC	ACG	CTC	TAC	GCC	CAG	AGC	CGA	1461
L	R	R	Q	G	I	F	C	I	H	P	L	R	I	N	L	G	G	K	L	507
CTG	CGG	AGA	CAG	GGC	ATT	TTC	TGC	ATC	CAC	CCA	CTG	CGC	ATC	AAC	CTG	GGG	GGC	AAG	CTG	1521
Q	L	V	C	F	D	K	T	G	T	L	T	E	D	G	L	D	V	M	G	527
CAG	CTG	GTG	TGT	TTC	GAC	AAG	ACG	GGC	ACC	CTC	ACT	GAG	GAC	GGC	TTA	GAC	GTG	ATG	GGG	1581
V	V	P	L	K	G	Q	A	F	L	P	L	V	P	E	P	R	R	L	P	547
GTG	GTG	CCC	CTG	AAG	GGG	CAG	GCA	TTC	CTG	CCC	CTG	GTG	CCA	GAG	CCT	CGC	CGC	CTG	CCT	1641
V	G	P	L	L	R	A	L	A	T	C	H	A	L	S	R	L	Q	D	T	567
GTG	GGG	CCC	CTG	CTC	CGA	GCA	CTG	GCC	ACC	TGC	CAT	GCC	CTC	AGC	CGG	CTC	CAG	GAC	ACC	1701
P	V	G	D	P	M	D	L	K	M	V	E	S	T	G	W	V	L	E	E	587
CCC	GTG	GGC	GAC	CCC	ATG	GAC	TTG	AAG	ATG	GTG	GAG	TCT	ACT	GGC	TGG	GTC	CTG	GAG	GAA	1761
E	P	A	A	D	S	A	F	G	T	Q	V	L	A	V	M	R	P	P	L	607
GAG	CCG	GCT	GCA	GAC	TCA	GCA	TTT	GGG	ACC	CAG	GTG	TTG	GCA	GTG	ATG	AGA	CCT	CCA	CTT	1821
W	E	P	Q	L	Q	A	M	E	E	P	P	V	P	V	S	V	L	H	R	627
TGG	GAG	CCC	CAG	CTG	CAG	GCA	ATG	GAG	GAG	CCC	CCG	GTG	CCA	GTC	AGC	GTC	CTC	CAC	CGC	1881
F	P	F	S	S	A	L	Q	R	M	S	V	V	V	A	W	P	G	A	T	647
TTC	CCC	TTC	TCT	TCG	GCT	CTG	CAG	CGC	ATG	AGT	GTG	GTG	GTG	GCG	TGG	CCA	GGG	GCC	ACT	1941
Q	P	E	A	Y	V	K	G	S	P	E	L	V	A	G	L	C	N	P	E	667
CAG	CCC	GAG	GCC	TAC	GTC	AAA	GGC	TCC	CCG	GAG	CTG	GTG	GCA	GGG	CTC	TGC	AAC	CCC	GAG	2001
T	V	P	T	D	F	A	Q	M	L	Q	S	Y	T	A	A	G	Y	R	V	687
ACA	GTG	CCC	ACC	GAC	TTC	GCC	CAG	ATG	CTG	CAG	AGC	TAT	ACA	GCT	GCT	GGC	TAC	CGT	GTC	2061

Fig. 1B

V	A	L	A	S	K	P	L	P	T	V	P	S	L	E	A	A	Q	Q	L	707
GTG	GCC	CTG	GCC	AGC	AAG	CCA	CTG	CCC	ACT	GTG	CCC	AGC	CTG	GAG	GCA	GCC	CAG	CAA	CTG	2121
T	R	D	T	V	E	G	D	L	S	L	L	G	L	L	V	M	R	N	L	727
ACG	AGG	GAC	ACT	GTG	GAA	GGA	GAC	CTG	AGC	CTC	CTG	GGG	CTG	CTG	GTC	ATG	AGG	AAC	CTA	2181
L	K	P	Q	T	T	P	V	I	Q	A	L	R	R	T	R	I	R	A	V	747
CTG	AAG	CCG	CAG	ACA	ACG	CCA	GTT	ATC	CAG	GCT	CTG	CGA	AGG	ACC	CGC	ATC	CGC	GCC	GTC	2241
M	V	T	G	D	N	L	Q	T	A	V	T	V	A	R	G	C	G	M	V	767
ATG	GTG	ACA	GGG	GAC	AAC	CTG	CAG	ACA	GCG	GTG	ACT	GTG	GCC	CGG	GGC	TGT	GGC	ATG	GTG	2301
A	P	Q	E	H	L	I	I	V	H	A	T	H	P	E	R	G	Q	P	A	787
GCC	CCC	CAG	GAG	CAT	CTG	ATC	ATC	GTC	CAC	GCC	ACC	CAC	CCT	GAG	CGG	GGT	CAG	CCT	GCC	2361
S	L	E	F	L	P	M	E	S	P	T	A	V	N	G	V	K	D	P	D	807
TCT	CTC	GAG	TTC	CTG	CCG	ATG	GAG	TCC	CCC	ACA	GCC	GTG	AAT	GGC	GTT	AAG	GAT	CCT	GAC	2421
Q	A	A	S	Y	T	V	E	P	D	P	R	S	R	H	L	A	L	S	G	827
CAG	GCT	GCA	AGC	TAC	ACC	GTG	GAG	CCA	GAC	CCC	CGA	TCC	AGG	CAC	CTG	GCC	CTC	AGC	GGG	2481
P	T	F	G	I	I	V	K	H	F	P	K	L	L	P	K	V	L	V	Q	847
CCC	ACC	TTT	GGT	ATC	ATT	GTG	AAG	CAC	TTC	CCC	AAG	CTG	CTG	CCC	AAG	GTC	CTG	GTC	CAG	2541
G	T	V	F	A	R	M	A	P	E	Q	K	T	E	L	V	C	E	L	Q	867
GGC	ACT	GTC	TTT	GCC	CGC	ATG	GCC	CCT	GAG	CAG	AAG	ACA	GAG	CTG	GTG	TGC	GAG	CTA	CAG	2601
K	L	Q	Y	C	V	G	M	C	G	D	G	A	N	D	C	G	A	L	K	887
AAG	CTT	CAG	TAC	TGC	GTG	GGC	ATG	TGC	GGA	GAC	GGT	GCC	AAT	GAC	TGT	GGG	GCC	CTG	AAG	2661
A	A	D	V	G	I	S	L	S	Q	A	E	A	S	V	V	S	P	F	T	907
GCG	GCT	GAT	GTC	GGC	ATC	TCG	CTG	TCC	CAG	GCA	GAA	GCC	TCA	GTG	GTC	TCA	CCC	TTC	ACC	2721
S	S	M	A	S	I	E	C	V	P	M	V	I	R	E	G	R	C	S	L	927
TCG	AGC	ATG	GCC	AGT	ATT	GAG	TGC	GTG	CCC	ATG	GTC	ATC	AGG	GAG	GGG	CGC	TGT	TCC	CTT	2781
D	T	S	F	S	V	F	K	Y	M	A	L	Y	S	L	T	Q	F	I	S	947
GAC	ACT	TCG	TTC	AGC	GTC	TTC	AAG	TAC	ATG	GCT	CTG	TAC	AGC	CTG	ACC	CAG	TTC	ATC	TCC	2841
V	L	I	L	Y	T	I	N	T	N	L	G	D	L	Q	F	L	A	I	D	967
GTC	CTG	ATC	CTC	TAC	ACG	ATC	AAC	ACC	AAC	CTG	GGT	GAC	CTG	CAG	TTC	CTG	GCC	ATC	GAC	2901
L	V	I	T	T	T	V	A	V	L	M	S	R	T	G	P	A	L	V	L	987
CTG	GTC	ATC	ACC	ACC	ACA	GTG	GCA	GTG	CTC	ATG	AGC	CGC	ACG	GGG	CCA	GCG	CTG	GTC	CTG	2961
G	R	V	R	P	P	G	A	L	L	S	V	P	V	L	S	S	L	L	L	1007
GGA	CGG	GTA	CGG	CCA	CCG	GGG	GCG	CTG	CTC	AGC	GTG	CCC	GTG	CTC	AGC	AGC	CTG	CTG	CTG	3021
Q	M	V	L	V	T	G	V	Q	L	G	G	Y	F	L	T	L	A	Q	P	1027
CAG	ATG	GTC	CTG	GTG	ACC	GGC	GTG	CAG	CTA	GGG	GGC	TAC	TTC	CTG	ACC	CTG	GCC	CAG	CCA	3081
W	F	V	P	L	N	R	T	V	A	A	P	D	N	L	P	N	Y	E	N	1047
TGG	TTC	GTG	CCT	CTG	AAC	AGG	ACA	GTG	GCC	GCA	CCA	GAC	AAC	CTG	CCC	AAC	TAC	GAG	AAC	3141

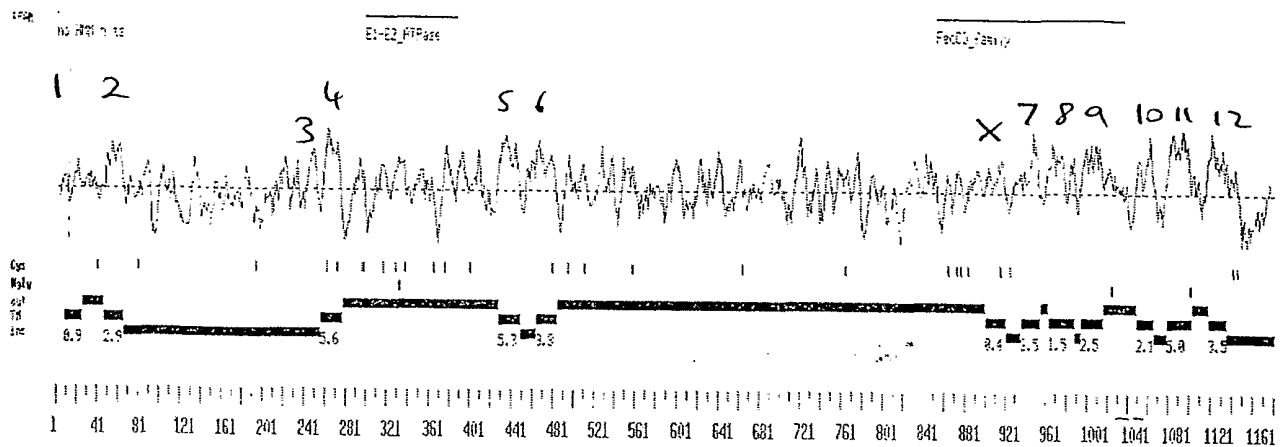
Fig. 1C

T	V	V	F	S	L	S	S	F	Q	Y	L	I	L	A	A	A	V	S	K	1067
ACC	GTG	GTC	TTC	TCT	CTG	TCC	AGC	TTC	CAG	TAC	CTC	ATC	CTG	GCT	GCA	GCC	GTG	TCC	AAG	3201
G	A	P	F	R	R	P	L	Y	T	N	V	P	F	L	V	A	L	A	L	1087
GGG	GCG	CCC	TTC	CGC	CGG	CCG	CTC	TAC	ACC	AAT	GTG	CCC	TTC	CTG	GTG	GCC	CTG	GCG	CTC	3261
L	S	S	V	L	V	G	L	V	L	V	P	G	L	L	Q	G	P	L	A	1107
CTG	AGC	TCC	GTC	CTG	GTG	GGC	CTT	GTC	CTG	GTC	CCC	GGC	CTC	CTG	CAG	GGG	CCG	CTG	GCG	3321
L	R	N	I	T	D	T	G	F	K	L	L	L	L	G	L	V	T	L	N	1127
CTG	AGG	AAC	ATC	ACT	GAC	ACC	GGC	TTC	AAG	CTG	CTG	CTG	CTG	GGT	CTG	GTC	ACC	CTC	AAC	3381
F	V	G	A	F	M	L	E	S	V	L	D	Q	C	L	P	A	C	L	R	1147
TTC	GTG	GGG	GCC	TTC	ATG	CTG	GAG	AGC	GTG	CTA	GAC	CAG	TGC	CTC	CCC	GCC	TGC	CTG	GCG	3441
R	L	R	P	K	R	A	S	K	K	R	F	K	Q	L	E	R	E	L	A	1167
CGC	CTC	CGG	CCC	AAG	CGG	GCC	TCC	AAG	AAG	CGC	TTC	AAG	CAG	CTG	GAA	CGA	GAG	CTG	GCC	3501
E	Q	P	W	P	P	L	P	A	G	P	L	R	*							1181
GAG	CAG	CCC	TGG	CCA	CCG	CTG	CCC	GCC	GGC	CCC	CTG	AGG	TAG							3543

TGCAGGCCCACGGGCACCCCAGACACTGGAACCTCCCTGCCTCTGAGCCACCAACTGGACCCCTCTCCAGCAACACCACC
GCCACCACCTCCCACATCCCTGAGGTTGGCGACTGTCTACACTCCTCCCCGAGACCACCCCCACCCCTGGGGAAGCGTT
GACTACTGTCCCCTACCTTGGACCATCCCGCGTAGGGGTGGCAGCCCCCAGCTCCCCTCAGTGCTGCTGTCTCAGTGCTAGC
AAATAAAGTCATGATATTTTCTGGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Fig. 1D

Analysis of Fbh49937 (1180 aa)



>Fbh49937

MSADSSPLVGSTPTGYGTLTIGTSIDPLSSSVSSVRLSGYCGSPWRVIGYHVVVMMAGI
 PLLLFRWKPLWGVRLRLRPCNLAAHAEITLVIEIRDKEDSSWQLFTVQVQTEAIGEGSLEPS
 PQSQAEDGRSQAAGAVPEGAWKDTAQLHKSEEAHSVGVQKRVLRYYLFQGGQRWIWIETQQ
 AFYQVSLLDHGRSCDDVHRSRHGLSLQDMVRKAIYGNVISIPVKSYPQLLVDEALNPY
 YGFQAFSIALWLADHYWYALCIFLISSISICLSLYKTRKQSQTLRDMVKLSMRVCVCRP
 GGEEWVDSSSELVPGDCLVLPQEGGLMPCDAALVAGECMVNESSLTGESIPVLKTALPEG
 LGPYCAETHRRHTLFCGTLILQARAYVGVHVLAVVTRTGCTAKGGLVSSILHPRPINFK
 FYKHSMKFVAALSVLALLGTIYSIFILYRNRVPLNEIVIRALDLVTVVVPPALPAAMTVC
 TLYAQSRRLRRQGFICIHPLRLNLGGKLQLVCFDKTGTLTEDGLDVMGVVPLKGQAFPLV
 PEPRLPVGPPLRALATCHALSRLQDTPVGDPMDLKMVESTGWVLEEEPAADSAGFTQVL
 AVMRPPLWEPQLQAMEEPPVPVSVLHRFPFSSALQRMSSVVVAWPGATQPEAYVKGSPVL
 AGLCNPETVPTDFAQMLQSYTAAGYRVVALASKPLPTVPSLEAAQQLTRDTVEGDLSSLG
 LLVMRNLLKPQTTPIQALRRTRIRAVMVTGDNLQTAVTVARGCGMVAPQEHLIIIVHATH
 PERGQPASLEFLPMESPTAVNGVKDPDQAASYTVEPDPRSRHLALSGPTFGIIVKHFPKL
 LPKVLVQGTVFARMAPEQKTELVCLEQLQYCVGMCGDGANDCGALKAADVGISLSQAEA
 SVVSPFTSSMASIECVPMVIREGRCSLDTFSVFKYMALYSLTQFISVLILYTINTNLGD
 LQFLAIDLVIITTTVAVLMSRTGPAVLGRVVRPPGALLSVPVLSLLLQMVLTGVQLGGY
 FLTLAQPFVPLNRTVAAPDNLPNYENTVVFSLSSFQYLILAAAVSKGAPFRRPLYTNVP
 FLVALALLSSVLVGLVLPGLLQGPLALRNITDTGFKLLLLGLVTLNFGAFMLESVLDQ
 CLPACLRRLRPKRASKKRFKQLERELAEQWPPLPAGPLR

FIG. 2

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM
 hmmpfam - search a single seq against HMM database
 HMMER 2.1.1 (Dec 1998)
 Copyright (C) 1992-1998 Washington University School of Medicine
 HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam5.4/Pfam
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.26407.seq

Query: Fbh49937

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
E1-E2_ATPase	E1-E2 ATPase	51.4	1.5e-14	1
FecCD_family	FecCD transport family	-218.8	7.5	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
E1-E2_ATPase	1/1	299	387	..	39 124 ..	51.4	1.5e-14
FecCD_family	1/1	852	1035	..	1 307 []	-218.8	7.5

Alignments of top-scoring domains:

E1-E2_ATPase: domain 1 of 1, from 299 to 387: score 51.4, E = 1.5e-14

```

      *->RdG.keeeipaeELvpGDIVev.kpGdrVPADGrvvegestLkVDES
      R G+++e ~++++LvpGD +++++G +P D+ +v+ge+ +V+ES
Fbh49937  299  RPGgEEEWVDSSELVPGDCLVLpQEGGLMPCDAALVAGEC--MVNES 343

```

```

      aLTGESlPVeKhsdvtp.....dnp.lerngmVfaGtlvvsGsItg<-*
      +LTGES+PV K   +++p++ ++ +++ +r+ ++f GTl+++ ++++
Fbh49937  344  SLTGESIPVLK---TALPeglgpYCAeTHRRHTLFCGTLILQARAYV 387

```

FecCD_family: domain 1 of 1, from 852 to 1035: score -218.8, E = 7.5

```

      *->GalsispadvlqalfgggtegevdeliidltlrRLPRvLlAlLVGA
      + +p + + + + el +++ +VG+
Fbh49937  852  ARM--APEQKTELVC-----ELQKLQY-----CVGM 875

```

```

      aLAVaGAILQgltrNPLASpgilGins.GAslgvvlaivlfpgsisalyI
      + + + + + G n+ GA ++ + i l ++ +
Fbh49937  876  C-----GD-----GANDcGALKAADVGISLSQA----EAS 901

```

```

      lpsfAfaGaliallVyllawkggrnglspvrLiLaGialsalfsAltll
      ++s f +++ V + ++ gr +l+++ ++ ++
Fbh49937  902  VVS-PFTSSMASIECVPMVIREGRCSLDTSF-----SVFKYM 937

```

```

      lllsddlqdaqalfWltGSlsgrnWedvklalpilliglplalllarqLn
      +l+s ++ + + +++ ++n+ d+ +++++ l+i++ a l+r
Fbh49937  938  ALYSLTQFISVLILYTI---NTNLGDLOFLAIDLVTITTTVAVLMSRTGP 983

```

```

      vLsLGddtAkgLGvnvervRllllllvvlLtGaaVAvAGpIgFVGLivPH
      +L LG + + + + l + +l+lL
Fbh49937  984  ALVLGR-----VRPPGALLSVPVLSSLL----- 1006

```

```

      iaRrLvGtdhrwLLPaSALlGAILLlADllARTlfaPiElPvGivTali
      L + +L+ + +l + l + +
Fbh49937  1007 -----LQM-----VLVTGVQLGGYFLTLAQ----- 1026

```

```

      GaPyFLYLLrr<-*
      P+F+ L r
Fbh49937  1027 --PWFVPLNRT 1035

```

FIG. 3

CLUSTAL W (1.1.4) multiple sequence alignment

```
Fbh49937      MSADSSPLVGSTPTGYGTLTIGTSDPLSSSVSSVRLSGYCGSPWRVIGYHVVVMMAGI
YH2Melegans    MTLES-----GDHTLT-----FAYRT---G-PFRITLFYALTIVLTIGI
               *: :*          *   ***:          : *      * *: * * : : : : * *

Fbh49937      PLLLFWRWKPLWGVRLRLRPCNLAHAETLVIEIRDKEDSSWQLFTVQVQTEAIGEGSLEPS
YH2Melegans    FRLILHWKQKWDVKMRMVPCTFEAAE--YIYIIDNHNVS-----ELQ-PVLRK-SNATI
               *:::* * *: * : * : * * * * : : * : : * :

Fbh49937      PQSQAEDGRSQAAVGAVPEGAWKDTAQLHKSEEAHSVQKRVLYRYLFQGQRYIWIETQQ
YH2Melegans    P---TENGER---KVPELRWFVYRKLEYVWIDDLNSDESVDSEISDNDN---CWKTS--
               *   :*: * :          *** *   :*.          : : *   : : *   :

Fbh49937      AFYQVSLLDHGRSCDDVHRSRHLGLSDQDMVRKAIYGPVVISIPVKSYPQLLVDEALNPY
YH2Melegans    --FEIANRIPCRSLLAVSESNFGLTLSEISRRLEFYGRNEIVVQLRPILYLLVMEVITPF
               : : :          * *   *   . : . * : : :          *   : * : : :

Fbh49937      YGFQAFSIALWLADHYWYALCIFLISSISICLSLYKTRKQSOTLRDMVKLSMRVCVCRP
YH2Melegans    YVFQIFSIVTVWYNDEYAYYASLIVLSLGSIVMDVYQIRTQEI RLRS MVHSTESVEVIRE
               * * * * : : *   * . * : * *   . : : * *   . : : * *   . : * * *

Fbh49937      GGESEWVDSSSELVPGDCLVLPQEGGLMPCDAALVAGECMVNESSLTGESIPVLKTALPEG
YH2Melegans    G-TEMTIGSDQLVPGDILLIPPHGCLMQCDSVLMNGTVIVNESVLTGESVPITKVALTDE
               *   *   . : . : * : : * : * * * : : *   : * : : * : : * : : :

Fbh49937      LGP--YCAETHRRHTLFCGTLILQARAYVGPVHLAVVTRTGCTAKGLVSSILHPRPIN
YH2Melegans    TNDSVFNIEKNSKNVLCFQGTQVLQTRFYRGKKVKAIVLRITAYSTLKGQLVRSIMYPKPD
               .   :   * : : : * : : : * : * * * : * * * * * : : : : :

Fbh49937      FKFYKHSMKFVAALSVLALLGTIYSIFILYRNRVPLNEIVIRALDLVTVVVPPALPAAMT
YH2Melegans    FRFTKDLFKFILFLACISGCGFIYTIIVMIMRGNTLRRIIVRSLDIITITVPPALPAAMS
               * : * *   . : : * : : : * * : : : .   . : . : * : : : : * : : : :

Fbh49937      VCTLYAQSRRLRRQIGFCIHPLRLNLGGKLQLVCFDKTGTLTEDGLDVMGVVPLKGQA---
YH2Melegans    VGIINAQLRLKKKEIFCISPTINTCGAINVVCFDKTGTLTEDGLDFHVVRPVMASVAVNQE
               *   : * * * : : * : * *   *   *   : : * : : * : : * : : * : :

Fbh49937      -----FL-PLVPEPR--LPV-GPLLRLATCHALSRLQDTPVGDPMDLKMW
YH2Melegans    IQVKLEKSNRTEFMGEMTELTSRNGLPFDGDLVKAATCHSLTRINGVLHGDPLDLILF
               * :   .   *   * *   * * : : * : : * : : : :   * : : * :

Fbh49937      ESTGWVLEEEPAADSAFGTQVLAVMRPPLWEPQLQAMEEPPVPVSVLHRFPFSSALQ RMS
YH2Melegans    QKTGWTMEEGIEGDIIEETQRFDNVQPSIIKP---TDDKSAEYSVIRQTFSSSLQ RMS
               : * : : * *   . *   * *   : : * : : * :   : :   * : : : * : : : :

Fbh49937      VVVAWPGATQP---EAYVKGSPELVAGLCNPETVPTDFAQMLQSYTAAGYRVVALASKPL
YH2Melegans    VIVFDPREDRPNMMLYSKGSPEMILSLCDPNTVPEDYLLQVNSYAQHGFRLLI AVARPL
               * : *   *   : *   * * : : : . : * : : * : * *   : : * : * : : * :

Fbh49937      PTVPSLEAAQQLTRDTVEGDLSSLGLLVMRNLLKPQTTPVIALRRTRIRAVMVTGDNLQ
YH2Melegans    DLN--FNKASKVKRDAVECDLEMLGLIVMENRVKPVTLGVINQLNRANIRTMVTGDNLL
               : :   * : : : * : * * : * : * *   * : * *   * : * : : * : : :

Fbh49937      TAVTVARGCGMVAQEH-LIIVHAT--HPERG-----QPAS--LEFLPMESPTAVN--
YH2Melegans    TGLSVARECGIIRPSKRAFLVEHVPGELEDEYGRTKIFVKQSVSSSDEVEDDASVSISM C
               * : : * * * * : * : : : : * .   * *   * . *   * :   : : : :

Fbh49937      -----GVKDPD--QAASYTVE-PDPR-----SRHLALSGPTFGIIVKHFPKLLP
```

Fig. 4A

YH2Melegans	SSTWKGSSEGDGFSPTNTEVETPNPVTADSLGH	SNHLAISGPTFAVIVHEYPVLVD
	* .: * ** * : *	* ***:*****.:***.:***:
Fbh49937	KVLVQGTVFARMAPEQKTELVCLOKLYCVGMC	QALKAADVGISLSQAEASV
YH2Melegans	QLCSVCDVFARMAPDQKQSLVEQLQIDYTVAM	CGDGCANDCAALKAHAGISLSDAEASI
	: : *****: ** . ** : ***: : * * . *****: *****: *****: *****:	
Fbh49937	VSPFTSSMASIECVPMVIREGRCSLDTSFSVFKY	MALYSLTQFISVLILYNTINTNLGDLQ
YH2Melegans	AAPFTSKVPDIRCVPTVISEGRAALVTSFGIFKY	MAGYSLTQFVTVMHLYWISNILT
	: : *****: . . . * * * * * . : * * * . : ***** *****: . : * * * . * * *	
Fbh49937	FLAIDLVTITTVAVLMSRTGPAV-LGRVRPPGALL	SVPVLSSLLQLMVLVTGVQLGGYF
YH2Melegans	FMYIDMFLITMFALLFGNT-PAFYRLAHTPPPTR	LLSIASMTSVVGQLIIIGVVQFIVFF
	* : * * . : * . : * . . * * : * . . * * * * . : : * : : * : : * : : *	
Fbh49937	LTLAQPFVPLNRTVAAPDNLP-NYENTVVFSLSS	FQYLILAAVSKGAPFRPLYTNVP
YH2Melegans	STSQQPWFTPYQPPVDDEVEDKRSMQGTALFCV	SMFQYIILALVYSGPPFRGNLWSNKP
	* *****: * : . * : . . : . : * . : * * * : * * * . * * * * * * : : * *	
Fbh49937	FLVALALLSSVLVGLVLPVPG--LLQGP-LALRN	ITDTGFKLLLLGLVTLNFGAFMLESV
YH2Melegans	IYKKKRSIEAIIID-YVPTTNSDHIRPSINGVT	SSRTESTLLSAEQQLHMTTSKNGKGG
	: : : : * . . . : : * : . : * . ** * : : . : . :	
Fbh49937	LDQCLPACLRRLRPKRASKKRFKQLERELAEQP	WPPLPAGPLR--
YH2Melegans	ENPHSSALFERLISRIGGEPTWLTNP	IPPHSLSEPEEPEKLERTY
	: . * : . ** . : . . : : . . * * *	

Fig. 4B

9/44

relative

Expression

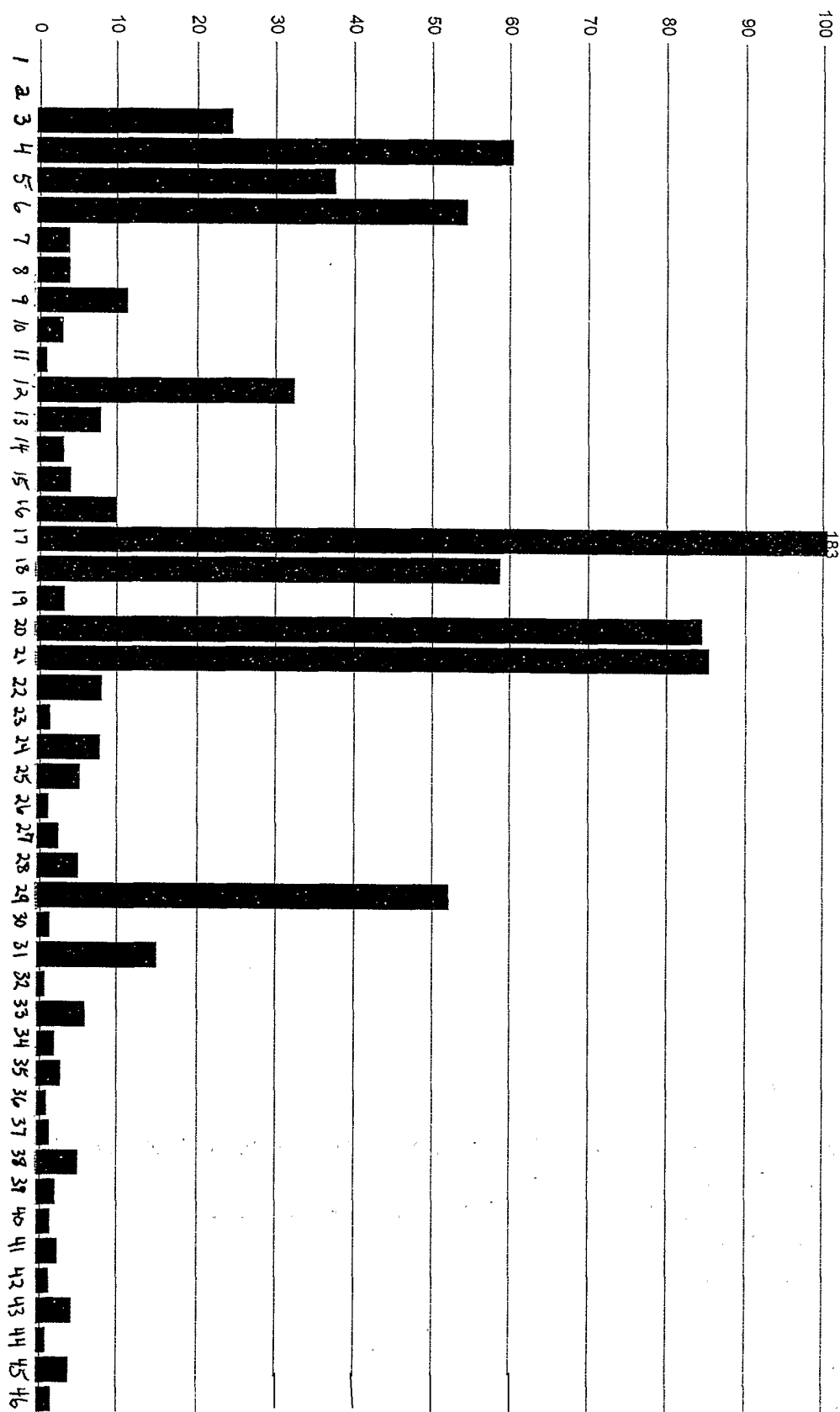


FIG. 5

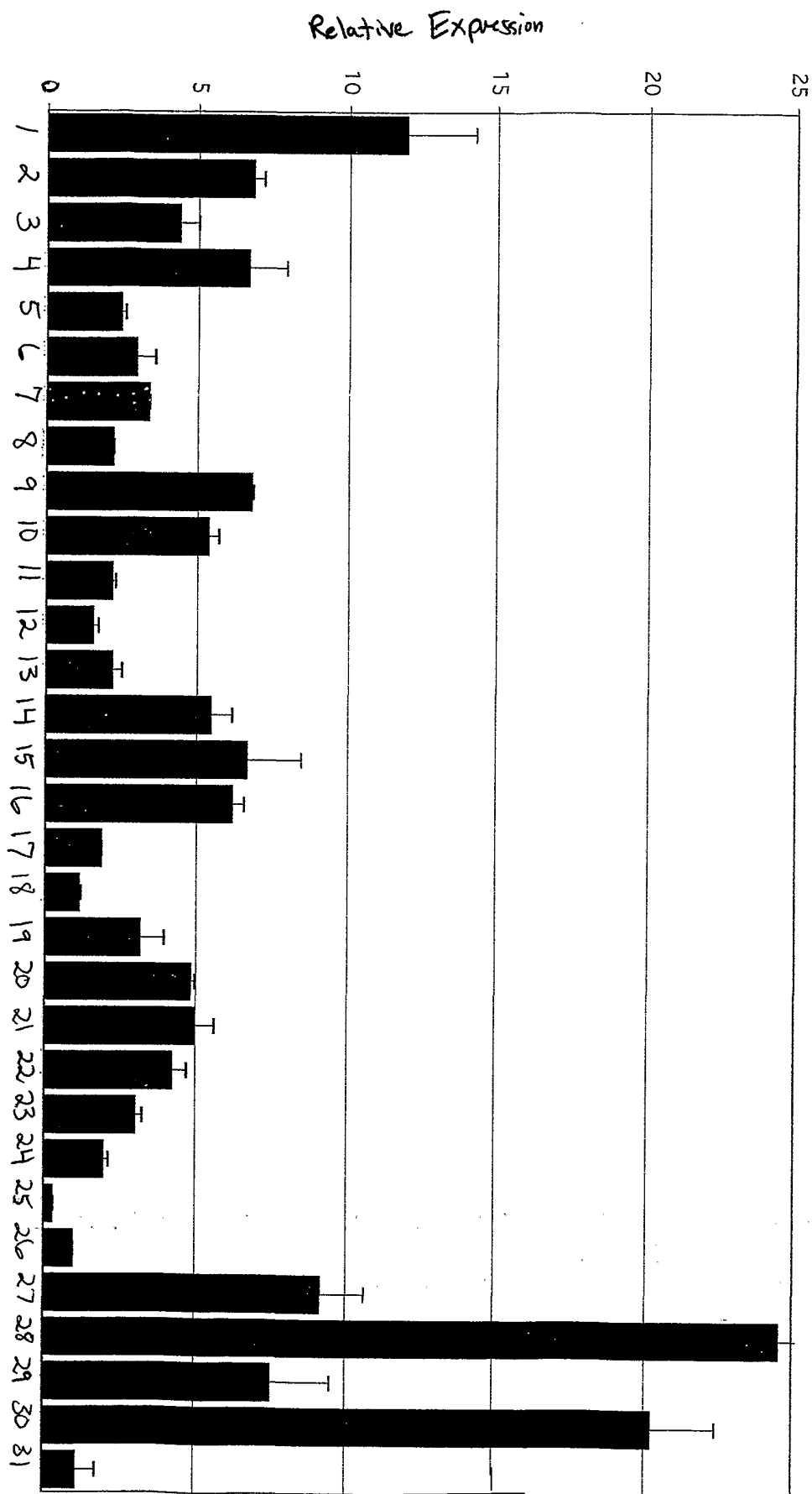


FIG. 6

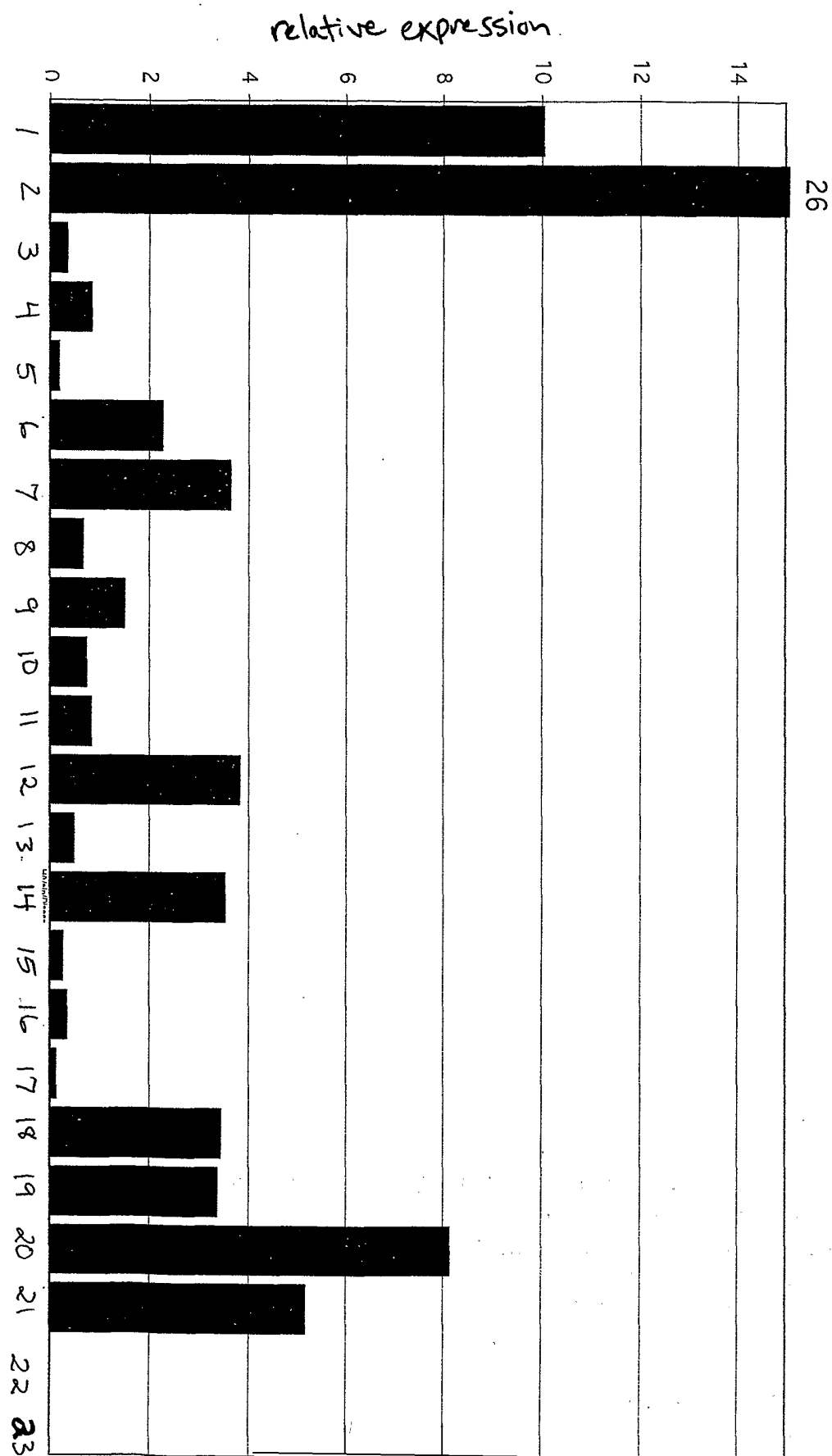


FIG. 7

49937

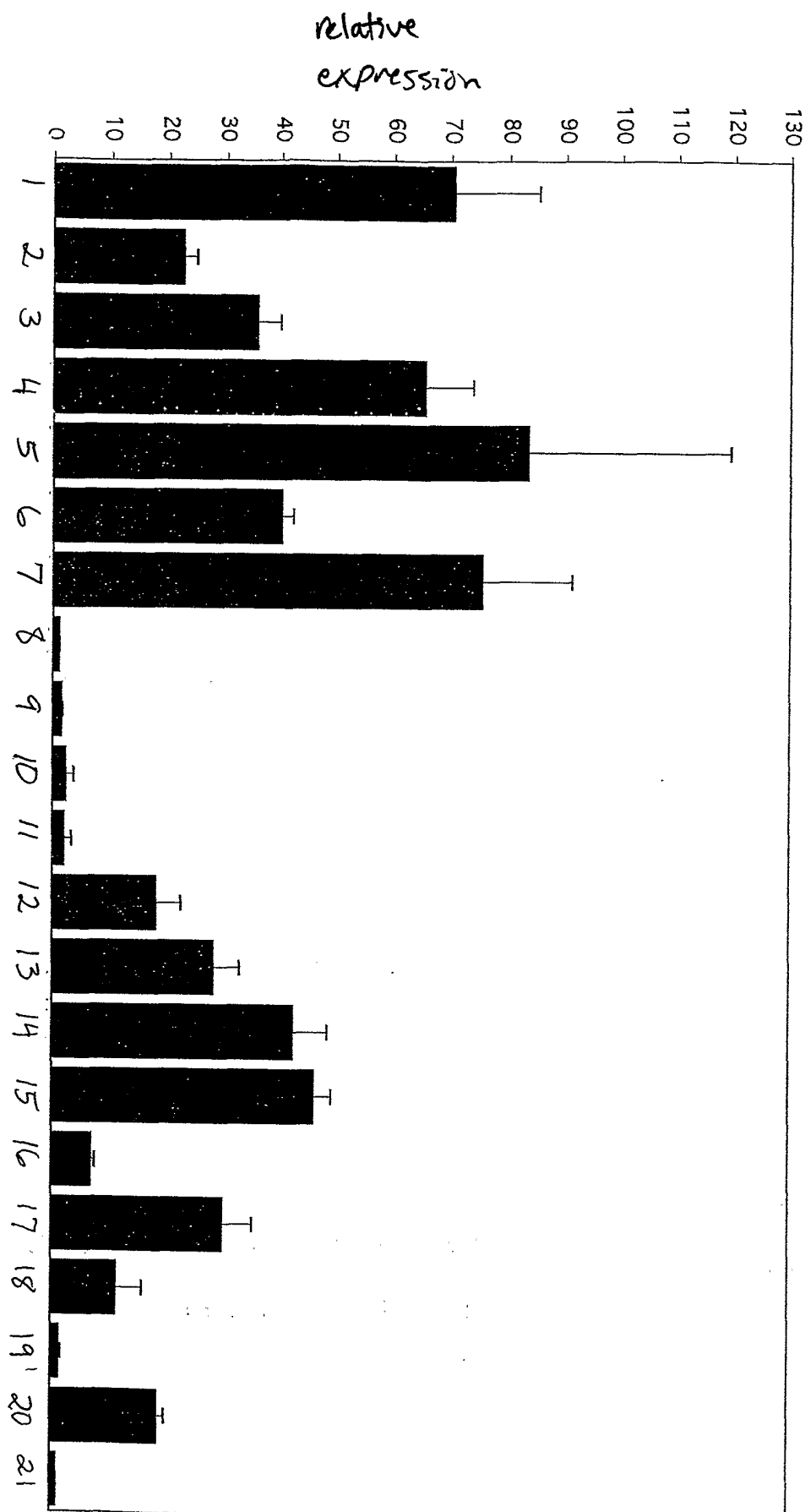


FIG. 8

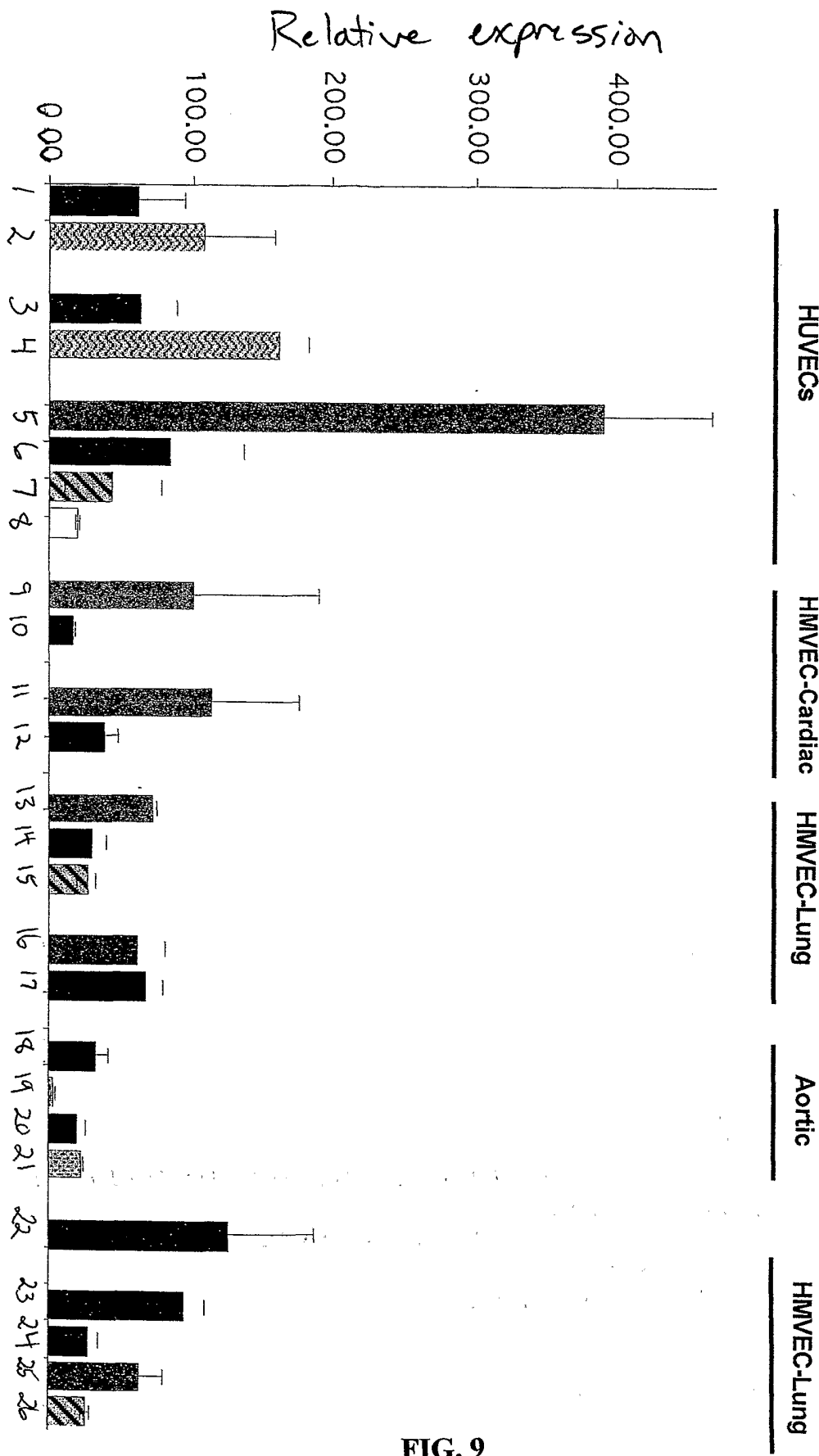
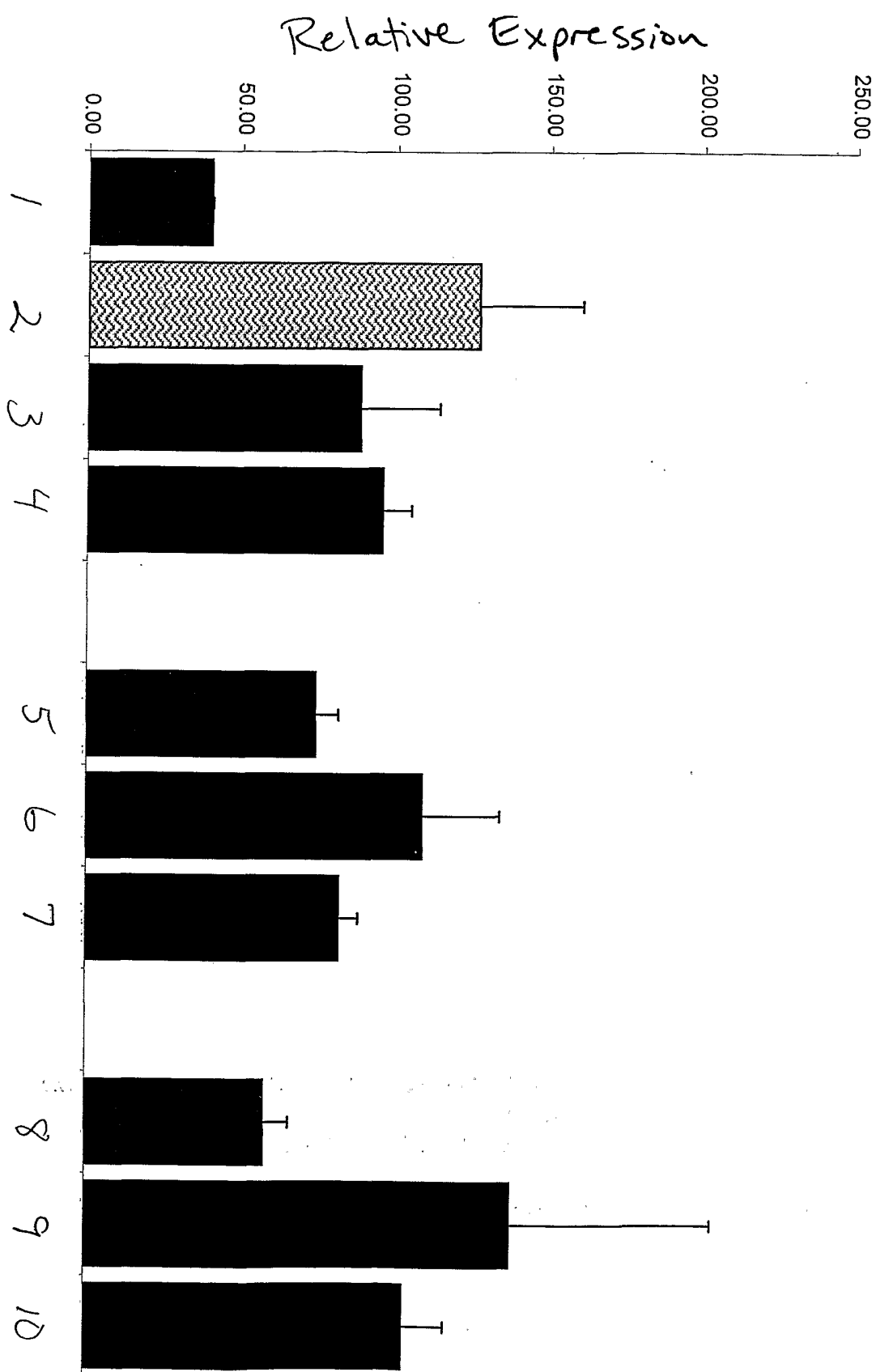


FIG. 9

**FIG. 10**

CACGCGTCCGGGCTGGGCTGAGGCGAGGCGGGCGGGCGACAGCGGCGGGCGGGTCCCCCGCGGGCCCTGGGGCTGGTC

CGGCCGCGAGGGAGGCGCGGAGGAGGCGGCGGGCGGGCGGCCAGTGAGCGGGCCCGATCTGACAGACATCCCTGAATC

TTGGTGT	TTTGGACATAGGAGTGATCTTCCATTGTTTGAAGCACTGGACCTTTAATCCACTGTAGGT	M D R	3
		ATG GAC AGG	9
E E R K T I N Q G Q E D E M E I Y G Y N			23
GAA GAA AGG AAG ACC ATC AAT CAG GGT CAA GAA GAT GAA ATG GAG ATT TAT GGT TAC AAT			69
L S R W K L A I V S L G V I C S D G F L			43
TTG AGT CGC TGG AAG CTT GCC ATA GTT TCT TTA GGA GTG ATT TGC TCT GAT GGG TTT CTC			129
L L L L Y W M P E W R V K A T C V R A A			63
CTC CTC CTC CTC TAT TGG ATG CCT GAG TGG CGG GTG AAA GCG ACC TGT GTC AGA GCT GCA			189
I K D C E V V L L R T T D E F K M W F C			83
ATT AAA GAC TGT GAA GTA GTG CTG CTG AGG ACT ACT GAT GAA TTC AAA ATG TGG TTT TGT			249
A K I R V L S L E T Y P V S S P K S M S			103
GCA AAA ATT CGC GTT CTT TCT TTG GAA ACT TAC CCA GTT TCA AGT CCA AAA TCT ATG TCT			309
N K L S N G H A V C L I E N P T E E N R			123
AAT AAG CTT TCA AAT GGC CAT GCA GTT TGT TTA ATT GAG AAT CCC ACT GAA GAA AAT AGG			369
H R I S K Y S Q T E S Q Q I R Y F T H H			143
CAC AGG ATC AGT AAA TAT TCA CAG ACT GAA TCA CAA CAG ATT CGT TAT TTC ACC CAC CAT			429
S V K Y F W N D T I H N F D F L K G L D			163
AGT GTA AAA TAT TTC TGG AAT GAT ACC ATT CAC AAT TTT GAT TTC TTA AAG GGA CTG GAT			489
E G V S C T S I Y E K H S A G L T K G M			183
GAA GGT GTT TCT TGT ACG TCA ATT TAT GAA AAG CAT AGT GCA GGA CTG ACA AAG GGG ATG			549
H A Y R K L L Y G V N E I A V K V P S V			203
CAT GCC TAC AGA AAA CTG CTT TAT GGA GTA AAT GAA ATT GCT GTA AAA GTG CCT TCT GTT			609
F K L L I K E V L N P F Y I F Q L F S V			223
TTT AAG CTT CTA ATT AAA GAG GTT CTC AAC CCA TTT TAC ATT TTC CAG CTG TTC AGT GTT			669
I L W S T D E Y Y Y Y A L A I V V M S I			243
ATA CTG TGG AGC ACT GAT GAA TAC TAT TAC TAT GCT CTA GCT ATT GTG GTT ATG TCC ATA			729
V S I V S S L Y S I R K Q Y V M L H D M			263
GTA TCA ATC GTA AGC TCA CTA TAT TCC ATT AGA AAG CAA TAT GTT ATG TTG CAT GAC ATG			789
V A T H S T V R V S V C R V N E E I E E			283
GTG GCA ACT CAT AGT ACC GTA AGA GTT TCA GTT TGT AGA GTA AAT GAA GAA ATA GAA GAA			849
I F S T D L V P G D V M V I P L N G T I			303
ATC TTT TCT ACC GAC CTT GTG CCA GGA GAT GTC ATG GTC ATT CCA TTA AAT GGG ACA ATA			909
M P C D A V L I N G T C I V N E S M L T			323
ATG CCT TGT GAT GCT GTG CTT ATT AAT GGT ACC TGC ATT GTA AAC GAA AGC ATG TTA ACA			969

Fig. 11A

G	E	S	V	P	V	T	K	T	N	L	P	N	P	S	V	D	V	K	G	343
GGA	GAA	AGT	GTT	CCA	GTG	ACA	AAG	ACT	AAT	TTG	CCA	AAT	CCT	TCA	GTG	GAT	GTG	AAA	GGA	1029
I	G	D	E	L	Y	N	P	E	T	H	K	R	H	T	L	F	C	G	T	363
ATA	GGA	GAT	GAA	TTA	TAT	AAT	CCA	GAA	ACA	CAT	AAA	CGA	CAT	ACT	TTG	TTT	TGT	GGG	ACA	1089
T	V	I	Q	T	R	F	Y	T	G	E	L	V	K	A	I	V	V	R	T	383
ACT	GTT	ATT	CAG	ACT	CGT	TTC	TAC	ACT	GGA	GAA	CTC	GTC	AAA	GCC	ATA	GTT	GTT	AGA	ACA	1149
G	F	S	T	S	K	G	Q	L	V	R	S	I	L	Y	P	K	P	T	D	403
GGA	TTT	AGT	ACT	TCC	AAA	GGA	CAG	CTT	GTT	CGT	TCC	ATA	TTG	TAT	CCC	AAA	CCA	ACT	GAT	1209
F	K	L	Y	R	D	A	Y	L	F	L	L	C	L	V	A	V	A	G	I	423
TTT	AAA	CTC	TAC	AGA	GAT	GCC	TAC	TTG	TTT	CTA	CTA	TGT	CTT	GTG	GCA	GTT	GCT	GGC	ATT	1269
G	F	I	Y	T	I	I	N	S	I	L	N	E	V	Q	V	G	V	I	I	443
GGG	TTT	ATC	TAC	ACT	ATT	ATT	AAT	AGC	ATT	TTA	AAT	GAG	GTA	CAA	GTT	GGG	GTC	ATA	ATT	1329
I	E	S	L	D	I	I	T	I	T	V	P	P	A	L	P	A	A	M	T	463
ATC	GAG	TCT	CTT	GAT	ATT	ATC	ACA	ATT	ACT	GTG	CCC	CCT	GCA	CTT	CCT	GCT	GCA	ATG	ACT	1389
A	G	I	V	Y	A	Q	R	R	L	K	K	I	G	I	F	C	I	S	P	483
GCT	GGT	ATT	GTG	TAT	GCT	CAG	AGA	AGA	CTG	AAA	AAA	ATC	GGT	ATT	TTC	TGT	ATC	AGT	CCT	1449
Q	R	I	N	I	C	G	Q	L	N	L	V	C	F	D	K	T	G	T	L	503
CAA	AGA	ATA	AAT	ATT	TGT	GGA	CAG	CTC	AAT	CTT	GTT	TGC	TTT	GAC	AAG	ACT	GGA	ACT	CTA	1509
T	E	D	G	L	D	L	W	G	I	Q	R	V	E	N	A	R	F	L	S	523
ACT	GAA	GAT	GGT	TTA	GAT	CTT	TGG	GGG	ATT	CAA	CGA	GTG	GAA	AAT	GCA	CGA	TTT	CTT	TCA	1569
P	E	E	N	V	C	N	E	M	L	V	K	S	Q	F	V	A	C	M	A	543
CCA	GAA	GAA	AAT	GTG	TGC	AAT	GAG	ATG	TTG	GTA	AAA	TCC	CAG	TTT	GTT	GCT	TGT	ATG	GCT	1629
T	C	H	S	L	T	K	I	E	G	V	L	S	G	D	P	L	D	L	K	563
ACT	TGT	CAT	TCA	CTT	ACA	AAA	ATT	GAA	GGA	GTG	CTC	TCT	GGT	GAT	CCA	CTT	GAT	CTG	AAA	1689
M	F	E	A	I	G	W	I	L	E	E	A	T	E	E	E	T	A	L	H	583
ATG	TTT	GAG	GCT	ATT	GGA	TGG	ATT	CTG	GAA	GAA	GCA	ACT	GAA	GAA	GAA	ACA	GCA	CTT	CAT	1749
N	R	I	M	P	T	V	V	R	P	P	K	Q	L	L	P	E	S	T	P	603
AAT	CGA	ATT	ATG	CCC	ACA	GTG	GTT	CGT	CCT	CCC	AAA	CAA	CTG	CTT	CCT	GAA	TCT	ACC	CCT	1809
A	G	N	Q	E	M	E	L	F	E	L	P	A	T	Y	E	I	G	I	V	623
GCA	GGA	AAC	CAA	GAA	ATG	GAG	CTG	TTT	GAA	CTT	CCA	GCT	ACT	TAT	GAG	ATA	GGA	ATT	GTT	1869
R	Q	F	P	F	S	S	A	L	Q	R	M	S	V	V	A	R	V	L	G	643
CGC	CAG	TTC	CCA	TTT	TCT	TCT	GCT	TTG	CAA	CGT	ATG	AGT	GTG	GTT	GCC	AGG	GTG	CTG	GGG	1929
D	R	K	M	D	A	Y	M	K	G	A	P	E	A	I	A	G	L	C	K	663
GAT	AGG	AAA	ATG	GAC	GCC	TAC	ATG	AAA	GGA	GCG	CCC	GAG	GCC	ATT	GCC	GGT	CTC	TGT	AAA	1989
P	E	T	V	P	V	D	F	Q	N	V	L	E	D	F	T	K	Q	G	F	683
CCT	GAA	ACA	GTT	CCT	GTC	GAT	TTT	CAA	AAC	GTT	TTG	GAA	GAC	TTC	ACT	AAA	CAG	GGC	TTC	2049
R	V	I	A	L	A	H	R	K	L	E	S	K	L	T	W	H	K	V	Q	703
CGT	GTG	ATT	GCT	CTT	GCA	CAC	AGA	AAA	TTG	GAG	TCA	AAA	CTG	ACA	TGG	CAT	AAA	GTA	CAG	2109

Fig. 11B

N	I	S	R	D	A	I	E	N	N	M	D	F	M	G	L	I	I	M	Q	723
AAT	ATT	AGC	AGA	GAT	GCA	ATT	GAG	AAC	AAC	ATG	GAT	TTT	ATG	GGA	TTA	ATT	ATA	ATG	CAG	2169
N	K	L	K	Q	E	T	P	A	V	L	E	D	L	H	K	A	N	I	R	743
AAC	AAA	TTA	AAG	CAA	GAA	ACC	CCT	GCA	GTA	CTT	GAA	GAT	TTG	CAT	AAA	GCC	AAC	ATT	CGC	2229
T	V	M	V	T	G	D	S	M	L	T	A	V	S	V	A	R	D	C	G	763
ACC	GTC	ATG	GTC	ACA	GGT	GAC	AGT	ATG	TTG	ACT	GCT	GTC	TCT	GTG	GCC	AGA	GAT	TGT	GGA	2289
M	I	L	P	Q	D	K	V	I	I	A	E	A	L	P	P	K	D	G	K	783
ATG	ATT	CTA	CCT	CAG	GAT	AAA	GTG	ATT	ATT	GCT	GAA	GCA	TTA	CCT	CCA	AAG	GAT	GGG	AAA	2349
V	A	K	I	N	W	H	Y	A	D	S	L	T	Q	C	S	H	P	S	A	803
GTT	GCC	AAA	ATA	AAT	TGG	CAT	TAT	GCA	GAC	TCC	CTC	ACG	CAG	TGC	AGT	CAT	CCA	TCA	GCA	2409
I	D	P	E	A	I	P	V	K	L	V	H	D	S	L	E	D	L	Q	M	823
ATT	GAC	CCA	GAG	GCT	ATT	CCG	GTT	AAA	TTG	GTC	CAT	GAT	AGC	TTA	GAG	GAT	CTT	CAA	ATG	2469
T	R	Y	H	F	A	M	N	G	K	S	F	S	V	I	L	E	H	F	Q	843
ACT	CGT	TAT	CAT	TTT	GCA	ATG	AAT	GGA	AAA	TCA	TTC	TCA	GTG	ATA	CTG	GAG	CAT	TTT	CAA	2529
D	L	V	P	K	L	M	L	H	G	T	V	F	A	R	M	A	P	D	Q	863
GAC	CTT	GTT	CCT	AAG	TTG	ATG	TTG	CAT	GGC	ACC	GTG	TTT	GCC	CGT	ATG	GCA	CCT	GAT	CAG	2589
K	T	Q	L	I	E	A	L	Q	N	V	D	Y	F	V	G	M	C	G	D	883
AAG	ACA	CAG	TTG	ATA	GAA	GCA	TTG	CAA	AAT	GTT	GAT	TAT	TTT	GTT	GGG	ATG	TGT	GGT	GAT	2649
G	A	N	D	C	G	A	L	K	R	A	H	G	G	I	S	L	S	E	L	903
GGC	GCA	AAT	GAT	TGT	GGT	GCT	TTG	AAG	AGG	GCA	CAC	GGA	GGC	ATT	TCC	TTA	TCG	GAG	CTC	2709
E	A	S	V	A	S	P	F	T	S	K	T	P	S	I	S	C	V	P	N	923
GAA	GCT	TCA	GTG	GCA	TCT	CCC	TTT	ACC	TCT	AAG	ACT	CCT	AGT	ATT	TCC	TGT	GTG	CCA	AAC	2769
L	I	R	E	G	R	A	A	L	I	T	S	F	C	V	F	K	F	M	A	943
CTT	ATC	AGG	GAA	GGC	CGT	GCT	GCT	TTA	ATA	ACT	TCC	TTC	TGT	GTG	TTT	AAA	TTC	ATG	GCA	2829
L	Y	S	I	I	Q	Y	F	S	V	T	L	L	Y	S	I	L	S	N	L	963
TTG	TAC	AGC	ATT	ATC	CAG	TAC	TTC	AGT	GTT	ACT	CTG	CTG	TAT	TCT	ATC	TTA	AGT	AAC	CTA	2889
G	D	F	Q	F	L	F	I	D	L	A	I	I	L	V	V	V	F	T	M	983
GGA	GAC	TTC	CAG	TTT	CTC	TTC	ATT	GAT	CTG	GCA	ATC	ATT	TTG	GTA	GTG	GTA	TTT	ACA	ATG	2949
S	L	N	P	A	W	K	E	L	V	A	Q	R	P	P	S	G	L	I	S	1003
AGT	TTA	AAT	CCT	GCC	TGG	AAA	GAA	CTT	GTG	GCA	CAA	AGA	CCA	CCT	TCG	GGT	CTT	ATA	TCT	3009
G	A	L	L	F	S	V	L	S	Q	I	I	I	C	I	G	F	Q	S	L	1023
GGG	GCC	CTT	CTC	TTC	TCC	GTT	TTG	TCT	CAG	ATT	ATC	ATC	TGC	ATT	GGA	TTT	CAA	TCT	TTG	3069
G	F	F	W	V	K	Q	Q	P	W	Y	E	V	W	H	P	K	S	D	A	1043
GGT	TTT	TTT	TGG	GTC	AAA	CAG	CAA	CCT	TGG	TAT	GAA	GTG	TGG	CAT	CCA	AAA	TCA	GAT	GCT	3129
C	N	T	T	G	S	G	F	W	N	S	S	H	V	D	N	E	T	E	L	1063
TGT	AAT	ACA	ACA	GGA	AGC	GGG	TTT	TGG	AAT	TCT	TCA	CAC	GTA	GAC	AAT	GAA	ACC	GAA	CTT	3189
D	E	H	N	I	Q	N	Y	E	N	T	T	V	F	F	I	S	S	F	Q	1083
GAT	GAA	CAT	AAT	ATA	CAA	AAT	TAT	GAA	AAT	ACC	ACA	GTG	TTT	TTT	ATT	TCC	AGT	TTT	CAG	3249

Fig. 11C

Y	L	I	V	A	I	A	F	S	K	G	K	P	F	R	Q	P	C	Y	K	1103
TAC	CTC	ATA	GTG	GCA	ATT	GCC	TTT	TCA	AAA	GGA	AAA	CCC	TTC	AGG	CAA	CCT	TGC	TAC	AAA	3309
N	Y	F	F	V	F	S	V	I	F	L	Y	I	F	I	L	F	I	M	L	1123
AAT	TAT	TTT	TTT	GTT	TTT	TCT	GTG	ATT	TTT	TTA	TAT	ATT	TTT	ATA	TTA	TTC	ATC	ATG	TTG	3369
Y	P	V	A	S	V	D	Q	V	L	Q	I	V	C	V	P	Y	Q	W	R	1143
TAT	CCA	GTT	GCC	TCT	GTT	GAC	CAG	GTT	CTT	CAG	ATA	GTG	TGT	GTA	CCA	TAT	CAG	TGG	CGT	3429
V	T	M	L	I	I	V	L	V	N	A	F	V	S	I	T	V	E	N	F	1163
GTA	ACT	ATG	CTC	ATC	ATT	GTT	CTT	GTC	AAT	GCC	TTT	GTG	TCT	ATC	ACA	GTG	GAG	AAC	TTC	3489
F	L	D	M	V	L	W	K	V	V	F	N	R	D	K	Q	G	E	Y	R	1183
TTC	CTT	GAC	ATG	GTC	CTT	TGG	AAA	GTT	GTG	TTC	AAC	CGA	GAC	AAA	CAA	GGA	GAG	TAT	CGG	3549
F	S	T	T	Q	P	P	Q	E	S	V	D	R	W	G	K	C	C	L	P	1203
TTC	AGC	ACC	ACA	CAG	CCA	CCG	CAG	GAG	TCA	GTG	GAT	CGG	TGG	GGA	AAA	TGC	TGC	TTA	CCC	3609
W	A	L	G	C	R	K	K	T	P	K	A	K	Y	M	Y	L	A	Q	E	1223
TGG	GCC	CTG	GGC	TGT	AGA	AAG	AAG	ACA	CCA	AAG	GCA	AAG	TAC	ATG	TAT	CTG	GCG	CAG	GAG	3669
L	L	V	D	P	E	W	P	P	K	P	Q	T	T	T	E	A	K	A	L	1243
CTC	TTG	GTT	GAT	CCA	GAA	TGG	CCA	CCA	AAA	CCT	CAG	ACA	ACC	ACA	GAA	GCT	AAA	GCT	TTA	3729
V	K	E	N	G	S	C	Q	I	I	T	I	T	*							1257
GTT	AAG	GAG	AAT	GGA	TCA	TGT	CAA	ATC	ATC	ACC	ATA	ACA	TAG							3771

CAGTGAATCAGTCTCAGTGGTATTGCTGATAGCAGTATTCAGGAATATGTGATTTTAGGAGTTTCTGATCCTGTGTGTC
AGAATGGCACTAGTTTCAGTTTATGTCCCTTCTGATATAGTAGCTTATTTGACAGCTTTGCTCTTCCTTAAAAATAAAAC
AGAAAAATATATCGTCCTAACAGTTAAATTAACAATCAATCCATAAAGTCCTATATCTTCATTCAGCAACCCAAATATT
ACATACATTTCCAGAATTTTCTTGATTGTTACTTTTCAGTGATATTCCTTTATATTGGGTACAGGAGAAGTTTGGTGTGTTG
GTAGGTTTTTCAACATTAGTTTTTGGAGACTAGTTTACCTCTTCACATTTATGCTCACAACCTCTTGTTAGAAAAGTCT
GTGTTTATATACAGGCTGTAAGTTTGTGATTGATAAAAAGAAGATGAGTGTAAATTAGCCTCCAGTGAAAATATACTGA
AAGCCTGTTTTCATTTGATTCCAATGTTTCTTCCAAAGAATTCTGTATAAACATATGCCAATTCCTTATGATGGTCTAG
AGTTAGGAATGAGTGTGTTATGGTGTGCTTATAGAACAACCTCAGGTAATCTCCATTTCTGGTTTATATTTTCTGTACA
AACTGCCTGGGTTTATTTTCTAATCAGCAAGGTGCTTCACTGCCTTCTTGAGACGCCTCTCAAAGCTCTTAAATGGC
TCCTGTGCTATGTGTGGTGTGGCAGTCTAATTTGCTTCTGTAAATGTTGTAGAACCTTTTTCAGTGGAAATAAGAT
TCATTTCTTTTCGGCAGTAGATGTAGATTCATCTTTTAACGTTTCTTCAAATTTGTTTCTGTGAGGCTTTGTGTTATTTT
AAATGGTTTTTTAAAAATTTCTTCTATGTTTCAATTACCTAAAGACATAGGATAATAGTTTTTTTTAAGTTAGAATTT
TACCTCATAAAATTTTTGAGGTTTGATGTATGTCTGTCTTATCAATAATGAGGCTTAAAAAATACTGGATTTGAAT
GGCTGCCGTTTTTTCAAAGCAATATGAATTTGATGAGTTTGTGTTTATGCCATTAGGTGGCGCCAGAGGTCAGAACATGT
CTATTTTGAATTGGATCGTTACAAATGAGCATATTTGATGCGGAAATTTCTGGGAGAAAAAAATGAGGAAATAAAGT

Fig. 11.D

TAAAAAATTGACATTCATTGAGCCAAAAGAGATGTGGAGAAACATTTTTTCACCTTTCTGTTTGGCCTGATTAACATTTA
AATTCCTTGCCAAAATTAAGGTAACCTTTTAAAAAACACCTTTTATAGGTGGATCCAGCAGTCTGGCAACGCCCACAGTTA
CCACAACACAGAAAACCTGATCGTGCTATAAAATGGACGCTAAACTATGAAAACAGTGTGACATTGTTCTCTGTTCTTCC
AGAGCCAGTAACATGCTTGCTCGTGCTTTCTACTTCTAGCTGATCATTCTTTTCCCAACATATATTTACAAATTACCAA
ATTTTACCTAGAATTTTAGGACCAAATGGTTCTCACTCTTTATGCTGCAAAGACCTGGATGATGTTTGGTAACTATAGA
AAAATAGAAATTACACTCAGGATCACTGTTACTGCTATTGCCACTGATGATTCTTGCAAAAATATAATCGAAGTTTTC
ATCAAAATGTATAATATGCTATTAATACACATTAGATGATAACAGTTGTTCCATGAATGATTCTATGAAGCTATGCATCT
TAGACCTCTTGAGCTGTGAATTAGCACTATTTTCTATAGTTACTTATTCTCTGGATCATTTTATAATTTCCATATTAAT
TTCAAATATGCTCGTGTTATTCTTCAGTGATTTCCACAATTGTGCATTTTATTCTTTGGTTTAAGTACTGAAGCATATA
ATGAAAGTAATTGCTAAGTAGCAGCTTAAAAATTCAATTATCCGATTGTATTTAACATCTTTAAGAGCATGATCATAAA
GAGCTATTTTGTACACCCCCCCCCCACTTTTTTAACATTTAGAGTTAGTAAGGGTTTTATATCTCTTCTGTCCATATTG
TTTTCAAAGGAATGAGGTGTTTAGGTGGCTGGAAAAGCATTTGTAGGAAGTTAGATTTGAATATAGACAAGGTGGGTTA
TTCACGTTGAGAATGTTATTTGAAGAATGCCTGTGAAGCCAGGTGTGGGTTCTACTCAGTGCCATAGATAGACTGAGTC
TTCTCTCGTAGGTTACCATTACATAGTAATTTTGATTCTGAATTACACATTAAATTATTTGAGTTTATACAGACCTAAA
TTTTAAATCTGTACATATATTTATTTTGATGTATTAAGATGAATATTGCTGATTTAAATTTTATTTATGCACATACTTA
AAGGACAGAAATGTCTGGGAAAGTAATTGTTAAATAATGATATGTAACCTTTTTAACTTTTTAAATAAAATAACAAGATTT
TTAATGTGTGTCTCCCTCAGGGTTGTTTAAAGTTTTTTTTCTCCCTCAAGTATAAATAGTGGTAACTATATGTTTTGTA
TCTTCTAGCACCAACTGCTGTAAAGCAATGCTGCAAATAATGCTTGAATACAAGTGGCTAAGCCAACAACAGAATAAAT
ACTTTTATAGTAGTTTATAATCCTGAAATTCGAAAGCTTTCCCAATTGCACCTGCATCTAAACAAAACCTGTTGCAGTT
TTTACTCTATTTATTTTGTTCCTCATGTTTATGAAAGTCCTGCACAGTTTCAAAGGCATGGTAAATAATATATCAATGT
TTATGTAGTCTGTTACAGAAACAGCTATAGATAACATTATCCAGTGAAGAGCAAAATTCAAGCTTTAGAAAATATTCAT
GCATGCAATTTTGACATATCTAAAAATAGGTTTTTGTATATTTATGGTGGGAGGTGGTTGGGAACTTTAAACAAAATGG
GGTGTAAATTTTGTACAGTCTGTGGGCATTTACACATTTTAAATGTATTAATAATTTGGTAATTATGTGTACATTAAAT
TAATAAAAGTTACTTCTAGTTATGATTTGTGAATTCCTTAAGACCTTGGATTTTTTTAAGTAACTTTATATCAGAAATG
ATACTGCATCTTTATATTTTAAATTTGTATTGCTGCTCAAGAATGGTACCCTCTTGTCAAAAAGGCATACATTCATAA
TTGTACATTCAGCATGTAAATAATCTTATGAAACCTTTTTTGATTGAAGCTATTCAAATAAAAAATTTAATGAATGA
AAAAAAAAAAAAAAAA

Fig. 11E

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM
 hmmpfam - search a single seq against HMM database
 HMMER 2.1.1 (Dec 1998)
 Copyright (C) 1992-1998 Washington University School of Medicine
 HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam5.4/Pfam
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.27934.seq

Query: Fbh49931FL

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
E1-E2_ATPase	E1-E2 ATPase	53.4	3.5e-15	1
Na_K_ATPase_C	Na+/K+ ATPase C-terminus	-106.2	4.2	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
E1-E2_ATPase	1/1	278	365	..	40 117 ..	53.4	3.5e-15
Na_K_ATPase_C	1/1	986	1219	..	1 190 []	-106.2	4.2

Alignments of top-scoring domains:

E1-E2_ATPase: domain 1 of 1, from 278 to 365: score 53.4, E = 3.5e-15

```

      *->dGkeeeipaeeLvpGDIVev.kpGdrVPADGrvvegestLkVDESaL
      + ++eei + +LvpGD+ +++ G ++P D++++ g++ V+ES+L
Fbh49931FL 278 NEEIEEIFSTDLVPGDVMVipLNGTIMPCDAVLINGTC--IVNESML 322

      TGESlPVeKhsdvvtp.....dnp.lerngmVfaGtlv<-*
      TGES+PV+K +++p+++ + ++ +++ np+ +++ ++f GT+v
Fbh49931FL 323 TGESVPVTK---TNLPnpsvdkvgigdelYNPeTHKRHTLFCGTTV 365
  
```

Na_K_ATPase_C: domain 1 of 1, from 986 to 1219: score -106.2, E = 4.2

```

      *->RkPRnpkkDkLVNer....LisyaYlQ..IGmiqAiaGF..ftYFvi
      np LV +r++++Lis a l++ + i+ ++GF+++ +F++
Fbh49931FL 986 ----NPAWKELVAQRppsgLISGALLFsvLSQIIICIGFqsLGFFWV 1028

      mae...nGfLP.sdLlglrkq..WdkkavnDLeDSyGqeWtyeqRkaLey
      ++ + + P+sd +++ + W + +v++ + e+++ ++Y
Fbh49931FL 1029 KQQpwyEVWHPkSDACNTTGSgfwNSSHVDNET-----ELDE--HNIQNY 1071

      tcyTAFfiSIVvvQwaDLiicKTRRnSlFqQGlmrNkvLnfglvfetiLa
      T+FFiS + +K F Q +N + +f+++f i +
Fbh49931FL 1072 ENTTVFFISSFOYLIVAIAPSKGKP---FRQPCYKNYFFVFSVIFLYIEI 1118

      aFLsYtPG..vdtiLrmqPlrfq.....
      +F P +vd +L ++ q + + + + + +
Fbh49931FL 1119 LFIMLYPVasVDQVLQIVCPYPQwrvmtliivlvnafvsvitvenffldmv 1168

      .....wWfvalPfsilIFvYDEiRKli
      + ++++++ + ++++++ ++w+ + lP ++ +RK
Fbh49931FL 1169 lwkvvfnrdkqgeyrfsttqppqesvdrWKGKCLPWALG-----CRK-- 1210

      iRryPggWlekEtY<-*
      + P+ k +Y
Fbh49931FL 1211 --KTPKA---KMY 1219
  
```

FIG. 12

22/44

CLUSTAL W (1.74) multiple sequence alignment

```

                                TM1
Fbh49931FL      MDREERKTINQGQEDEMEYGYNLSRWKLAIVSLGVICSDGFLLLLLLYWMPEWRVKATCV
YH2Melegans     -----MTLESGDHT-LTLFAYRTGPFRTILFYALTVLTLGIFRLILHWKQKWDVKMRMV
                  *::*: : :*: : : : : : : *::*:*:*: * * * *

Fbh49931FL      RAAIKDCEVVLRLRTTDEFKMWFCAKIRVLSLETYPVSSPKSMSNKLSNGHAVCLIENPTE
YH2Melegans     PCTFEAAEYIYI--IDN--H-----NVS-ELQPVLRKSNATIPTENG---EMRKVP--
                  .:::.* : : * : : * * * * . : . * * : : *

Fbh49931FL      ENRHRISKYSQTESQQIR-YFTHHSVKYFW-NDTIHNFDFLKGLDEGVSCTSIYEKHSA-
YH2Melegans     ELRWFV--YRKLEYVWIDDLNSDESVDEISDNDNCWKTSFE--IANRIPCRSLLAVSESN
                  * * : * : * * :...*: : * * : * : : : * * : :

Fbh49931FL      -GLTKGMHAYRKLLYGVNEIAVKVPSVFKLLIKEVLNPFYIFQLLWSTDEYYYVAL
YH2Melegans     FGLTLSEISRLEFYGRNEIVVQLRPILYLLVMEVITPFYVFQIFSVTVWYNDEYAYYAS
                  *** : * : ** * * : : : * : * : * : * : * : * : * : *

                                TM2
Fbh49931FL      AIVVMSIVSIVSSLYSIIRKQYVMLHDMVATHSTVRVSVCRVNEEIEEIFSTD-LVPGDM
YH2Melegans     LIVILSLGSIVMDVYQIRTQEIRLSMV--HSTESVEVIR--EGTEMTIGSDQLVPGDIL
                  *::*: * : : * * : : * : * * * * * * * : : * * : : *

                                TM3
Fbh49931FL      VIPLNGTIMPCDAVLINGTCIVNESMLTGESVPVTKTNLPNPSVDVKGIGDELYNPETHK
YH2Melegans     LIPPHGCLMQCDSVLMNGTVIVNESVLTGESVPITKVALTDETN-----DSVFNIEKNS
                  : * : * : * * : * : * * : * * : * : * : * : * : * : *

Fbh49931FL      RHTLFCGTTVIQTRFYTGELVKAIVVRTGFSTSKGQLVRSILYPKPTDFKLYRDAYLFLL
YH2Melegans     KNVLFCGTQVLQTRFYRGKKVKAIVLRTAYSTLKGQLVRSIMYPKPVDFRFTKDLFKFIL
                  : : * * * * * : * * * * * : * * * * * : * * * * * : * : * : *

                                TM4
Fbh49931FL      CLVAVAGIGFIYTIINSINEVQVGVIIESLDIITITVPPALPAAMTAGIVYAQRRLKK
YH2Melegans     FLACISGCGFIYTIIVMIMRGNTLRRIIVRSLDIITITVPPALPAAMSVGIINAQLRLKK
                  * : : : * * * * * * : : : * * * * * * * : * : * * * *

                                TM5
Fbh49931FL      IGIFCISPQRINICGQLNLVCFDKTGTLTEDGLDL-----WGIQRVEN---AR-
YH2Melegans     KEIFCISPSTINTCGAINVVCFDKGTLTEDGLDFHVVRPMSAVNQEIQVKLEKSNRT
                  * * * * * : * * * * * : * * * * * : * * * * *

                                TM6
Fbh49931FL      -FLSP--EENVCNEMLVKSQFVACMATCHSLTKIEGVLSGDPLDLKMFEAIGWILEEATE
YH2Melegans     EFMGEMTELTSRNGLPFDGDLVKAIATCHSLTRINGVLHGDPLDLILFQKTGWTMEEGIE
                  * : . * * : : : : * : * * * * : * * * * * : * : * * : *

Fbh49931FL      EETALHNRIMPTVVRPPKQLPESTPAGNQEMELFELPATYEIGIVRQPFFSSALQRMSV
YH2Melegans     GDIEEETQRFDN-VQP--SIIK--PTDDKSAEY-----S-VIRQFTFSSSLQRMSV
                  : : : : : * : * : : : * : : : : : * : : : : * : * * * *

Fbh49931FL      VARVLG-DRK--MDAYMKGAPEAIAGLCKPETVPVDFQNVLEDFTKQCFRVIALAHRKLE
YH2Melegans     IVFDPREDRPDNMMLYSKGSPEMILSLCDPNTVPEDYLLQVNSYAQHGFRLIAVARRPLD
                  : . * * * * * : * * * * * : : : : : * : * : * : *

Fbh49931FL      SKLTWHKVQNISRDAIENNMDFMGLIIMQNKLKQETPAVLEDLHKANIRTVMTGDSMLT
YH2Melegans     --LNFNKASKVKRDAVECDLEMLGLIVMENRVKPVTLGVINQLNRANIRTVMTGDNLLT
                  * : : * : : * : * : : * : * : * : * : * : : : * : * : * : *

Fbh49931FL      AVSVARDCGMILPODKVIIAEALP-PKD--GK-----VAKINWHYAD----SLTQCS
YH2Melegans     GLSVARECGIIRPSKRAFLVEHVPGELDEYGRTKIFVKQSVSSSDEVIDEDDASVSISMCS
                  : * * * * * : * : : : * : * : : : * : : : : * : * : * : *

```

Fig. 14A

Fbh49931FL HP----SAID---P---EA-IPVKLVHDSLEDLQMTRYHFAMNGKSFVILEHFQDLVPK
YH2Melegans STWKGSSSEGDGFSPTNTEVETENPVTADSLGHLIASSYHLAISGPTFAVIVHEYPELVDO
* * * * * : * * * * : * * * * : * * * * : * * * * : * * * *

Fbh49931FL LMLHGTVFARMAPDQKTQLIEALQNVDYFVGMCGDGANDCGALKRAHGGISLSELEASVA
YH2Melegans LCSVCDVFARMAPDQKQSLVEQLQQIDYTVAMCGDGANDCAALKAHAGISLSDAEASIA
* * * * * : * * * * : * * * * : * * * * : * * * * : * * * *

Fbh49931FL SPFTSKTPSISCVPNLIREGRAALITSEFCVFKFMALYSTIIQYFSVTLLYSILSNLGDFQF
YH2Melegans APFTSKVPDIRCVPTVISEGRAALVTSFGIFKYMAGYSLTQFVTVMHLYWISNILDGQF
: * * * * : * * * * : * * * * : * * * * : * * * * : * * * * : * * * *

Fbh49931FL LFIDLAILLVVETMSINPAWKELVAQRPPSGLISGALLFSVLSQLIICIGFQSLGFFFWV
YH2Melegans MYIDMFLITMFAILLFGNTPAFYRLAHTPPPTRLSTIASMTSVVGQLIIIGVVQFIVFFST
: * * * : * * : * * : * * : * * : * * : * * : * * : * * : * * : * *

Fbh49931FL KQQPWYEVVHWPKSDACNTTGSFGWNSSHVDNETELDEHNIQNYENTTVFFISSFQYLIVA
YH2Melegans SQQPWFPTYQPP-----VDDEVE-DKRSMQG---TALFCVSMFQYIILA
: * * * : * * : * * : * * : * * : * * : * * : * * : * * : * * : * *

Fbh49931FL IAFSKGKPFROPCKYKVFVFSVIFLYIFILFIMLYPVASVDQVLQIVCVPYQWRVTMLI
YH2Melegans LVYSKGPPFRGNLWSNKPPIYKKKRSIEAIIIDYV---PTTNSDHIRR---P---SINGV
: * * * * : * * : * * : * * : * * : * * : * * : * * : * * : * * : *

Fbh49931FL IVLVNAFVSITVENFFLDMLVKVFNDRDKQGEYRFSTTOPPQESVDRWGKCLP-WALG
YH2Melegans TSSRTESTLLSAEGQQLHMTTSK---N-GKGENPHSSALFER-LISRIGGE--PTWLT-
: * * : * * : * * : * * : * * : * * : * * : * * : * * : *

Fbh49931FL CRKKTTPKAKYMYLAQELLVDPEWPPKPKQTTEAKALVKENGSCQIITIT
YH2Melegans ----NP-----IPPHSLSEPEEPEKLERTY-----
: * : * : * : * : * : * : * : * : * : * : *

Fig. 14B

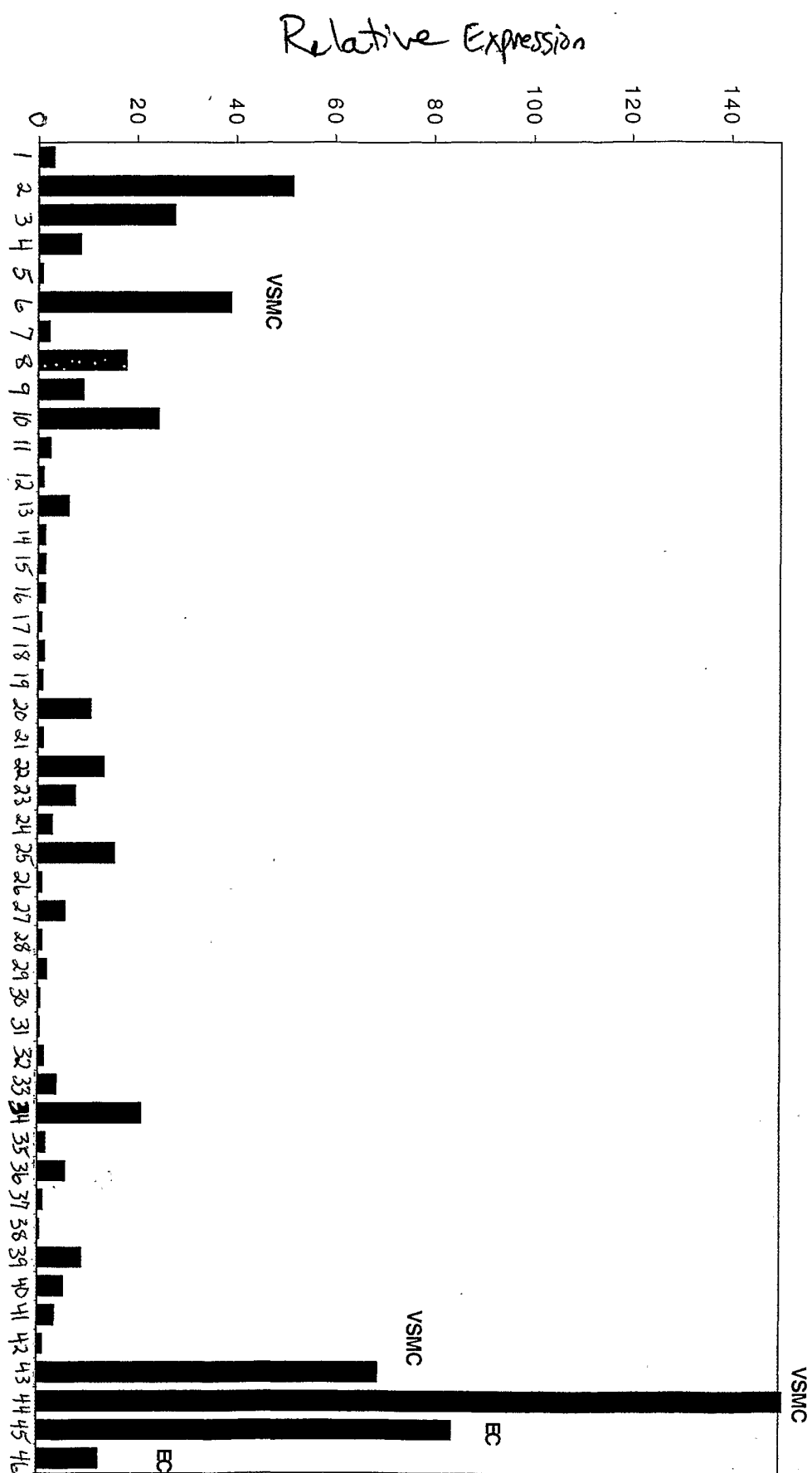


FIG. 15

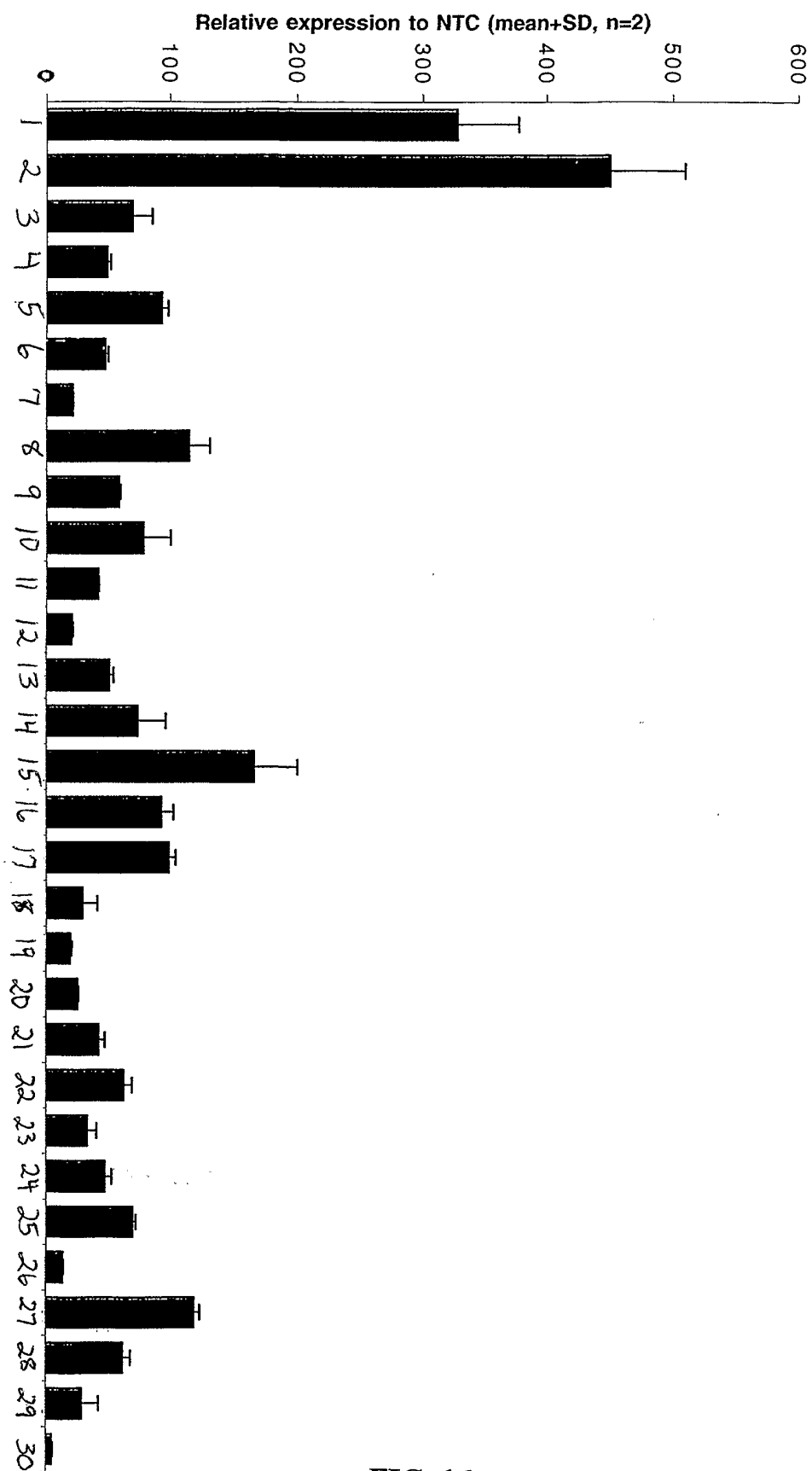


FIG. 16

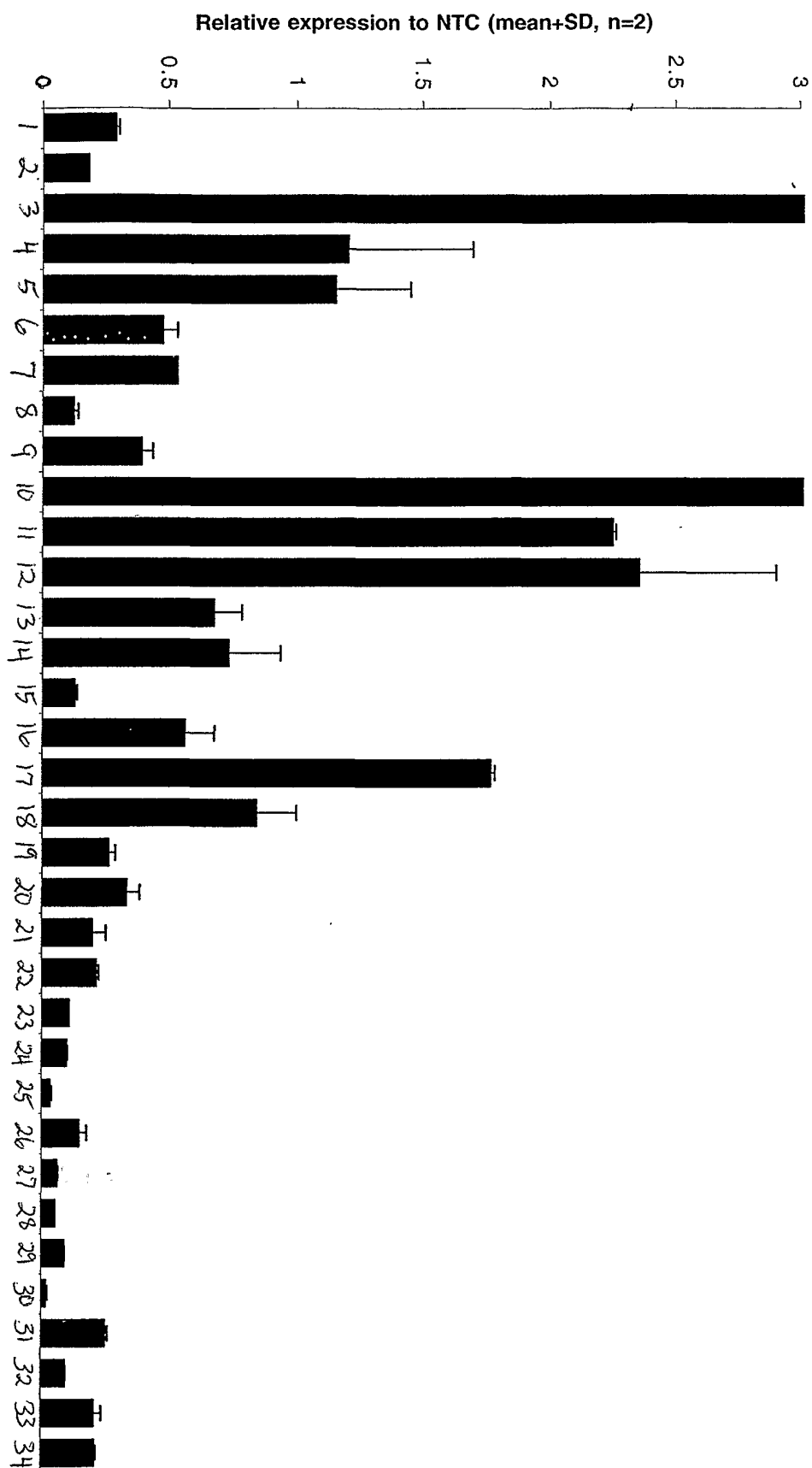


FIG. 17

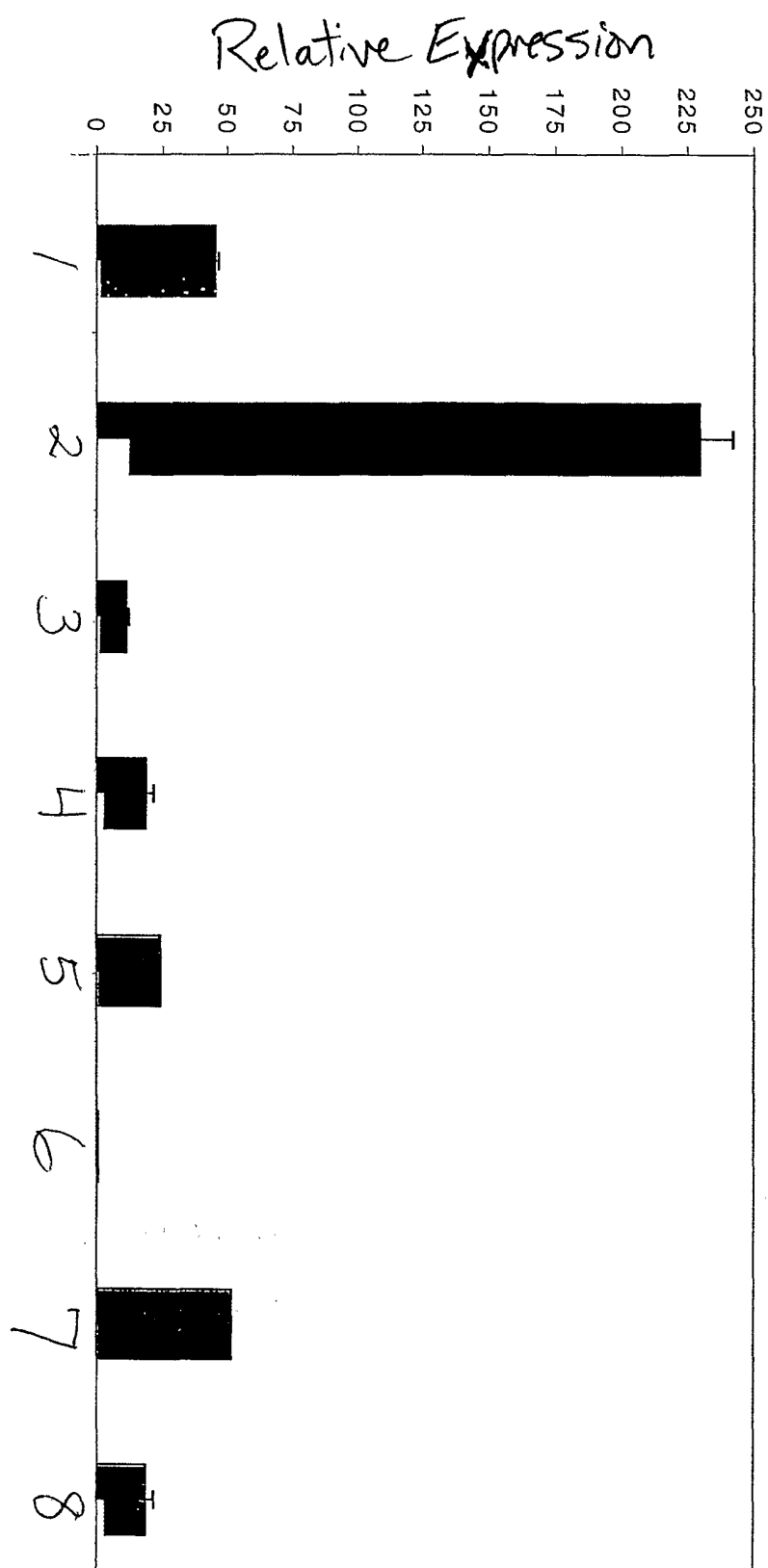
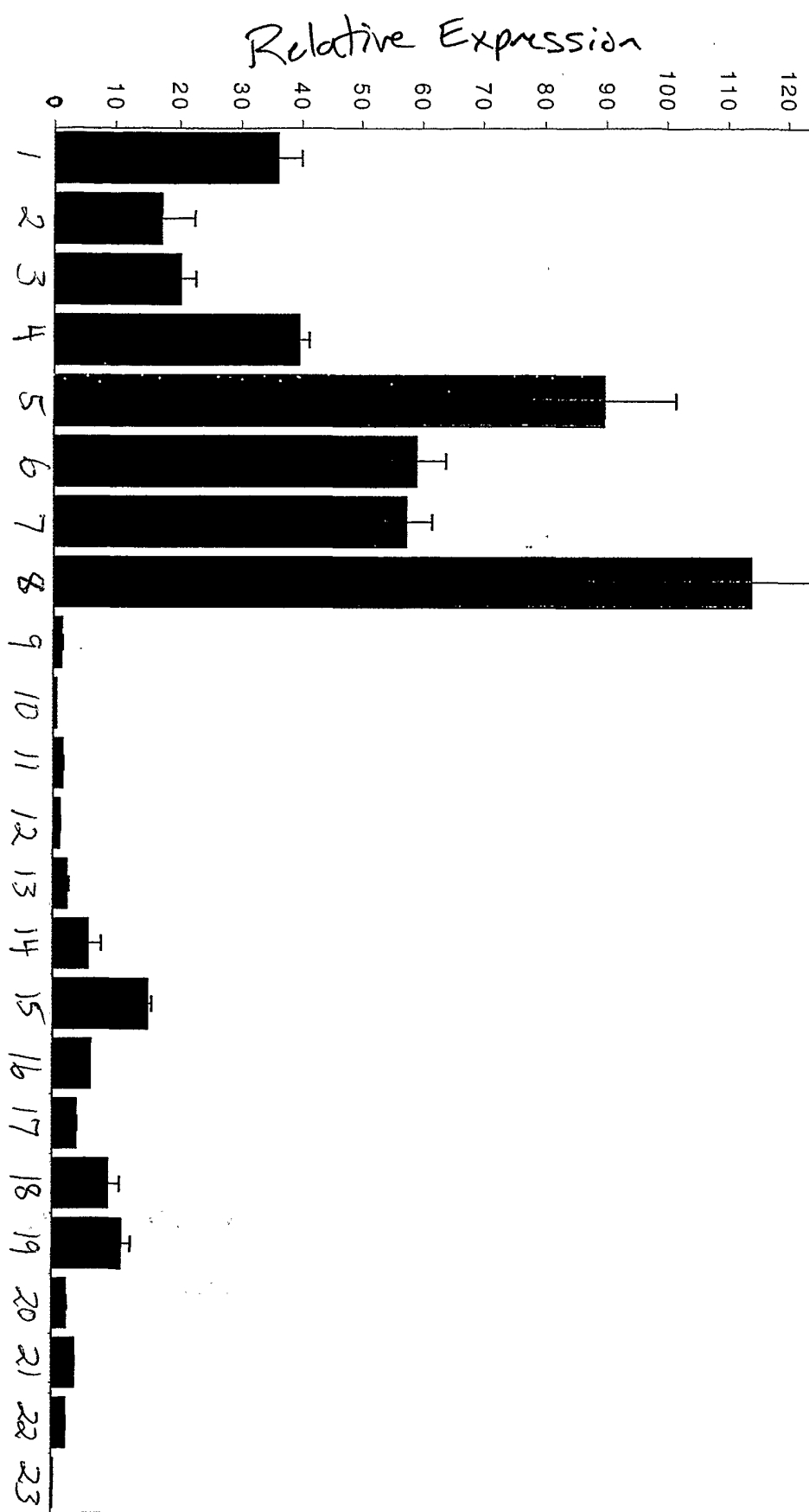


FIG. 18



49931

FIG. 19

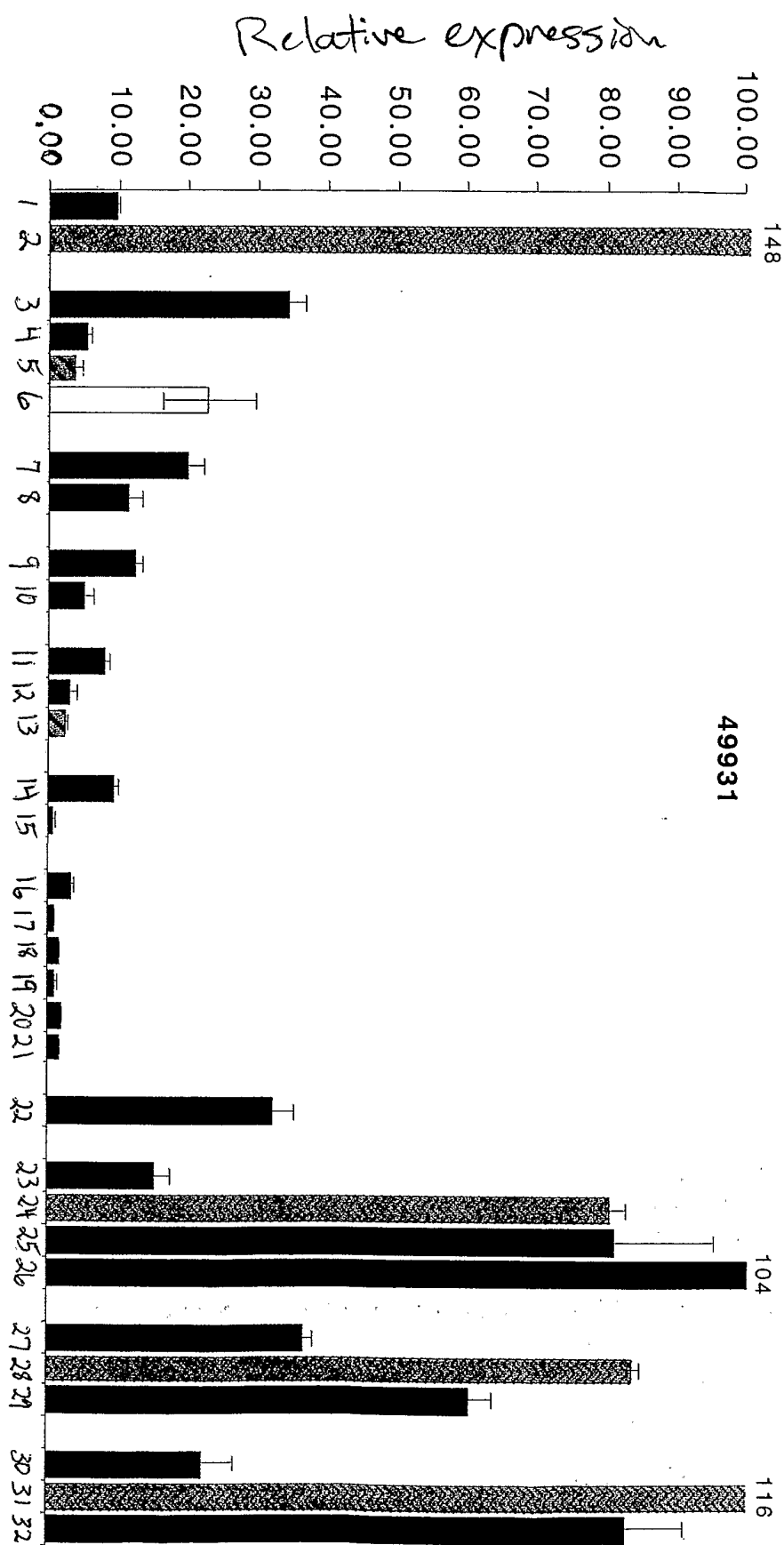


FIG. 20

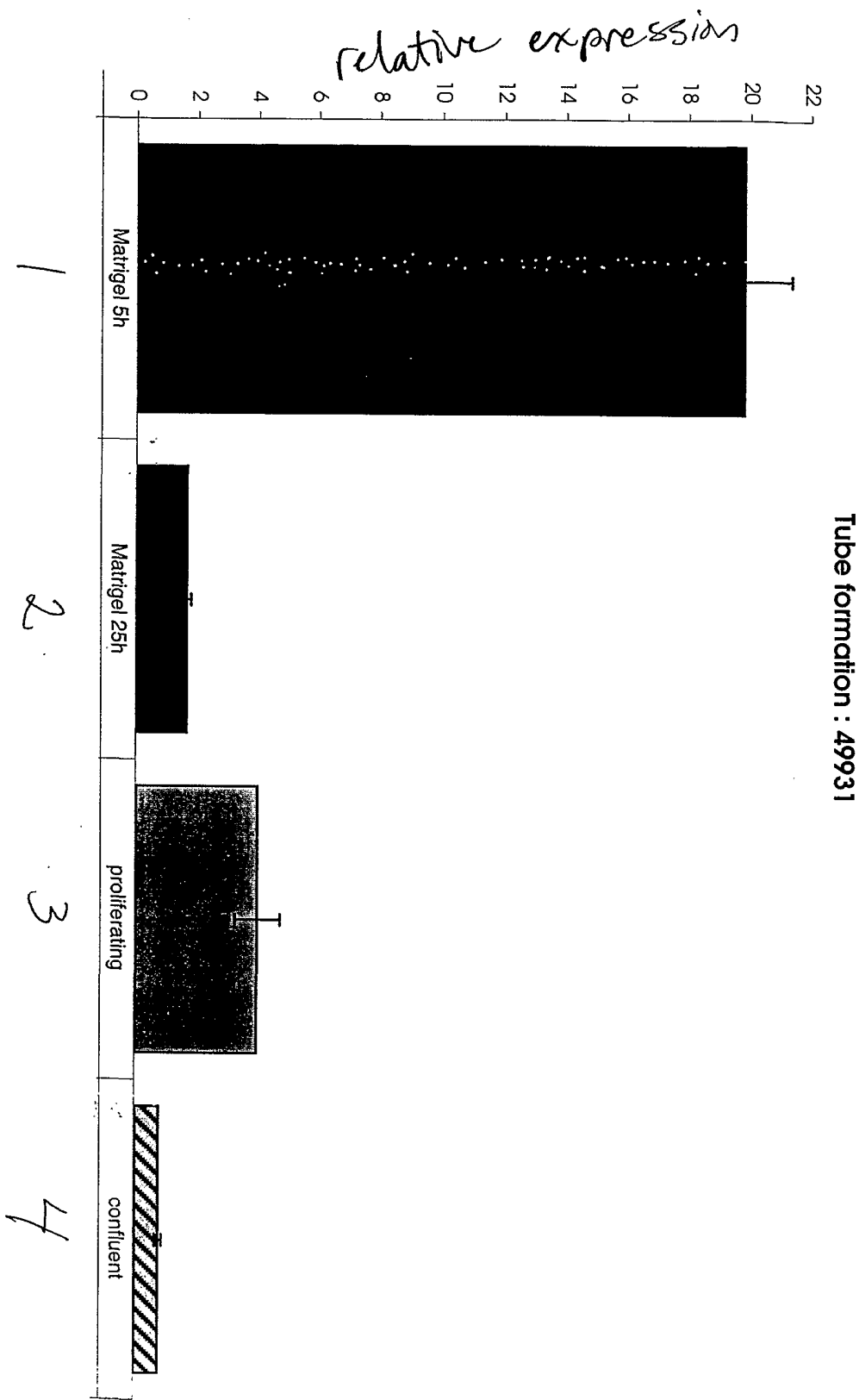


FIG. 21

TTACCGGAAGTAAAACTTCGGAAGTGAGGCGTTCCCTCTGCCCCGGAAGTGAGCGCCGCGCTAGGAAAAG	M	A	A	3
ATG GCG GCA				9
A A A V G N A V P C G A R P C G V R P D				23
GCG GCG GCG GTG GGC AAC GCG GTG CCC TGC GGG GCC CGG CCT TGC GGG GTC CGG CCT GAC				69
G Q P K P G P Q P R A L L A A G P A L I				43
GGG CAG CCC AAG CCC GGG CCG CAG CCG CGC GCG CTC CTT GCC GCC GGG CCG GCG CTC ATA				129
A N G D E L V A A V W P Y R R L A L L R				63
GCG AAC GGT GAC GAG CTG GTG GCT GCC GTG TGG CCG TAC CGG CGG TTG GCG CTG TTG CGG				189
R L T V L P F A G L L Y P A W L G A A A				83
CGC CTC ACG GTG CTG CCA TTC GCC GGG CTG CTT TAC CCG GCC TGG TTG GGT GCC GCA GCC				249
A G C W G W G S S W V Q I P E A A L L V				103
GCT GGC TGC TGG GGC TGG GGC AGC AGT TGG GTG CAG ATC CCC GAA GCT GCG CTG CTC GTG				309
L A T I C L A H A L T V L S G H W S V H				123
CTT GCC ACC ATC TGC CTC GCG CAC GCG CTC ACT GTC CTC TCG GGG CAT TGG TCT GTG CAC				369
A H C A L T C T P E Y D P S K A T F V K				143
GCG CAT TGC GCG CTC ACC TGC ACC CCG GAG TAC GAC CCC AGC AAA GCG ACC TTT GTG AAG				429
V A P T P N N G S T E L V A L H R N E G				163
GTG GCG CCA ACC CCC AAC AAT GGC TCC ACG GAG CTC GTG GCC CTG CAC CGC AAT GAG GGC				489
E D G L E V L S F E F Q K I K Y S Y D A				183
GAA GAC GGG CTT GAG GTG CTG TCC TTC GAA TTC CAG AAG ATC AAG TAT TCC TAC GAT GCC				549
L E K K Q F L P V A F P V G N A F S Y Y				203
CTG GAG AAG AAG CAG TTT CTC CCC GTG GCC TTT CCT GTG GGA AAC GCC TTC TCA TAC TAT				609
Q S N R G F Q E D S E I R A A E K K F G				223
CAG AGC AAC AGA GGC TTC CAG GAA GAC TCA GAG ATC CGA GCA GCT GAG AAG AAA TTT GGG				669
S N K A E M V V P D F S E L F K E R A T				243
AGC AAC AAG GCC GAG ATG GTG GTG CCT GAC TTC TCG GAG CTT TTC AAG GAG AGA GCC ACA				729
A P F F V F Q V F C V G L W C L D E Y W				263
GCC CCC TTC TTT GTA TTT CAG GTG TTC TGT GTG GGG CTC TGG TGC CTG GAT GAG TAC TGG				789
Y Y S V F T L S M L V A F E A S L V Q Q				283
TAC TAC AGC GTC TTT ACG CTA TCC ATG CTG GTG GCG TTC GAG GCC TCG CTG GTG CAG CAG				849
Q M R N M S E I R K M G N K P H M I Q V				303
CAG ATG CGG AAC ATG TCG GAG ATC CGG AAG ATG GGC AAC AAG CCC CAC ATG ATC CAG GTC				909
Y R S R K W R P I A S D E I V P G D I V				323
TAC CGA AGC CGC AAG TGG AGG CCC ATT GCC AGT GAT GAG ATC GTA CCA GGG GAC ATC GTC				969
S I G R S P Q E N L V P C D V L L L R G				343
TCC ATC GGC CGC TCC CCA CAG GAG AAC CTG GTG CCA TGT GAC GTG CTT CTG CTG CGA GGC				1029

Fig. 22A

R	C	I	V	D	E	A	M	L	T	G	E	S	V	P	Q	M	K	E	P	363
CGC	TGC	ATC	GTA	GAC	GAG	GCC	ATG	CTC	ACG	GGG	GAG	TCC	GTG	CCA	CAG	ATG	AAG	GAG	CCC	1089
I	E	D	L	S	P	D	R	V	L	D	L	Q	A	D	S	R	L	H	V	383
ATC	GAA	GAC	CTC	AGC	CCA	GAC	CGG	GTG	CTG	GAC	CTC	CAG	GCT	GAT	TCC	CGG	CTG	CAC	GTC	1149
I	F	G	G	T	K	V	V	Q	H	I	P	P	Q	K	A	T	T	G	L	403
ATC	TTC	GGG	GGC	ACC	AAG	GTG	GTG	CAG	CAC	ATC	CCC	CCA	CAG	AAA	GCC	ACC	ACG	GGC	CTG	1209
K	P	V	D	S	G	C	V	A	Y	V	L	R	T	G	F	N	T	S	Q	423
AAG	CCG	GTT	GAC	AGC	GGG	TGC	GTG	GCC	TAC	GTC	CTG	CGG	ACC	GGA	TTC	AAC	ACA	TCT	CAG	1269
G	K	L	L	R	T	I	L	F	G	V	K	R	V	T	A	N	N	L	E	443
GGC	AAG	CTG	CTG	CGC	ACC	ATC	CTC	TTC	GGG	GTC	AAG	AGG	GTG	ACT	GCG	AAC	AAC	CTG	GAG	1329
T	F	I	F	I	L	F	L	L	V	F	A	I	A	A	A	A	Y	V	W	463
ACC	TTC	ATC	TTC	ATC	CTC	TTC	CTC	CTG	GTG	TTT	GCC	ATC	GCT	GCA	GCT	GCC	TAT	GTA	TGG	1389
I	E	G	T	K	D	P	S	R	N	R	Y	K	L	F	L	E	C	T	L	483
ATT	GAA	GGT	ACC	AAG	GAC	CCC	AGC	CGG	AAC	CGC	TAC	AAG	CTG	TTT	CTG	GAG	TGC	ACC	CTG	1449
I	L	T	S	V	V	P	P	E	L	P	I	E	L	S	L	A	V	N	T	503
ATC	CTC	ACC	TCG	GTC	GTG	CCT	CCT	GAG	CTG	CCC	ATC	GAG	CTG	TCC	CTG	GCC	GTC	AAC	ACC	1509
S	L	I	A	L	A	K	L	Y	M	Y	C	T	E	P	F	R	I	P	F	523
TCC	CTC	ATC	GCC	CTG	GCC	AAG	CTC	TAC	ATG	TAC	TGC	ACA	GAG	CCC	TTC	CGG	ATC	CCC	TTT	1569
A	G	K	V	E	V	C	C	F	D	K	T	G	T	L	T	S	D	S	L	543
GCT	GGC	AAG	GTC	GAG	GTG	TGC	TGC	TTT	GAC	AAG	ACG	GGG	ACG	TTG	ACC	AGT	GAC	AGC	CTG	1629
V	V	R	G	V	A	G	L	R	D	G	K	E	V	T	P	V	S	S	I	563
GTG	GTG	CGC	GGT	GTG	GCC	GGG	CTG	AGA	GAC	GGG	AAG	GAG	GTG	ACC	CCA	GTG	TCC	AGC	ATC	1689
P	V	E	T	H	R	A	L	A	S	C	H	S	L	M	Q	L	D	D	G	583
CCT	GTA	GAA	ACA	CAC	CGG	GCC	CTG	GCC	TCG	TGC	CAC	TCG	CTC	ATG	CAG	CTG	GAC	GAC	GGC	1749
T	L	V	G	D	P	L	E	K	A	M	L	T	A	V	D	W	T	L	T	603
ACC	CTC	GTG	GGT	GAC	CCT	CTA	GAG	AAG	GCC	ATG	CTG	ACG	GCC	GTG	GAC	TGG	ACG	CTG	ACC	1809
K	D	E	K	V	F	P	R	S	I	K	T	Q	G	L	K	I	H	Q	R	623
AAA	GAT	GAG	AAA	GTA	TTC	CCC	CGA	AGT	ATT	AAA	ACT	CAG	GGG	CTG	AAA	ATT	CAC	CAG	CGC	1869
F	H	F	A	S	A	L	K	R	M	S	V	L	A	S	Y	E	K	L	G	643
TTT	CAT	TTT	GCC	AGT	GCC	CTG	AAG	CGA	ATG	TCC	GTG	CTT	GCC	TCG	TAT	GAG	AAG	CTG	GGC	1929
S	T	D	L	C	Y	I	A	A	V	K	G	A	P	E	T	L	H	S	M	663
TCC	ACC	GAC	CTC	TGC	TAC	ATC	GCG	GCC	GTG	AAG	GGG	GCC	CCC	GAA	ACT	CTG	CAC	TCC	ATG	1989
F	S	Q	C	P	P	D	Y	H	H	I	H	T	E	I	S	R	E	G	A	683
TTC	TCC	CAG	TGC	CCG	CCC	GAC	TAC	CAC	CAC	ATC	CAC	ACC	GAG	ATC	TCC	CCG	GAA	GGA	GCC	2049
R	V	L	A	L	G	Y	K	E	L	G	H	L	T	H	Q	Q	A	R	E	703
CGC	GTC	CTG	GCG	CTG	GGG	TAC	AAG	GAG	CTG	GGA	CAC	CTC	ACT	CAC	CAG	CAG	GCC	CGG	GAG	2109
V	K	R	E	A	L	E	C	S	L	K	F	V	G	F	I	V	V	S	C	723
GTC	AAG	CGG	GAG	GCC	CTG	GAG	TGC	AGC	CTC	AAG	TTC	GTC	GGC	TTC	ATT	GTG	GTC	TCC	TGC	2169

Fig. 22B

P	L	K	A	D	S	K	A	V	I	R	E	I	Q	N	A	S	H	R	V	743
CCG	CTC	AAG	GCT	GAC	TCC	AAG	GCC	GTG	ATC	CGG	GAG	ATC	CAG	AAT	GCG	TCC	CAC	CGG	GTG	2229
V	M	I	T	G	D	N	P	L	T	A	C	H	V	A	Q	E	L	H	F	763
GTC	ATG	ATC	ACG	GGA	GAC	AAC	CCG	CTC	ACT	GCA	TGC	CAC	GTG	GCC	CAG	GAG	CTG	CAC	TTC	2289
I	E	K	A	H	T	L	I	L	Q	P	P	S	E	K	G	R	Q	C	E	783
ATT	GAA	AAG	GCC	CAC	ACG	CTG	ATC	CTG	CAG	CCT	CCC	TCC	GAG	AAA	GGC	CGG	CAG	TGC	GAG	2349
W	R	S	I	D	G	S	I	V	L	P	L	A	R	G	S	P	K	A	L	803
TGG	CGC	TCC	ATT	GAC	GGC	AGC	ATC	GTG	CTG	CCC	CTG	GCC	CGG	GGC	TCC	CCA	AAG	GCA	CTG	2409
A	L	E	Y	A	L	C	L	T	G	D	G	L	A	H	L	Q	A	T	D	823
GCC	CTG	GAG	TAC	GCA	CTG	TGC	CTC	ACA	GGC	GAC	GGC	TTG	GCC	CAC	CTG	CAG	GCC	ACC	GAC	2469
P	Q	Q	L	L	R	L	I	P	H	V	Q	V	F	A	R	V	A	P	K	843
CCC	CAG	CAG	CTG	CTC	CGC	CTC	ATC	CCC	CAT	GTG	CAG	GTG	TTC	GCC	CGT	GTG	GCT	CCC	AAG	2529
Q	K	E	F	V	I	T	S	L	K	E	L	G	Y	V	T	L	M	C	G	863
CAG	AAG	GAG	TTT	GTC	ATC	ACC	AGC	CTG	AAG	GAG	CTG	GGC	TAC	GTG	ACC	CTC	ATG	TGT	GGG	2589
D	G	T	N	D	V	G	A	L	K	H	A	D	V	G	V	A	L	L	A	883
GAT	GGC	ACC	AAC	GAC	GTG	GGC	GCC	CTG	AAG	CAT	GCT	GAC	GTG	GGT	GTG	GCG	CTC	TTG	GCC	2649
N	A	P	E	R	V	V	E	R	R	R	R	P	R	D	S	P	T	L	S	903
AAT	GCC	CCT	GAG	CGG	GTT	GTC	GAG	CGG	CGA	CGG	CGG	CCC	CGG	GAC	AGC	CCA	ACC	CTG	AGC	2709
N	S	G	I	R	A	T	S	R	T	A	K	Q	R	S	G	L	P	P	S	923
AAC	AGT	GGC	ATC	AGA	GCC	ACC	TCC	AGG	ACA	GCC	AAG	CAG	CGG	TGC	GGG	CTC	CCT	CCC	TCC	2769
E	E	Q	P	T	S	Q	R	D	R	L	S	Q	V	L	R	D	L	E	D	943
GAG	GAG	CAG	CCA	ACC	TCC	CAG	AGG	GAC	CGC	CTG	AGC	CAG	GTG	CTG	CGA	GAC	CTC	GAG	GAC	2829
E	S	T	P	I	V	K	L	G	D	A	S	I	A	A	P	F	T	S	K	963
GAG	AGT	ACG	CCC	ATT	GTG	AAA	CTG	GGG	GAT	GCC	AGC	ATC	GCA	GCA	CCC	TTC	ACC	TCC	AAG	2889
L	S	S	I	Q	C	I	C	H	V	I	K	Q	G	R	C	T	L	V	T	983
CTC	TCA	TCC	ATC	CAG	TGC	ATC	TGC	CAC	GTG	ATC	AAG	CAG	GGC	CGC	TGC	ACG	CTG	GTG	ACC	2949
T	L	Q	M	F	K	I	L	A	L	N	A	L	I	L	A	Y	S	Q	S	1003
ACG	CTA	CAG	ATG	TTC	AAG	ATC	CTG	GCG	CTC	AAT	GCC	CTC	ATC	CTG	GCC	TAC	AGC	CAG	AGC	3009
V	L	Y	L	E	G	V	K	F	S	D	F	Q	A	T	L	Q	G	L	L	1023
GTC	CTC	TAC	CTG	GAG	GGA	GTC	AAG	TTC	AGT	GAC	TTC	CAG	GCC	ACC	CTA	CAG	GGG	CTG	CTG	3069
L	A	G	C	F	L	F	I	S	R	S	K	P	L	K	T	L	S	R	E	1043
CTG	GCC	GGC	TGC	TTC	CTC	TTC	ATC	TCC	CGT	TCC	AAG	CCC	CTC	AAG	ACC	CTC	TCC	CGA	GAA	3129
R	P	L	P	N	I	F	N	L	Y	T	I	L	T	V	M	L	Q	F	F	1063
CGG	CCC	CTG	CCC	AAC	ATC	TTC	AAC	CTG	TAC	ACC	ATC	CTC	ACC	GTC	ATG	CTC	CAG	TTC	TTT	3189
V	H	F	L	S	L	V	Y	L	Y	R	E	A	Q	A	R	S	P	E	K	1083
GTG	CAC	TTC	CTG	AGC	CTT	GTC	TAC	CTG	TAC	CGT	GAG	GCC	CAG	GCC	CGG	AGC	CCC	GAG	AAG	3249
Q	E	Q	F	V	D	L	Y	K	E	F	E	P	S	L	V	N	S	T	V	1103
CAG	GAG	CAG	TTC	GTG	GAC	TTG	TAC	AAG	GAG	TTT	GAG	CCA	AGC	CTG	GTC	AAC	AGC	ACC	GTC	3309

Fig. 22C

Y	I	M	A	M	A	M	Q	M	A	T	F	A	I	N	Y	K	G	P	P	1123
TAC	ATC	ATG	GCC	ATG	GCC	ATG	CAG	ATG	GCC	ACC	TTC	GCC	ATC	AAT	TAC	AAA	GGC	CCG	CCC	3369
F	M	E	S	L	P	E	N	K	P	L	V	W	S	L	A	V	S	L	L	1143
TTC	ATG	GAG	AGC	CTG	CCC	GAG	AAC	AAG	CCC	CTG	GTG	TGG	AGT	CTG	GCA	GTT	TCA	CTC	CTG	3429
A	I	I	G	L	L	L	G	S	S	P	D	F	N	S	Q	F	G	L	V	1163
GCC	ATC	ATT	GGC	CTG	CTC	CTC	GGC	TCC	TCG	CCC	GAC	TTC	AAC	AGC	CAG	TTT	GGC	CTC	GTG	3489
D	I	P	V	E	F	K	L	V	I	A	Q	V	L	L	L	D	F	C	L	1183
GAC	ATC	CCT	GTG	GAG	TTC	AAG	CTG	GTC	ATT	GCC	CAG	GTC	CTG	CTC	CTG	GAC	TTC	TGC	CTG	3549
A	L	L	A	D	R	V	L	Q	F	F	L	G	T	P	K	L	K	V	P	1203
GCG	CTC	CTG	GCC	GAC	CGC	GTC	CTG	CAG	TTC	TTC	CTG	GGG	ACC	CCG	AAG	CTG	AAA	GTG	CCT	3609
S	*																			1205
TCC	TGA																			3615

GATGGCAGTGCTGGTACCCACTGCCCCACCC'TGGCTGCCGCTGGGCGGGAACCCCAACAGGGCCCCGGGAGGGAACCC'TG
CCCCCAACCCCCACAGCAAGGCTGTACAGTCTCGCCCTTGGAAGACTGAGCTGGGACCCCCACAGCCATCCGCTGGCT
TGGCCAGCAGAACCAGCCCCAAGCCAGCACCTTTGGTAAATAAAGCAGCATCTGAGATTTTAAAAAAAAAAAAAAAAAAAA

Fig. 22D

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM
 hmmpfam - search a single seq against HMM database
 HMMER 2.1.1 (Dec 1998)
 Copyright (C) 1992-1998 Washington University School of Medicine
 HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam5.4/Pfam
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.13531.seq

Query: Fbh49933FL1

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
E1-E2_ATPase	E1-E2 ATPase	37.0	3.4e-10	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
E1-E2_ATPase	1/1	302	392	..	36 119 ..	37.0	3.4e-10

Alignments of top-scoring domains:

E1-E2_ATPase: domain 1 of 1, from 302 to 392: score 37.0, E = 3.4e-10
 *->tVlRdGkeeeipaeelvpGDIVev.kpGd..rVPADGrvvegestLk
 +V+R k++ i ++e+vpGDIV +++ ++ VP D+ ++ g++
 Fbh49933FL 302 QVYRSRKWRPIASDEIVPGDIVSIgRSPQenLVPCDVLLRGRC--I 346

VDESaLTGESlPVeKhsdvtp.....dnp.lernmVfaGtlvvs<-
 VDE++LTGES+P++K +++++ +++++ d r f+GT vv+
 Fbh49933FL 347 VDEAMLTGESVPQMK--EPIEDlspdrvlDLQaDSRLHVIFGGTKVVQ 392

Fbh49933FL

FIG. 23

Analysis of Fbh49933FL1 (1204 aa)

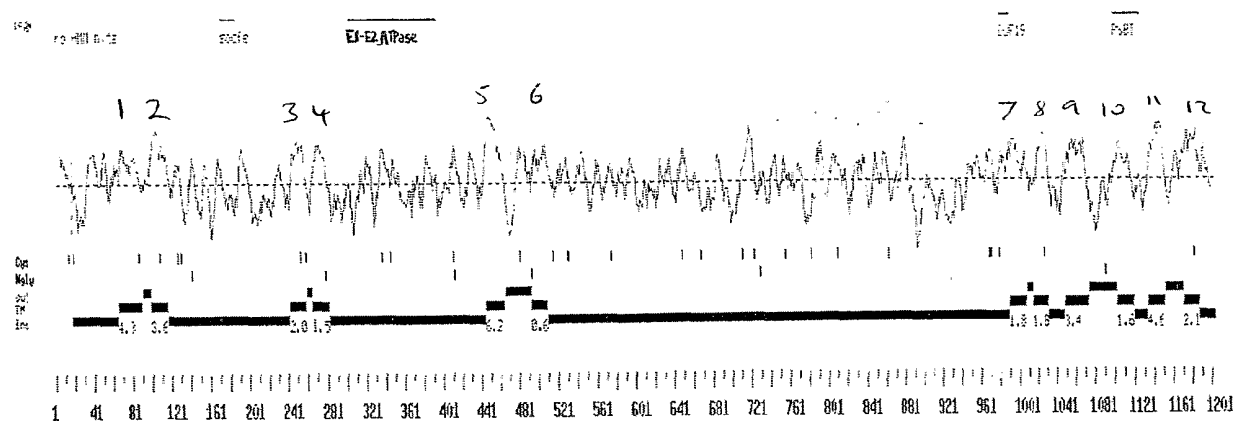


FIG. 24

CLUSTAL W (1.74) multiple sequence alignment

```

Fbh49933FL1      MAAAAVGNVPCGARPCGVRFDGQPKPGPQPRALLAAGPALIANGDELVAAVWPYRRLA
YE56elegans      -----MGVDQLVETIIPYNLRS
                                   . : ** : : ** . :

                                   TM1
Fbh49933FL1      LLRRITVLPFAGLLYPAWLGAAAAGCWGWGSSWVQIPEAALLVLATICLAHALTVLSGHW
YE56elegans      IATHLYVPPFT-IITAIWT-----YVWLNIFGYEYYELGMLGYAAIFVILALVLLFCHW
:   : * * * : : : . *           *       . : : * . : * : * : : * : *
                                   TM2
Fbh49933FL1      SVHAHCALTCTPEYDPSKATFVKVAPTPNNGSTELVALHRNEGEDGLEVLSEFEQKIKYS
YE56elegans      MMPVRCFLMCSKQEDVRIASHVCVIPTQNNGWPELVKLMRTT-RDKQTKLWFEFQRVHYT
:   . : * * * : : *   * : . * * * * * . * * * * * : : * :

Fbh49933FL1      YDALEKKQFLPVAFFVGNAFSYYQSNRGFQEDSEIRAAEKKFGSNAEMVVPDFSELFKE
YE56elegans      WDE-ESREFQTKTLDTAKPMVFFQKSHGFVEEHVKDAKYLLGDNKTEMIVPQFLEMFIE
: *   . : * . : : . : : : : * . : : : : * : : * : * : * : * : *
                                   TM3
Fbh49933FL1      RATAPFFVFQVFCVGLWCLEDEYVYVSFTLSMLVAFEASLVQQQMRNMSEIRKMGNKPHM
YE56elegans      RATAPFFVFQVFCVGLWCLEDMWYSLFTLFMLMTFEATLVKQQMKNMSEIRNMGNKTYM
***** : : * : * * * : : * : * : * : * : * : * : * : * : *
                                   TM4
Fbh49933FL1      IQVYRSRKWRPIASDEIVPGDIVSIGRSPQENLVPCDVLLLRGRCIVDEAMLTGESVPQM
YE56elegans      INVLRGKKWQKIKIEELVAGDIVSIGRGAEEECVPCDLLLRGPCIVDESMLTGESVPQM
* : * . : * : *   : * : . * : * : * : * : * : * : * : * : * : *

Fbh49933FL1      KEPIEDLSPDRVLDLQADSRLHVFEGGTVVQHI PPQKATTGLKPVDSG-CVAYVLRGTG
YE56elegans      KEPIEDVEKDKIFDIETDSRLHVFEGGTVVQHTAPGKAAEGMVKSPDGNCICYVIRGTG
***** . : : : : : * : * : * : * : * : * : * : * : * : *
                                   TM5
Fbh49933FL1      NTSQGKLLRTILFGVKRV TANNLETFIFILFLLVFAIAAAAYVWIEGTDPSRNRYKLF
YE56elegans      NTSQGKLLRTIMFGVKKATANNLETFIFILFLLVFAIAAAAYLWIKGSVDETRSKYKLF
***** : * : * : * : * : * : * : * : * : * : * : * : * : *
                                   TM6
Fbh49933FL1      ECTLILTSVVPPELPIELSLAVNTSLIALAKLYMYCTEPFRIPFAGKVEVCCFDKTGTLT
YE56elegans      ECTLILTSVIPPPELPIELSLAVNSSIMALQKLGIFCTEPFRIPFAGKVDICCFDKTGTLT
***** : * : * : * : * : * : * : * : * : * : * : * : * : *
                                   Phosphorylation site
Fbh49933FL1      SDSLVRGVAGLRDGKE--VTPVSSI PVETHRALASCHSLMQLDGTLVGDPLEKAMLTA
YE56elegans      TDNLVVEGVA-LNNQKEGMIRNAEDLPHESLQVLASCHSLVRFEED-LVGDPLEKACLSW
: * . * . * . * . : * *   :   . : * * : : * : * : * : * : * : * : *

Fbh49933FL1      VDWTLTKEKVFPRSIKTQ---GLKIHQRHFHFAALKRMSVLASYEKLGSTDLCYIAAVK
YE56elegans      CGWNLTGDAVMPPKTAAGKISGIKIFHRYHFSSAMKRMTVVAGYQSPGTSDTTFIVAVK
. * . * . : * : * . : :   * : * . : * : * : * : * : * : * : * : *

Fbh49933FL1      GAPETLHSMFSQCPPDYHHIHTETISREGARVLALGYKELGHLTHQQAREVKREALECSLK
YE56elegans      GAPEVLRNMYADLPDYDETYTRLTRQGSRLAMGIRKLGETRVGELRDKKRENFENDLA
***** : * : * : * : * : * : * : * : * : * : * : * : * : *

```

Fig. 25A

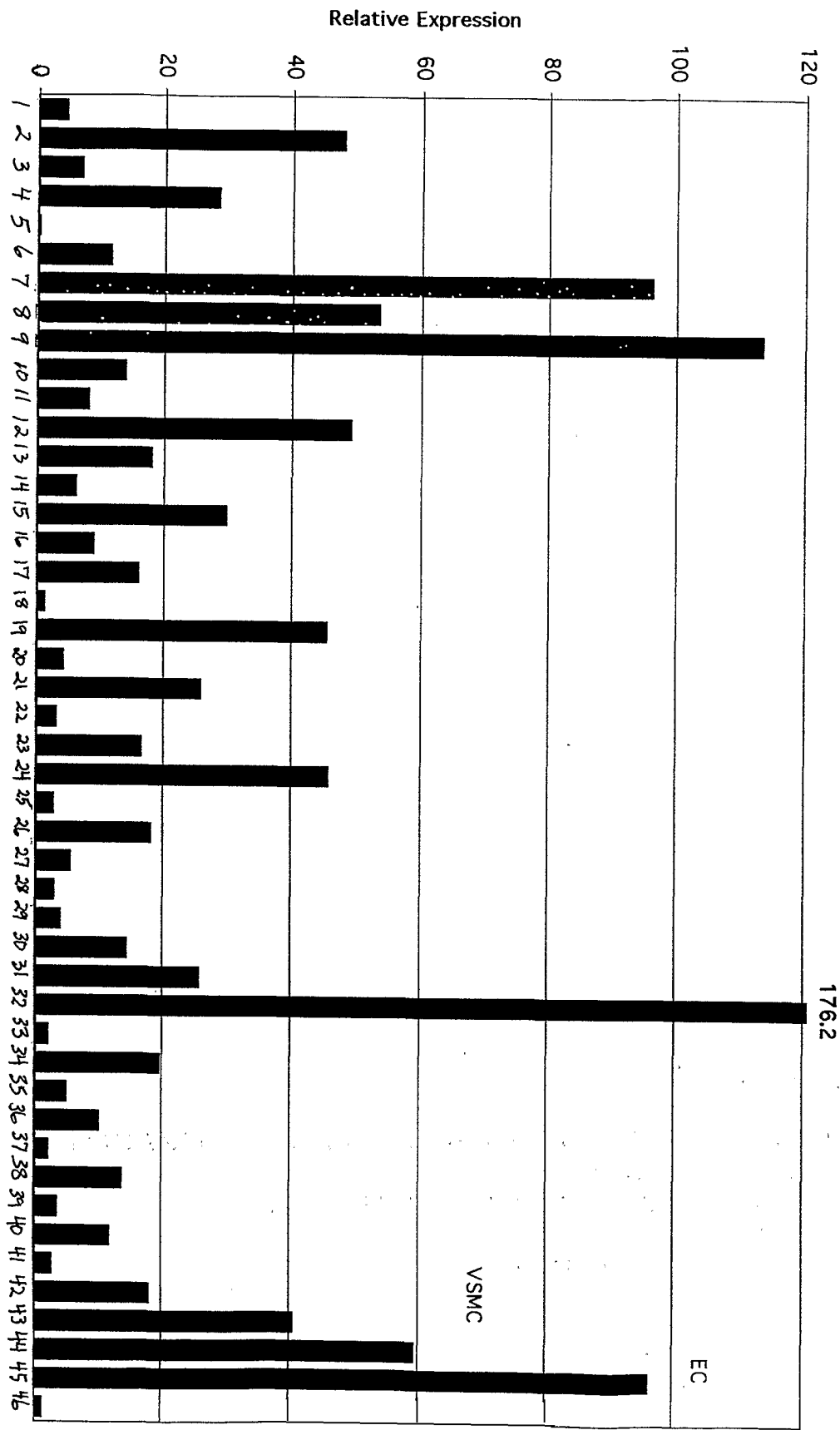
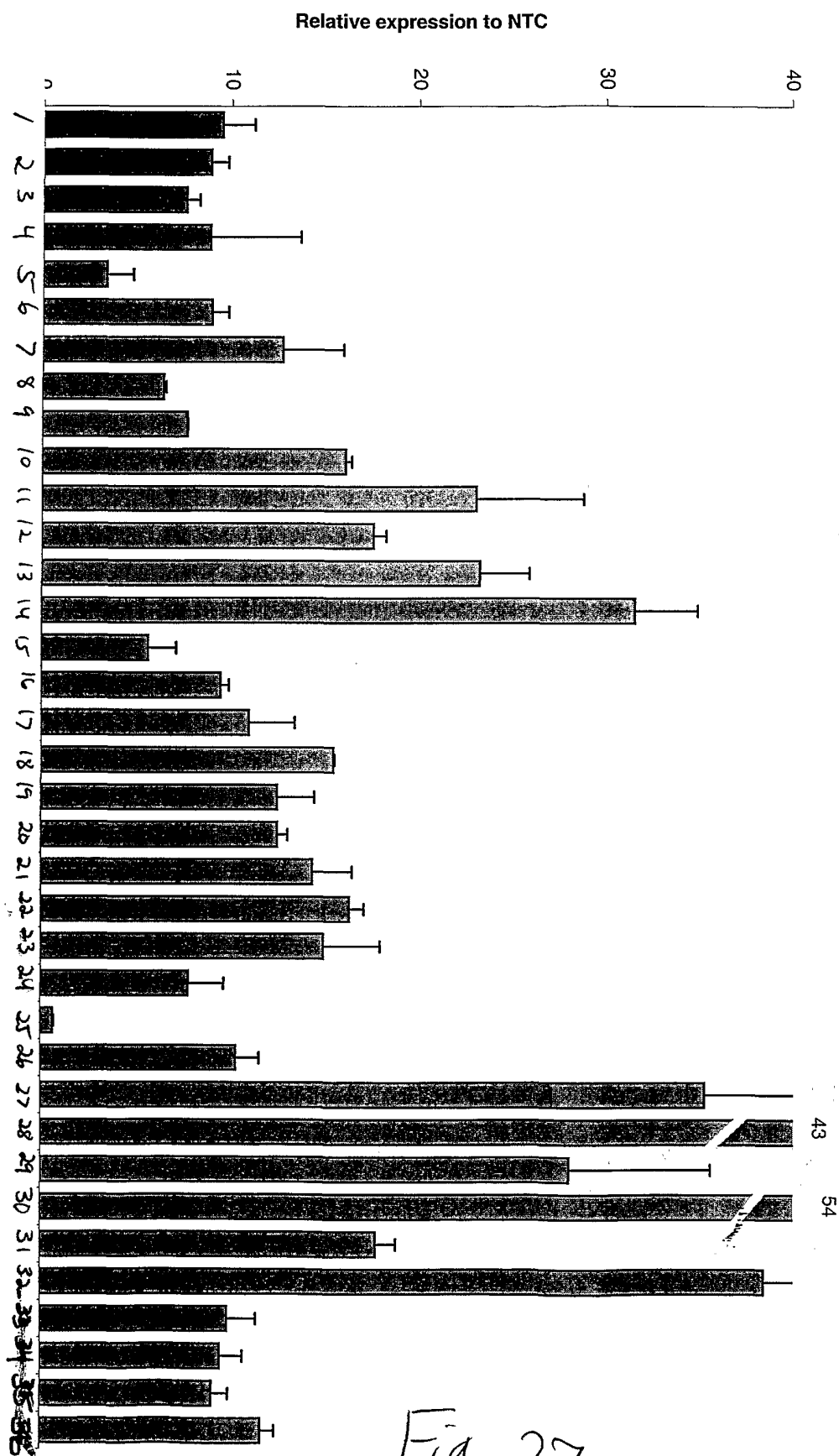


FIG. 26



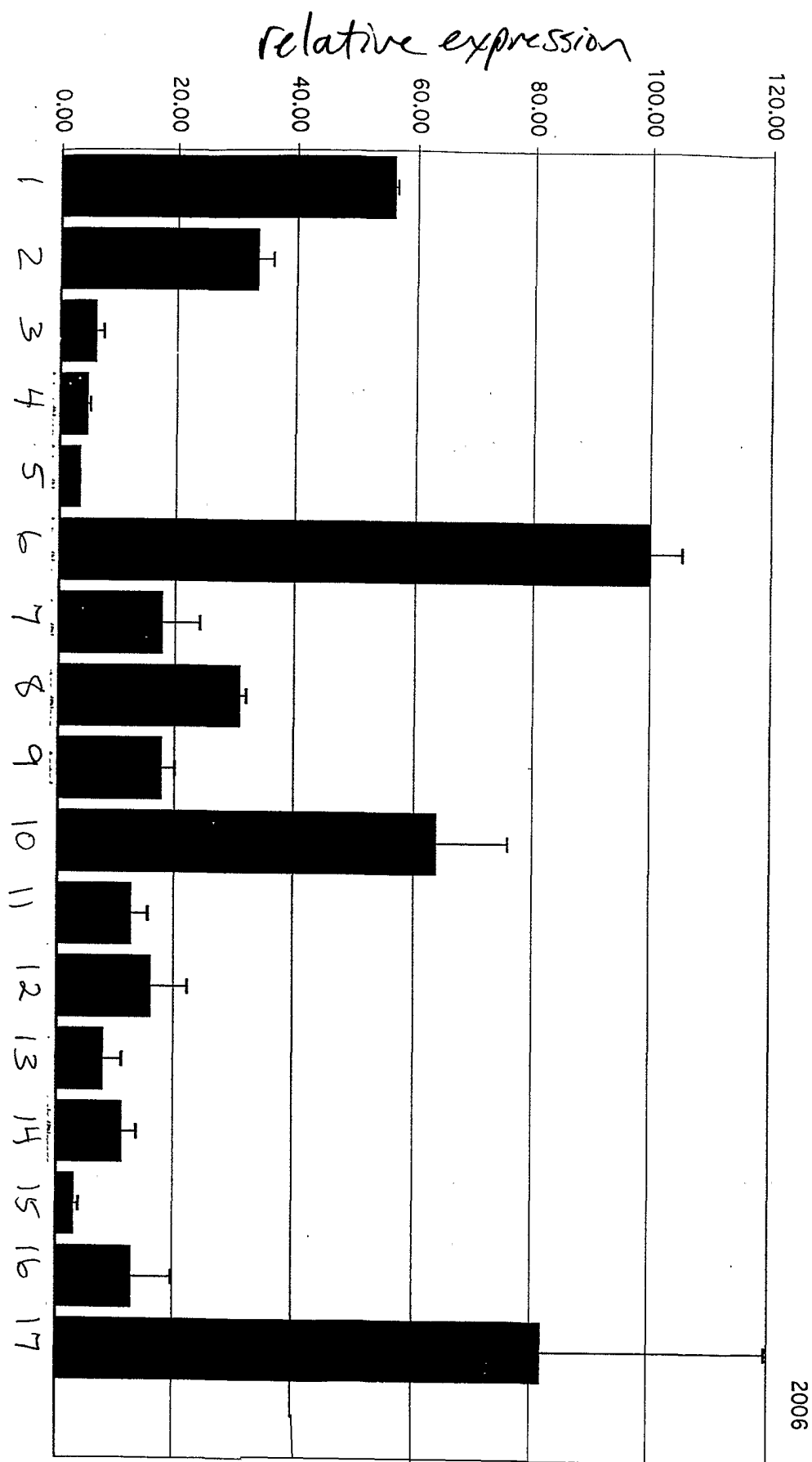


FIG. 28

49933

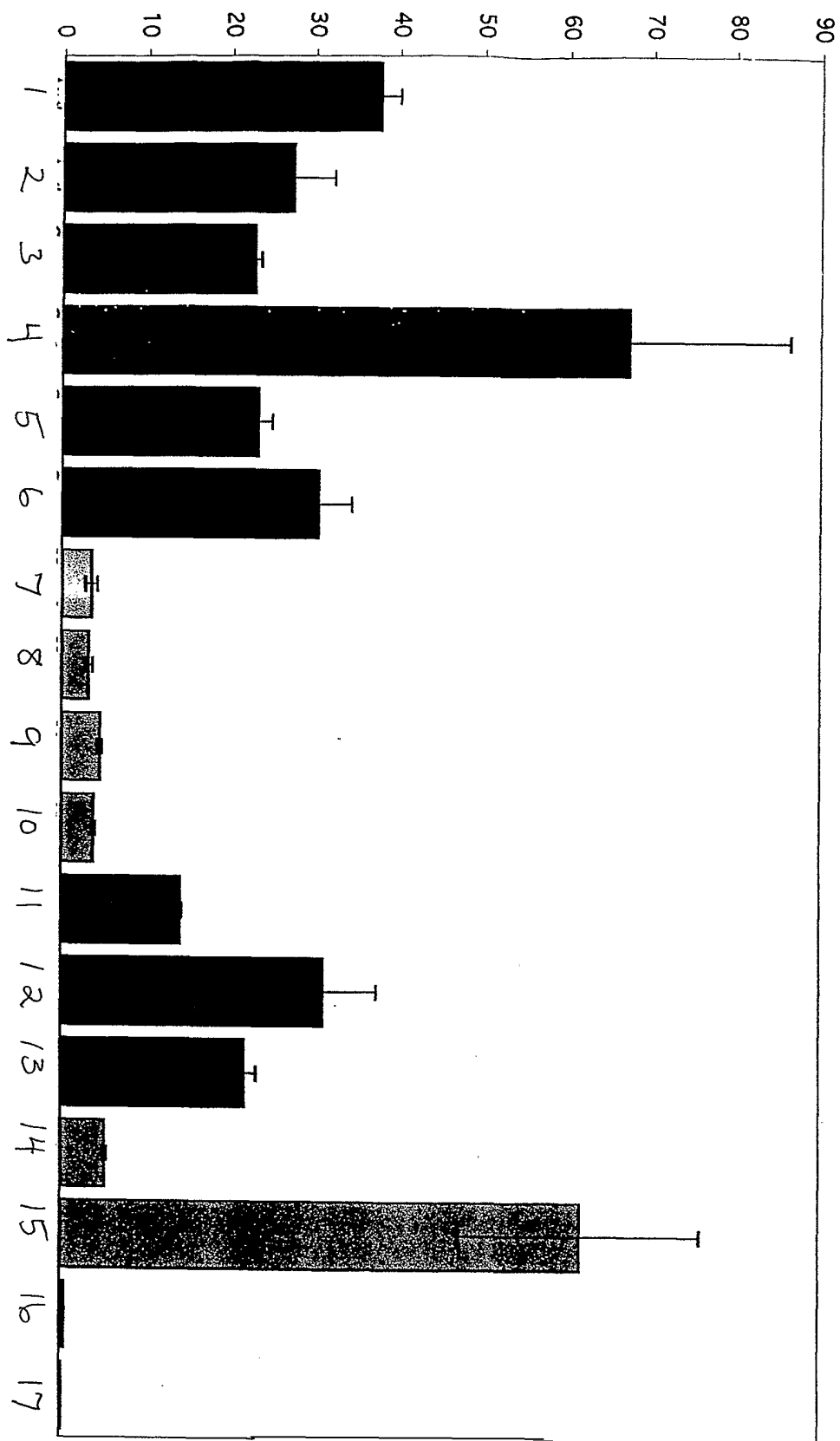


FIG. 29

	GenBank Accession No.		SEQ ID NO:	
SERCA mutagenesis		IPEGLPA	16	
Fbh49937FL (HEAT-1)		DLVTVVVPPALPAAMTVCTLYAQSRLRR	17	
Fbh49931FL (HEAT-2)		DIITITVPPALPAAMTAGIVYAQRRLKK	18	
Fbh49933FL (HEAT-3)		LILTSVVPPELPIELSLAVNTSLIALAK	19	
ATC9_YEAST_yor291	Q12697	DIITIVVPPALPATLTIGTNFALSRLKEK	20	Type V
ATC6_YEAST_SPF1_yel031w	P39986	LIITSVVPPELPMELTMAVNSSLAALAK	21	
ATCY_SCHPO_sp_0140220	O14022	VLTIIVVPPALPATLSVGIANSIARLSRA	22	
gp_7324471_CE	AAF59622	DLVTIVVPPALPAVMGIGIFYAQRRLRQK	23	
sp_Q21286_YBF7_CAEEL	Q21286	DIITIVVPPALPAAMSVGIINANSRLKKK	24	
YH2M_CAEEL_sp_q27533	Q27533	DIITITVPPALPAAMSVGIINAQLRLKKK	25	
YE56_CAEEL_CAB05683	CAB05683	LILTSVIPPELPIELSLAVNSSLMALQKL	26	
ATP2A1_H_AAB53112	AAB53112	ALAVAAIPEGLPAVITTCALGTRRMAKK	27	Ca ²⁺
rabbitSERCA1	P04191	ALAVAAIPEGLPAVITTCALGTRRMAKK	28	
ATCB_CHICK_SERCA1	P13585	ALAVAAIPEGLPAVITTCALGTRRMAKK	29	
ATC2_FELCA_SERCA2	Q00779	ALAVAAIPEGLPAVITTCALGTRRMAKK	30	
SERCA_Procambarus_clarkii	AAB82291	ALAVAAIPEGLPAVITTCALGTRRMAKK	31	
hSERCA2	P16615	ALAVAAIPEGLPAVITTCALGTRRMAKK	32	
hSERCA3	Q93084	ALAVAAIPEGLPAVITTCALGTRRMAKK	33	
ATCB_DROME_P22700_158416	P22700	AVAVAAIPEGLPAVITTCALGTRRMAKK	34	
ATC1_YEAST_PMR1	P13586	SLAVAAIPEGLPIIVTVTLALGVLRMAKR	35	
Y_PMC1	P38929	TVIVVAVPEGLPLAVTLALAFATTRMTKD	36	
hPMCA1	P20020	TVLVVAVPEGLPLAVTISLAYSVKKMMKD	37	
hPMCA2	Q01814	TVLVVAVPEGLPLAVTISLAYSVKKMMKD	38	
ratPMCA1		TVLVVAVPEGLPLAVTISLAYSVKKMMKD	39	
PMCA3_H	Q16720	TVLVVAVPEGLPLAVTISLAYSVKKMMKD	40	
PMCA4_H	P23634	TVLVVAVPEGLPLAVTISLAYSVKKMMKD	41	
Pcalp_YEAST	P38360	TVLIVSCPCVIGLAVPIVFVIASGVAAKR	42	Cu ²⁺
AT7A_HUMAN	Q04656	TVLCIACPCSLGLATPTAVMVGTVGAQN	43	
AT7B_HUMAN	P35670	TVLCIACPCSLGLATPTAVMVGTVGAQN	44	
ATNA_DROME_P13607_732656	P13607	GIIIVANVPEGLLATVTVCLTLTAKRMAK	45	Na ⁺ /K ⁺
ATNA_HYDAT	P35317	GIIIVANVPEGLLATVTVCLTLTAKKMAK	46	
ATN1_BUFMA	P30714	GIIIVANVPEGLLATVTVCLTLTAKRMAK	47	
ATN1_HUMAN	P05023	GIIIVANVPEGLLATVTVCLTLTAKRMAK	48	
ATN2_HUMAN	P50993	GIIIVANVPEGLLATVTVCLTLTAKRMAK	49	
ATN3_HUMAN	P13637	GIIIVANVPEGLLATVTVCLTVTAKRMAK	50	
ATPP2_H_AAD34706	AAD34706	ILFNNLIPISLLVTLEVVKFTQAYFINWD	51	PL
DRS2_ATC4_YEAST	P39524	ILFSNLVPISLFVTVELIKYYQAFMIGSD	52	

Fig. 30

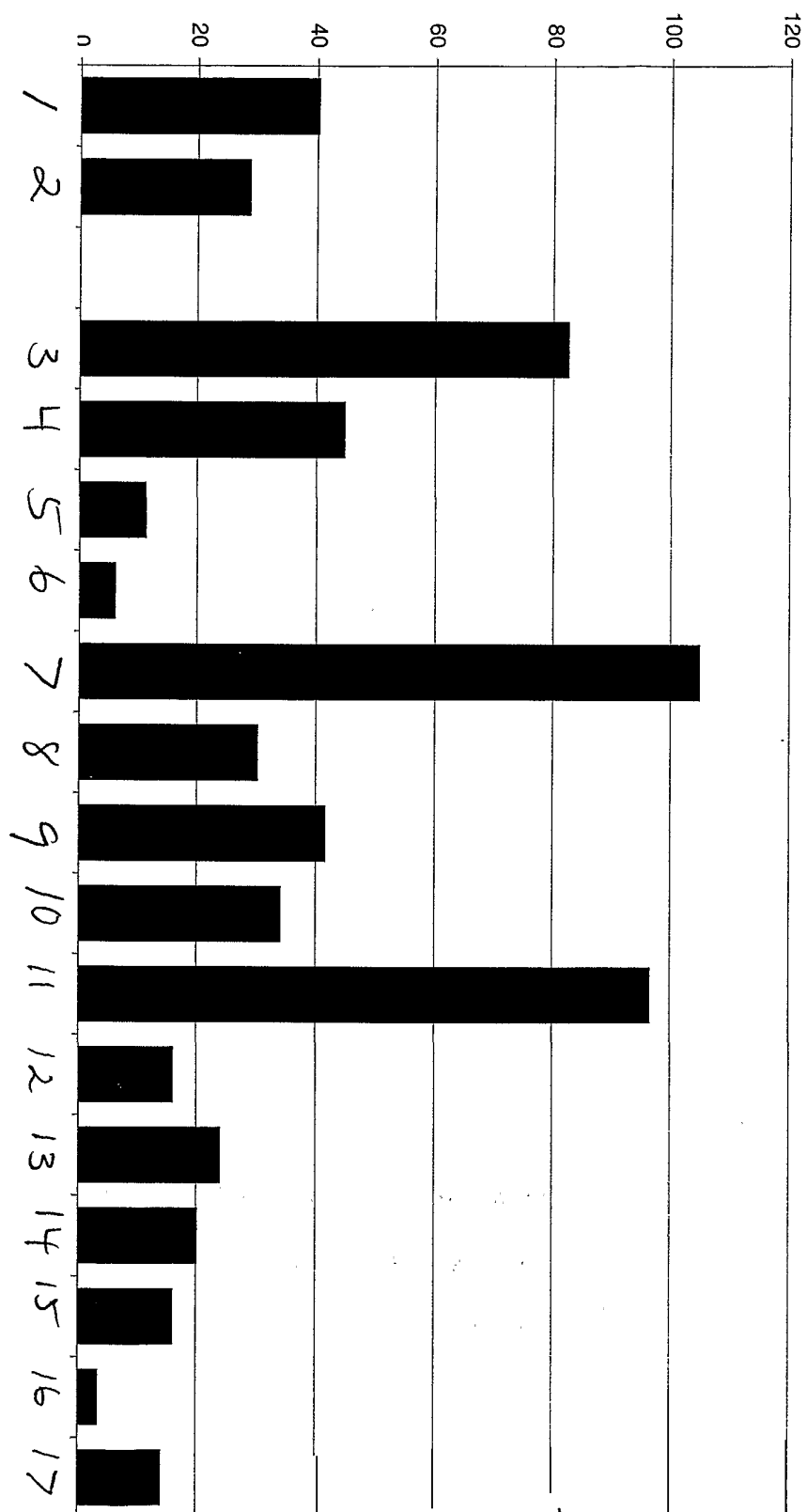


Fig. 31

<110> Millennium Pharmaceuticals, Inc. et al.

<130> MNI-184PC

```
<150> 60/226,504
<151> 2000-08-21
```

<160> 52

<170> PatentIn Ver. 2.0

```
<210> 1
<211> 4055
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> (210)..(3749)
```

```
<400> 1
tactataggg agtcgaacca cgcgtccggc cgcgccgagc. ctggtggccc aggtgccccg 60
cccgcgtcag ccctgctcca gcccgcgct agcccagcgc ccctgcctcc gggccgtccg 120
gaccgcgccc ccgcccaggg ccttgcgcac gccggggccc aggccgaggg ccgcagcgcc 180
ggggccggcg atgagcgcga ggagccggc atg agc gca gac agc agc cct ctc      233
                               Met Ser Ala Asp Ser Ser Pro Leu
                               1               5
```

gtg ggc agc acg ccc acc ggt tat ggg acc ctg acg ata ggg aca tca 281
Val Gly Ser Thr Pro Thr Gly Tyr Gly Thr Leu Thr Ile Gly Thr Ser
10 15 20

ata gat ccc ctc agc tcc tca gtt tca tcc gtg agg ctc agc ggc tac 329
Ile Asp Pro Leu Ser Ser Ser Val Ser Ser Val Arg Leu Ser Gly Tyr
25 30 35 40

tgt ggc agt cca tgg agg gtc atc ggc tat cac gtc gtg gtc tgg atg 377
Cys Gly Ser Pro Trp Arg Val Ile Gly Tyr His Val Val Val Trp Met
45 50 55

atg gct ggg atc cct ttg ctg ctc ttc cgt tgg aag ccc ctg tgg ggg 425
Met Ala Gly Ile Pro Leu Leu Leu Phe Arg Trp Lys Pro Leu Trp Gly
60 65 70

gtg cgg ctg cgg ctc cgg ccc tgc aac ctg gcc cac gcc gaa aca ctc 473
Val Arg Leu Arg Leu Arg Pro Cys Asn Leu Ala His Ala Glu Thr Leu
75 80 85

gtt atc gaa ata aga gac aaa gag gat agt tcc tgg cag ctc ttc act	521
Val Ile Glu Ile Arg Asp Lys Glu Asp Ser Ser Trp Gln Leu Phe Thr	
90 95 100	
gtc cag gtg cag act gag gcc atc ggc gag ggc agc ctg gag ccg tcc	569
Val Gln Val Gln Thr Glu Ala Ile Gly Glu Gly Ser Leu Glu Pro Ser	
105 110 115 120	
cca cag tcc cag gca gag gat ggc cgg agc cag gcg gca gtt ggg gcg	617
Pro Gln Ser Gln Ala Glu Asp Gly Arg Ser Gln Ala Ala Val Gly Ala	
125 130 135	
gta cca gag ggt gcc tgg aag gat acg gcc cag ctc cac aag agc gag	665
Val Pro Glu Gly Ala Trp Lys Asp Thr Ala Gln Leu His Lys Ser Glu	
140 145 150	
gag gcg gtg agt gtc gga cag aag cgg gtg ctg cgg tat tac ctc ttc	713
Glu Ala Val Ser Val Gly Gln Lys Arg Val Leu Arg Tyr Tyr Leu Phe	
155 160 165	
cag ggc cag cgc tat atc tgg atc gag acc cag caa gcc ttc tac cag	761
Gln Gly Gln Arg Tyr Ile Trp Ile Glu Thr Gln Gln Ala Phe Tyr Gln	
170 175 180	
gtc agc ctc ctg gac cat ggc cgc tct tgt gac gac gtc cac cgc tcc	809
Val Ser Leu Leu Asp His Gly Arg Ser Cys Asp Asp Val His Arg Ser	
185 190 195 200	
cgc cat ggc ctc agc ctc cag gac caa atg gtg agg aag gcc att tac	857
Arg His Gly Leu Ser Leu Gln Asp Gln Met Val Arg Lys Ala Ile Tyr	
205 210 215	
ggc ccc aac gtg atc agc ata ccg gtc aag tcc tac ccc cag ctg ctg	905
Gly Pro Asn Val Ile Ser Ile Pro Val Lys Ser Tyr Pro Gln Leu Leu	
220 225 230	
gtg gac gag gca ctg aac ccc tac tat ggg ttc cag gcc ttc agc atc	953
Val Asp Glu Ala Leu Asn Pro Tyr Tyr Gly Phe Gln Ala Phe Ser Ile	
235 240 245	
gcg ctg tgg ctg gct gac cac tac tac tgg tac gcc ctg tgc atc ttc	1001
Ala Leu Trp Leu Ala Asp His Tyr Tyr Trp Tyr Ala Leu Cys Ile Phe	
250 255 260	
ctc att tcc tcc atc tcc atc tgc ctg tcg ctg tac aag acc aga aag	1049
Leu Ile Ser Ser Ile Ser Ile Cys Leu Ser Leu Tyr Lys Thr Arg Lys	
265 270 275 280	
caa agc cag act cta agg gac atg gtc aag ttg tcc atg cgg gtg tgc	1097
Gln Ser Gln Thr Leu Arg Asp Met Val Lys Leu Ser Met Arg Val Cys	
285 290 295	
gtg tgc cgg cca ggg gga gag gaa gag tgg gtg gac tcc agt gag cta	1145
Val Cys Arg Pro Gly Gly Glu Glu Glu Trp Val Asp Ser Ser Glu Leu	
300 305 310	
gtg ccc gga gac tgc ctg gtg ctg ccc cag gag ggt ggg ctg atg ccc	1193
Val Pro Gly Asp Cys Leu Val Leu Pro Gln Glu Gly Gly Leu Met Pro	
315 320 325	

- 3 -

tgt gat gcc gcc ctg gtg gcc ggc gag tgc atg gtg aat gag agc tct	1241
Cys Asp Ala Ala Leu Val Ala Gly Glu Cys Met Val Asn Glu Ser Ser	
330 335 340	
ctg aca gga gag agc att cca gtg ctg aag acg gca ctg ccg gag ggg	1289
Leu Thr Gly Glu Ser Ile Pro Val Leu Lys Thr Ala Leu Pro Glu Gly	
345 350 355 360	
ctg ggg ccc tac tgt gca gag aca cac cgg cgg cac aca ctc ttc tgc	1337
Leu Gly Pro Tyr Cys Ala Glu Thr His Arg Arg His Thr Leu Phe Cys	
365 370 375	
ggg acc ctc atc ttg cag gcc cgg gcc tat gtg gga ccg cac gtc ctg	1385
Gly Thr Leu Ile Leu Gln Ala Arg Ala Tyr Val Gly Pro His Val Leu	
380 385 390	
gca gtg gtg acc cgc aca ggg ttc tgc acg gca aaa ggg ggc ctg gtg	1433
Ala Val Val Thr Arg Thr Gly Phe Cys Thr Ala Lys Gly Gly Leu Val	
395 400 405	
agc tcc atc ttg cac ccc cgg ccc atc aac ttc aag ttc tat aaa cac	1481
Ser Ser Ile Leu His Pro Arg Pro Ile Asn Phe Lys Phe Tyr Lys His	
410 415 420	
agc atg aag ttt gtg gct gcc ctc tct gtc ctg gct ctc ctc ggc acc	1529
Ser Met Lys Phe Val Ala Ala Leu Ser Val Leu Ala Leu Leu Gly Thr	
425 430 435 440	
atc tac agc atc ttc atc ctc tac cga aac cgg gtg cct ctg aat gag	1577
Ile Tyr Ser Ile Phe Ile Leu Tyr Arg Asn Arg Val Pro Leu Asn Glu	
445 450 455	
att gta atc cgg gct ctc gac ctg gtg acc gtg gtg gtg cca cct gcc	1625
Ile Val Ile Arg Ala Leu Asp Leu Val Thr Val Val Val Pro Pro Ala	
460 465 470	
ctg cct gct gcc atg act gtg tgc acg ctc tac gcc cag agc cga ctg	1673
Leu Pro Ala Ala Met Thr Val Cys Thr Leu Tyr Ala Gln Ser Arg Leu	
475 480 485	
cgg aga cag ggc att ttc tgc atc cac cca ctg cgc atc aac ctg ggg	1721
Arg Arg Gln Gly Ile Phe Cys Ile His Pro Leu Arg Ile Asn Leu Gly	
490 495 500	
ggc aag ctg cag ctg gtg tgt ttc gac aag acg ggc acc ctc act gag	1769
Gly Lys Leu Gln Leu Val Cys Phe Asp Lys Thr Gly Thr Leu Thr Glu	
505 510 515 520	
gac ggc tta gac gtg atg ggg gtg gtg ccc ctg aag ggg cag gca ttc	1817
Asp Gly Leu Asp Val Met Gly Val Val Pro Leu Lys Gly Gln Ala Phe	
525 530 535	
ctg ccc ctg gtc cca gag cct cgc cgc ctg cct gtg ggg ccc ctg ctc	1865
Leu Pro Leu Val Pro Glu Pro Arg Arg Leu Pro Val Gly Pro Leu Leu	
540 545 550	
cga gca ctg gcc acc tgc cat gcc ctc agc cgg ctc cag gac acc ccc	1913
Arg Ala Leu Ala Thr Cys His Ala Leu Ser Arg Leu Gln Asp Thr Pro	
555 560 565	

gtg ggc gac ccc atg gac ttg aag atg gtg gag tct act ggc tgg gtc 1961 Val Gly Asp Pro Met Asp Leu Lys Met Val Glu Ser Thr Gly Trp Val 570 575 580
ctg gag gaa gag ccg gct gca gac tca gca ttt ggg acc cag gtc ttg 2009 Leu Glu Glu Glu Pro Ala Ala Asp Ser Ala Phe Gly Thr Gln Val Leu 585 590 595 600
gca gtg atg aga cct cca ctt tgg gag ccc cag ctg cag gca atg gag 2057 Ala Val Met Arg Pro Pro Leu Trp Glu Pro Gln Leu Gln Ala Met Glu 605 610 615
gag ccc ccg gtg cca gtc agc gtc ctc cac cgc ttc ccc ttc tct tcg 2105 Glu Pro Pro Val Pro Val Ser Val Leu His Arg Phe Pro Phe Ser Ser 620 625 630
gct ctg cag cgc atg agt gtg gtg gtg gcg tgg cca ggg gcc act cag 2153 Ala Leu Gln Arg Met Ser Val Val Val Ala Trp Pro Gly Ala Thr Gln 635 640 645
ccc gag gcc tac gtc aaa ggc tcc ccg gag ctg gtg gca ggg ctc tgc 2201 Pro Glu Ala Tyr Val Lys Gly Ser Pro Glu Leu Val Ala Gly Leu Cys 650 655 660
aac ccc gag aca gtg ccc acc gac ttc gcc cag atg ctg cag agc tat 2249 Asn Pro Glu Thr Val Pro Thr Asp Phe Ala Gln Met Leu Gln Ser Tyr 665 670 675 680
aca gct gct ggc tac cgt gtc gtg gcc ctg gcc agc aag cca ctg ccc 2297 Thr Ala Ala Gly Tyr Arg Val Val Ala Leu Ala Ser Lys Pro Leu Pro 685 690 695
act gtg ccc agc ctg gag gca gcc cag caa ctg acg agg gac act gtg 2345 Thr Val Pro Ser Leu Glu Ala Ala Gln Gln Leu Thr Arg Asp Thr Val 700 705 710
gaa gga gac ctg agc ctc ctg ggg ctg ctg gtc atg agg aac cta ctg 2393 Glu Gly Asp Leu Ser Leu Leu Gly Leu Leu Val Met Arg Asn Leu Leu 715 720 725
aag ccg cag aca acg cca gtt atc cag gct ctg cga agg acc cgc atc 2441 Lys Pro Gln Thr Thr Pro Val Ile Gln Ala Leu Arg Arg Thr Arg Ile 730 735 740
cgc gcc gtc atg gtg aca ggg gac aac ctg cag aca gcg gtg act gtg 2489 Arg Ala Val Met Val Thr Gly Asp Asn Leu Gln Thr Ala Val Thr Val 745 750 755 760
gcc cgg ggc tgt ggc atg gtg gcc ccc cag gag cat ctg atc atc gtc 2537 Ala Arg Gly Cys Gly Met Val Ala Pro Gln Glu His Leu Ile Ile Val 765 770 775
cac gcc acc cac cct gag cgg ggt cag cct gcc tct ctc gag ttc ctg 2585 His Ala Thr His Pro Glu Arg Gly Gln Pro Ala Ser Leu Glu Phe Leu 780 785 790
ccg atg gag tcc ccc aca gcc gtg aat ggc gtt aag gat cct gac cag 2633 Pro Met Glu Ser Pro Thr Ala Val Asn Gly Val Lys Asp Pro Asp Gln 795 800 805

gct gca agc tac acc gtg gag cca gac ccc cga tcc agg cac ctg gcc	2681
Ala Ala Ser Tyr Thr Val Glu Pro Asp Pro Arg Ser Arg His Leu Ala	
810 815 820	
ctc agc ggg ccc acc ttt ggt atc att gtg aag cac ttc ccc aag ctg	2729
Leu Ser Gly Pro Thr Phe Gly Ile Ile Val Lys His Phe Pro Lys Leu	
825 830 835 840	
ctg ccc aag gtc ctg gtc cag ggc act gtc ttt gcc cgc atg gcc cct	2777
Leu Pro Lys Val Leu Val Gln Gly Thr Val Phe Ala Arg Met Ala Pro	
845 850 855	
gag cag aag aca gag ctg gtg tgc gag cta cag aag ctt cag tac tgc	2825
Glu Gln Lys Thr Glu Leu Val Cys Glu Leu Gln Lys Leu Gln Tyr Cys	
860 865 870	
gtg ggc atg tgc gga gac ggt gcc aat gac tgt ggg gcc ctg aag gcg	2873
Val Gly Met Cys Gly Asp Gly Ala Asn Asp Cys Gly Ala Leu Lys Ala	
875 880 885	
gct gat gtc ggc atc tcg ctg tcc cag gca gaa gcc tca gtg gtc tca	2921
Ala Asp Val Gly Ile Ser Leu Ser Gln Ala Glu Ala Ser Val Val Ser	
890 895 900	
ccc ttc acc tcg agc atg gcc agt att gag tgc gtg ccc atg gtc atc	2969
Pro Phe Thr Ser Ser Met Ala Ser Ile Glu Cys Val Pro Met Val Ile	
905 910 915 920	
agg gag ggg cgc tgt tcc ctt gac act tcg ttc agc gtc ttc aag tac	3017
Arg Glu Gly Arg Cys Ser Leu Asp Thr Ser Phe Ser Val Phe Lys Tyr	
925 930 935	
atg gct ctg tac agc ctg acc cag ttc atc tcc gtc ctg atc ctc tac	3065
Met Ala Leu Tyr Ser Leu Thr Gln Phe Ile Ser Val Leu Ile Leu Tyr	
940 945 950	
acg atc aac acc aac ctg ggt gac ctg cag ttc ctg gcc atc gac ctg	3113
Thr Ile Asn Thr Asn Leu Gly Asp Leu Gln Phe Leu Ala Ile Asp Leu	
955 960 965	
gtc atc acc acc aca gtg gca gtg ctc atg agc cgc acg ggg cca gcg	3161
Val Ile Thr Thr Thr Val Ala Val Leu Met Ser Arg Thr Gly Pro Ala	
970 975 980	
ctg gtc ctg gga cgg gta cgg cca ccg ggg gcg ctg ctc agc gtg ccc	3209
Leu Val Leu Gly Arg Val Arg Pro Pro Gly Ala Leu Leu Ser Val Pro	
985 990 995 1000	
gtg ctc agc agc ctg ctg ctg cag atg gtc ctg gtg acc ggc gtg cag	3257
Val Leu Ser Ser Leu Leu Leu Gln Met Val Leu Val Thr Gly Val Gln	
1005 1010 1015	
cta ggg ggc tac ttc ctg acc ctg gcc cag cca tgg ttc gtg cct ctg	3305
Leu Gly Gly Tyr Phe Leu Thr Leu Ala Gln Pro Trp Phe Val Pro Leu	
1020 1025 1030	
aac agg aca gtg gcc gca cca gac aac ctg ccc aac tac gag aac acc	3353
Asn Arg Thr Val Ala Ala Pro Asp Asn Leu Pro Asn Tyr Glu Asn Thr	
1035 1040 1045	

```

gtg gtc ttc tct ctg tcc agc ttc cag tac ctc atc ctg gct gca gcc 3401
Val Val Phe Ser Leu Ser Ser Phe Gln Tyr Leu Ile Leu Ala Ala Ala
      1050                      1055                      1060

gtg tcc aag ggg gcg ccc ttc cgc cgg ccg ctc tac acc aat gtg ccc 3449
Val Ser Lys Gly Ala Pro Phe Arg Arg Pro Leu Tyr Thr Asn Val Pro
1065                      1070                      1075                      1080

ttc ctg gtg gcc ctg gcg ctc ctg agc tcc gtc ctg gtg ggc ctt gtc 3497
Phe Leu Val Ala Leu Ala Leu Leu Ser Ser Val Leu Val Gly Leu Val
      1085                      1090                      1095

ctg gtc ccc ggc ctc ctg cag ggg ccg ctg gcg ctg agg aac atc act 3545
Leu Val Pro Gly Leu Leu Gln Gly Pro Leu Ala Leu Arg Asn Ile Thr
      1100                      1105                      1110

gac acc ggc ttc aag ctg ctg ctg ctg ggt ctg gtc acc ctc aac ttc 3593
Asp Thr Gly Phe Lys Leu Leu Leu Leu Gly Leu Val Thr Leu Asn Phe
      1115                      1120                      1125

gtg ggg gcc ttc atg ctg gag agc gtg cta gac cag tgc ctc ccc gcc 3641
Val Gly Ala Phe Met Leu Glu Ser Val Leu Asp Gln Cys Leu Pro Ala
      1130                      1135                      1140

tgc ctg cgc cgc ctc cgg ccc aag cgg gcc tcc aag aag cgc ttc aag 3689
Cys Leu Arg Arg Leu Arg Pro Lys Arg Ala Ser Lys Lys Arg Phe Lys
1145                      1150                      1155                      1160

cag ctg gaa cga gag ctg gcc gag cag ccc tgg cca ccg ctg ccc gcc 3737
Gln Leu Glu Arg Glu Leu Ala Glu Gln Pro Trp Pro Pro Leu Pro Ala
      1165                      1170                      1175

ggc ccc ctg agg tagtgcaggc ccacgggcac cccagacact ggaactccct 3789
Gly Pro Leu Arg
      1180

gcctctgagc caccaactgg acccctctcc agcaacacca ccgccaccac ctcccacatc 3849

cctgagggttg gcgactgtct aactcctcc cccgagacca cccccaccct ggggaagcgt 3909

tgactactgt cccctacctt ggaccatccc gcgtaggggt ggcagcccc agctcccctc 3969

agtgtgtgtg tcaagtgtagc aaataaagtc atgatatttt cctggcaaaa aaaaaaaaaa 4029
aaaaaaaaaa aaaaaaaaaa aaaaaa
4055

<210> 2
<211> 1180
<212> PRT
<213> Homo sapiens

<400> 2
Met Ser Ala Asp Ser Ser Pro Leu Val Gly Ser Thr Pro Thr Gly Tyr
  1          5          10          15

Gly Thr Leu Thr Ile Gly Thr Ser Ile Asp Pro Leu Ser Ser Ser Val
      20          25          30

Ser Ser Val Arg Leu Ser Gly Tyr Cys Gly Ser Pro Trp Arg Val Ile
      35          40          45

```

- 7 -

Gly Tyr His Val Val Val Trp Met Met Ala Gly Ile Pro Leu Leu Leu
 50 55 60
 Phe Arg Trp Lys Pro Leu Trp Gly Val Arg Leu Arg Leu Arg Pro Cys
 65 70 75 80
 Asn Leu Ala His Ala Glu Thr Leu Val Ile Glu Ile Arg Asp Lys Glu
 85 90 95
 Asp Ser Ser Trp Gln Leu Phe Thr Val Gln Val Gln Thr Glu Ala Ile
 100 105 110
 Gly Glu Gly Ser Leu Glu Pro Ser Pro Gln Ser Gln Ala Glu Asp Gly
 115 120 125
 Arg Ser Gln Ala Ala Val Gly Ala Val Pro Glu Gly Ala Trp Lys Asp
 130 135 140
 Thr Ala Gln Leu His Lys Ser Glu Glu Ala Val Ser Val Gly Gln Lys
 145 150 155 160
 Arg Val Leu Arg Tyr Tyr Leu Phe Gln Gly Gln Arg Tyr Ile Trp Ile
 165 170 175
 Glu Thr Gln Gln Ala Phe Tyr Gln Val Ser Leu Leu Asp His Gly Arg
 180 185 190
 Ser Cys Asp Asp Val His Arg Ser Arg His Gly Leu Ser Leu Gln Asp
 195 200 205
 Gln Met Val Arg Lys Ala Ile Tyr Gly Pro Asn Val Ile Ser Ile Pro
 210 215 220
 Val Lys Ser Tyr Pro Gln Leu Leu Val Asp Glu Ala Leu Asn Pro Tyr
 225 230 235 240
 Tyr Gly Phe Gln Ala Phe Ser Ile Ala Leu Trp Leu Ala Asp His Tyr
 245 250 255
 Tyr Trp Tyr Ala Leu Cys Ile Phe Leu Ile Ser Ser Ile Ser Ile Cys
 260 265 270
 Leu Ser Leu Tyr Lys Thr Arg Lys Gln Ser Gln Thr Leu Arg Asp Met
 275 280 285
 Val Lys Leu Ser Met Arg Val Cys Val Cys Arg Pro Gly Gly Glu Glu
 290 295 300
 Glu Trp Val Asp Ser Ser Glu Leu Val Pro Gly Asp Cys Leu Val Leu
 305 310 315 320
 Pro Gln Glu Gly Gly Leu Met Pro Cys Asp Ala Ala Leu Val Ala Gly
 325 330 335
 Glu Cys Met Val Asn Glu Ser Ser Leu Thr Gly Glu Ser Ile Pro Val
 340 345 350
 Leu Lys Thr Ala Leu Pro Glu Gly Leu Gly Pro Tyr Cys Ala Glu Thr
 355 360 365

His	Arg	Arg	His	Thr	Leu	Phe	Cys	Gly	Thr	Leu	Ile	Leu	Gln	Ala	Arg		
370						375					380						
Ala	Tyr	Val	Gly	Pro	His	Val	Leu	Ala	Val	Val	Thr	Arg	Thr	Gly	Phe		
385					390					395					400		
Cys	Thr	Ala	Lys	Gly	Gly	Leu	Val	Ser	Ser	Ile	Leu	His	Pro	Arg	Pro		
				405					410					415			
Ile	Asn	Phe	Lys	Phe	Tyr	Lys	His	Ser	Met	Lys	Phe	Val	Ala	Ala	Leu		
			420					425						430			
Ser	Val	Leu	Ala	Leu	Leu	Gly	Thr	Ile	Tyr	Ser	Ile	Phe	Ile	Leu	Tyr		
		435					440						445				
Arg	Asn	Arg	Val	Pro	Leu	Asn	Glu	Ile	Val	Ile	Arg	Ala	Leu	Asp	Leu		
	450					455					460						
Val	Thr	Val	Val	Val	Pro	Pro	Ala	Leu	Pro	Ala	Ala	Met	Thr	Val	Cys		
465					470					475					480		
Thr	Leu	Tyr	Ala	Gln	Ser	Arg	Leu	Arg	Arg	Gln	Gly	Ile	Phe	Cys	Ile		
				485					490					495			
His	Pro	Leu	Arg	Ile	Asn	Leu	Gly	Gly	Lys	Leu	Gln	Leu	Val	Cys	Phe		
			500					505						510			
Asp	Lys	Thr	Gly	Thr	Leu	Thr	Glu	Asp	Gly	Leu	Asp	Val	Met	Gly	Val		
		515					520						525				
Val	Pro	Leu	Lys	Gly	Gln	Ala	Phe	Leu	Pro	Leu	Val	Pro	Glu	Pro	Arg		
	530					535						540					
Arg	Leu	Pro	Val	Gly	Pro	Leu	Leu	Arg	Ala	Leu	Ala	Thr	Cys	His	Ala		
545					550					555					560		
Leu	Ser	Arg	Leu	Gln	Asp	Thr	Pro	Val	Gly	Asp	Pro	Met	Asp	Leu	Lys		
				565					570					575			
Met	Val	Glu	Ser	Thr	Gly	Trp	Val	Leu	Glu	Glu	Glu	Pro	Ala	Ala	Asp		
			580					585						590			
Ser	Ala	Phe	Gly	Thr	Gln	Val	Leu	Ala	Val	Met	Arg	Pro	Pro	Leu	Trp		
		595					600						605				
Glu	Pro	Gln	Leu	Gln	Ala	Met	Glu	Glu	Pro	Pro	Val	Pro	Val	Ser	Val		
	610					615						620					
Leu	His	Arg	Phe	Pro	Phe	Ser	Ser	Ala	Leu	Gln	Arg	Met	Ser	Val	Val		
625					630					635					640		
Val	Ala	Trp	Pro	Gly	Ala	Thr	Gln	Pro	Glu	Ala	Tyr	Val	Lys	Gly	Ser		
				645					650					655			
Pro	Glu	Leu	Val	Ala	Gly	Leu	Cys	Asn	Pro	Glu	Thr	Val	Pro	Thr	Asp		
			660					665						670			
Phe	Ala	Gln	Met	Leu	Gln	Ser	Tyr	Thr	Ala	Ala	Gly	Tyr	Arg	Val	Val		
		675					680					685					

Ala Leu Ala Ser Lys Pro Leu Pro Thr Val Pro Ser Leu Glu Ala Ala
 690 695 700
 Gln Gln Leu Thr Arg Asp Thr Val Glu Gly Asp Leu Ser Leu Leu Gly
 705 710 715 720
 Leu Leu Val Met Arg Asn Leu Leu Lys Pro Gln Thr Thr Pro Val Ile
 725 730 735
 Gln Ala Leu Arg Arg Thr Arg Ile Arg Ala Val Met Val Thr Gly Asp
 740 745 750
 Asn Leu Gln Thr Ala Val Thr Val Ala Arg Gly Cys Gly Met Val Ala
 755 760 765
 Pro Gln Glu His Leu Ile Ile Val His Ala Thr His Pro Glu Arg Gly
 770 775 780
 Gln Pro Ala Ser Leu Glu Phe Leu Pro Met Glu Ser Pro Thr Ala Val
 785 790 795 800
 Asn Gly Val Lys Asp Pro Asp Gln Ala Ala Ser Tyr Thr Val Glu Pro
 805 810 815
 Asp Pro Arg Ser Arg His Leu Ala Leu Ser Gly Pro Thr Phe Gly Ile
 820 825 830
 Ile Val Lys His Phe Pro Lys Leu Leu Pro Lys Val Leu Val Gln Gly
 835 840 845
 Thr Val Phe Ala Arg Met Ala Pro Glu Gln Lys Thr Glu Leu Val Cys
 850 855 860
 Glu Leu Gln Lys Leu Gln Tyr Cys Val Gly Met Cys Gly Asp Gly Ala
 865 870 875 880
 Asn Asp Cys Gly Ala Leu Lys Ala Ala Asp Val Gly Ile Ser Leu Ser
 885 890 895
 Gln Ala Glu Ala Ser Val Val Ser Pro Phe Thr Ser Ser Met Ala Ser
 900 905 910
 Ile Glu Cys Val Pro Met Val Ile Arg Glu Gly Arg Cys Ser Leu Asp
 915 920 925
 Thr Ser Phe Ser Val Phe Lys Tyr Met Ala Leu Tyr Ser Leu Thr Gln
 930 935 940
 Phe Ile Ser Val Leu Ile Leu Tyr Thr Ile Asn Thr Asn Leu Gly Asp
 945 950 955 960
 Leu Gln Phe Leu Ala Ile Asp Leu Val Ile Thr Thr Thr Val Ala Val
 965 970 975
 Leu Met Ser Arg Thr Gly Pro Ala Leu Val Leu Gly Arg Val Arg Pro
 980 985 990
 Pro Gly Ala Leu Leu Ser Val Pro Val Leu Ser Ser Leu Leu Leu Gln
 995 1000 1005

Met Val Leu Val Thr Gly Val Gln Leu Gly Gly Tyr Phe Leu Thr Leu
 1010 1015 1020

Ala Gln Pro Trp Phe Val Pro Leu Asn Arg Thr Val Ala Ala Pro Asp
 1025 1030 1035 1040

Asn Leu Pro Asn Tyr Glu Asn Thr Val Val Phe Ser Leu Ser Ser Phe
 1045 1050 1055

Gln Tyr Leu Ile Leu Ala Ala Ala Val Ser Lys Gly Ala Pro Phe Arg
 1060 1065 1070

Arg Pro Leu Tyr Thr Asn Val Pro Phe Leu Val Ala Leu Ala Leu Leu
 1075 1080 1085

Ser Ser Val Leu Val Gly Leu Val Leu Val Pro Gly Leu Leu Gln Gly
 1090 1095 1100

Pro Leu Ala Leu Arg Asn Ile Thr Asp Thr Gly Phe Lys Leu Leu Leu
 1105 1110 1115 1120

Leu Gly Leu Val Thr Leu Asn Phe Val Gly Ala Phe Met Leu Glu Ser
 1125 1130 1135

Val Leu Asp Gln Cys Leu Pro Ala Cys Leu Arg Arg Leu Arg Pro Lys
 1140 1145 1150

Arg Ala Ser Lys Lys Arg Phe Lys Gln Leu Glu Arg Glu Leu Ala Glu
 1155 1160 1165

Gln Pro Trp Pro Pro Leu Pro Ala Gly Pro Leu Arg
 1170 1175 1180

<210> 3

<211> 3540

<212> DNA

<213> Homo sapiens

<400> 3

atgagcgcag acagcagccc tctcgtgggc agcacgccca ccggttatgg gaccctgacg 60

ataggacat caatagatcc cctcagctcc tcagtttcat ccgtgaggct cagcggctac 120

tgtggcagtc catggagggt catcggctat cacgtcgtgg tctggatgat ggctgggata 180

cctttgctgc tcttccgttg gaagcccctg tgggggggtgc ggctgcggct ccggccctgc 240

aacctggccc acgccgaaac actcgttatc gaaataagag acaaagagga tagttccttg 300

cagctcttca ctgtccaggt gcagactgag gccatcggcg agggcagcct ggagccgtcc 360

ccacagtccc aggcagagga tggccggagc caggcggcag ttggggcggt accagagggt 420

gcctggaagg atacggccca gctccacaag agcgaggagg cggtgagtgt cggacagaag 480

cggtgtctgc ggtattacct cttccagggc cagcgtata tctggatcga gaccagcaa 540

gccttctacc aggtcagcct cctggaccat ggccgctctt gtgacgacgt ccaccgctcc 600

cgccatggcc tcagcctcca ggaccaaagt gtgaggaagg ccatttacgg cccaacgtg 660

atcagcatatc cggtcaagtc ctacccccag ctgctgggtg acgaggcact gaacccctac 720
tatgggttcc aggccttcag catcgcgctg tggctggctg accactacta ctggtacgcc 780
ctgtgcatct tcttcatttc ctccatctcc atctgcctgt cgctgtacaa gaccagaaag 840
caaagccaga ctctaaggga catggtcaag ttgtccatgc ggggtgtcgt gtgccggcca 900
gggggagagg aagagtgggt ggactccagt gagctagtgc ccggagactg cctgggtgctg 960
ccccaggagg gtgggctgat gccctgtgat gccgccctgg tggccggcga gtgcatgggtg 1020
aatgagagct ctctgacagg agagagcatt ccagtgtga agacggcact gccggagggg 1080
ctggggccct actgtgcaga gacacaccgg cggcacacac tcttctgcgg gaccctcatc 1140
ttgcaggccc gggcctatgt gggaccgcac gtcttgccag tggtgaccg cacagggttc 1200
tgcacggcaa aagggggcct ggtgagctcc atcttgacc ccgggccat caacttcaag 1260
ttctataaac acagcatgaa gtttgtggct gccctctctg tcttggtct cctcggcacc 1320
atctacagca tcttcactct ctaccgaaac cgggtgcctc tgaatgagat tgtaatccgg 1380
gctctcgacc tggtgaccgt ggtgggtgcca cctgccctgc ctgctgcat gactgtgtgc 1440
acgtctatcg ccagagccg actgcggaga cagggcattt tctgcatcca cccactgcgc 1500
atcaacctgg ggggcaagct gcagctggtg tgtttcgaca agacgggcac cctcactgag 1560
gacggcttag acgtgatggg ggtgggtgcc ctgaagggg aggcattcct gcccttggtc 1620
ccagagcctc gccgcctgcc tgtggggccc ctgctccgag cactggccac ctgccatgcc 1680
ctcagccggc tccaggacac ccccggtggg gaccccatgg acttgaagat ggtggagtct 1740
actggctggg tcttgaggga agagccggt gcagactcag catttgggac ccaggctctt 1800
gcagtgatga gacctccact ttgggagccc cagctgcagg caatggagga gccccgggtg 1860
ccagtcagcg tctccaccg cttcccttc tcttcggctc tgcagcgcat gagtgtggtg 1920
gtggcgtggc cagggggcac tcagcccgag gcctacgtca aaggctcccc ggagctggtg 1980
gcagggctct gcaaccccg gacagtgcc accgacttcg ccagatgct gcagagctat 2040
acagctgctg gctaccgtgt cgtggccctg gccagcaagc cactgcccac tgtgccagc 2100
ctggaggcag ccagcaact gacgaggac actgtggaag gagacctgag cctcctggg 2160
ctgctggtca tgaggaacct actgaagccg cagacaacgc cagttatcca ggctctgcga 2220
aggacccgca tccgcgccgt catggtgaca ggggacaacc tgcagacagc ggtgactgtg 2280
gcccggggct gtggcatggt gggccccag gagcatctga tcatcgtcca cgccaccac 2340
cctgagcggg gtcagcctgc ctctctcgag ttctgcca tggagtcccc cacagccgtg 2400
aatggcggtta aggatcctga ccaggctgca agctacaccg tggagccaga ccccgatcc 2460

- 12 -

aggcacctgg ccctcagcgg gccacacctt ggtatcattg tgaagcactt cccaagctg 2520
 ctgccaaggg tcctggtcca gggcactgtc tttgcccgca tggcccctga gcagaagaca 2580
 gagctggtgt gcgagctaca gaagcttcag tactgcgtgg gcatgtgcgg agacgggtgcc 2640
 aatgactgtg gggccctgaa ggcggctgat gtcggcatct cgtgttccca ggcagaagcc 2700
 tcagtgggtct cacccttcac ctcgagcatg gccagtattg agtgcgtgcc catggtcatc 2760
 agggagggggc gctgttccct tgacacttcg ttcagcgtct tcaagtacat ggctctgtac 2820
 agcctgaccc agttcatctc cgtcctgatc ctctacacga tcaacaccaa cctgggtgac 2880
 ctgcagttcc tggccatcga cctggtcatc accaccacag tggcagtgtc catgagccgc 2940
 acggggccag cgtggttctt gggacgggta cggccaccgg gggcgctgct cagcgtgccc 3000
 gtgctcagca gcctgctgct gcagatggtc ctggtgaccg gcgtgcagct agggggctac 3060
 ttctgaccc tggcccagcc atggttcgtg cctctgaaca ggacagtggc cgcaccagac 3120
 aacctgccc actacgagaa caccgtggtc ttctctctgt ccagcttcca gtacctcatc 3180
 ctggctgcag ccgtgtccaa gggggcgccc ttccgcccgc cgtctacac caatgtgccc 3240
 ttctggtgg ccctggcgct cctgagctcc gtcctggtgg gccttgctct ggtccccggc 3300
 ctctgcagg ggccgctggc gctgaggaac atcactgaca ccggcttcaa gctgctgctg 3360
 ctgggtctgg tcacctcaa ctctgtggg gccttcatgc tggagagcgt gctagaccag 3420
 tgccctcccg cctgcctgcg ccgcctccgg cccaagcggg cctccaagaa gcgcttcaag 3480
 cagctggaac gagagctggc cgagcagccc tggccaccgc tgcccgccgg cccctgagg 3540

<210> 4

<211> 1187

<212> PRT

<213> *Caenorhabditis elegans*

<400> 4

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

Glu Leu Arg Trp Phe Val Tyr Arg Lys Leu Glu Tyr Val Trp Ile Asp
 100 105 110
 Asp Leu Asn Ser Asp Glu Ser Val Asp Glu Ile Ser Asp Asn Asp Asn
 115 120 125
 Cys Trp Lys Thr Ser Phe Glu Ile Ala Asn Arg Ile Pro Cys Arg Ser
 130 135 140
 Leu Leu Ala Val Ser Glu Ser Asn Phe Gly Leu Thr Leu Ser Glu Ile
 145 150 155 160
 Ser Arg Arg Leu Glu Phe Tyr Gly Arg Asn Glu Ile Val Val Gln Leu
 165 170 175
 Arg Pro Ile Leu Tyr Leu Leu Val Met Glu Val Ile Thr Pro Phe Tyr
 180 185 190
 Val Phe Gln Ile Phe Ser Val Thr Val Trp Tyr Asn Asp Glu Tyr Ala
 195 200 205
 Tyr Tyr Ala Ser Leu Ile Val Ile Leu Ser Leu Gly Ser Ile Val Met
 210 215 220
 Asp Val Tyr Gln Ile Arg Thr Gln Glu Ile Arg Leu Arg Ser Met Val
 225 230 235 240
 His Ser Thr Glu Ser Val Glu Val Ile Arg Glu Gly Thr Glu Met Thr
 245 250 255
 Ile Gly Ser Asp Gln Leu Val Pro Gly Asp Ile Leu Leu Ile Pro Pro
 260 265 270
 His Gly Cys Leu Met Gln Cys Asp Ser Val Leu Met Asn Gly Thr Val
 275 280 285
 Ile Val Asn Glu Ser Val Leu Thr Gly Glu Ser Val Pro Ile Thr Lys
 290 295 300
 Val Ala Leu Thr Asp Glu Thr Asn Asp Ser Val Phe Asn Ile Glu Lys
 305 310 315 320
 Asn Ser Lys Asn Val Leu Phe Cys Gly Thr Gln Val Leu Gln Thr Arg
 325 330 335
 Phe Tyr Arg Gly Lys Lys Val Lys Ala Ile Val Leu Arg Thr Ala Tyr
 340 345 350
 Ser Thr Leu Lys Gly Gln Leu Val Arg Ser Ile Met Tyr Pro Lys Pro
 355 360 365
 Val Asp Phe Arg Phe Thr Lys Asp Leu Phe Lys Phe Ile Leu Phe Leu
 370 375 380
 Ala Cys Ile Ser Gly Cys Gly Phe Ile Tyr Thr Ile Ile Val Met Ile
 385 390 395 400
 Met Arg Gly Asn Thr Leu Arg Arg Ile Ile Val Arg Ser Leu Asp Ile
 405 410 415

- 14 -

Ile Thr Ile Thr Val Pro Pro Ala Leu Pro Ala Ala Met Ser Val Gly
 420 425 430
 Ile Ile Asn Ala Gln Leu Arg Leu Lys Lys Lys Glu Ile Phe Cys Ile
 435 440 445
 Ser Pro Ser Thr Ile Asn Thr Cys Gly Ala Ile Asn Val Val Cys Phe
 450 455 460
 Asp Lys Thr Gly Thr Leu Thr Glu Asp Gly Leu Asp Phe His Val Val
 465 470 475 480
 Arg Pro Val Met Ser Ala Val Asn Gln Glu Ile Gln Lys Val Lys Leu
 485 490 495
 Glu Lys Ser Asn Arg Thr Glu Phe Met Gly Glu Met Thr Glu Leu Thr
 500 505 510
 Ser Arg Asn Gly Leu Pro Phe Asp Gly Asp Leu Val Lys Ala Ile Ala
 515 520 525
 Thr Cys His Ser Leu Thr Arg Ile Asn Gly Val Leu His Gly Asp Pro
 530 535 540
 Leu Asp Leu Ile Leu Phe Gln Lys Thr Gly Trp Thr Met Glu Glu Gly
 545 550 555 560
 Ile Glu Gly Asp Ile Glu Glu Glu Thr Gln Arg Phe Asp Asn Val Gln
 565 570 575
 Pro Ser Ile Ile Lys Pro Thr Asp Asp Lys Ser Ala Glu Tyr Ser Val
 580 585 590
 Ile Arg Gln Phe Thr Phe Ser Ser Ser Leu Gln Arg Met Ser Val Ile
 595 600 605
 Val Phe Asp Pro Arg Glu Asp Arg Pro Asp Asn Met Met Leu Tyr Ser
 610 615 620
 Lys Gly Ser Pro Glu Met Ile Leu Ser Leu Cys Asp Pro Asn Thr Val
 625 630 635 640
 Pro Glu Asp Tyr Leu Leu Gln Val Asn Ser Tyr Ala Gln His Gly Phe
 645 650 655
 Arg Leu Ile Ala Val Ala Arg Arg Pro Leu Asp Leu Asn Phe Asn Lys
 660 665 670
 Ala Ser Lys Val Lys Arg Asp Ala Val Glu Cys Asp Leu Glu Met Leu
 675 680 685
 Gly Leu Ile Val Met Glu Asn Arg Val Lys Pro Val Thr Leu Gly Val
 690 695 700
 Ile Asn Gln Leu Asn Arg Ala Asn Ile Arg Thr Val Met Val Thr Gly
 705 710 715 720
 Asp Asn Leu Leu Thr Gly Leu Ser Val Ala Arg Glu Cys Gly Ile Ile
 725 730 735

- 15 -

Arg	Pro	Ser	Lys	Arg	Ala	Phe	Leu	Val	Glu	His	Val	Pro	Gly	Glu	Leu	740	745	750
Asp	Glu	Tyr	Gly	Arg	Thr	Lys	Ile	Phe	Val	Lys	Gln	Ser	Val	Ser	Ser	755	760	765
Ser	Asp	Glu	Val	Ile	Glu	Asp	Asp	Ala	Ser	Val	Ser	Ile	Ser	Met	Cys	770	775	780
Ser	Ser	Thr	Trp	Lys	Gly	Ser	Ser	Glu	Gly	Asp	Gly	Phe	Ser	Pro	Thr	785	790	795
Asn	Thr	Glu	Val	Glu	Thr	Pro	Asn	Pro	Val	Thr	Ala	Asp	Ser	Leu	Gly	805	810	815
His	Leu	Ile	Ala	Ser	Ser	Tyr	His	Leu	Ala	Ile	Ser	Gly	Pro	Thr	Phe	820	825	830
Ala	Val	Ile	Val	His	Glu	Tyr	Pro	Glu	Leu	Val	Asp	Gln	Leu	Cys	Ser	835	840	845
Val	Cys	Asp	Val	Phe	Ala	Arg	Met	Ala	Pro	Asp	Gln	Lys	Gln	Ser	Leu	850	855	860
Val	Glu	Gln	Leu	Gln	Gln	Ile	Asp	Tyr	Thr	Val	Ala	Met	Cys	Gly	Asp	865	870	875
Gly	Ala	Asn	Asp	Cys	Ala	Ala	Leu	Lys	Ala	Ala	His	Ala	Gly	Ile	Ser	885	890	895
Leu	Ser	Asp	Ala	Glu	Ala	Ser	Ile	Ala	Ala	Pro	Phe	Thr	Ser	Lys	Val	900	905	910
Pro	Asp	Ile	Arg	Cys	Val	Pro	Thr	Val	Ile	Ser	Glu	Gly	Arg	Ala	Ala	915	920	925
Leu	Val	Thr	Ser	Phe	Gly	Ile	Phe	Lys	Tyr	Met	Ala	Gly	Tyr	Ser	Leu	930	935	940
Thr	Gln	Phe	Val	Thr	Val	Met	His	Leu	Tyr	Trp	Ile	Ser	Asn	Ile	Leu	945	950	955
Thr	Asp	Gly	Gln	Phe	Met	Tyr	Ile	Asp	Met	Phe	Leu	Ile	Thr	Met	Phe	965	970	975
Ala	Leu	Leu	Phe	Gly	Asn	Thr	Pro	Ala	Phe	Tyr	Arg	Leu	Ala	His	Thr	980	985	990
Pro	Pro	Pro	Thr	Arg	Leu	Leu	Ser	Ile	Ala	Ser	Met	Thr	Ser	Val	Val	995	1000	1005
Gly	Gln	Leu	Ile	Ile	Ile	Gly	Val	Val	Gln	Phe	Ile	Val	Phe	Phe	Ser	1010	1015	1020
Thr	Ser	Gln	Gln	Pro	Trp	Phe	Thr	Pro	Tyr	Gln	Pro	Pro	Val	Asp	Asp	1025	1030	1035
Glu	Val	Glu	Asp	Lys	Arg	Ser	Met	Gln	Gly	Thr	Ala	Leu	Phe	Cys	Val	1045	1050	1055

- 16 -

Ser Met Phe Gln Tyr Ile Ile Leu Ala Leu Val Tyr Ser Lys Gly Pro
 1060 1065 1070

Pro Phe Arg Gly Asn Leu Trp Ser Asn Lys Pro Ile Tyr Lys Lys Lys
 1075 1080 1085

Arg Ser Ile Glu Ala Ile Ile Asp Tyr Val Pro Thr Thr Asn Ser Asp
 1090 1095 1100

His Ile Arg Arg Pro Ser Ile Asn Gly Val Thr Ser Ser Arg Thr Glu
 1105 1110 1115 1120

Ser Thr Leu Leu Ser Ala Glu Gly Gln Gln Leu His Met Thr Thr Ser
 1125 1130 1135

Lys Asn Gly Lys Gly Gly Glu Asn Pro His Ser Ser Ala Leu Phe Glu
 1140 1145 1150

Arg Leu Ile Ser Arg Ile Gly Gly Glu Pro Thr Trp Leu Thr Asn Pro
 1155 1160 1165

Ile Pro Pro His Ser Leu Ser Glu Pro Glu Glu Pro Glu Lys Leu Glu
 1170 1175 1180

Arg Thr Tyr
 1185

<210> 5
 <211> 7249
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (225)..(3995)

<400> 5
 cacgcgtccg ggtctgggctg aggcgaggcg gcggcggcga cagcggcggc cgggtcccc 60
 gcggccccctg gggctggtcc ggccgcgagg gaggccgcgg aggaggcggc gcggcggcgg 120
 ccagtgcgacg gccccgatct gacagacatc cctgaatctt ggtgtttgga cataggagtg 180
 atcttccatt gtttgaagca ctggaccttt aatccactgt aggt atg gac agg gaa 236
 Met Asp Arg Glu
 1

gaa agg aag acc atc aat cag ggt caa gaa gat gaa atg gag att tat 284
 Glu Arg Lys Thr Ile Asn Gln Gly Gln Glu Asp Glu Met Glu Ile Tyr
 5 10 15 20

ggt tac aat ttg agt cgc tgg aag ctt gcc ata gtt tct tta gga gtg 332
 Gly Tyr Asn Leu Ser Arg Trp Lys Leu Ala Ile Val Ser Leu Gly Val
 25 30 35

att tgc tct gat ggg ttt ctc ctc ctc ctc ctc tat tgg atg cct gag 380
 Ile Cys Ser Asp Gly Phe Leu Leu Leu Leu Tyr Trp Met Pro Glu
 40 45 50

tgg	cgg	gtg	aaa	gcg	acc	tgt	gtc	aga	gct	gca	att	aaa	gac	tgt	gaa	428
Trp	Arg	Val	Lys	Ala	Thr	Cys	Val	Arg	Ala	Ala	Ile	Lys	Asp	Cys	Glu	
		55					60					65				
gta	gtg	ctg	ctg	agg	act	act	gat	gaa	ttc	aaa	atg	tgg	ttt	tgt	gca	476
Val	Val	Leu	Leu	Arg	Thr	Thr	Asp	Glu	Phe	Lys	Met	Trp	Phe	Cys	Ala	
		70				75					80					
aaa	att	cgc	gtt	ctt	tct	ttg	gaa	act	tac	cca	gtt	tca	agt	cca	aaa	524
Lys	Ile	Arg	Val	Leu	Ser	Leu	Glu	Thr	Tyr	Pro	Val	Ser	Ser	Pro	Lys	
	85				90					95					100	
tct	atg	tct	aat	aag	ctt	tca	aat	ggc	cat	gca	gtt	tgt	tta	att	gag	572
Ser	Met	Ser	Asn	Lys	Leu	Ser	Asn	Gly	His	Ala	Val	Cys	Leu	Ile	Glu	
				105					110					115		
aat	ccc	act	gaa	gaa	aat	agg	cac	agg	atc	agt	aaa	tat	tca	cag	act	620
Asn	Pro	Thr	Glu	Glu	Asn	Arg	His	Arg	Ile	Ser	Lys	Tyr	Ser	Gln	Thr	
			120					125					130			
gaa	tca	caa	cag	att	cgt	tat	ttc	acc	cac	cat	agt	gta	aaa	tat	ttc	668
Glu	Ser	Gln	Gln	Ile	Arg	Tyr	Phe	Thr	His	His	Ser	Val	Lys	Tyr	Phe	
		135					140					145				
tgg	aat	gat	acc	att	cac	aat	ttt	gat	ttc	tta	aag	gga	ctg	gat	gaa	716
Trp	Asn	Asp	Thr	Ile	His	Asn	Phe	Asp	Phe	Leu	Lys	Gly	Leu	Asp	Glu	
	150					155					160					
ggg	gtt	tct	tgt	acg	tca	att	tat	gaa	aag	cat	agt	gca	gga	ctg	aca	764
Gly	Val	Ser	Cys	Thr	Ser	Ile	Tyr	Glu	Lys	His	Ser	Ala	Gly	Leu	Thr	
	165				170					175					180	
aag	ggg	atg	cat	gcc	tac	aga	aaa	ctg	ctt	tat	gga	gta	aat	gaa	att	812
Lys	Gly	Met	His	Ala	Tyr	Arg	Lys	Leu	Leu	Tyr	Gly	Val	Asn	Glu	Ile	
				185				190						195		
gct	gta	aaa	gtg	cct	tct	gtt	ttt	aag	ctt	cta	att	aaa	gag	gtt	ctc	860
Ala	Val	Lys	Val	Pro	Ser	Val	Phe	Lys	Leu	Leu	Ile	Lys	Glu	Val	Leu	
			200					205					210			
aac	cca	ttt	tac	att	ttc	cag	ctg	ttc	agt	gtt	ata	ctg	tgg	agc	act	908
Asn	Pro	Phe	Tyr	Ile	Phe	Gln	Leu	Phe	Ser	Val	Ile	Leu	Trp	Ser	Thr	
		215				220						225				
gat	gaa	tac	tat	tac	tat	gct	cta	gct	att	gtg	gtt	atg	tcc	ata	gta	956
Asp	Glu	Tyr	Tyr	Tyr	Tyr	Ala	Leu	Ala	Ile	Val	Val	Met	Ser	Ile	Val	
		230				235					240					
tca	atc	gta	agc	tca	cta	tat	tcc	att	aga	aag	caa	tat	gtt	atg	ttg	1004
Ser	Ile	Val	Ser	Ser	Leu	Tyr	Ser	Ile	Arg	Lys	Gln	Tyr	Val	Met	Leu	
		245			250					255					260	
cat	gac	atg	gtg	gca	act	cat	agt	acc	gta	aga	gtt	tca	gtt	tgt	aga	1052
His	Asp	Met	Val	Ala	Thr	His	Ser	Thr	Val	Arg	Val	Ser	Val	Cys	Arg	
				265					270					275		
gta	aat	gaa	gaa	ata	gaa	gaa	atc	ttt	tct	acc	gac	ctt	gtg	cca	gga	1100
Val	Asn	Glu	Glu	Ile	Glu	Glu	Ile	Phe	Ser	Thr	Asp	Leu	Val	Pro	Gly	
			280					285					290			

- 18 -

gat gtc atg gtc att cca tta aat ggg aca ata atg cct tgt gat gct	1148
Asp Val Met Val Ile Pro Leu Asn Gly Thr Ile Met Pro Cys Asp Ala	
295 300 305	
gtg ctt att aat ggt acc tgc att gta aac gaa agc atg tta aca gga	1196
Val Leu Ile Asn Gly Thr Cys Ile Val Asn Glu Ser Met Leu Thr Gly	
310 315 320	
gaa agt gtt cca gtg aca aag act aat ttg cca aat cct tca gtg gat	1244
Glu Ser Val Pro Val Thr Lys Thr Asn Leu Pro Asn Pro Ser Val Asp	
325 330 335 340	
gtg aaa gga ata gga gat gaa tta tat aat cca gaa aca cat aaa cga	1292
Val Lys Gly Ile Gly Asp Glu Leu Tyr Asn Pro Glu Thr His Lys Arg	
345 350 355	
cat act ttg ttt tgt ggg aca act gtt att cag act cgt ttc tac act	1340
His Thr Leu Phe Cys Gly Thr Thr Val Ile Gln Thr Arg Phe Tyr Thr	
360 365 370	
gga gaa ctc gtc aaa gcc ata gtt gtt aga aca gga ttt agt act tcc	1388
Gly Glu Leu Val Lys Ala Ile Val Val Arg Thr Gly Phe Ser Thr Ser	
375 380 385	
aaa gga cag ctt gtt cgt tcc ata ttg tat ccc aaa cca act gat ttt	1436
Lys Gly Gln Leu Val Arg Ser Ile Leu Tyr Pro Lys Pro Thr Asp Phe	
390 395 400	
aaa ctc tac aga gat gcc tac ttg ttt cta cta tgt ctt gtg gca gtt	1484
Lys Leu Tyr Arg Asp Ala Tyr Leu Phe Leu Cys Leu Val Ala Val	
405 410 415 420	
gct ggc att ggg ttt atc tac act att att aat agc att tta aat gag	1532
Ala Gly Ile Gly Phe Ile Tyr Thr Ile Ile Asn Ser Ile Leu Asn Glu	
425 430 435	
gta caa gtt ggg gtc ata att atc gag tct ctt gat att atc aca att	1580
Val Gln Val Gly Val Ile Ile Ile Glu Ser Leu Asp Ile Ile Thr Ile	
440 445 450	
act gtg ccc cct gca ctt cct gct gca atg act gct ggt att gtg tat	1628
Thr Val Pro Pro Ala Leu Pro Ala Ala Met Thr Ala Gly Ile Val Tyr	
455 460 465	
gct cag aga aga ctg aaa aaa atc ggt att ttc tgt atc agt cct caa	1676
Ala Gln Arg Arg Leu Lys Lys Ile Gly Ile Phe Cys Ile Ser Pro Gln	
470 475 480	
aga ata aat att tgt gga cag ctc aat ctt gtt tgc ttt gac aag act	1724
Arg Ile Asn Ile Cys Gly Gln Leu Asn Leu Val Cys Phe Asp Lys Thr	
485 490 495 500	
gga act cta act gaa gat ggt tta gat ctt tgg ggg att caa cga gtg	1772
Gly Thr Leu Thr Glu Asp Gly Leu Asp Leu Trp Gly Ile Gln Arg Val	
505 510 515	
gaa aat gca cga ttt ctt tca cca gaa gaa aat gtg tgc aat gag atg	1820
Glu Asn Ala Arg Phe Leu Ser Pro Glu Glu Asn Val Cys Asn Glu Met	
520 525 530	

ttg gta aaa tcc cag ttt gtt gct tgt atg gct act tgt cat tca ctt	1868
Leu Val Lys Ser Gln Phe Val Ala Cys Met Ala Thr Cys His Ser Leu	
535 540 545	
aca aaa att gaa gga gtg ctc tct ggt gat cca ctt gat ctg aaa atg	1916
Thr Lys Ile Glu Gly Val Leu Ser Gly Asp Pro Leu Asp Leu Lys Met	
550 555 560	
ttt gag gct att gga tgg att ctg gaa gaa gca act gaa gaa gaa aca	1964
Phe Glu Ala Ile Gly Trp Ile Leu Glu Glu Ala Thr Glu Glu Glu Thr	
565 570 575 580	
gca ctt cat aat cga att atg ccc aca gtg gtt cgt cct ccc aaa caa	2012
Ala Leu His Asn Arg Ile Met Pro Thr Val Val Arg Pro Pro Lys Gln	
585 590 595	
ctg ctt cct gaa tct acc cct gca gga aac caa gaa atg gag ctg ttt	2060
Leu Leu Pro Glu Ser Thr Pro Ala Gly Asn Gln Glu Met Glu Leu Phe	
600 605 610	
gaa ctt cca gct act tat gag ata gga att gtt cgc cag ttc cca ttt	2108
Glu Leu Pro Ala Thr Tyr Glu Ile Gly Ile Val Arg Gln Phe Pro Phe	
615 620 625	
tct tct gct ttg caa cgt atg agt gtg gtt gcc agg gtg ctg ggg gat	2156
Ser Ser Ala Leu Gln Arg Met Ser Val Val Ala Arg Val Leu Gly Asp	
630 635 640	
agg aaa atg gac gcc tac atg aaa gga gcg ccc gag gcc att gcc ggt	2204
Arg Lys Met Asp Ala Tyr Met Lys Gly Ala Pro Glu Ala Ile Ala Gly	
645 650 655 660	
ctc tgt aaa cct gaa aca gtt cct gtc gat ttt caa aac gtt ttg gaa	2252
Leu Cys Lys Pro Glu Thr Val Pro Val Asp Phe Gln Asn Val Leu Glu	
665 670 675	
gac ttc act aaa cag ggc ttc cgt gtg att gct ctt gca cac aga aaa	2300
Asp Phe Thr Lys Gln Gly Phe Arg Val Ile Ala Leu Ala His Arg Lys	
680 685 690	
ttg gag tca aaa ctg aca tgg cat aaa gta cag aat att agc aga gat	2348
Leu Glu Ser Lys Leu Thr Trp His Lys Val Gln Asn Ile Ser Arg Asp	
695 700 705	
gca att gag aac aac atg gat ttt atg gga tta att ata atg cag aac	2396
Ala Ile Glu Asn Asn Met Asp Phe Met Gly Leu Ile Ile Met Gln Asn	
710 715 720	
aaa tta aag caa gaa acc cct gca gta ctt gaa gat ttg cat aaa gcc	2444
Lys Leu Lys Gln Glu Thr Pro Ala Val Leu Glu Asp Leu His Lys Ala	
725 730 735 740	
aac att cgc acc gtc atg gtc aca ggt gac agt atg ttg act gct gtc	2492
Asn Ile Arg Thr Val Met Val Thr Gly Asp Ser Met Leu Thr Ala Val	
745 750 755	
tct gtg gcc aga gat tgt gga atg att cta cct cag gat aaa gtg att	2540
Ser Val Ala Arg Asp Cys Gly Met Ile Leu Pro Gln Asp Lys Val Ile	
760 765 770	

att gct gaa gca tta cct cca aag gat ggg aaa gtt gcc aaa ata aat	2588
Ile Ala Glu Ala Leu Pro Pro Lys Asp Gly Lys Val Ala Lys Ile Asn	
775 780 785	
tggt cat tat gca gac tcc ctc acg cag tgc agt cat cca tca gca att	2636
Trp His Tyr Ala Asp Ser Leu Thr Gln Cys Ser His Pro Ser Ala Ile	
790 795 800	
gac cca gag gct att ccg gtt aaa ttg gtc cat gat agc tta gag gat	2684
Asp Pro Glu Ala Ile Pro Val Lys Leu Val His Asp Ser Leu Glu Asp	
805 810 815 820	
ctt caa atg act cgt tat cat ttt gca atg aat gga aaa tca ttc tca	2732
Leu Gln Met Thr Arg Tyr His Phe Ala Met Asn Gly Lys Ser Phe Ser	
825 830 835	
gtg ata ctg gag cat ttt caa gac ctt gtt cct aag ttg atg ttg cat	2780
Val Ile Leu Glu His Phe Gln Asp Leu Val Pro Lys Leu Met Leu His	
840 845 850	
ggc acc gtg ttt gcc cgt atg gca cct gat cag aag aca cag ttg ata	2828
Gly Thr Val Phe Ala Arg Met Ala Pro Asp Gln Lys Thr Gln Leu Ile	
855 860 865	
gaa gca ttg caa aat gtt gat tat ttt gtt ggg atg tgt ggt gat ggc	2876
Glu Ala Leu Gln Asn Val Asp Tyr Phe Val Gly Met Cys Gly Asp Gly	
870 875 880	
gca aat gat tgt ggt gct ttg aag agg gca cac gga ggc att tcc tta	2924
Ala Asn Asp Cys Gly Ala Leu Lys Arg Ala His Gly Gly Ile Ser Leu	
885 890 895 900	
tcg gag ctc gaa gct tca gtg gca tct ccc ttt acc tct aag act cct	2972
Ser Glu Leu Glu Ala Ser Val Ala Ser Pro Phe Thr Ser Lys Thr Pro	
905 910 915	
agt att tcc tgt gtg cca aac ctt atc agg gaa ggc cgt gct gct tta	3020
Ser Ile Ser Cys Val Pro Asn Leu Ile Arg Glu Gly Arg Ala Ala Leu	
920 925 930	
ata act tcc ttc tgt gtg ttt aaa ttc atg gca ttg tac agc att atc	3068
Ile Thr Ser Phe Cys Val Phe Lys Phe Met Ala Leu Tyr Ser Ile Ile	
935 940 945	
cag tac ttc agt gtt act ctg ctg tat tct atc tta agt aac cta gga	3116
Gln Tyr Phe Ser Val Thr Leu Leu Tyr Ser Ile Leu Ser Asn Leu Gly	
950 955 960	
gac ttc cag ttt ctc ttc att gat ctg gca atc att ttg gta gtg gta	3164
Asp Phe Gln Phe Leu Phe Ile Asp Leu Ala Ile Ile Leu Val Val Val	
965 970 975 980	
ttt aca atg agt tta aat cct gcc tgg aaa gaa ctt gtg gca caa aga	3212
Phe Thr Met Ser Leu Asn Pro Ala Trp Lys Glu Leu Val Ala Gln Arg	
985 990 995	
cca cct tcg ggt ctt ata tct ggg gcc ctt ctc ttc tcc gtt ttg tot	3260
Pro Pro Ser Gly Leu Ile Ser Gly Ala Leu Leu Phe Ser Val Leu Ser	
1000 1005 1010	

cag att atc atc tgc att gga ttt caa tct ttg ggt ttt ttt tgg gtc 3308
 Gln Ile Ile Ile Cys Ile Gly Phe Gln Ser Leu Gly Phe Phe Trp Val
 1015 1020 1025

aaa cag caa cct tgg tat gaa gtg tgg cat cca aaa tca gat gct tgt 3356
 Lys Gln Gln Pro Trp Tyr Glu Val Trp His Pro Lys Ser Asp Ala Cys
 1030 1035 1040

aat aca aca gga agc ggg ttt tgg aat tct tca cac gta gac aat gaa 3404
 Asn Thr Thr Gly Ser Gly Phe Trp Asn Ser Ser His Val Asp Asn Glu
 1045 1050 1055 1060

acc gaa ctt gat gaa cat aat ata caa aat tat gaa aat acc aca gtg 3452
 Thr Glu Leu Asp Glu His Asn Ile Gln Asn Tyr Glu Asn Thr Thr Val
 1065 1070 1075

ttt ttt att tcc agt ttt cag tac ctc ata gtg gca att gcc ttt tca 3500
 Phe Phe Ile Ser Ser Phe Gln Tyr Leu Ile Val Ala Ile Ala Phe Ser
 1080 1085 1090

aaa gga aaa ccc ttc agg caa cct tgc tac aaa aat tat ttt ttt gtt 3548
 Lys Gly Lys Pro Phe Arg Gln Pro Cys Tyr Lys Asn Tyr Phe Phe Val
 1095 1100 1105

ttt tct gtg att ttt tta tat att ttt ata tta ttc atc atg ttg tat 3596
 Phe Ser Val Ile Phe Leu Tyr Ile Phe Ile Leu Phe Ile Met Leu Tyr
 1110 1115 1120

cca gtt gcc tct gtt gac cag gtt ctt cag ata gtg tgt gta cca tat 3644
 Pro Val Ala Ser Val Asp Gln Val Leu Gln Ile Val Cys Val Pro Tyr
 1125 1130 1135 1140

cag tgg cgt gta act atg ctc atc att gtt ctt gtc aat gcc ttt gtg 3692
 Gln Trp Arg Val Thr Met Leu Ile Ile Val Leu Val Asn Ala Phe Val
 1145 1150 1155

tct atc aca gtg gag aac ttc ttc ctt gac atg gtc ctt tgg aaa gtt 3740
 Ser Ile Thr Val Glu Asn Phe Phe Leu Asp Met Val Leu Trp Lys Val
 1160 1165 1170

gtg ttc aac cga gac aaa caa gga gag tat cgg ttc agc acc aca cag 3788
 Val Phe Asn Arg Asp Lys Gln Gly Glu Tyr Arg Phe Ser Thr Thr Gln
 1175 1180 1185

cca ccg cag gag tca gtg gat cgg tgg gga aaa tgc tgc tta ccc tgg 3836
 Pro Pro Gln Glu Ser Val Asp Arg Trp Gly Lys Cys Cys Leu Pro Trp
 1190 1195 1200

gcc ctg ggc tgt aga aag aag aca cca aag gca aag tac atg tat ctg 3884
 Ala Leu Gly Cys Arg Lys Lys Thr Pro Lys Ala Lys Tyr Met Tyr Leu
 1205 1210 1215 1220

gcg cag gag ctc ttg gtt gat cca gaa tgg cca cca aaa cct cag aca 3932
 Ala Gln Glu Leu Leu Val Asp Pro Glu Trp Pro Pro Lys Pro Gln Thr
 1225 1230 1235

acc aca gaa gct aaa gct tta gtt aag gag aat gga tca tgt caa atc 3980
 Thr Thr Glu Ala Lys Ala Leu Val Lys Glu Asn Gly Ser Cys Gln Ile
 1240 1245 1250

atc acc ata aca tag cagtgaatca gtctcagtgg tattgctgat agcagtattc 4035
 Ile Thr Ile Thr
 1255

aggaatatgt gatttttagga gtttctgato ctgtgtgtca gaatggcact agttcagttt 4095
 atgtcccttc tgatatagta gcttatttga cagctttgct cttccttaaa ataaaaacag 4155
 aaaaatatat cgtcctaaca gttaaattaa caatcaatcc ataaagtcct atatcttcat 4215
 tcagcaaccc aaatattaca tacatttcca gaattttctt gattgttact ttcagtgata 4275
 ttctttatat tgggtacagg agaagtttgg tgtttggtag gtttttcaac attagttttt 4335
 gagactagtt tacctcttca catttatgct cacaaccctc ttgttagaaa agtctgtgtt 4395
 tatatacagg ctgtaagttt gtgattgata aaaagaagat gagtgttaat tagcctccag 4455
 tgaaaatata ctgaaagcct gttttcattt gattccaatg tttcttccaa agaattctgt 4515
 ataaacatat gccaatcccc tatgatggtc tagagttagg aatgagtgtt tatggtgttg 4575
 cttatagaac aactcaggta atctccattt ctggttttat attttctgta caaactgcct 4635
 gggttttatt tttctaatca gcaagggtgt tcaactgcctt cttgagacgc ctotcaaagc 4695
 tcttaaatgg ctctgtgtct atgtgtggtg ttggcagtct aatttgcttc tgttaaatgt 4755
 tgtagaacct ttttcaactag gaaataagat tcatttcttt cggcagtaga tgtagattca 4815
 tcttttaacg tttcttcaaa tttgtttctg tcaggctttg tgttatttta aatggttttt 4875
 taaaattttc ttctatgttt tcaattacct aaagacatag gataatagtt ttttttaagt 4935
 tagaatttta cctcataaaa ttttttgagg tttgatgtat gtctctgtct tatcaataat 4995
 gaggcttaaa aaatactgga tttgaatggc tgccgttttt tcaaagcaat atgaatttga 5055
 tgagtttggt ttatgccatt aggtggcgcc agaggtcaga acatgtctat tttgaattgg 5115
 atcgttacaa atgagcatat ttgatcgga aatttctggg agaaaaaaaa ttgaggaaat 5175
 aaagttaaaa aattgacatt cattgagcca aaagagatgt ggagaaacat ttttcacctt 5235
 tctgtttggc ctgattaaca tttaaattct tgccaaaatt aaggtaactt ttaaaaaaca 5295
 ccttttatag gtggatccag cagtctggca acgcccacag ttaccacaac acagaaaact 5355
 gatcgtgcta taaaatggac gctaaactat gaaaacagtg tgacattgtt ctctgttctt 5415
 ccagagccag taacatgctt gctcgtgctt tctacttcta gctgatcatt cttttcccaa 5475
 catatattta caaattacca aattttacct agaatttttag gaccaaattg ttctcaactct 5535
 ttatgctgca aagacctgga tgatgtttgg taactataga aaaatagaaa ttacactcag 5595
 gatcactgtt actgctattg ccaactgatga ttcttgcaaa aatataatcg aagttttcca 5655
 tcaaattgat aatatgctat taatacacat tagatgataa cagttgttcc atgaatgatt 5715
 ctatgaagct atgcatctta gacctcttga gctgtgaatt agcactattt tctatagtta 5775

cttattctct ggatcatttt ataatttcca tattaatttc aaatatgctc gtgttattct 5835
 tcagtgatct ccacaattgt gcattttatt ctttggttta agtactgaag catataatga 5895
 aagtaattgc taagtagcag cttaaaaatt caattatccg attgtattta acatctttta 5955
 gagcatgac ataaagagct atttttgaca ccccccccc acttttttaa catttagagt 6015
 tagtaagggt tttatatctc ttctgtccat attgttttca aaggaatgag gtgttttaggt 6075
 ggctggaaaa gcattttagt gaagttagat ttgaatatag acaagggtgg ttattcacgt 6135
 tgagaatgtt atttgaagaa tgccgtgtaa gccagggtgt ggttctactc agtgccatag 6195
 atagactgag tcttctctcg taggttaacca ttacatagta attttgattc tgaattacac 6255
 attaaattat ttgagtttat acagacctaa attttaaaat ctgtacatat attattttga 6315
 tgtattaaga tgaatattgc tgatttaaat tttatttatg cacatactta aaggacagaa 6375
 atgtctggga aagtaattgt taaataatga tatgtaactt ttttaacttt taaataaata 6435
 acaagatttt taatgtgtgt ctccctcagg gttgtttaaa gttttttttc tccctcaagt 6495
 ataaatagtg gtaactatat gttttgtatc ttctagcacc aactgctgta aagcaatgct 6555
 gcaaataatg cttgaataca agtggctaag ccaacaacag aataaatact tttatagtag 6615
 ttttataatc ctgaaattcg aaagctttcc caattgcact tgcactctaa caaaaactgt 6675
 gcagttttta ctctatttat tttgttcccc atgtttatga aagtcctgca cagtttcaaa 6735
 ggcatggtaa ataatatatc aatgtttatg tagtctgtta cagaaacagc tatagataac 6795
 attatccagt gaagagcaaa attcaagctt tagaaaatat tcatgcatgc aattttgaca 6855
 tatctaaaaa taggtttttg tatatttatg gtgggaggtg gttgggaact ttttaacaaa 6915
 tgggggtgtta atttttgtac agtctgtggg catttacaca tttttaatgt attaaaattt 6975
 ggtaattatg tgtacattaa attaataaaa gttacttcta gttatgattt gtgaattccc 7035
 taagaccttg gatTTTTTTT agtaacttta tatcagaaat gatactgcat ctttatattt 7095
 ttaaaattgt attgctgctc aagaatggta ccctctgtc aaaaaggcat acattcataa 7155
 ttgtacattc agcattgtaa ataatttat gaaacctttt ttgattgaag ctattcaaaa 7215
 taaaaatttt aatgaatgaa aaaaaaaaaa aaaa 7249

<210> 6
 <211> 1256
 <212> PRT
 <213> Homo sapiens

<400> 6
 Met Asp Arg Glu Glu Arg Lys Thr Ile Asn Gln Gly Gln Glu Asp Glu
 1 5 10 15

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

- 25 -

Pro Ser Val Asp Val Lys Gly Ile Gly Asp Glu Leu Tyr Asn Pro Glu
 340 345 350
 Thr His Lys Arg His Thr Leu Phe Cys Gly Thr Thr Val Ile Gln Thr
 355 360 365
 Arg Phe Tyr Thr Gly Glu Leu Val Lys Ala Ile Val Val Arg Thr Gly
 370 375 380
 Phe Ser Thr Ser Lys Gly Gln Leu Val Arg Ser Ile Leu Tyr Pro Lys
 385 390 395 400
 Pro Thr Asp Phe Lys Leu Tyr Arg Asp Ala Tyr Leu Phe Leu Leu Cys
 405 410 415
 Leu Val Ala Val Ala Gly Ile Gly Phe Ile Tyr Thr Ile Ile Asn Ser
 420 425 430
 Ile Leu Asn Glu Val Gln Val Gly Val Ile Ile Ile Glu Ser Leu Asp
 435 440 445
 Ile Ile Thr Ile Thr Val Pro Pro Ala Leu Pro Ala Ala Met Thr Ala
 450 455 460
 Gly Ile Val Tyr Ala Gln Arg Arg Leu Lys Lys Ile Gly Ile Phe Cys
 465 470 475 480
 Ile Ser Pro Gln Arg Ile Asn Ile Cys Gly Gln Leu Asn Leu Val Cys
 485 490 495
 Phe Asp Lys Thr Gly Thr Leu Thr Glu Asp Gly Leu Asp Leu Trp Gly
 500 505 510
 Ile Gln Arg Val Glu Asn Ala Arg Phe Leu Ser Pro Glu Glu Asn Val
 515 520 525
 Cys Asn Glu Met Leu Val Lys Ser Gln Phe Val Ala Cys Met Ala Thr
 530 535 540
 Cys His Ser Leu Thr Lys Ile Glu Gly Val Leu Ser Gly Asp Pro Leu
 545 550 555 560
 Asp Leu Lys Met Phe Glu Ala Ile Gly Trp Ile Leu Glu Glu Ala Thr
 565 570 575
 Glu Glu Glu Thr Ala Leu His Asn Arg Ile Met Pro Thr Val Val Arg
 580 585 590
 Pro Pro Lys Gln Leu Leu Pro Glu Ser Thr Pro Ala Gly Asn Gln Glu
 595 600 605
 Met Glu Leu Phe Glu Leu Pro Ala Thr Tyr Glu Ile Gly Ile Val Arg
 610 615 620
 Gln Phe Pro Phe Ser Ser Ala Leu Gln Arg Met Ser Val Val Ala Arg
 625 630 635 640
 Val Leu Gly Asp Arg Lys Met Asp Ala Tyr Met Lys Gly Ala Pro Glu
 645 650 655

Ala Ile Ala Gly Leu Cys Lys Pro Glu Thr Val Pro Val Asp Phe Gln
 660 665 670
 Asn Val Leu Glu Asp Phe Thr Lys Gln Gly Phe Arg Val Ile Ala Leu
 675 680 685
 Ala His Arg Lys Leu Glu Ser Lys Leu Thr Trp His Lys Val Gln Asn
 690 695 700
 Ile Ser Arg Asp Ala Ile Glu Asn Asn Met Asp Phe Met Gly Leu Ile
 705 710 715 720
 Ile Met Gln Asn Lys Leu Lys Gln Glu Thr Pro Ala Val Leu Glu Asp
 725 730 735
 Leu His Lys Ala Asn Ile Arg Thr Val Met Val Thr Gly Asp Ser Met
 740 745 750
 Leu Thr Ala Val Ser Val Ala Arg Asp Cys Gly Met Ile Leu Pro Gln
 755 760 765
 Asp Lys Val Ile Ile Ala Glu Ala Leu Pro Pro Lys Asp Gly Lys Val
 770 775 780
 Ala Lys Ile Asn Trp His Tyr Ala Asp Ser Leu Thr Gln Cys Ser His
 785 790 795 800
 Pro Ser Ala Ile Asp Pro Glu Ala Ile Pro Val Lys Leu Val His Asp
 805 810 815
 Ser Leu Glu Asp Leu Gln Met Thr Arg Tyr His Phe Ala Met Asn Gly
 820 825 830
 Lys Ser Phe Ser Val Ile Leu Glu His Phe Gln Asp Leu Val Pro Lys
 835 840 845
 Leu Met Leu His Gly Thr Val Phe Ala Arg Met Ala Pro Asp Gln Lys
 850 855 860
 Thr Gln Leu Ile Glu Ala Leu Gln Asn Val Asp Tyr Phe Val Gly Met
 865 870 875 880
 Cys Gly Asp Gly Ala Asn Asp Cys Gly Ala Leu Lys Arg Ala His Gly
 885 890 895
 Gly Ile Ser Leu Ser Glu Leu Glu Ala Ser Val Ala Ser Pro Phe Thr
 900 905 910
 Ser Lys Thr Pro Ser Ile Ser Cys Val Pro Asn Leu Ile Arg Glu Gly
 915 920 925
 Arg Ala Ala Leu Ile Thr Ser Phe Cys Val Phe Lys Phe Met Ala Leu
 930 935 940
 Tyr Ser Ile Ile Gln Tyr Phe Ser Val Thr Leu Leu Tyr Ser Ile Leu
 945 950 955 960
 Ser Asn Leu Gly Asp Phe Gln Phe Leu Phe Ile Asp Leu Ala Ile Ile
 965 970 975

Leu Val Val Val Phe Thr Met Ser Leu Asn Pro Ala Trp Lys Glu Leu
 980 985 990
 Val Ala Gln Arg Pro Pro Ser Gly Leu Ile Ser Gly Ala Leu Leu Phe
 995 1000 1005
 Ser Val Leu Ser Gln Ile Ile Ile Cys Ile Gly Phe Gln Ser Leu Gly
 1010 1015 1020
 Phe Phe Trp Val Lys Gln Gln Pro Trp Tyr Glu Val Trp His Pro Lys
 1025 1030 1035 1040
 Ser Asp Ala Cys Asn Thr Thr Gly Ser Gly Phe Trp Asn Ser Ser His
 1045 1050 1055
 Val Asp Asn Glu Thr Glu Leu Asp Glu His Asn Ile Gln Asn Tyr Glu
 1060 1065 1070
 Asn Thr Thr Val Phe Phe Ile Ser Ser Phe Gln Tyr Leu Ile Val Ala
 1075 1080 1085
 Ile Ala Phe Ser Lys Gly Lys Pro Phe Arg Gln Pro Cys Tyr Lys Asn
 1090 1095 1100
 Tyr Phe Phe Val Phe Ser Val Ile Phe Leu Tyr Ile Phe Ile Leu Phe
 1105 1110 1115 1120
 Ile Met Leu Tyr Pro Val Ala Ser Val Asp Gln Val Leu Gln Ile Val
 1125 1130 1135
 Cys Val Pro Tyr Gln Trp Arg Val Thr Met Leu Ile Ile Val Leu Val
 1140 1145 1150
 Asn Ala Phe Val Ser Ile Thr Val Glu Asn Phe Phe Leu Asp Met Val
 1155 1160 1165
 Leu Trp Lys Val Val Phe Asn Arg Asp Lys Gln Gly Glu Tyr Arg Phe
 1170 1175 1180
 Ser Thr Thr Gln Pro Pro Gln Glu Ser Val Asp Arg Trp Gly Lys Cys
 1185 1190 1195 1200
 Cys Leu Pro Trp Ala Leu Gly Cys Arg Lys Lys Thr Pro Lys Ala Lys
 1205 1210 1215
 Tyr Met Tyr Leu Ala Gln Glu Leu Leu Val Asp Pro Glu Trp Pro Pro
 1220 1225 1230
 Lys Pro Gln Thr Thr Thr Glu Ala Lys Ala Leu Val Lys Glu Asn Gly
 1235 1240 1245
 Ser Cys Gln Ile Ile Thr Ile Thr
 1250 1255

<210> 7

<211> 3768

<212> DNA

<213> Homo sapiens

<400> 7

atggacaggg aagaaaggaa gaccatcaat caggggtcaag aagatgaaat ggagatttat 60
ggttacaatt tgagtcgctg gaagcttgcc atagtttctt taggagtgat ttgctctgat 120
gggtttctcc tctctctcct ctattggatg cctgagtggc ggggtgaaagc gacctgtgtc 180
agagctgcaa ttaaagactg tgaagtagtg ctgctgagga ctactgatga attcaaaatg 240
tggttttgtg caaaaattcg cgttctttct ttggaaactt acccagtttc aagtccaaaa 300
tctatgtota ataagctttc aaatggccat gcagtttggt taattgagaa tccactgaa 360
gaaaataggg acaggatcag taaatattca cagactgaat cacaacagat tcgttatttc 420
accaccata gtgtaaaata tttctggaat gataccattc acaattttga tttcttaaag 480
ggactggatg aagggtgtttc ttgtacgtca atttatgaaa agcatagtgc aggactgaca 540
aaggggatgc atgcctacag aaaactgctt tatggagtaa atgaaattgc tgtaaaagtg 600
ccttctgttt ttaagcttct aattaaagag gttctcaacc cattttacat tttccagctg 660
ttcagtgtta tactgtggag cactgatgaa tactattact atgctctagc tattgtgggt 720
atgtccatag tatcaatcgt aagctcacta tattccatta gaaagcaata tgttatgttg 780
catgacatgg tggcaactca tagtaccgta agagtttcag tttgtagagt aaatgaagaa 840
atagaagaaa tcttttctac cgaccttggt ccaggagatg toatggtcac tccattaaat 900
gggacaataa tgccttggtg tgctgtgctt attaatggta cctgcattgt aaacgaaagc 960
atgttaacag gagaaagtgt tccagtgaca aagactaatt tgccaaatcc ttcagtggat 1020
gtgaaaggaa taggagatga attatataat ccagaaacac ataaacgaca tactttgttt 1080
tgtgggacaa ctgttattca gactcgtttc tacaactggag aactcgtcaa agccatagtt 1140
gttagaacag gatttagtac ttcaaagga cagcttggtc gttccatatt gtatcccaaa 1200
ccaactgatt ttaaactcta cagagatgoc tacttgtttc tactatgtct tgtggcagtt 1260
gctggcattg ggtttatcta cactattatt aatagcattt taaatgaggt acaagttggg 1320
gtcataatta tcgagtctct tgatattatc acaattactg tgccccctgc acttcctgct 1380
gcaatgactg ctgggtattgt gtatgctcag agaagactga aaaaaatcgg tattttctgt 1440
atcagtcctc aaagaataaa tatttggtga cagctcaatc ttgtttgctt tgacaagact 1500
ggaactctaa ctgaagatgg tttagatctt tgggggatcc aacgagtggg aaatgcacga 1560
tttctttcac cagaagaaaa tgtgtgcaat gagatgttgg taaaatccca gtttggtgct 1620
tgtatggcta cttgtcattc acttacaaaa attgaaggag tgctctctgg tgatccactt 1680
gatctgaaaa tgtttgaggc tattggatgg attctggaag aagcaactga agaagaaaca 1740
gcacttcata atcgaattat gccacagtg gttcgtcctc ccaaacaact gcttcctgaa 1800

tctaccctg caggaaacca agaaatggag ctgtttgaac ttccagctac ttatgagata 1860
ggaattgttc gccagttccc attttcttct gctttgcaac gtatgagtgt ggttgccagg 1920
gtgctggggg ataggaaaat ggacgcctac atgaaaggag cgcccgaggc cattgccggg 1980
ctctgtaaac ctgaaacagt tcctgtcgat tttcaaaacg ttttggaaga cttcactaaa 2040
cagggcttcc gtgtgattgc tcttgcacac agaaaattgg agtcaaaact gacatggcat 2100
aaagtacaga atattagcag agatgcaatt gagaacaaca tggattttat gggattaatt 2160
ataatgcaga acaaattaaa gcaagaaacc cctgcagtac ttgaagattt gcataaagcc 2220
aacattcgca ccgtcatggt cacagggtgac agtatgttga ctgctgtctc tgtggccaga 2280
gattgtggaa tgattctacc tcaggataaa gtgattattg ctgaagcatt acctccaaag 2340
gatgggaaag ttgccaaaat aaattggcat tatgcagact ccctcacgca gtgcagtcac 2400
ccatcagcaa ttgaccaga ggctattccg gttaaattgg tccatgatag cttagaggat 2460
cttcaaatga ctcgttatca ttttgcaatg aatggaaaat cattctcagt gatactggag 2520
cattttcaag accttgttcc taagttgatg ttgcatggca ccgtgtttgc ccgtatggca 2580
cctgatcaga agacacagtt gatagaagca ttgcaaaatg ttgattattt tgttgggatg 2640
tgtgggtgatg gcgcaaatga ttgtggtgct ttgaagaggg cacacggagg catttcctta 2700
tcggagctcg aagcttcagt ggcatctccc tttaacctta agactcctag tatttcctgt 2760
gtgccaaacc ttatcagga aggccgtgct gctttaataa cttccttctg tgtgtttaaa 2820
ttcatggcat tgtacagcat tatccagtac ttcagtgtta ctctgctgta ttctatctta 2880
agtaacctag gagacttcca gtttctcttc attgatctgg caatcatttt ggtagtggtta 2940
tttacaatga gtttaaatcc tgccctggaa gaacttgtgg cacaagacc accttcgggt 3000
cttatatctg gggcccttct cttctccgtt ttgtctcaga ttatcatctg cattggattt 3060
caatcttttg gttttttttg ggtcaaacag caaccttggg atgaagtgtg gcatccaaaa 3120
tcagatgctt gtaatacaac aggaagcggg ttttggaatt cttcacacgt agacaatgaa 3180
accgaacttg atgaacataa tatacaaaat tatgaaaata ccacagtgtt ttttatttcc 3240
agttttcagt acctcatagt ggcaattgcc ttttcaaaag gaaaaccctt caggcaacct 3300
tgctacaaaa attatttttt tgttttttct gtgatttttt tataatattt tataattattc 3360
atcatgttgt atccagttgc ctctgttgac caggttcttc agatagtgtg tgtaccatat 3420
cagtggcgtg taactatgct catcattgtt cttgtcaatg cctttgtgtc tatcacagtg 3480
gagaacttct tccttgacat ggtcctttgg aaagtgtgtg tcaaccgaga caaacaagga 3540
gagtatcggg tcagcaccac acagccaccg caggagtcag tggatcgggt gggaaaatgc 3600
tgcttaccct gggccctggg ctgtagaaag aagacaccaa aggcaaagta catgtatctg 3660

gcgcaggagc tcttggttga tccagaatgg ccacaaaaac ctcagacaac cacagaagct 3720
aaagcttttag ttaaggagaa tggatcatgt caaatcatca ccataaca 3768

```
<210> 8
<211> 3919
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> (68)..(3679)
```

<400> 8																
ttaccggaag	taaaacttcg	gaagtgaggc	gttcctctgc	ccggaagtga	gcgcgcgcgt	60										
aggaaaag	atg	gcg	gca	gcg	gcg	gcg	gtg	ggc	aac	gcg	gtg	ccc	tgc	ggg	109	
	Met	Ala	Ala	Ala	Ala	Ala	Val	Gly	Asn	Ala	Val	Pro	Cys	Gly		
	1					5				10						
gcc	cgg	cct	tgc	ggg	gtc	cgg	cct	gac	ggg	cag	ccc	aag	ccc	ggg	157	
Ala	Arg	Pro	Cys	Gly	Val	Arg	Pro	Asp	Gly	Gln	Pro	Lys	Pro	Gly	Pro	
15					20					25					30	
cag	ccg	cgc	gcg	ctc	ctt	gcc	gcc	ggg	ccg	gcg	ctc	ata	gcg	aac	ggg	205
Gln	Pro	Arg	Ala	Leu	Leu	Ala	Ala	Gly	Pro	Ala	Leu	Ile	Ala	Asn	Gly	
				35					40					45		
gac	gag	ctg	gtg	gct	gcc	gtg	tgg	ccg	tac	cgg	cgg	ttg	gcg	ctg	ttg	253
Asp	Glu	Leu	Val	Ala	Ala	Val	Trp	Pro	Tyr	Arg	Arg	Leu	Ala	Leu	Leu	
			50					55					60			
cgg	cgc	ctc	acg	gtg	ctg	cca	ttc	gcc	ggg	ctg	ctt	tac	ccg	gcc	tgg	301
Arg	Arg	Leu	Thr	Val	Leu	Pro	Phe	Ala	Gly	Leu	Leu	Tyr	Pro	Ala	Trp	
		65					70					75				
ttg	ggt	gcc	gca	gcc	gct	ggc	tgc	tgg	ggc	tgg	ggc	agc	agt	tgg	gtg	349
Leu	Gly	Ala	Ala	Ala	Ala	Gly	Cys	Trp	Gly	Trp	Gly	Ser	Ser	Trp	Val	
	80					85					90					
cag	atc	ccc	gaa	gct	gcg	ctg	ctc	gtg	ctt	gcc	acc	atc	tgc	ctc	gcg	397
Gln	Ile	Pro	Glu	Ala	Ala	Leu	Leu	Val	Leu	Ala	Thr	Ile	Cys	Leu	Ala	
95					100					105					110	
cac	gcg	ctc	act	gtc	ctc	tcg	ggg	cat	tgg	tct	gtg	cac	gcg	cat	tgc	445
His	Ala	Leu	Thr	Val	Leu	Ser	Gly	His	Trp	Ser	Val	His	Ala	His	Cys	
				115					120					125		
gcg	ctc	acc	tgc	acc	ccg	gag	tac	gac	ccc	agc	aaa	gcg	acc	ttt	gtg	493
Ala	Leu	Thr	Cys	Thr	Pro	Glu	Tyr	Asp	Pro	Ser	Lys	Ala	Thr	Phe	Val	
			130					135					140			
aag	gtg	gcg	cca	acc	ccc	aac	aat	ggc	tcc	acg	gag	ctc	gtg	gcc	ctg	541
Lys	Val	Ala	Pro	Thr	Pro	Asn	Asn	Gly	Ser	Thr	Glu	Leu	Val	Ala	Leu	
		145					150					155				
cac	cgc	aat	gag	ggc	gaa	gac	ggg	ctt	gag	gtg	ctg	tcc	ttc	gaa	ttc	589
His	Arg	Asn	Glu	Gly	Glu	Asp	Gly	Leu	Glu	Val	Leu	Ser	Phe	Glu	Phe	
	160					165					170					

cag aag atc aag tat tcc tac gat gcc ctg gag aag aag cag ttt ctc	637
Gln Lys Ile Lys Tyr Ser Tyr Asp Ala Leu Glu Lys Lys Gln Phe Leu	
175 180 185 190	
ccc gtg gcc ttt cct gtg gga aac gcc ttc tca tac tat cag agc aac	685
Pro Val Ala Phe Pro Val Gly Asn Ala Phe Ser Tyr Tyr Gln Ser Asn	
195 200 205	
aga ggc ttc cag gaa gac tca gag atc cga gca gct gag aag aaa ttt	733
Arg Gly Phe Gln Glu Asp Ser Glu Ile Arg Ala Ala Glu Lys Lys Phe	
210 215 220	
ggg agc aac aag gcc gag atg gtg gtg cct gac ttc tcg gag ctt ttc	781
Gly Ser Asn Lys Ala Glu Met Val Val Pro Asp Phe Ser Glu Leu Phe	
225 230 235	
aag gag aga gcc aca gcc ccc ttc ttt gta ttt cag gtg ttc tgt gtg	829
Lys Glu Arg Ala Thr Ala Pro Phe Phe Val Phe Gln Val Phe Cys Val	
240 245 250	
ggg ctc tgg tgc ctg gat gag tac tgg tac tac agc gtc ttt acg cta	877
Gly Leu Trp Cys Leu Asp Glu Tyr Trp Tyr Tyr Ser Val Phe Thr Leu	
255 260 265 270	
tcc atg ctg gtg gcg ttc gag gcc tcg ctg gtg cag cag cag atg cgg	925
Ser Met Leu Val Ala Phe Glu Ala Ser Leu Val Gln Gln Gln Met Arg	
275 280 285	
aac atg tcg gag atc cgg aag atg ggc aac aag ccc cac atg atc cag	973
Asn Met Ser Glu Ile Arg Lys Met Gly Asn Lys Pro His Met Ile Gln	
290 295 300	
gtc tac cga agc cgc aag tgg agg ccc att gcc agt gat gag atc gta	1021
Val Tyr Arg Ser Arg Lys Trp Arg Pro Ile Ala Ser Asp Glu Ile Val	
305 310 315	
cca ggg gac atc gtc tcc atc ggc cgc tcc cca cag gag aac ctg gtg	1069
Pro Gly Asp Ile Val Ser Ile Gly Arg Ser Pro Gln Glu Asn Leu Val	
320 325 330	
cca tgt gac gtg ctt ctg ctg cga ggc cgc tgc atc gta gac gag gcc	1117
Pro Cys Asp Val Leu Leu Leu Arg Gly Arg Cys Ile Val Asp Glu Ala	
335 340 345 350	
atg ctc acg ggg gag tcc gtg cca cag atg aag gag ccc atc gaa gac	1165
Met Leu Thr Gly Glu Ser Val Pro Gln Met Lys Glu Pro Ile Glu Asp	
355 360 365	
ctc agc cca gac cgg gtg ctg gac ctc cag gct gat tcc cgg ctg cac	1213
Leu Ser Pro Asp Arg Val Leu Asp Leu Gln Ala Asp Ser Arg Leu His	
370 375 380	
gtc atc ttc ggg ggc acc aag gtg gtg cag cac atc ccc cca cag aaa	1261
Val Ile Phe Gly Gly Thr Lys Val Val Gln His Ile Pro Pro Gln Lys	
385 390 395	
gcc acc acg ggc ctg aag ccg gtt gac agc ggg tgc gtg gcc tac gtc	1309
Ala Thr Thr Gly Leu Lys Pro Val Asp Ser Gly Cys Val Ala Tyr Val	
400 405 410	

ctg cgg acc gga ttc aac aca tct cag ggc aag ctg ctg cgc acc atc	1357
Leu Arg Thr Gly Phe Asn Thr Ser Gln Gly Lys Leu Leu Arg Thr Ile	
415 420 425 430	
ctc ttc ggg gtc aag agg gtg act gcg aac aac ctg gag acc ttc atc	1405
Leu Phe Gly Val Lys Arg Val Thr Ala Asn Asn Leu Glu Thr Phe Ile	
435 440 445	
ttc atc ctc ttc ctc ctg gtg ttt gcc atc gct gca gct gcc tat gta	1453
Phe Ile Leu Phe Leu Leu Val Phe Ala Ile Ala Ala Ala Tyr Val	
450 455 460	
tgg att gaa ggt acc aag gac ccc agc cgg aac cgc tac aag ctg ttt	1501
Trp Ile Glu Gly Thr Lys Asp Pro Ser Arg Asn Arg Tyr Lys Leu Phe	
465 470 475	
ctg gag tgc acc ctg atc ctc acc tcg gtc gtg cct cct gag ctg ccc	1549
Leu Glu Cys Thr Leu Ile Leu Thr Ser Val Val Pro Pro Glu Leu Pro	
480 485 490	
atc gag ctg tcc ctg gcc gtc aac acc tcc ctc atc gcc ctg gcc aag	1597
Ile Glu Leu Ser Leu Ala Val Asn Thr Ser Leu Ile Ala Leu Ala Lys	
495 500 505 510	
ctc tac atg tac tgc aca gag ccc ttc cgg atc ccc ttt gct ggc aag	1645
Leu Tyr Met Tyr Cys Thr Glu Pro Phe Arg Ile Pro Phe Ala Gly Lys	
515 520 525	
gtc gag gtg tgc tgc ttt gac aag acg ggg acg ttg acc agt gac agc	1693
Val Glu Val Cys Cys Phe Asp Lys Thr Gly Thr Leu Thr Ser Asp Ser	
530 535 540	
ctg gtg gtg cgc ggt gtg gcc ggg ctg aga gac ggg aag gag gtg acc	1741
Leu Val Val Arg Gly Val Ala Gly Leu Arg Asp Gly Lys Glu Val Thr	
545 550 555	
cca gtg tcc agc atc cct gta gaa aca cac cgg gcc ctg gcc tcg tgc	1789
Pro Val Ser Ser Ile Pro Val Glu Thr His Arg Ala Leu Ala Ser Cys	
560 565 570	
cac tcg ctc atg cag ctg gac gac ggc acc ctc gtg ggt gac cct cta	1837
His Ser Leu Met Gln Leu Asp Asp Gly Thr Leu Val Gly Asp Pro Leu	
575 580 585 590	
gag aag gcc atg ctg acg gcc gtg gac tgg acg ctg acc aaa gat gag	1885
Glu Lys Ala Met Leu Thr Ala Val Asp Trp Thr Leu Thr Lys Asp Glu	
595 600 605	
aaa gta ttc ccc cga agt att aaa act cag ggg ctg aaa att cac cag	1933
Lys Val Phe Pro Arg Ser Ile Lys Thr Gln Gly Leu Lys Ile His Gln	
610 615 620	
cgc ttt cat ttt gcc agt gcc ctg aag cga atg tcc gtg ctt gcc tcg	1981
Arg Phe His Phe Ala Ser Ala Leu Lys Arg Met Ser Val Leu Ala Ser	
625 630 635	
tat gag aag ctg ggc tcc acc gac ctc tgc tac atc gcg gcc gtg aag	2029
Tyr Glu Lys Leu Gly Ser Thr Asp Leu Cys Tyr Ile Ala Ala Val Lys	
640 645 650	

ggg gcc ccc gaa act ctg cac tcc atg ttc tcc cag tgc ccg ccc gac	2077
Gly Ala Pro Glu Thr Leu His Ser Met Phe Ser Gln Cys Pro Pro Asp	
655 660 665 670	
tac cac cac atc cac acc gag atc tcc cgg gaa gga gcc cgc gtc ctg	2125
Tyr His His Ile His Thr Glu Ile Ser Arg Glu Gly Ala Arg Val Leu	
675 680 685	
gcg ctg ggg tac aag gag ctg gga cac ctc act cac cag cag gcc cgg	2173
Ala Leu Gly Tyr Lys Glu Leu Gly His Leu Thr His Gln Gln Ala Arg	
690 695 700	
gag gtc aag cgg gag gcc ctg gag tgc agc ctc aag ttc gtc ggc ttc	2221
Glu Val Lys Arg Glu Ala Leu Glu Cys Ser Leu Lys Phe Val Gly Phe	
705 710 715	
att gtg gtc tcc tgc ccg ctc aag gct gac tcc aag gcc gtg atc cgg	2269
Ile Val Val Ser Cys Pro Leu Lys Ala Asp Ser Lys Ala Val Ile Arg	
720 725 730	
gag atc cag aat gcg tcc cac cgg gtg gtc atg atc acg gga gac aac	2317
Glu Ile Gln Asn Ala Ser His Arg Val Val Met Ile Thr Gly Asp Asn	
735 740 745 750	
ccg ctc act gca tgc cac gtg gcc cag gag ctg cac ttc att gaa aag	2365
Pro Leu Thr Ala Cys His Val Ala Gln Glu Leu His Phe Ile Glu Lys	
755 760 765	
gcc cac acg ctg atc ctg cag cct ccc tcc gag aaa ggc cgg cag tgc	2413
Ala His Thr Leu Ile Leu Gln Pro Pro Ser Glu Lys Gly Arg Gln Cys	
770 775 780	
gag tgg cgc tcc att gac ggc agc atc gtg ctg ccc ctg gcc cgg ggc	2461
Glu Trp Arg Ser Ile Asp Gly Ser Ile Val Leu Pro Leu Ala Arg Gly	
785 790 795	
tcc cca aag gca ctg gcc ctg gag tac gca ctg tgc ctc aca ggc gac	2509
Ser Pro Lys Ala Leu Ala Leu Glu Tyr Ala Leu Cys Leu Thr Gly Asp	
800 805 810	
ggc ttg gcc cac ctg cag gcc acc gac ccc cag cag ctg ctc cgc ctc	2557
Gly Leu Ala His Leu Gln Ala Thr Asp Pro Gln Gln Leu Leu Arg Leu	
815 820 825 830	
atc ccc cat gtg cag gtg ttc gcc cgt gtg gct ccc aag cag aag gag	2605
Ile Pro His Val Gln Val Phe Ala Arg Val Ala Pro Lys Gln Lys Glu	
835 840 845	
ttt gtc atc acc agc ctg aag gag ctg ggc tac gtg acc ctc atg tgt	2653
Phe Val Ile Thr Ser Leu Lys Glu Leu Gly Tyr Val Thr Leu Met Cys	
850 855 860	
ggg gat ggc acc aac gac gtg ggc gcc ctg aag cat gct gac gtg ggt	2701
Gly Asp Gly Thr Asn Asp Val Gly Ala Leu Lys His Ala Asp Val Gly	
865 870 875	
gtg gcg ctc ttg gcc aat gcc cct gag cgg gtt gtc gag cgg cga cgg	2749
Val Ala Leu Leu Ala Asn Ala Pro Glu Arg Val Val Glu Arg Arg Arg	
880 885 890	

- 35 -

```

gtg tgg agt ctg gca gtt tca ctc ctg gcc atc att ggc ctg ctc ctc 3517
Val Trp Ser Leu Ala Val Ser Leu Leu Ala Ile Ile Gly Leu Leu Leu
1135          1140          1145          1150

ggc tcc tcg ccc gac ttc aac agc cag ttt ggc ctc gtg gac atc cct 3565
Gly Ser Ser Pro Asp Phe Asn Ser Gln Phe Gly Leu Val Asp Ile Pro
          1155          1160          1165

gtg gag ttc aag ctg gtc att gcc cag gtc ctg ctc ctg gac ttc tgc 3613
Val Glu Phe Lys Leu Val Ile Ala Gln Val Leu Leu Leu Asp Phe Cys
          1170          1175          1180

ctg gcg ctc ctg gcc gac cgc gtc ctg cag ttc ttc ctg ggg acc ccg 3661
Leu Ala Leu Leu Ala Asp Arg Val Leu Gln Phe Phe Leu Gly Thr Pro
          1185          1190          1195

aag ctg aaa gtg cct tcc tgagatggca gtgctggtac ccactgccca 3709
Lys Leu Lys Val Pro Ser
1200

ccctggctgc cgctgggagg gaaccccaac agggccccgg gaggggaacc tgccccaac 3769

ccccacagc aaggctgtac agtctcgccc ttggaagact gagctgggac cccacagcc 3829

atcogctggc ttggccagca gaaccagccc caagccagca cctttggtaa ataaagcagc 3889

atctgagatt ttaaaaaaaaa aaaaaaaaaa 3919

```

<210> 9

<211> 1204

<212> PRT

<213> Homo sapiens

<400> 9

```

Met Ala Ala Ala Ala Val Gly Asn Ala Val Pro Cys Gly Ala Arg
  1          5          10          15

Pro Cys Gly Val Arg Pro Asp Gly Gln Pro Lys Pro Gly Pro Gln Pro
          20          25          30

Arg Ala Leu Leu Ala Ala Gly Pro Ala Leu Ile Ala Asn Gly Asp Glu
          35          40          45

Leu Val Ala Ala Val Trp Pro Tyr Arg Arg Leu Ala Leu Leu Arg Arg
          50          55          60

Leu Thr Val Leu Pro Phe Ala Gly Leu Leu Tyr Pro Ala Trp Leu Gly
          65          70          75          80

Ala Ala Ala Ala Gly Cys Trp Gly Trp Gly Ser Ser Trp Val Gln Ile
          85          90          95

Pro Glu Ala Ala Leu Leu Val Leu Ala Thr Ile Cys Leu Ala His Ala
          100          105          110

Leu Thr Val Leu Ser Gly His Trp Ser Val His Ala His Cys Ala Leu
          115          120          125

Thr Cys Thr Pro Glu Tyr Asp Pro Ser Lys Ala Thr Phe Val Lys Val
          130          135          140

```

- 36 -

Ala	Pro	Thr	Pro	Asn	Asn	Gly	Ser	Thr	Glu	Leu	Val	Ala	Leu	His	Arg	145	150	155	160
Asn	Glu	Gly	Glu	Asp	Gly	Leu	Glu	Val	Leu	Ser	Phe	Glu	Phe	Gln	Lys	165	170		175
Ile	Lys	Tyr	Ser	Tyr	Asp	Ala	Leu	Glu	Lys	Lys	Gln	Phe	Leu	Pro	Val	180	185		190
Ala	Phe	Pro	Val	Gly	Asn	Ala	Phe	Ser	Tyr	Tyr	Gln	Ser	Asn	Arg	Gly	195	200		205
Phe	Gln	Glu	Asp	Ser	Glu	Ile	Arg	Ala	Ala	Glu	Lys	Lys	Phe	Gly	Ser	210	215		220
Asn	Lys	Ala	Glu	Met	Val	Val	Pro	Asp	Phe	Ser	Glu	Leu	Phe	Lys	Glu	225	230		240
Arg	Ala	Thr	Ala	Pro	Phe	Phe	Val	Phe	Gln	Val	Phe	Cys	Val	Gly	Leu	245	250		255
Trp	Cys	Leu	Asp	Glu	Tyr	Trp	Tyr	Tyr	Ser	Val	Phe	Thr	Leu	Ser	Met	260	265		270
Leu	Val	Ala	Phe	Glu	Ala	Ser	Leu	Val	Gln	Gln	Gln	Met	Arg	Asn	Met	275	280		285
Ser	Glu	Ile	Arg	Lys	Met	Gly	Asn	Lys	Pro	His	Met	Ile	Gln	Val	Tyr	290	295		300
Arg	Ser	Arg	Lys	Trp	Arg	Pro	Ile	Ala	Ser	Asp	Glu	Ile	Val	Pro	Gly	305	310		320
Asp	Ile	Val	Ser	Ile	Gly	Arg	Ser	Pro	Gln	Glu	Asn	Leu	Val	Pro	Cys	325	330		335
Asp	Val	Leu	Leu	Leu	Arg	Gly	Arg	Cys	Ile	Val	Asp	Glu	Ala	Met	Leu	340	345		350
Thr	Gly	Glu	Ser	Val	Pro	Gln	Met	Lys	Glu	Pro	Ile	Glu	Asp	Leu	Ser	355	360		365
Pro	Asp	Arg	Val	Leu	Asp	Leu	Gln	Ala	Asp	Ser	Arg	Leu	His	Val	Ile	370	375		380
Phe	Gly	Gly	Thr	Lys	Val	Val	Gln	His	Ile	Pro	Pro	Gln	Lys	Ala	Thr	385	390		400
Thr	Gly	Leu	Lys	Pro	Val	Asp	Ser	Gly	Cys	Val	Ala	Tyr	Val	Leu	Arg	405	410		415
Thr	Gly	Phe	Asn	Thr	Ser	Gln	Gly	Lys	Leu	Leu	Arg	Thr	Ile	Leu	Phe	420	425		430
Gly	Val	Lys	Arg	Val	Thr	Ala	Asn	Asn	Leu	Glu	Thr	Phe	Ile	Phe	Ile	435	440		445
Leu	Phe	Leu	Leu	Val	Phe	Ala	Ile	Ala	Ala	Ala	Ala	Tyr	Val	Trp	Ile	450	455		460

- 37 -

Glu	Gly	Thr	Lys	Asp	Pro	Ser	Arg	Asn	Arg	Tyr	Lys	Leu	Phe	Leu	Glu	465	470	475	480
Cys	Thr	Leu	Ile	Leu	Thr	Ser	Val	Val	Pro	Pro	Glu	Leu	Pro	Ile	Glu	485	490		495
Leu	Ser	Leu	Ala	Val	Asn	Thr	Ser	Leu	Ile	Ala	Leu	Ala	Lys	Leu	Tyr	500	505	510	
Met	Tyr	Cys	Thr	Glu	Pro	Phe	Arg	Ile	Pro	Phe	Ala	Gly	Lys	Val	Glu	515	520	525	
Val	Cys	Cys	Phe	Asp	Lys	Thr	Gly	Thr	Leu	Thr	Ser	Asp	Ser	Leu	Val	530	535	540	
Val	Arg	Gly	Val	Ala	Gly	Leu	Arg	Asp	Gly	Lys	Glu	Val	Thr	Pro	Val	545	550	555	560
Ser	Ser	Ile	Pro	Val	Glu	Thr	His	Arg	Ala	Leu	Ala	Ser	Cys	His	Ser	565	570		575
Leu	Met	Gln	Leu	Asp	Asp	Gly	Thr	Leu	Val	Gly	Asp	Pro	Leu	Glu	Lys	580	585		590
Ala	Met	Leu	Thr	Ala	Val	Asp	Trp	Thr	Leu	Thr	Lys	Asp	Glu	Lys	Val	595	600	605	
Phe	Pro	Arg	Ser	Ile	Lys	Thr	Gln	Gly	Leu	Lys	Ile	His	Gln	Arg	Phe	610	615	620	
His	Phe	Ala	Ser	Ala	Leu	Lys	Arg	Met	Ser	Val	Leu	Ala	Ser	Tyr	Glu	625	630	635	640
Lys	Leu	Gly	Ser	Thr	Asp	Leu	Cys	Tyr	Ile	Ala	Ala	Val	Lys	Gly	Ala	645	650		655
Pro	Glu	Thr	Leu	His	Ser	Met	Phe	Ser	Gln	Cys	Pro	Pro	Asp	Tyr	His	660	665		670
His	Ile	His	Thr	Glu	Ile	Ser	Arg	Glu	Gly	Ala	Arg	Val	Leu	Ala	Leu	675	680	685	
Gly	Tyr	Lys	Glu	Leu	Gly	His	Leu	Thr	His	Gln	Gln	Ala	Arg	Glu	Val	690	695	700	
Lys	Arg	Glu	Ala	Leu	Glu	Cys	Ser	Leu	Lys	Phe	Val	Gly	Phe	Ile	Val	705	710	715	720
Val	Ser	Cys	Pro	Leu	Lys	Ala	Asp	Ser	Lys	Ala	Val	Ile	Arg	Glu	Ile	725	730		735
Gln	Asn	Ala	Ser	His	Arg	Val	Val	Met	Ile	Thr	Gly	Asp	Asn	Pro	Leu	740	745	750	
Thr	Ala	Cys	His	Val	Ala	Gln	Glu	Leu	His	Phe	Ile	Glu	Lys	Ala	His	755	760	765	
Thr	Leu	Ile	Leu	Gln	Pro	Pro	Ser	Glu	Lys	Gly	Arg	Gln	Cys	Glu	Trp	770	775	780	

- 38 -

Arg Ser Ile Asp Gly Ser Ile Val Leu Pro Leu Ala Arg Gly Ser Pro
 785 790 795 800
 Lys Ala Leu Ala Leu Glu Tyr Ala Leu Cys Leu Thr Gly Asp Gly Leu
 805 810 815
 Ala His Leu Gln Ala Thr Asp Pro Gln Gln Leu Leu Arg Leu Ile Pro
 820 825 830
 His Val Gln Val Phe Ala Arg Val Ala Pro Lys Gln Lys Glu Phe Val
 835 840 845
 Ile Thr Ser Leu Lys Glu Leu Gly Tyr Val Thr Leu Met Cys Gly Asp
 850 855 860
 Gly Thr Asn Asp Val Gly Ala Leu Lys His Ala Asp Val Gly Val Ala
 865 870 875 880
 Leu Leu Ala Asn Ala Pro Glu Arg Val Val Glu Arg Arg Arg Arg Pro
 885 890 895
 Arg Asp Ser Pro Thr Leu Ser Asn Ser Gly Ile Arg Ala Thr Ser Arg
 900 905 910
 Thr Ala Lys Gln Arg Ser Gly Leu Pro Pro Ser Glu Glu Gln Pro Thr
 915 920 925
 Ser Gln Arg Asp Arg Leu Ser Gln Val Leu Arg Asp Leu Glu Asp Glu
 930 935 940
 Ser Thr Pro Ile Val Lys Leu Gly Asp Ala Ser Ile Ala Ala Pro Phe
 945 950 955 960
 Thr Ser Lys Leu Ser Ser Ile Gln Cys Ile Cys His Val Ile Lys Gln
 965 970 975
 Gly Arg Cys Thr Leu Val Thr Thr Leu Gln Met Phe Lys Ile Leu Ala
 980 985 990
 Leu Asn Ala Leu Ile Leu Ala Tyr Ser Gln Ser Val Leu Tyr Leu Glu
 995 1000 1005
 Gly Val Lys Phe Ser Asp Phe Gln Ala Thr Leu Gln Gly Leu Leu Leu
 1010 1015 1020
 Ala Gly Cys Phe Leu Phe Ile Ser Arg Ser Lys Pro Leu Lys Thr Leu
 1025 1030 1035 1040
 Ser Arg Glu Arg Pro Leu Pro Asn Ile Phe Asn Leu Tyr Thr Ile Leu
 1045 1050 1055
 Thr Val Met Leu Gln Phe Phe Val His Phe Leu Ser Leu Val Tyr Leu
 1060 1065 1070
 Tyr Arg Glu Ala Gln Ala Arg Ser Pro Glu Lys Gln Glu Gln Phe Val
 1075 1080 1085
 Asp Leu Tyr Lys Glu Phe Glu Pro Ser Leu Val Asn Ser Thr Val Tyr
 1090 1095 1100

Ile Met Ala Met Ala Met Gln Met Ala Thr Phe Ala Ile Asn Tyr Lys
 105 1110 1115 1120

Gly Pro Pro Phe Met Glu Ser Leu Pro Glu Asn Lys Pro Leu Val Trp
 1125 1130 1135

Ser Leu Ala Val Ser Leu Leu Ala Ile Ile Gly Leu Leu Leu Gly Ser
 1140 1145 1150

Ser Pro Asp Phe Asn Ser Gln Phe Gly Leu Val Asp Ile Pro Val Glu
 1155 1160 1165

Phe Lys Leu Val Ile Ala Gln Val Leu Leu Leu Asp Phe Cys Leu Ala
 1170 1175 1180

Leu Leu Ala Asp Arg Val Leu Gln Phe Phe Leu Gly Thr Pro Lys Leu
 185 1190 1195 1200

Lys Val Pro Ser

<210> 10

<211> 3612

<212> DNA

<213> Homo sapiens

<400> 10

atggcggcag cggcggcggt gggcaacgcg gtgccctgcg gggcccgcc ttgcgggggtc 60
 cggcctgacg ggcagcccaa gcccgggccg cagccgcgcg cgctccttgc cgccggggccg 120
 gcgctcatag cgaacgggtga cgagctgggtg gctgccgtgt ggccgtaccg gcggttggcg 180
 ctgttgccgc gcctcacggg gctgccattc gccgggctgc tttaccggc ctggttgggt 240
 gccgcagccg ctggctgctg gggctggggc agcagttggg tgcagatccc cgaagctgcg 300
 ctgctcgtgc ttgccaccat ctgcctcgcg cagcgcgtca ctgtcctctc ggggcattgg 360
 tctgtgcacg cgcattgcgc gctcacctgc acccggagt acgacccag caaagcgacc 420
 tttgtgaagg tggcgccaac cccaacaat ggctccacgg agctcgtggc cctgcacgcg 480
 aatgagggcg aagacgggct tgaggtgctg tccttcgaat tccagaagat caagtattcc 540
 tacgatgcc tggagaagaa gcagtttctc ccctggcct ttcctgtggg aaacgccttc 600
 tcatactatc agagcaacag aggcttcag gaagactcag agatccgagc agctgagaag 660
 aaatttgga gcaacaaggc cgagatgggt gtgcctgact tctcggagct tttcaaggag 720
 agagccacag ccccttctt tgtatttcag gtgttctgtg tggggctctg gtgcctggat 780
 gagtactggt actacagcgt ctttacgcta tccatgctgg tggcgttcga ggcctcgcgt 840
 gtgcagcagc agatgcggaa catgtcggag atccggaaga tgggcaaca gcccacatg 900
 atccaggtct accgaagccg caagtggagg ccattgccg gtgatgagat cgtaccaggg 960
 gacatcgtct ccacggccg ctcccacag gagaacctgg tgccatgtga cgtgcttctg 1020

ctgcgaggcc gctgcatcgt agacgaggcc atgctcacgg gggagtcctg gccacagatg 1080
aaggagccca tcgaagacct cagcccagac cgggtgctgg acctccaggc tgattcccgg 1140
ctgcacgtca tcttcggggg caccaaggtg gtgcagcaca tccccccaca gaaagccacc 1200
acggggcctga agccggttga cagcgggtgc gtggcctacg tcctgcgga cggattcaac 1260
acatctcagg gcaagctgct gcgcaccatc ctcttcgggg tcaagagggt gactgcgaac 1320
aacctggaga ccttcattct catcctcttc ctctgggtgt ttgccatcgc tgcagctgcc 1380
tatgtatgga ttgaaggtag caaggacccc agccggaacc gctacaagct gtttctggag 1440
tgcaccctga tcctcacctc ggtcgtgcct cctgagctgc ccatcgagct gtccctggcc 1500
gtcaacacct ccctcatcgc cctggccaag ctctacatgt actgcacaga gcccttccgg 1560
atcccccttg ctggcaaggc cgagggtgtg tgctttgaca agacggggac gttgaccagt 1620
gacagcctgg tgggtgcgcg tgtggccggg ctgagagacg ggaaggaggt gaccccagt 1680
tccagcatcc ctgtagaac acaccgggccc ctggcctcgt gccactcgt catgcagctg 1740
gacgaaggca ccctcgtggg tgaccctcta gagaaggcca tgctgacggc cgtggactgg 1800
acgtgacca aagatgagaa agtattcccc cgaagtatta aaactcagg gctgaaaatt 1860
caccagcgt ttcattttgc cagtgcctg aagcgaatgt ccgtgcttgc ctctatgag 1920
aagctgggct ccaccgacct ctgctacatc gcggccgtga agggggcccc cgaaactctg 1980
cactccatgt tctcccagt cccgcccagc taccaccaca tccacaccga gatctcccgg 2040
gaaggagccc gcgtcctggc gctggggtac aaggagctgg gacacctcac tcaccagcag 2100
gcccgggagg tcaagcggga ggccctggag tgcagcctca agttcgtcgg cttcattgtg 2160
gtctcctgcc cgctcaaggc tgactccaag gccgtgatcc gggagatcca gaatgcgtcc 2220
caccgggtgg tcatgatcac gggagacaac ccgtcactg catgocacgt ggcccaggag 2280
ctgcacttca ttgaaaaggc ccacacgtg atcctgcagc ctccctccga gaaaggccgg 2340
cagtgcgagt ggcgtccat tgacggcagc atcgtgctgc ccctggccc gggctcccca 2400
aaggcactgg ccctggagta cgcactgtgc ctacaggcg acggcttggc ccacctgcag 2460
gccaccgacc ccagcagct gctccgcctc atccccatg tgcagggtgt cggccgtgtg 2520
gtccccaagc agaaggagtt tgtcatcacc agcctgaagg agctgggcta cgtgaccctc 2580
atgtgtggg atggcaccaa cgacgtgggc gccctgaagc atgtgacgt ggggtgtggc 2640
ctcttgcca atgcccctga gcgggtgtg gagcgcgac ggcggcccc ggacagccca 2700
accctgagca acagtggcat cagagccacc tccaggacag ccaagcagcg gtcggggctc 2760
cctccctccg aggagcagc aacctcccag agggaccgcc tgagccaggt gctgcgagac 2820

ctcgaggacg agagtacgcc cattgtgaaa ctgggggatg ccagcatcgc agcacccttc 2880
acctccaagc tctcatccat ccagtgcacg tgccacgtga tcaagcaggg ccgctgcacg 2940
ctgggtgacca cgctacagat gttcaagatc ctgggogtca atgccctcat cctggcctac 3000
agccagagcg tcctctacct ggagggagtc aagttcagtg acttccaggc caccctacag 3060
gggctgctgc tggccggctg cttcctcttc atctcccgtt ccaagcccct caagaccctc 3120
tcccgagaac ggcccctgcc caacatcttc aaactgtaca ccacctcac cgtcatgctc 3180
cagttctttg tgcacttcct gagccttgtc tacctgtacc gtgaggccca ggcccggagc 3240
cccgagaagc aggagcagtt cgtggacttg tacaaggagt ttgagccaag cctgggtcaac 3300
agcaccgtct acatcatggc catggccatg cagatggcca ccttcgccat caattacaaa 3360
ggcccgcctt tcatggagag cctgcccag aacaagcccc tgggtgtggag tctggcagtt 3420
tcactcctgg ccatcattgg cctgctctc ggctcctcgc ccgacttcaa cagccagttt 3480
ggcctcgtgg acatccctgt ggagttcaag ctgggtcattg cccaggtcct gctcctggac 3540
ttctgctgg cgctcctggc cgaccgcgtc ctgcagttct tcctggggac cccgaagctg 3600
aaagtcctt cc 3612

<210> 11

<211> 7

<212> PRT

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 6

<223> Xaa may be Leu, Ile, Val, or Met

<220>

<223> Description of Artificial Sequence:consensus
sequence

<220>

<221> misc_feature

<222> 7

<223> Xaa may be Thr or Ile

<400> 11

Asp Lys Thr Gly Thr Xaa Xaa
1 5

<210> 12

<211> 1157

<212> PRT

<213> Caenorhabditis elegans

<400> 12

Met Gly Val Asp Gln Leu Val Glu Thr Ile Ile Pro Tyr Asn Leu Arg
1 5 10 15

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

Thr Lys Ile Val Gln His Thr Ala Pro Gly Lys Ala Ala Glu Gly Met
 340 345 350
 Val Lys Ser Pro Asp Gly Asn Cys Ile Cys Tyr Val Ile Arg Thr Gly
 355 360 365
 Phe Asn Thr Ser Gln Gly Lys Leu Leu Arg Thr Ile Met Phe Gly Val
 370 375 380
 Lys Lys Ala Thr Ala Asn Asn Leu Glu Thr Phe Cys Phe Ile Leu Phe
 385 390 395 400
 Leu Leu Ile Phe Ala Ile Ala Ala Ala Tyr Leu Trp Ile Lys Gly
 405 410 415
 Ser Val Asp Glu Thr Arg Ser Lys Tyr Lys Leu Phe Leu Glu Cys Thr
 420 425 430
 Leu Ile Leu Thr Ser Val Ile Pro Pro Glu Leu Pro Ile Glu Leu Ser
 435 440 445
 Leu Ala Val Asn Ser Ser Leu Met Ala Leu Gln Lys Leu Gly Ile Phe
 450 455 460
 Cys Thr Glu Pro Phe Arg Ile Pro Phe Ala Gly Lys Val Asp Ile Cys
 465 470 475 480
 Cys Phe Asp Lys Thr Gly Thr Leu Thr Thr Asp Asn Leu Val Val Glu
 485 490 495
 Gly Val Ala Leu Asn Asn Gln Lys Glu Gly Met Ile Arg Asn Ala Glu
 500 505 510
 Asp Leu Pro His Glu Ser Leu Gln Val Leu Ala Ser Cys His Ser Leu
 515 520 525
 Val Arg Phe Glu Glu Asp Leu Val Gly Asp Pro Leu Glu Lys Ala Cys
 530 535 540
 Leu Ser Trp Cys Gly Trp Asn Leu Thr Lys Gly Asp Ala Val Met Pro
 545 550 555 560
 Pro Lys Thr Ala Ala Lys Gly Ile Ser Gly Ile Lys Ile Phe His Arg
 565 570 575
 Tyr His Phe Ser Ser Ala Met Lys Arg Met Thr Val Val Ala Gly Tyr
 580 585 590
 Gln Ser Pro Gly Thr Ser Asp Thr Thr Phe Ile Val Ala Val Lys Gly
 595 600 605
 Ala Pro Glu Val Leu Arg Asn Met Tyr Ala Asp Leu Pro Ser Asp Tyr
 610 615 620
 Asp Glu Thr Tyr Thr Arg Leu Thr Arg Gln Gly Ser Arg Val Leu Ala
 625 630 635 640
 Met Gly Ile Arg Lys Leu Gly Glu Thr Arg Val Gly Glu Leu Arg Asp
 645 650 655

Lys Lys Arg Glu Asn Phe Glu Asn Asp Leu Ala Phe Ala Gly Phe Val
 660 665 670
 Val Ile Ser Cys Pro Leu Lys Ser Asp Thr Lys Thr Met Ile Arg Glu
 675 680 685
 Ile Met Asp Ser Ser His Val Val Ala Met Ile Thr Gly Asp Asn Pro
 690 695 700
 Leu Thr Ala Cys His Val Ser Lys Val Leu Lys Phe Thr Lys Lys Ser
 705 710 715 720
 Leu Pro Thr Leu Val Leu Asp Glu Pro Ala Asp Gly Val Asp Trp Met
 725 730 735
 Trp Lys Ser Val Asp Gly Thr Ile Glu Leu Pro Leu Lys Pro Glu Thr
 740 745 750
 Lys Asn Lys Met Glu Arg Lys Ala Phe Phe Asn Ser His Glu Phe Cys
 755 760 765
 Leu Thr Gly Ser Ala Phe His His Leu Val His Asn Glu His Thr Phe
 770 775 780
 Leu Arg Glu Leu Ile Leu His Val Lys Val Phe Ala Arg Met Ala Pro
 785 790 795 800
 Lys Gln Lys Glu Arg Ile Ile Asn Glu Leu Lys Ser Leu Gly Lys Val
 805 810 815
 Thr Leu Met Cys Gly Asp Gly Thr Asn Asp Val Gly Ala Leu Lys His
 820 825 830
 Ala Asn Val Gly Val Ala Leu Leu Thr Asn Pro Tyr Asp Ala Glu Lys
 835 840 845
 Ala Ala Glu Lys Glu Lys Glu Lys Lys Ala Lys Ile Glu Glu Ala Arg
 850 855 860
 Ser Leu Val Arg Ser Gly Ala Gln Leu Pro Gln Arg Pro Gly Ala Pro
 865 870 875 880
 Gly Ala Pro Pro Ala Ala Asn Ala Ala Arg Pro Arg Leu Asp Asn Leu
 885 890 895
 Met Lys Glu Leu Glu Glu Glu Glu Lys Ala Gln Val Ile Lys Leu Gly
 900 905 910
 Asp Ala Ser Ile Ala Ala Pro Phe Thr Ser Lys Tyr Thr Ser Ile Ala
 915 920 925
 Ser Ile Cys His Val Ile Lys Gln Gly Arg Cys Thr Leu Val Thr Thr
 930 935 940
 Leu Gln Met Phe Lys Ile Leu Ala Leu Asn Ala Leu Val Ser Ala Tyr
 945 950 955 960
 Ser Leu Ser Ala Leu Tyr Leu Asp Gly Val Lys Phe Ser Asp Thr Gln
 965 970 975

Ala Thr Ile Gln Gly Leu Leu Leu Ala Ala Cys Phe Leu Phe Ile Ser
 980 985 990
 Lys Ser Lys Pro Leu Lys Thr Leu Ser Arg Gln Arg Pro Met Ala Asn
 995 1000 1005
 Ile Phe Asn Ala Tyr Thr Leu Leu Thr Val Thr Leu Gln Phe Ile Val
 1010 1015 1020
 His Phe Ser Cys Leu Leu Tyr Ile Val Gly Leu Ala His Glu Ala Asn
 1025 1030 1035 1040
 Thr Glu Lys Ala Pro Val Asp Leu Glu Ala Lys Phe Thr Pro Asn Ile
 1045 1050 1055
 Leu Asn Thr Thr Val Tyr Ile Ile Ser Met Ala Leu Gln Val Cys Thr
 1060 1065 1070
 Phe Ala Val Asn Tyr Arg Gly Arg Pro Phe Met Glu Ser Leu Phe Glu
 1075 1080 1085
 Asn Lys Ala Met Leu Tyr Ser Ile Met Phe Ser Gly Gly Ala Val Phe
 1090 1095 1100
 Thr Leu Ala Ser Gly Gln Ala Thr Asp Leu Met Ile Gln Phe Glu Leu
 1105 1110 1115 1120
 Val Val Leu Pro Glu Ala Leu Arg Asn Ala Leu Leu Met Cys Val Thr
 1125 1130 1135
 Ala Asp Leu Val Ile Cys Tyr Ile Ile Asp Arg Gly Leu Asn Phe Leu
 1140 1145 1150
 Leu Gly Asp Met Phe
 1155

<210> 13

<211> 9

<212> PRT

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1

<223> Xaa may be Asp, Asn, or Ser

<220>

<221> misc_feature

<222> 2

<223> Xaa may be Gln, Glu, Asn, or Arg

<220>

<221> misc_feature

<222> 3

<223> Xaa may be Ser or Ala

<220>
<221> misc_feature
<222> 4
<223> Xaa may be Leu, Ile, Val, Ser, Ala
or Asn

<220>
<221> misc_feature
<222> 5
<223> Xaa may be Leu, Ile, or Val

<220>
<221> misc_feature
<222> 6
<223> Xaa may be Thr, Ser, or Asn

<220>
<221> misc_feature
<222> 9
<223> Xaa may be Ser or Asn

<220>
<223> Description of Artificial Sequence:consensus
sequence

<400> 13
Xaa Xaa Xaa Xaa Xaa Xaa Gly Glu Xaa
1 5

<210> 14
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<221> misc_feature
<222> 1
<223> Xaa may be Leu, Ile or Val

<220>
<221> misc_feature
<222> 2
<223> Xaa may be Cys, Ala, Met, or Leu

<220>
<221> misc_feature
<222> 3
<223> Xaa may be Ser, Thr, Phe or Leu

<220>
<221> misc_feature
<222> 9
<223> Xaa may be Leu or Ile

<220>
<223> Description of Artificial Sequence:consensus
sequence

- 47 -

<400> 14

Xaa Xaa Xaa Asp Lys Thr Gly Thr Xaa Thr
1 5 10

<210> 15

<211> 11

<212> PRT

<213> Artificial Sequence

<220>

<221> misc_feature

<222> 1

<223> Xaa may be Thr, Ile, or Val

<220>

<221> misc_feature

<222> 5

<223> Xaa may be any amino acid

<220>

<221> misc_feature

<222> 8

<223> Xaa may be Ala, Ser or Gly

<220>

<221> misc_feature

<222> 10

<223> Xaa may be Ala, Ser, or Val

<220>

<223> Description of Artificial Sequence:consensus
sequence

<400> 15

Xaa Gly Asp Gly Xaa Asn Asp Xaa Pro Xaa Leu
1 5 10

<210> 16

<211> 7

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:amino acid
residues important for calcium transport

<400> 16

Ile Pro Glu Gly Leu Pro Ala
1 5

<210> 17

<211> 28

<212> PRT

<213> Homo sapiens

<400> 17

Asp Leu Val Thr Val Val Val Pro Pro Ala Leu Pro Ala Ala Met Thr
1 5 10 15

Val Cys Thr Leu Tyr Ala Gln Ser Arg Leu Arg Arg
20 25

<210> 18
<211> 28
<212> PRT
<213> Homo sapiens

<400> 18
Asp Ile Ile Thr Ile Thr Val Pro Pro Ala Leu Pro Ala Ala Met Thr
1 5 10 15

Ala Gly Ile Val Tyr Ala Gln Arg Arg Leu Lys Lys
20 25

<210> 19
<211> 28
<212> PRT
<213> Homo sapiens

<400> 19
Leu Ile Leu Thr Ser Val Val Pro Pro Glu Leu Pro Ile Glu Leu Ser
1 5 10 15

Leu Ala Val Asn Thr Ser Leu Ile Ala Leu Ala Lys
20 25

<210> 20
<211> 29
<212> PRT
<213> Saccharomyces cerevisiae

<400> 20
Asp Ile Ile Thr Ile Val Val Pro Pro Ala Leu Pro Ala Thr Leu Thr
1 5 10 15

Ile Gly Thr Asn Phe Ala Leu Ser Arg Leu Lys Glu Lys
20 25

<210> 21
<211> 28
<212> PRT
<213> Saccharomyces cerevisiae

<400> 21
Leu Ile Ile Thr Ser Val Val Pro Pro Glu Leu Pro Met Glu Leu Thr
1 5 10 15

Met Ala Val Asn Ser Ser Leu Ala Ala Leu Ala Lys
20 25

<210> 22
<211> 28
<212> PRT
<213> Schizosaccharomyces pombe

<400> 22

Val Leu Thr Ile Leu Val Pro Pro Ala Leu Pro Ala Thr Leu Ser Val
1 5 10 15

Gly Ile Ala Asn Ser Ile Ala Arg Leu Ser Arg Ala
20 25

<210> 23

<211> 29

<212> PRT

<213> *Caenorhabditis elegans*

<400> 23

Asp Leu Val Thr Ile Val Val Pro Pro Ala Leu Pro Ala Val Met Gly
1 5 10 15

Ile Gly Ile Phe Tyr Ala Gln Arg Arg Leu Arg Gln Lys
20 25

<210> 24

<211> 29

<212> PRT

<213> *Caenorhabditis elegans*

<400> 24

Asp Ile Ile Thr Ile Val Val Pro Pro Ala Leu Pro Ala Ala Met Ser
1 5 10 15

Val Gly Ile Ile Asn Ala Asn Ser Arg Leu Lys Lys Lys
20 25

<210> 25

<211> 29

<212> PRT

<213> *Caenorhabditis elegans*

<400> 25

Asp Ile Ile Thr Ile Thr Val Pro Pro Ala Leu Pro Ala Ala Met Ser
1 5 10 15

Val Gly Ile Ile Asn Ala Gln Leu Arg Leu Lys Lys Lys
20 25

<210> 26

<211> 29

<212> PRT

<213> *Caenorhabditis elegans*

<400> 26

Leu Ile Leu Thr Ser Val Ile Pro Pro Glu Leu Pro Ile Glu Leu Ser
1 5 10 15

Leu Ala Val Asn Ser Ser Leu Met Ala Leu Gln Lys Leu
20 25

- 50 -

<210> 27
<211> 29
<212> PRT
<213> Homo sapiens

<400> 27
Ala Leu Ala Val Ala Ala Ile Pro Glu Gly Leu Pro Ala Val Ile Thr
1 5 10 15
Thr Cys Leu Ala Leu Gly Thr Arg Arg Met Ala Lys Lys
20 25

<210> 28
<211> 29
<212> PRT
<213> Oryctolagus cuniculus

<400> 28
Ala Leu Ala Val Ala Ala Ile Pro Glu Gly Leu Pro Ala Val Ile Thr
1 5 10 15
Thr Cys Leu Ala Leu Gly Thr Arg Arg Met Ala Lys Lys
20 25

<210> 29
<211> 29
<212> PRT
<213> Gallus gallus

<400> 29
Ala Leu Ala Val Ala Ala Ile Pro Glu Gly Leu Pro Ala Val Ile Thr
1 5 10 15
Thr Cys Leu Ala Leu Gly Thr Arg Arg Met Ala Lys Lys
20 25

<210> 30
<211> 29
<212> PRT
<213> Felis catus

<400> 30
Ala Leu Ala Val Ala Ala Ile Pro Glu Gly Leu Pro Ala Val Ile Thr
1 5 10 15
Thr Cys Leu Ala Leu Gly Thr Arg Arg Met Ala Lys Lys
20 25

<210> 31
<211> 29
<212> PRT
<213> Procambarus clarkii

<400> 31
Ala Leu Ala Val Ala Ala Ile Pro Glu Gly Leu Pro Ala Val Ile Thr
1 5 10 15

- 51 -

Thr Cys Leu Ala Leu Gly Thr Arg Arg Met Ala Lys Lys
20 25

<210> 32
<211> 29
<212> PRT
<213> Homo sapiens

<400> 32
Ala Leu Ala Val Ala Ala Ile Pro Glu Gly Leu Pro Ala Val Ile Thr
1 5 10 15

Thr Cys Leu Ala Leu Gly Thr Arg Arg Met Ala Lys Lys
20 25

<210> 33
<211> 29
<212> PRT
<213> Homo sapiens

<400> 33
Ala Leu Ala Val Ala Ala Ile Pro Glu Gly Leu Pro Ala Val Ile Thr
1 5 10 15

Thr Cys Leu Ala Leu Gly Thr Arg Arg Met Ala Arg Lys
20 25

<210> 34
<211> 29
<212> PRT
<213> Drosophila melanogaster

<400> 34
Ala Val Ala Val Ala Ala Ile Pro Glu Gly Leu Pro Ala Val Ile Thr
1 5 10 15

Thr Cys Leu Ala Leu Gly Thr Arg Arg Met Ala Lys Lys
20 25

<210> 35
<211> 29
<212> PRT
<213> Saccharomyces cerevisiae

<400> 35
Ser Leu Ala Val Ala Ala Ile Pro Glu Gly Leu Pro Ile Ile Val Thr
1 5 10 15

Val Thr Leu Ala Leu Gly Val Leu Arg Met Ala Lys Arg
20 25

<210> 36
<211> 29
<212> PRT
<213> Saccharomyces cerevisiae

<400> 36

Thr Val Ile Val Val Ala Val Pro Glu Gly Leu Pro Leu Ala Val Thr
1 5 10 15

Leu Ala Leu Ala Phe Ala Thr Thr Arg Met Thr Lys Asp
20 25

<210> 37

<211> 29

<212> PRT

<213> Homo sapiens

<400> 37

Thr Val Leu Val Val Ala Val Pro Glu Gly Leu Pro Leu Ala Val Thr
1 5 10 15

Ile Ser Leu Ala Tyr Ser Val Lys Lys Met Met Lys Asp
20 25

<210> 38

<211> 29

<212> PRT

<213> Homo sapiens

<400> 38

Thr Val Leu Val Val Ala Val Pro Glu Gly Leu Pro Leu Ala Val Thr
1 5 10 15

Ile Ser Leu Ala Tyr Ser Val Lys Lys Met Met Lys Asp
20 25

<210> 39

<211> 29

<212> PRT

<213> Rattus norvegicus

<400> 39

Thr Val Leu Val Val Ala Val Pro Glu Gly Leu Pro Leu Ala Val Thr
1 5 10 15

Ile Ser Leu Ala Tyr Ser Val Lys Lys Met Met Lys Asp
20 25

<210> 40

<211> 29

<212> PRT

<213> Homo sapiens

<400> 40

Thr Val Leu Val Val Ala Val Pro Glu Gly Leu Pro Leu Ala Val Thr
1 5 10 15

Ile Ser Leu Ala Tyr Ser Val Lys Lys Met Met Lys Asp
20 25

<210> 41
<211> 29
<212> PRT
<213> Homo sapiens

<400> 41
Thr Val Leu Val Val Ala Val Pro Glu Gly Leu Pro Leu Ala Val Thr
1 5 10 15
Ile Ser Leu Ala Tyr Ser Val Lys Lys Met Met Lys Asp
20 25

<210> 42
<211> 29
<212> PRT
<213> Saccharomyces cerevisiae

<400> 42
Thr Val Leu Ile Val Ser Cys Pro Cys Val Ile Gly Leu Ala Val Pro
1 5 10 15
Ile Val Phe Val Ile Ala Ser Gly Val Ala Ala Lys Arg
20 25

<210> 43
<211> 29
<212> PRT
<213> Homo sapiens

<400> 43
Thr Val Leu Cys Ile Ala Cys Pro Cys Ser Leu Gly Leu Ala Thr Pro
1 5 10 15
Thr Ala Val Met Val Gly Thr Gly Val Gly Ala Gln Asn
20 25

<210> 44
<211> 29
<212> PRT
<213> Homo sapiens

<400> 44
Thr Val Leu Cys Ile Ala Cys Pro Cys Ser Leu Gly Leu Ala Thr Pro
1 5 10 15
Thr Ala Val Met Val Gly Thr Gly Val Ala Ala Gln Asn
20 25

<210> 45
<211> 29
<212> PRT
<213> Drosophila melanogaster

<400> 45
Gly Ile Ile Val Ala Asn Val Pro Glu Gly Leu Leu Ala Thr Val Thr
1 5 10 15

- 54 -

Val Cys Leu Thr Leu Thr Ala Lys Arg Met Ala Ser Lys
20 25

<210> 46
<211> 29
<212> PRT
<213> Hydra vulgaris

<400> 46
Gly Ile Ile Val Ala Asn Val Pro Glu Gly Leu Leu Ala Thr Val Thr
1 5 10 15

Val Cys Leu Thr Leu Thr Ala Lys Lys Met Ala Lys Lys
20 25

<210> 47
<211> 29
<212> PRT
<213> Bufo marinus

<400> 47
Gly Ile Ile Val Ala Asn Val Pro Glu Gly Leu Leu Ala Thr Val Thr
1 5 10 15

Val Cys Leu Thr Leu Thr Ala Lys Arg Met Ala Arg Lys
20 25

<210> 48
<211> 29
<212> PRT
<213> Homo sapiens

<400> 48
Gly Ile Ile Val Ala Asn Val Pro Glu Gly Leu Leu Ala Thr Val Thr
1 5 10 15

Val Cys Leu Thr Leu Thr Ala Lys Arg Met Ala Arg Lys
20 25

<210> 49
<211> 29
<212> PRT
<213> Homo sapiens

<400> 49
Gly Ile Ile Val Ala Asn Val Pro Glu Gly Leu Leu Ala Thr Val Thr
1 5 10 15

Val Cys Leu Thr Leu Thr Ala Lys Arg Met Ala Arg Lys
20 25

<210> 50
<211> 29
<212> PRT
<213> Homo sapiens

<400> 50

Gly Ile Ile Val Ala Asn Val Pro Glu Gly Leu Leu Ala Thr Val Thr
1 5 10 15

Val Cys Leu Thr Val Thr Ala Lys Arg Met Ala Arg Lys
20 25

<210> 51

<211> 29

<212> PRT

<213> Homo sapiens

<400> 51

Ile Leu Phe Asn Asn Leu Ile Pro Ile Ser Leu Leu Val Thr Leu Glu
1 5 10 15

Val Val Lys Phe Thr Gln Ala Tyr Phe Ile Asn Trp Asp
20 25

<210> 52

<211> 29

<212> PRT

<213> Saccharomyces cerevisiae

<400> 52

Ile Leu Phe Ser Asn Leu Val Pro Ile Ser Leu Phe Val Thr Val Glu
1 5 10 15

Leu Ile Lys Tyr Tyr Gln Ala Phe Met Ile Gly Ser Asp
20 25