

WO 02/06302 A2



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).

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.

Published:

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

16816 AND 16839, NOVEL HUMAN PHOSPHOLIPASE C MOLECULES AND USES THEREFOR

Background of the Invention

5 Phospholipase C (PLC) belongs to a family of enzymes, also known as disulfide isomerases, which play an important role in mediating signal transduction pathways. Many extracellular signaling molecules including hormones, growth factors, neurotransmitters, and immunoglobulins bind to their respective cell surface receptors and activate PLCs. Activated PLCs then catalyze the hydrolysis of phosphatidyl-inositol-
10 4,5-bisphosphate (PIP₂), a component of the plasma membrane, to produce diacylglycerol and inositol 1,4,5-trisphosphate (IP₃).

 In their respective biochemical pathways, IP₃ and diacylglycerol serve as second messengers and trigger a series of intracellular responses. IP₃ induces the release of calcium from internal cellular storage, and diacylglycerol activates protein kinase C
15 (PKC). Both pathways are part of transmembrane signal transduction mechanisms, which regulate numerous cellular processes, including secretion, neural activity, metabolism, and proliferation.

 PLC molecules have been found in a broad spectrum of organisms including bacteria, simple eukaryotes, plants and animals (Munnik *et al.*, Biochim. Biophys. Acta.
20 1389:222-272, (1998)). Several distinct isoforms of PLC have been identified in animals and are categorized as PLC-beta, PLC-gamma, and PLC-delta. Subtypes are designated by adding Arabic numbers after the Greek letters, e.g., PLC-beta-1. PLCs have a molecular mass of 62-68 kDa, and their amino acid sequences show two regions of significant similarity.

25

Summary of the Invention

 The present invention is based, in part, on the discovery of a novel human phospholipase C, referred to herein as "16816 or 16839". The nucleotide sequence of a cDNA encoding 16816 or 16839 is shown in SEQ ID NO:1 or SEQ ID NO:4, and the
30 amino acid sequence of a 16816 or 16839 polypeptide is shown in SEQ ID NO:2 or SEQ ID NO:5. In addition, the nucleotide sequence of the coding region is depicted in SEQ ID NO:3 or SEQ ID NO:6.

 Accordingly, in one aspect, the invention features a nucleic acid molecule which encodes a 16816 or 16839 protein or polypeptide, e.g., a biologically active portion of the

16816 or 16839 protein. In a preferred embodiment, the isolated nucleic acid molecule encodes a polypeptide having the amino acid sequence of SEQ ID NO:2. In other embodiments, the invention provides an isolated 16816 or 16839 nucleic acid molecule having the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:4, or the sequence
5 of the DNA insert of the plasmid deposited with ATCC Accession Number _____. In still other embodiments, the invention provides nucleic acid molecules that are substantially identical (e.g., naturally occurring allelic variants) to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:4, or the sequence of the DNA insert of the plasmid deposited with ATCC Accession Number _____. In other embodiments, the invention
10 provides a nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:4, or the sequence of the DNA insert of the plasmid deposited with ATCC Accession Number _____, wherein the nucleic acid encodes a full length 16816 or 16839 protein or an active fragment thereof.

15 In a related aspect, the invention further provides nucleic acid constructs which include a 16816 or 16839 nucleic acid molecule described herein. In certain embodiments, the nucleic acid molecules of the invention are operatively linked to native or heterologous regulatory sequences. Also included, are vectors and host cells containing the 16816 or 16839 nucleic acid molecules of the invention e.g., vectors and
20 host cells suitable for producing 16816 or 16839 nucleic acid molecules and polypeptides.

In another related aspect, the invention provides nucleic acid fragments suitable as primers or hybridization probes for the detection of 16816 or 16839-encoding nucleic acids.

In still another related aspect, isolated nucleic acid molecules that are antisense to
25 a 16816 or 16839 encoding nucleic acid molecule are provided.

In another aspect, the invention features 16816 or 16839 polypeptides, and biologically active or antigenic fragments thereof that are useful, e.g., as reagents or targets in assays applicable to treatment and diagnosis of 16816 or 16839-mediated or -related disorders. In another embodiment, the invention provides 16816 or 16839
30 polypeptides having a 16816 or 16839 activity. Preferred polypeptides are 16816 or 16839 proteins including at least one phospholipase C domain, and, preferably, having a 16816 or 16839 activity, e.g., a 16816 or 16839 activity as described herein.

In other embodiments, the invention provides 16816 or 16839 polypeptides, e.g., a 16816 or 16839 polypeptide having the amino acid sequence shown in SEQ ID NO:2 or

SEQ ID NO:5; the amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC Accession Number ____; an amino acid sequence that is substantially identical to the amino acid sequence shown in SEQ ID NO:2 or SEQ ID NO:5; or an amino acid sequence encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:4, or the sequence of the DNA insert of the plasmid deposited with ATCC Accession Number ____, wherein the nucleic acid encodes a full length 16816 or 16839 protein or an active fragment thereof.

10 In a related aspect, the invention further provides nucleic acid constructs which include a 16816 or 16839 nucleic acid molecule described herein.

In a related aspect, the invention provides 16816 or 16839 polypeptides or fragments operatively linked to non-16816 or 16839 polypeptides to form fusion proteins.

15 In another aspect, the invention features antibodies and antigen-binding fragments thereof, that react with, or more preferably specifically bind 16816 or 16839 polypeptides.

In another aspect, the invention provides methods of screening for compounds that modulate the expression or activity of the 16816 or 16839 polypeptides or nucleic acids.

20 In still another aspect, the invention provides a process for modulating 16816 or 16839 polypeptide or nucleic acid expression or activity, e.g. using the screened compounds. In certain embodiments, the methods involve treatment of conditions related to aberrant activity or expression of the 16816 or 16839 polypeptides or nucleic acids, such as conditions involving aberrant or deficient cellular proliferation or differentiation.

25 In another aspect, the invention features, 16816 or 16839 polypeptides, and biologically active or antigenic fragments thereof that are useful, e.g., as reagents or targets in assays applicable to treatment and diagnosis of 16816 or 16839 mediated or related disorders. In another embodiment, the invention provides 16816 or 16839 polypeptides having a 16816 or 16839 activity. Preferred polypeptides are 16816 or 16839 proteins including at least one domain, e.g., a PLC-X domain (from about amino acids 291-436 of SEQ ID NO:2), an EF hand domain (from about 138-166 and 174-202 of SEQ ID NO:2), a PLC-Y domain (from about amino acids 492-609 of SEQ ID NO:2), a calcium binding (C2) domain (from about amino acids 629-719 of SEQ ID NO:2), or a pleckstrin homology (PH) domain (from about amino acids 17-124 of SEQ ID NO:2), and, preferably, having a 16816 or 16839 activity, e.g., an activity as described herein,

e.g., the ability to catalyze the hydrolysis of phosphatidyl-inositol-4,5-bisphosphate (PIP₂) producing diacylglycerol and inositol 1,4,5-trisphosphate (IP₃).

The invention also provides assays for determining the activity of or the presence or absence of 16816 or 16839 polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

In further aspect the invention provides assays for determining the presence or absence of a genetic alteration in a 16816 or 16839 polypeptide or nucleic acid molecule, including for disease diagnosis.

Brief Description of the Drawings

Figures 1A-C depicts a cDNA sequence (SEQ ID NO:1) and predicted amino acid sequence (SEQ ID NO:2) of human 16816. The methionine-initiated open reading frame of human 16816 (without the 5' and 3' untranslated regions) extends from nucleotide position 1 to position 2289 of SEQ ID NO:3, not including the terminal codon.

Figure 2 depicts a hydropathy plot of human 16816. Relatively hydrophobic residues are shown above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The numbers corresponding to the amino acid sequence of human 16816 are indicated. Polypeptides of the invention include fragments which include: all or part of a hydrophobic sequence, e.g., a sequence above the dashed line, e.g., the sequence from about amino acid 485 to 500, from about 650 to 660, and from about 685 to 700 of SEQ ID NO:2; all or part of a hydrophilic sequence, e.g., a sequence below the dashed line, e.g., the sequence from about amino acid 125 to 150, from about 465 to 480, and from about 665 to 680 of SEQ ID NO:2; a sequence which includes a Cys, or a glycosylation site.

Figure 3 depicts an alignment of the PH domain of human 16816 with a consensus amino acid sequence derived from a hidden Markov model (HMM) from PFAM. The upper sequences are the consensus amino acid sequence (SEQ ID NO:7), while the lower amino acid sequences correspond to amino acids 17 to 124 of SEQ ID NO:2.

Figures 4a-b depict an alignment of the EF hand domain of human 16816 with a consensus amino acid sequence derived from a hidden Markov model (HMM) from PFAM. The upper sequences are the consensus amino acid sequence (SEQ ID NOs:8 and

9), while the lower amino acid sequences correspond to amino acids 138 to 166 and 174 to 202 of SEQ ID NO:2.

5 *Figure 5* depicts an alignment of the phosphatidylinositol-specific phospholipase C, X domain of human 16816 with a consensus amino acid sequence derived from a hidden Markov model (HMM) from PFAM. The upper sequences are the consensus amino acid sequence (SEQ ID NO:10), while the lower amino acid sequences correspond to amino acids 291 to 436 of SEQ ID NO:2.

10 *Figure 6* depicts an alignment of the phosphatidylinositol-specific phospholipase C, Y domain of human 16816 with a consensus amino acid sequence derived from a hidden Markov model (HMM) from PFAM. The upper sequences are the consensus amino acid sequence (SEQ ID NO:11), while the lower amino acid sequences correspond to amino acids 492 to 609 of SEQ ID NO:2.

15 *Figure 7* depicts an alignment of the C2 domain of human 16816 with a consensus amino acid sequence derived from a hidden Markov model (HMM) from PFAM. The upper sequences are the consensus amino acid sequence (SEQ ID NO:12), while the lower amino acid sequences correspond to amino acids 629 to 719 of SEQ ID NO:2.

20 *Figure 8* depicts a BLAST alignment of human 16816 with a consensus amino acid sequence derived from a ProDomain “phospholipase phosphodiesterase hydrolase phosphoinositide-specific 1-phosphatidylinositol-45-bisphosphate degradation transducer lipid beta” (Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>). The lower sequence is amino acid residues 2 to 171 of the 170 amino acid consensus sequence (SEQ ID NO:13), while the upper amino acid sequence corresponds to the “phospholipase phosphodiesterase hydrolase phosphoinositide-specific 1-phosphatidylinositol-45-bisphosphate degradation transducer lipid beta” domain of human 16816, amino acid
25 residues 275 to 436 of SEQ ID NO:2.

30 *Figure 9* depicts a BLAST alignment of human 16816 with a consensus amino acid sequence derived from a ProDomain “phospholipase C delta calcium-binding PLC-III hydrolase phosphodiesterase lipid PLC-delta-1 1-phosphatidylinositol-45-bisphosphate” (Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>). The lower sequence is amino acid residues 2 to 202 of the 202 amino acid consensus sequence (SEQ ID NO:14), while the upper amino acid sequence corresponds to the “phospholipase C delta calcium-binding PLC-III hydrolase phosphodiesterase lipid PLC-delta-1 1-phosphatidylinositol-45-bisphosphate” domain of human 16816, amino acid residues 1 to 191 of SEQ ID NO:2.

Figure 10 depicts a BLAST alignment of human 16816 with a consensus amino acid sequence derived from a ProDomain “phospholipase phosphodiesterase hydrolase phosphoinositide-specific 1-phosphatidylinositol-45-bisphosphate degradation lipid transducer beta” (Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>). The lower sequence is amino acid residues 11 to 129 of the 119 amino acid consensus sequence (SEQ ID NO:15), while the upper amino acid sequence corresponds to the “phospholipase phosphodiesterase hydrolase phosphoinositide-specific 1-phosphatidylinositol-45-bisphosphate degradation lipid transducer beta” domain of human 16816, amino acid residues 491 to 608 of SEQ ID NO:2.

Figure 11 depicts a BLAST alignment of human 16816 with a consensus amino acid sequence derived from a ProDomain “C phospholipase delta-4 delta4 phospholipase” (Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>). The lower sequence is amino acid residues 1 to 40 of the 40 amino acid consensus sequence (SEQ ID NO:16), while the upper amino acid sequence corresponds to the “C phospholipase delta-4 delta4 phospholipase” domain of human 16816, amino acid residues 722 to 761 of SEQ ID NO:2.

Figure 12 depicts a BLAST alignment of human 16816 with a consensus amino acid sequence derived from a ProDomain “1-phosphatidylinositol-4 phosphodiesterase-like bisphosphate” (Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>). The lower sequence is amino acid residues 9 to 71 of the 63 amino acid consensus sequence (SEQ ID NO:17), while the upper amino acid sequence corresponds to the “1-phosphatidylinositol-4 phosphodiesterase-like bisphosphate” domain of human 16816, amino acid residues 562 to 621 of SEQ ID NO:2.

Figures 13A-B depict a cDNA sequence (SEQ ID NO:4) and predicted amino acid sequence (SEQ ID NO:5) of human 16839. The methionine-initiated open reading frame of human 16839 (without the 5' and 3' untranslated regions) extends from nucleotide position 1 to position 1827 of SEQ ID NO:6, not including the terminal codon.

Figure 14 depicts a hydropathy plot of human 16839. Relatively hydrophobic residues are shown above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The numbers corresponding to the amino acid sequence of human 16816 are indicated. Polypeptides of the invention include fragments which include: all or part of a hydrophobic sequence, e.g., a sequence above the dashed line, e.g., the sequence from about amino acid 340 to 350, from about 480 to

490, and from about 540 to 560 of SEQ ID NO:2; all or part of a hydrophilic sequence, e.g., a sequence below the dashed line, e.g., the sequence from about amino acid 300 to 325, from about 360 to 390, and from about 405 to 420 of SEQ ID NO:2; a sequence which includes a Cys, or a glycosylation site.

5 *Figure 15* depicts an alignment of the EF hand domain of human 16839 with a consensus amino acid sequence derived from a hidden Markov model (HMM) from PFAM. The upper sequences are the consensus amino acid sequence (SEQ ID NO:18), while the lower amino acid sequences correspond to amino acids 39 to 67 of SEQ ID NO:2.

10 *Figure 16* depicts an alignment of the phosphatidylinositol-specific phospholipase C, X domain of human 16839 with a consensus amino acid sequence derived from a hidden Markov model (HMM) from PFAM. The upper sequences are the consensus amino acid sequence (SEQ ID NO:19), while the lower amino acid sequences correspond to amino acids 156 to 300 of SEQ ID NO:2.

15 *Figure 17* depicts an alignment of the phosphatidylinositol-specific phospholipase C, Y domain of human 16839 with a consensus amino acid sequence derived from a hidden Markov model (HMM) from PFAM. The upper sequences are the consensus amino acid sequence (SEQ ID NO:20), while the lower amino acid sequences correspond to amino acids 348 to 465 of SEQ ID NO:2.

20 *Figure 18* depicts an alignment of the C2 domain of human 16839 with a consensus amino acid sequence derived from a hidden Markov model (HMM) from PFAM. The upper sequences are the consensus amino acid sequence (SEQ ID NO:21), while the lower amino acid sequences correspond to amino acids 484 to 572 of SEQ ID NO:2.

25 *Figure 19* depicts a BLAST alignment of human 16839 with a consensus amino acid sequence derived from a ProDomain “phospholipase phosphodiesterase hydrolase phosphoinositide-specific 1-phosphatidylinositol-45-bisphosphate degradation transducer lipid beta” (Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>). The lower sequence is amino acid residues 2 to 192 of the 191 amino acid consensus sequence (SEQ
30 ID NO:22), while the upper amino acid sequence corresponds to the “phospholipase phosphodiesterase hydrolase phosphoinositide-specific 1-phosphatidylinositol-45-bisphosphate degradation transducer lipid beta” domain of human 16839, amino acid residues 140 to 324 of SEQ ID NO:2.

Figure 20 depicts a BLAST alignment of human 16839 with a consensus amino acid sequence derived from a ProDomain “phospholipase phosphodiesterase hydrolase phosphoinositide-specific 1-phosphatidylinositol-45-bisphosphate degradation transducer lipid beta” (Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>). The lower sequence is amino acid residues 14 to 141 of the 128 amino acid consensus sequence (SEQ ID NO:23), while the upper amino acid sequence corresponds to the “phospholipase phosphodiesterase hydrolase phosphoinositide-specific 1-phosphatidylinositol-45-bisphosphate degradation transducer lipid beta” domain of human 16839, amino acid residues 350 to 473 of SEQ ID NO:2.

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

Detailed Description

Human 16816

The human 16816 sequence (Figure 1A-C; SEQ ID NO:1), which is approximately 2629 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 2289 nucleotides (nucleotides 257-2545 of SEQ ID NO:1; SEQ ID NO:3), not including the terminal codon. The coding sequence encodes a 762 amino acid protein (SEQ ID NO:2).

This mature protein form is approximately 762 amino acid residues in length (from about amino acid 1 to amino acid 762 of SEQ ID NO:2). Human 16816 contains the following regions or other structural features:

one predicted phosphatidylinositol-specific phospholipase C domain X (PFAM Accession Number PF00388) located at about amino acid residues 291-436;

one predicted phosphatidylinositol-specific phospholipase C domain Y (PFAM Accession Number PF00387) located at about amino acid residues 492-609 of SEQ ID NO:2;

two predicted EF hand domains (PFAM Accession Number PF00036) located at about amino acid residues 138-166 and 174-202 of SEQ ID NO:2;

one predicted C2 domain (PFAM Accession Number PF00168) located at about amino acid residues 291-436 and 492-609 of SEQ ID NO:2;

two cAMP- and cGMP-dependent protein kinase phosphorylation sites (PS00004) located at about amino acids 435-438 and 482-485 of SEQ ID NO:2;

nine predicted protein kinase C phosphorylation sites (PS00005) located at about amino acids 31-33, 56-58, 68-70, 203-205, 257-259, 355-357, 504-506, 666-668 and 741-743 of SEQ ID NO:2;

5 seventeen predicted casein kinase II phosphorylation sites (PS00006) located at about amino 11-14, 62-65, 80-83, 100-103, 108-111, 127-130, 155-158, 223-226, 318-321, 410-413, 422-425, 438-441, 463-466, 467-470, 522-525, 649-652 and 710-713 of SEQ ID NO:2;

four predicted N-myristoylation sites (PS00008) located at about amino acids 188-193, 219-224, 414-419 and 684-689 of SEQ ID NO:2;

10 two predicted amidation sites (PS00009) located at about amino acids 96-99 and 433-436 of SEQ ID NO:2;

one RGD cell attachment sequence (PS00016) located at about amino acids 145-147 of SEQ ID NO:2; and/or

15 two EF hand calcium-binding domains (PS00018) located at about amino acids 147-159 and 183-195 of SEQ ID NO:2.

In one embodiment, a 16816 family member can include at least one phosphatidylinositol-specific phospholipase C domain X (PFAM Accession Number PF00388); at least one phosphatidylinositol-specific phospholipase C domain Y (PFAM Accession Number PF00387); at least one preferably two EF hand domains (PFAM Accession Number PF00036 or PS00018); at least one predicted C2 domain (PFAM Accession Number PF00168). Furthermore, a 16816 family member can include at least one and preferably two cAMP- and cGMP-dependent protein kinase phosphorylation sites; at least one, two, three, four, five, six, seven, eight, and preferably nine protein kinase C phosphorylation sites (PS00005); at least one, two, three, four, five, six, seven, 25 eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen and preferably seventeen casein kinase II phosphorylation sites (PS00006); at least one, two, three, and preferably four N-myristoylation sites (PS00008); at least one and preferably two predicted amidation sites; at least one RGD cell attachment sequence.

30 Human 16839

The human 16839 sequence (Figure 13A-B; SEQ ID NO:4), which is approximately 2171 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 1827 nucleotides (nucleotides 232-2058 of

SEQ ID NO:4; SEQ ID NO:6), not including the terminal codon. The coding sequence encodes a 608 amino acid protein (SEQ ID NO:5).

This mature protein form is approximately 608 amino acid residues in length (from about amino acid 1 to amino acid 608 of SEQ ID NO:2). Human 16839 contains the following regions or other structural features:

one EF hand domain (PFAM Accession Number PF00036) located at about amino acids 39 to 67 of SEQ ID NO:2;

one phosphatidylinositol-specific phospholipase C domain X (PFAM Accession Number PF00388) located at about amino acids 156 to 300 of SEQ ID NO:2;

one phosphatidylinositol-specific phospholipase C domain Y (PFAM Accession Number PF00387) located at about amino acids 348 to 465 of SEQ ID NO:2;

one C2 domain (PFAM Accession Number PF00168) located at about amino acids 484 to 572 of SEQ ID NO:2;

two N-glycosylation sites (PS00001) located at about amino acids 376-379 and 537-540 of SEQ ID NO:5;

three cAMP- and cGMP-dependent protein kinase phosphorylation sites (PS00004) located at about amino acids 310-313, 337-340 and 385-388 of SEQ ID NO:5;

ten predicted protein kinase C phosphorylation sites (PS00005) located at about amino acids 24-26, 68-70, 220-222, 303-305, 313-315, 340-342, 399-401, 485-487, 501-503 and 533-535 of SEQ ID NO:5;

eight predicted casein kinase II phosphorylation sites (PS00006) located at about amino 56-59, 68-71, 79-82, 267-270, 303-306, 356-359, 378-381 and 411-414 of SEQ ID NO:5; and

three predicted N-myristoylation sites (PS00008) located at about amino acids 16-21, 479-484 and 560-565 of SEQ ID NO:5.

In one embodiment, a 16839 family member can include at least one EF hand domain (PFAM Accession Number PF00036); at least one phosphatidylinositol-specific phospholipase C domain X (PFAM Accession Number PF00388); at least one phosphatidylinositol-specific phospholipase C domain Y (PFAM Accession Number PF00387); at least one C2 domain (PFAM Accession Number PF00168). Furthermore, a 16839 family member can include at least one N-glycosylation site (PS00001); at least one, two and preferably three cAMP- and cGMP-dependent protein kinase phosphorylation sites; at least one, two, three, four, five, six, seven, eight, nine and

preferably ten protein kinase C phosphorylation sites (PS00005); at least one, two, three, four, five, six, seven, and preferably eight casein kinase II phosphorylation sites (PS00006); at least one, two, and preferably three N-myristoylation sites (PS00008).

For general information regarding PFAM identifiers, PS prefix and PF prefix
5 domain identification numbers, refer to Sonnhammer *et al.* (1997) *Protein* 28:405-420 and <http://www.psc.edu/general/software/packages/pfam/pfam.html>.

Plasmids containing the nucleotide sequences encoding human 16816 and 16839 were deposited with American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Numbers
10 _____. These deposits 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. These deposits 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 16816 and 16839 protein contain a significant number of structural
15 characteristics in common with members of the phospholipase C family. The term "family" when referring to the protein and nucleic acid molecules of the invention means 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
20 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.

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as "phospholipase C" or "16816" or "16839" nucleic acid
25 and polypeptide molecules, which play a role in or function in modulating signal transduction pathways.

PLC molecules have been found in a broad spectrum of organisms including bacteria, simple eukaryotes, plants and animals. Members of a family can also have common functional characteristics. Members of the PLC family share one or more
30 common domains such as a pleckstrin homology domain, an EF hand domain, a phosphatidylinositol-specific phospholipase domain X (PLC-X) domain, a phosphatidylinositol-specific phospholipase domain Y (PLC-Y) domain or a C2 domain. Members of this family can also have common functional characteristics, e.g., the ability to hydrolyze phosphatidylinositols.

A 16816 polypeptide can include a "pleckstrin homology (PH) domain" or regions homologous with a "PH domain". As used herein, the term "PH domain" refers to a protein domain having an amino acid sequence of about 10 to 200, preferably about 50 to 150, more preferably about 108 amino acid residues. By "PH domain" is meant a domain
 5 that can function as a recognition site for a phosphatidylinositol, e.g., a 3,4,5-trisphosphate (PIP3) or another kinase ligand product, and can function as a means to localize PLC to the cytoplasmic face of the plasma membrane.

As used herein, the term "PH domain" includes an amino acid sequence of about 108 amino acid residues in length and having a bit score for the alignment of the sequence
 10 to the PH domain (HMM) of at least 10. Preferably, a PH domain includes at least about 10-200 amino acids, more preferably about 50-150 amino acid residues, or about 75-110 amino acids and has a bit score for the alignment of the sequence to the PH domain (HMM) of at least 20, 30, or greater. An alignment of the PH domain (amino acids 17-124 of SEQ ID NO:2) of human PH with a consensus amino acid sequence derived from
 15 a hidden Markov model is depicted in Figure 3.

In a preferred embodiment a 16816 polypeptide or protein has a "PH domain" or a region which includes at least about 10-200 amino acids, more preferably about 50-150 amino acid residues, or about 107 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a "PH domain," e.g., the PH domain of
 20 human 16816 (e.g., residues 17-124 of SEQ ID NO:2).

A 16816 or 16839 polypeptide can also include an "EF hand domain" or regions homologous with an "EF hand domain". As used herein, the term "EF hand domain" refers to a protein domain having an amino acid sequence of about 5 to 50, preferably about 5 to 40, more preferably about 28-29 amino acid residues. By "EF hand domain" is
 25 meant a type of calcium-binding domain that consists of a twelve residue loop flanked on both sides by a twelve residue alpha-helical domain. In an EF-hand loop the calcium ion is coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12 provides two oxygens for liganding Ca
 30 (bidentate ligand). Preferably, the EF hand domain includes the following amino acid consensus sequence having Prosite signatures as PS00018, or sequences homologous thereto. D-x-[DNS]-{ILVFIYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[DE]-[LIVMFIYW] (SEQ ID NO:24). In the above conserved motif, and other motifs described herein, the standard IUPAC one-letter code for the

amino acids is used. Each element in the pattern is separated by a dash (-); square brackets ([]) indicate the particular residues that are accepted at that position; x indicates that any residue is accepted at that position; and numbers in parentheses () indicate the number of residues represented by the accompanying amino acid. The EF hand domains
 5 are located in mostly hydrophilic regions of the molecule of human 16816 or 16839 polypeptide and which corresponds to about amino acids 138-166 and 172-202 of SEQ ID NO:2; or amino acids 39-67 of SEQ ID NO:5. The domain (HMM) has been assigned the PFAM Accession Number PF00036 (<http://genome.wustl.edu/Pfam/.html>).

The "EF hand domain" includes an amino acid sequence of about 28 amino acid
 10 residues in length and can have a bit score for the alignment of the sequence to the EF hand (HMM) of at least 5. Preferably, an EF hand domain includes at least about 5-50 amino acids, or at least about 5-40, or about 28 amino acids and has a bit score for the alignment of the sequence to the EF hand (HMM) of at least 5, 10, 15, 20, or greater. An alignment of the EF hand domain (amino acids 138-166 and 172-202 of SEQ ID NO:2; or
 15 amino acids 39-67 of SEQ ID NO:5) of human 16816 or 16839 with a consensus amino acid sequence derived from a hidden Markov model is depicted in Figures 4 or 15.

In a preferred embodiment, the 16816 or 16839 polypeptide or protein has an "EF hand domain" or a region which includes at least about 5-50, more preferably about 5-40 or 28-29 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or
 20 100% homology with an "EF hand domain," e.g., the EF hand domain of human 16816 or 16839 (e.g., residues 138-166 and 172-202 of SEQ ID NO:2; or amino acids 39-67 of SEQ ID NO:5).

A 16816 or 16839 polypeptide can also include a "phosphatidylinositol-specific phospholipase C domain X (referred to herein as "PLC-X domain")" or regions
 25 homologous with a "PLC-X domain". As used herein, the term "PLC-X domain" refers to a protein domain having an amino acid sequence of about 8 to 200, preferably about 15 to 170, more preferably about 145 amino acid residues. By "PLC-X domain" is meant a subdomain that composes the catalytic site of the phospholipase, e.g., PLC-X subdomain can fold together with another subdomain, e.g., phosphatidylinositol-specific
 30 phospholipase C domain Y such that a functioning catalytic site that hydrolyzes a phosphatidylinositol is formed, e.g., phosphatidylinositol 4, 5-bisphosphate, is formed.

The "PLC-X domain" includes an amino acid sequence of about 145 amino acid residues in length and can have a bit score for the alignment of the sequence to the phosphatidylinositol-specific phospholipase-C domain X (HMM) of at least 50.

Preferably, a PLC-X domain includes at least about 15-170 amino acids, or at least about 20-150, or about 145 amino acids and has a bit score for the alignment of the sequence to the phosphatidylinositol-specific phospholipase-C domain X (HMM) of at least 60, 70, 80, 90, 100, 150, 200, 250, or greater. An alignment of the PLC-X domain (amino acids 291-436 of SEQ ID NO:2; or amino acids 156-300 of SEQ ID NO:5) of human 16816 or 16839 with a consensus amino acid sequence derived from a hidden Markov model is depicted in Figures 5 or 16.

In a preferred embodiment, the 16816 or 16839 polypeptide or protein has a "PLC-X" or a region which includes at least about 8-200, more preferably about 15-170 or 20-150 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a "PLC-X domain," e.g., the PLC-X domain of human 16816 or 16839 (e.g., residues 291-436 of SEQ ID NO:2; or residues 156-300 of SEQ ID NO:5).

A 16816 or 16839 polypeptide can include a "phosphatidylinositol-specific phospholipase C domain Y (referred to herein as PLC-Y domain)" or regions homologous with a "PLC-Y domain". As used herein, the term "PLC-Y domain" refers to a protein domain having an amino acid sequence of about 8 to 200, preferably about 15 to 170, more preferably about 117 amino acid residues. By "PLC-Y domain" is meant a subdomain that composes the catalytic site of the phospholipase, e.g., the subdomain can fold together with another subdomain, e.g., PLC-X domain such that a functioning catalytic site that hydrolyzes a phosphatidylinositol, e.g., phosphatidylinositol 4, 5-bisphosphate, is formed.

The "PLC-Y domain" includes an amino acid sequence of about 117 amino acid residues in length and can have a bit score for the alignment of the sequence to the PLC-Y domain (HMM) of at least 50. Preferably, a PLC-Y domain includes at least about 15-170 amino acids, or at least about 20-150, or about 117 amino acids and has a bit score for the alignment of the sequence to the PLC-Y domain (HMM) of at least 60, 70, 80, 90, 100, 110, 120, 140, 160, 180, or greater. An alignment of the PLC-Y domain (amino acids 492-609 of SEQ ID NO:2; or amino acids 348-465 of SEQ ID NO:5) of human 16816 or 16839 with a consensus amino acid sequence derived from a hidden Markov model depicted in Figure 6 or 17.

In a preferred embodiment 16816 or 16839 polypeptide or protein has a "PLC-Y domain" or a region which includes at least about 8-200, more preferably about 15-170 or 20-150 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or

100% homology with a "PLC-Y domain" e.g., PLC-Y domain of human 16816 or 16839 (e.g., residues 492-609 of SEQ ID NO:2; or residues 348-465 of SEQ ID NO:5).

5 A 16816 or 16839 polypeptide can include a "calcium (Ca²⁺) binding domain (referred to as C2 domain") or regions homologous with a "C2 domain". As used herein, the term "C2" refers to a protein domain having an amino acid sequence of about 8 to 200, preferably about 15 to 170, more preferably about 20 to 100, or still more preferably about 90 amino acid residues. By "C2 domain" is meant a domain that can mediate interaction with calcium or phospholipids.

10 The "C2 domain" includes an amino acid sequence of about 90 amino acid residues in length and can have a bit score for the alignment of the sequence to the C2 domain (HMM) of at least 50. Preferably, a C2 includes at least about 8-200, or at least about 15-170, or at least 20-100, or about 90 amino acids and has a bit score for the alignment of the sequence to the C2 domain (HMM) of at least 60, 70, 80, 85, or greater. An alignment of the C2 domain (amino acids 629-719 of SEQ ID NO:2; or amino acids
15 484-572 of SEQ ID NO:5) of human 16816 or 16839 with a consensus amino acid sequence derived from a hidden Markov model is depicted in Figure 7 or 18.

In a preferred embodiment, a 16816 or 16839 polypeptide or protein has a "C2" or a region which includes at least about 10-200, more preferably about 15-170 or 20-100 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100%
20 homology with a "C2," e.g., the C2 domain of human 16816 or 16839 (e.g., residues 629-719 of SEQ ID NO:2; or residues 484-572 of SEQ ID NO:5).

To identify the presence of a "PH domain," "EF hand," "PLC-X domain," "PLC-Y domain," or a "C2 domain" in a 16816 or 16839 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino
25 acid sequence of the protein can be searched against a database of HMMs (e.g., the Pfam database, release 2.1) using the default parameters (http://www.sanger.ac.uk/Software/Pfam/HMM_search). For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for MILPAT0063 and a score of 15 is the default
30 threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (e.g., to 8 bits). A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28(3):405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.* (1990) *Meth. 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.

A search was performed against the HMM database resulting in the identification of a "PH domain" domain in the amino acid sequence of human 16816 at about residues 17-124 of SEQ ID NO:2 (see Figure 3); an
5 "EF domain" in the amino acid sequence of human 16816 or 16839 at about residues 138-166 and 172-202 of SEQ ID NO:2 (see Figure 4) or amino acids 39-67 of SEQ ID NO:5 (see Figure 15) respectively; a "C2 domain" in the amino acid sequence of human 16816 or 16839 at about residues 629-719 of SEQ ID NO:2 (see Figure 7) and residues 484-572 of SEQ ID NO:5 (see Figure 18) respectively; a "PLC-Y domain" in the amino acid
10 sequence of human 16816 or 16839 at about residues 492-609 of SEQ ID NO:2 (see Figure 6) and residues 348-465 of SEQ ID NO:5 (see Figure 17); and a "PLC-X domain" in the amino acid sequence of human 16816 or 16839 at about residues 291-436 of SEQ ID NO:2 (see Figure 5) and residues 156-300 of SEQ ID NO:5 (see Figure 16).

To identify the presence of a "phospholipase C" domain in a 16816 or 16839
15 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against a database of domains, e.g., the ProDom database (Corpet *et al.* (1999), *Nucl. Acids Res.* 27:263-267). The ProDom protein domain database consists of an automatic compilation of homologous domains. Current versions of ProDom are built using recursive PSI-
20 BLAST searches (Altschul SF *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402; Gouzy *et al.* (1999) *Computers and Chemistry* 23:333-340) of the SWISS-PROT 38 and TREMBL protein databases. The database automatically generates a consensus sequence for each domain. A BLAST search was performed against the HMM database resulting in the identification of a "phospholipase C" domain in the amino acid sequence of human 16816
25 or 16839.

The phospholipase C domain is homologous to ProDom family PD001214 ("phospholipase phosphodiesterase hydrolase phosphoinositide-specific 1-phosphatidylinositol-45-bisphosphate degradation transducer lipid beta" SEQ ID NO:13, ProDomain Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>). An alignment of
30 the phospholipase C domain (amino acids 275 to 436 of SEQ ID NO:2) of human 16816 with a consensus amino acid sequence (SEQ ID NO:13) derived from a hidden Markov model is depicted in Figure 8. The consensus sequence for SEQ ID NO:13 is 54% identical over amino acids 275 to 436 of SEQ ID NO:2 as shown in Figure 8.

The phospholipase C domain is also homologous to ProDom family PD186804 (“phospholipase C delta calcium-binding PLC-III hydrolase phosphodiesterase lipid PLC-delta-1 1-phosphatidylinositol-45-bisphosphate” SEQ ID NO:14, ProDomain Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>. An alignment of the phospholipase C domain (amino acids 1 to 191 of SEQ ID NO:2) of human 16816 with a consensus amino acid sequence (SEQ ID NO:14) derived from a hidden Markov model is depicted in Figure 9. The consensus sequence for SEQ ID NO:14 is 44% identical over amino acids 1 to 191 of SEQ ID NO:2 as shown in Figure 9.

The phospholipase C domain is also homologous to ProDom family PD001202 (“phospholipase phosphodiesterase hydrolase phosphoinositide-specific 1-phosphatidylinositol-45-bisphosphate degradation lipid transducer beta” SEQ ID NOs:15 and 23, ProDomain Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>. An alignment of the phospholipase C domain (amino acids 491 to 608 of SEQ ID NO:2) of human 16816 with a consensus amino acid sequence (SEQ ID NO:15) derived from a hidden Markov model is depicted in Figure 10. The consensus sequence for SEQ ID NO:15 is 57% identical over amino acids 491 to 608 of SEQ ID NO:2 as shown in Figure 10. An alignment of the phospholipase C domain (amino acids 350 to 473 of SEQ ID NO:2) of human 16839 with a consensus amino acid sequence (SEQ ID NO:23) derived from a hidden Markov model is depicted in Figure 20. The consensus sequence for SEQ ID NO:23 is 47% identical over amino acids 350 to 473 of SEQ ID NO:2 as shown in Figure 20.

The phospholipase C domain is also homologous to ProDom family PD033204 (“C phospholipase delta-4 delta4 phospholipase” SEQ ID NO:16, ProDomain Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>. An alignment of the phospholipase C domain (amino acids 722 to 761 of SEQ ID NO:2) of human 16816 with a consensus amino acid sequence (SEQ ID NO:16) derived from a hidden Markov model is depicted in Figure 11. The consensus sequence for SEQ ID NO:16 is 70% identical over amino acids 722 to 761 of SEQ ID NO:2 as shown in Figure 11.

The phospholipase C domain is also homologous to ProDom family PD270355 (“1-phosphatidylinositol-4 phosphodiesterase-like bisphosphate” SEQ ID NO:17, ProDomain Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>. An alignment of the phospholipase C domain (amino acids 562 to 621 of SEQ ID NO:2) of human 16816 with a consensus amino acid sequence (SEQ ID NO:17) derived from a hidden Markov

model is depicted in Figure 12. The consensus sequence for SEQ ID NO:17 is 46% identical over amino acids 562 to 621 of SEQ ID NO:2 as shown in Figure 12.

The phospholipase C domain is also homologous to ProDom family PD001214 (“phospholipase phosphodiesterase hydrolase phosphoinositide-specific 1-phosphatidylinositol-45-bisphosphate degradation transducer lipid beta” SEQ ID NO:22, ProDomain Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>. An alignment of the phospholipase C domain (amino acids 140 to 324 of SEQ ID NO:2) of human 16839 with a consensus amino acid sequence (SEQ ID NO:22) derived from a hidden Markov model is depicted in Figure 19. The consensus sequence for SEQ ID NO:22 is 48% identical over amino acids 140 to 324 of SEQ ID NO:2 as shown in Figure 19.

As the 16816 or 16839 polypeptide of the invention may modulate 16816 or 16839-mediated activity, they may be useful as of for developing novel diagnostic and therapeutic agents for 16816 or 16839-mediated or related disorders, as described below. As used herein, a “16816 or 16839 activity”, “biological activity of 16816 or 16839” or “functional activity of 16816 or 16839”, refers to an activity exerted by a 16816 or 16839 protein, polypeptide or nucleic acid molecule on e.g., a 16816 or 16839-responsive cell or on a 16816 or 16839 substrate, e.g., a protein substrate, as determined *in vivo* or *in vitro*. In one embodiment, a 16816 or 16839 activity is a direct activity, such as an association with a 16816 or 16839 target molecule. A “target molecule” or “binding partner” is a molecule with which a 16816 or 16839 protein binds or interacts with in nature. In an exemplary embodiment, it is a receptor, e.g., a tyrosine kinase receptor. In another embodiment, 16816 or 16839 can associate with a second messenger molecule such as a specialized adaptor molecule; with inositol phosphates and inositol lipids; membrane proteins; or with a guanine nucleotide binding-regulatory protein (G-protein). A 16816 or 16839 activity can also be an indirect activity, e.g., a cellular signaling activity mediated by interaction of the 16816 or 16839 protein with a receptor or another signaling molecule. For example, the 16816 or 16839 proteins of the present invention can have one or more of the following activities: (1) transduction of transmembrane signals; (2) lipid-metabolizing activity, e.g., 16816 or 16839 can catalyze the hydrolysis of phosphatidyl-inositol-4,5-bisphosphate (PIP2) producing diacylglycerol and inositol 1,4,5-trisphosphate; (3) the regulation of transmission of signals from cellular receptors such as hormones such as serotonin, growth factors such as platelet-derived growth factor (PDGF), fibroblast growth factor (FGF), and nerve growth factor (NGF), neurotransmitters and immunoglobulins; (4) modulation of cell proliferation; (5)

modulation of cell differentiation; (6) modulation of cell migration; (7) modulation of fertilization; and (8) modulation of hypertension.

Based on the above-described sequence similarities, the 16816 or 16839 molecules of the present invention are predicted to have similar biological activities as members of the PLC family. Members of the PLC family play a very important role in transmembrane signal transduction. Extracellular signaling molecules including hormones, growth factors, neurotransmitters, and immunoglobulins bind to their respective cell surface receptors and activate phospholipase-C. PLC molecules have many functions including: glycogenolysis in liver cells, histamine secretion by mast cells, serotonin release by blood platelets, aggregation by blood platelets, insulin release by pancreatic islet cells, epinephrine secretion by adrenal chromaffin cells, and smooth muscle contraction. In general, biological systems that are activated by receptor tyrosine kinase cause the activation of phospholipase-C. The role of an activated PLC is to catalyze the hydrolysis of phosphatidyl-inositol-4,5-bisphosphate (PIP₂), a minor component of the plasma membrane, to produce diacylglycerol and inositol 1,4,5-trisphosphate (IP₃). Inositol trisphosphate releases calcium from intracellular stores and increases the influx of calcium from the extracellular fluid. The calcium ions directly regulate target enzymes and indirectly affect other enzymes by functioning as a second messenger and interacting with calcium-binding proteins, such as troponin C and calmodulin. For example, calcium ions regulate muscle contraction, glycogen breakdown and exocytosis. Diacylglycerol, a product of the hydrolysis by PLCs, acts as a second messenger by activating protein kinase C. Activated protein kinase C phosphorylates a great number of intracellular proteins at the serine and threonine residues and modulates different signaling pathways. For example, the phosphorylation of glycogen synthase by protein kinase C stops the synthesis of glycogen. Moreover, protein kinase C controls cell division and proliferation. Both pathways are part of transmembrane signal transduction mechanisms, which regulate cellular processes, which include secretion, neural activity, metabolism, differentiation and proliferation.

Both 16816 and 16839 proteins are homologous to the phospholipase C molecule, PLC1. Stimulation of 16816 or 16839 activity is desirable in situations in which 16816 or 16839 is abnormally downregulated and/or in which increased 16816 or 16839 activity is likely to have a beneficial effect. For example, research on chromosome 20q has associated this gene locus with tumor suppressor activity. Deletions and mutations of the 20q chromosome have been associated with myelodysplasia and myeloproliferative

disorders. PLC1 is one of the genes present at this locus, and has been found to be deleted in these cases. (Asimakopoulos *et al.* (1994) *Blood* 84(9):3086-94). As such, 16816 and 16839 may play a role in preventing or treating myeloid disorders.

Likewise, inhibition of 16816 or 16839 activity is desirable in situations in which 5 16816 or 16839 is abnormally upregulated and/or in which decreased 16816 or 16839 activity is likely to have a beneficial effect. It has been shown that PLC1 overexpression is associated with hepatocellular carcinoma. One antibody, k-PLC1, was shown to react with PLC1 (Wiedmann *et al.* (1987) *Hepatology* 7(3):543-50). As such, inhibitors such as 16816- or 16839-specific antibodies may be useful to reduce the quantity of PLC1 in 10 such situations, and potentially decrease the severity and/or occurrence of hepatocellular carcinoma. Thus, the 16816 or 16839 molecules can act as novel diagnostic targets and therapeutic agents for controlling disorders caused by abnormal or aberrant PLC activity. Evidence indicates that a high percentage of primary human mammary carcinomas concomitantly show abnormally high levels of PLC-gamma-1 (Kassis *et al.*, 15 *Clin Cancer Res.*, Aug; 5(8):2251-60, 1999). Likewise, studies on spontaneous hypertensive rats have suggested that one of the main causes for the hypertension is an abnormal activation of PLC-delta-1 resulting from point mutations in the X and Y regions of the PLC amino acid sequence (Sanada *et al.*, *Hypertension* 33(4):1036-42, 1999). Therefore, the 16816 or 16839 molecules can act as novel diagnostic targets and 20 therapeutic agents for controlling one or more of cellular proliferative and/or differentiative disorders, disorders associated with bone metabolism, immune disorders, hematopoietic disorders, cardiovascular disorders, liver disorders, viral diseases, pain or metabolic disorders.

As the 16816 or 16839 polypeptides of the invention may modulate 16816 or 25 16839-mediated activities, they may be useful for developing novel diagnostic and therapeutic agents for 16816 or 16839-mediated or related disorders, as described below.

Accordingly, 16816 or 16839 protein may mediate various disorders, including cellular proliferative and/or differentiative disorders, brain disorders, heart disorders, blood vessel disorders, and platelet disorders.

30 Examples of cellular proliferative and/or differentiative disorders include cancer, e.g., carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic disorders, e.g., leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of prostate, colon, lung, breast and liver origin.

As used herein, the terms “cancer”, “hyperproliferative” and “neoplastic” refer to cells having the capacity for autonomous growth, i.e., an abnormal state or condition characterized by rapidly proliferating cell growth. Hyperproliferative and neoplastic disease states may be categorized as pathologic, i.e., characterizing or constituting a disease state, or may be categorized as non-pathologic, i.e., a deviation from normal but not associated with a disease state. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. “Pathologic hyperproliferative” cells occur in disease states characterized by malignant tumor growth. Examples of non-pathologic hyperproliferative cells include proliferation of cells associated with wound repair.

The terms “cancer” or “neoplasms” include malignancies of the various organ systems, such as affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genitourinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus.

The term “carcinoma” is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term also includes carcinosarcomas, e.g., which include malignant tumors composed of carcinomatous and sarcomatous tissues. An “adenocarcinoma” refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures.

The term “sarcoma” is art recognized and refers to malignant tumors of mesenchymal derivation.

The 16816 or 16839 nucleic acid and protein of the invention can be used to treat and/or diagnose a variety of proliferative disorders. E.g., such disorders include hematopoietic neoplastic disorders. As used herein, the term “hematopoietic neoplastic disorders” includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, e.g., arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Preferably, the diseases arise from poorly differentiated acute leukemias, e.g., erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary

myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus, L., (1991) *Crit. Rev. in Oncol./Hematol.* 11:267-97); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), 5 prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large 10 granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

Disorders involving the brain include, but are not limited to, disorders involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, 15 forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states-- global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) 20 hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive cerebrovascular disease, including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain abscess, subdural empyema, and extradural 25 abscess, chronic bacterial meningitis, including tuberculosis and mycobacterioses, neurosyphilis, and neuroborreliosis (Lyme disease), viral meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2, *Varicella-zoster* virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including 30 HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral neuropathy, and AIDS in children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple

sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute necrotizing hemorrhagic encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer disease and Pick disease, degenerative diseases of basal ganglia and brain stem, including Parkinsonism, idiopathic Parkinson disease (paralysis agitans), progressive supranuclear palsy, corticobasal degeneration, multiple system atrophy, including striatonigral degeneration, Shy-Drager syndrome, and oligopontocerebellar atrophy, and Huntington disease; spinocerebellar degeneration, including spinocerebellar ataxias, including Friedreich ataxia, and ataxia-telanglectasia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; inborn errors of metabolism, such as leukodystrophies, including Krabbe disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, and Canavan disease, mitochondrial encephalomyopathies, including Leigh disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic diseases, including vitamin deficiencies such as thiamin (vitamin B₁) deficiency and vitamin B₁₂ deficiency, neurologic sequelae of metabolic disturbances, including hypoglycemia, hyperglycemia, and hepatic encephalopathy, toxic disorders, including carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate and radiation-induced injury; tumors, such as gliomas, including astrocytoma, including fibrillary (diffuse) astrocytoma and glioblastoma multiform, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma, oligodendroglioma, and ependymoma and related paraventricular mass lesions, neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors, including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including neurofibromatosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease.

Disorders involving the heart, include but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death;

hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcific aortic stenosis, calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, and disorders involving cardiac transplantation.

Disorders involving blood vessels include, but are not limited to, responses of vascular cell walls to injury, such as endothelial dysfunction and endothelial activation and intimal thickening; vascular diseases including, but not limited to, congenital anomalies, such as arteriovenous fistula, atherosclerosis, and hypertensive vascular disease, such as hypertension; inflammatory disease--the vasculitides, such as giant cell (temporal) arteritis, Takayasu arteritis, polyarteritis nodosa (classic), Kawasaki syndrome (mucocutaneous lymph node syndrome), microscopic polyanglitis (microscopic polyarteritis, hypersensitivity or leukocytoclastic angitis), Wegener granulomatosis, thromboanglitis obliterans (Buerger disease), vasculitis associated with other disorders, and infectious arteritis; Raynaud disease; aneurysms and dissection, such as abdominal aortic aneurysms, syphilitic (luetie) aneurysms, and aortic dissection (dissecting hematoma); disorders of veins and lymphatics, such as varicose veins, thrombophlebitis and phlebothrombosis, obstruction of superior vena cava (superior vena cava syndrome),

obstruction of inferior vena cava (inferior vena cava syndrome), and lymphangitis and lymphedema; tumors, including benign tumors and tumor-like conditions, such as hemangioma, lymphangioma, glomus tumor (glomangioma), vascular ectasias, and bacillary angiomatosis, and intermediate-grade (borderline low-grade malignant) tumors, such as Kaposi sarcoma and hemangioendothelioma, and malignant tumors, such as angiosarcoma and hemangiopericytoma; and pathology of therapeutic interventions in vascular disease, such as balloon angioplasty and related techniques and vascular replacement, such as coronary artery bypass graft surgery.

The 16816 or 16839 protein, fragments thereof, and derivatives and other variants of the sequence in SEQ ID NO:2 are collectively referred to as “polypeptides or proteins of the invention” or “16816 or 16839 polypeptides or proteins”. Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as “nucleic acids of the invention” or “16816 or 16839 nucleic acids.” 16816 or 16839 molecules refer to 16816 or 16839 nucleic acids, polypeptides, and antibodies.

As used herein, the term “nucleic acid molecule” includes DNA molecules (e.g., a cDNA or genomic DNA) and RNA molecules (e.g., an mRNA) and analogs of the DNA or RNA generated, e.g., by the use of nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term “isolated or purified 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 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/or 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 nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of 5’ and/or 3’ 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 other chemicals when chemically synthesized.

As used herein, the term “hybridizes under stringent conditions” describes conditions for hybridization and washing. Stringent conditions are known to those skilled

in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Aqueous and nonaqueous methods are described in that reference and either can be used. A preferred, example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Preferably, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. Particularly preferred stringency conditions (and the conditions that should be used if the practitioner is uncertain about what conditions should be applied to determine if a molecule is within a hybridization limitation of the invention) are 0.5M Sodium Phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1 or SEQ ID NO:4, corresponds to a naturally-occurring nucleic acid molecule.

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

As used herein, the terms “gene” and “recombinant gene” refer to nucleic acid molecules which include an open reading frame encoding a 16816 or 16839 protein, preferably a mammalian 16816 or 16839 protein, and can further include non-coding regulatory sequences, and introns.

An “isolated” or “purified” polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language “substantially free” means preparation of 16816 or 16839 protein having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-16816 or 16839 protein (also referred to herein as a “contaminating protein”), or of chemical precursors or non-16816 or 16839 chemicals.

When the 16816 or 16839 protein or biologically 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 invention includes
5 isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

A “non-essential” amino acid residue is a residue that can be altered from the wild-type sequence of 16816 or 16839(e.g., the sequence of SEQ ID NO:1 SEQ ID NO:4, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____) without abolishing or more preferably, without substantially
10 altering a biological activity, whereas an “essential” amino acid residue results in such a change. For example, amino acid residues that are conserved among the polypeptides of the present invention, e.g., those present in the phospholipase C domain, are predicted to be particularly unamenable to alteration.

A “conservative amino acid substitution” is one in which the amino acid residue is
15 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), nonpolar side chains (e.g., alanine,
20 valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), 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 16816 or 16839 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment,
25 mutations can be introduced randomly along all or part of a 16816 or 16839 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 16816 or 16839 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1 or SEQ ID NO:4, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, the encoded
30 protein can be expressed recombinantly and the activity of the protein can be determined.

As used herein, a “biologically active portion” of a 16816 or 16839 protein includes a fragment of a 16816 or 16839 protein which participates in an interaction between a 16816 or 16839 molecule and a non-16816 or 16839 molecule. Biologically active portions of a 16816 or 16839 protein include peptides comprising amino acid

sequences sufficiently homologous to or derived from the amino acid sequence of the 16816 or 16839 protein, e.g., the amino acid sequence shown in SEQ ID NO:2 or SEQ ID NO:5, which include less amino acids than the full length 16816 or 16839 proteins, and exhibit at least one activity of a 16816 or 16839 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the 16816 or 16839 protein, e.g., phospholipase C activity. A biologically active portion of a 16816 or 16839 protein can be a polypeptide which is, for example, 10, 25, 50, 100, 200 or more amino acids in length. Biologically active portions of a 16816 or 16839 protein can be used as targets for developing agents which modulate a 16816 or 16839 mediated activity, e.g., phospholipase C activity.

Calculations of homology or sequence identity between sequences (the terms are used interchangeably herein) are performed as follows.

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%, 90%, 100% of the length of the reference sequence (e.g., when aligning a second sequence to the 16816 amino acid sequence of SEQ ID NO:2 having 762 amino acid residues, at least 229, preferably at least 305, more preferably at least 382, even more preferably at least 458, and even more preferably at least 534, 610, 687 or 762 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 <http://www.gcg.com>), 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 <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity or homology limitation of the invention) is using a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.*, (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 16816 or 16839 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 16816 or 16839 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 <http://www.ncbi.nlm.nih.gov>.

"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: expression at non-wild type levels, i.e., 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 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 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.

“Subject”, as used herein, can refer to a mammal, e.g., a human, or to an experimental or animal or disease model. The subject can also be a non-human animal, e.g., a horse, cow, goat, or other domestic animal.

A “purified preparation of cells”, as used herein, refers to, in the case of plant or animal cells, an in vitro preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

Various aspects of the invention are described in further detail below.

Isolated Nucleic Acid Molecules

In one aspect, the invention provides, an isolated or purified, nucleic acid molecule that encodes a 16816 or 16839 polypeptide described herein, e.g., a full length 16816 or 16839 protein or a fragment thereof, e.g., a biologically active portion of 16816 or 16839 protein. Also included is a nucleic acid fragment suitable for use as a hybridization probe, which can be used, e.g., to identify nucleic acid molecule encoding a polypeptide of the invention, 16816 or 16839 mRNA, and fragments suitable for use as primers, e.g., PCR primers for the amplification or mutation of nucleic acid molecules.

In one embodiment, an isolated nucleic acid molecule of the invention includes the nucleotide sequence shown in SEQ ID NO:1, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion of any of these nucleotide sequences. In one embodiment, the nucleic acid molecule includes sequences encoding the human 16816 or 16839 protein (i.e., “the coding region”, from nucleotides 257-2546 of SEQ ID NO:1, or 223-2050 of SEQ ID NO:4; not including the terminal codons), as well as 5’ untranslated sequences (nucleotides 1-256 of SEQ ID NO:1 or 1-222 of SEQ ID NO:4). Alternatively, the nucleic acid molecule can include

only the coding region of SEQ ID NO:1 (e.g., nucleotides 257-2545 of SEQ ID NO:1, or 232-2058 of SEQ ID NO:4, corresponding to SEQ ID NO:3 or SEQ ID NO:6) and, e.g., no flanking sequences which normally accompany the subject sequence. In another embodiment, the nucleic acid molecule encodes a sequence corresponding to the mature protein of SEQ ID NO:2 or SEQ ID NO:5.

In another embodiment, an isolated nucleic acid molecule of the invention includes a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:4, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion of any of these nucleotide sequences. In other embodiments, the nucleic acid molecules of the invention are sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:4, or the nucleotide sequences of the DNA insert of the plasmid deposited with ATCC as Accession Numbers _____ such that they can hybridize to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:4, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, thereby forming a stable duplex.

In one embodiment, an isolated nucleic acid molecule of the present invention includes a nucleotide sequence which is at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:4, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In the case of an isolated nucleic acid molecule which is longer than or equivalent in length to the reference sequence, e.g., SEQ ID NO:1 or SEQ ID NO:4, the comparison is made with the full length of the reference sequence. Where the isolated nucleic acid molecule is shorter than the reference sequence, e.g., shorter than SEQ ID NO:1 or SEQ ID NO:4, the comparison is made to a segment of the reference sequence of the same length (excluding any loop required by the homology calculation).

16816 or 16839 Nucleic Acid Fragments

A nucleic acid molecule of the invention can include only a portion of the nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:4, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. For example, such a nucleic acid molecule can include a fragment which can be used as a probe or primer or a fragment encoding a portion of a 16816 or 16839 protein, e.g., an

immunogenic or biologically active portion of a 16816 or 16839 protein. A fragment can comprise: nucleotides 870-1305 or 1479-1827 of SEQ ID NO:3, or nucleotides 465-897 or 1047-1395 of SEQ ID NO:6, which encode phospholipase C domains of human 16816 or 16839. The nucleotide sequence determined from the cloning of the 16816 or 16839 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other 16816 or 16839 family members, or fragments thereof, as well as 16816 or 16839 homologues, or fragments thereof, from other species.

In another embodiment, a nucleic acid includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5' or 3' noncoding region. Other embodiments include a fragment which includes a nucleotide sequence encoding an amino acid fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof, particularly fragments thereof which are at least 150 amino acids in length. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments should not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

A nucleic acid fragment can include a sequence corresponding to a domain, region, or functional site described herein. A nucleic acid fragment can also include one or more domain, region, or functional site described herein. Thus, for example, the nucleic acid fragment can include an phospholipase C domain. In a preferred embodiment the fragment is at least, 50, 100, 200, 300, 400, 500, 600, 700, or 900 base pairs in length.

16816 or 16839 probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, 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 or antisense sequence of SEQ ID NO:1 or SEQ ID NO:4, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1 or SEQ ID NO:4, or the nucleotide sequences of the DNA inserts of the plasmids deposited with ATCC as Accession Numbers _____.

In a preferred embodiment the nucleic acid is a probe which is at least 5 or 10, and less than 200, more preferably less than 100, or less than 50, base pairs in length. It should be identical, or differ by 1, or less than in 5 or 10 bases, from a sequence disclosed herein. If alignment is needed for this comparison the sequences should be aligned for

maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

A probe or primer can be derived from the sense or anti-sense strand of a nucleic acid which encode phospholipase C domains (e.g., about amino acid residues 290-435 or
5 493-609 of SEQ ID NO:2; or about amino acid residues 155-299 or 349-465 of SEQ ID NO:5).

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 16816 or 16839 sequence, e.g., a region described herein. The primers should be at least 5, 10, or 50 base pairs in
10 length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differ by one base from a sequence disclosed herein or from a naturally occurring variant. E.g., primers suitable for amplifying all or a portion of any of the following regions are provided: an phospholipase C domain (e.g., about amino acid residues 290-435 or 493-609 of SEQ ID NO:2; or about amino acid residues 155-299 and
15 349-465 of SEQ ID NO:5).

A nucleic acid fragment can encode an epitope bearing region of a polypeptide described herein.

A nucleic acid fragment encoding a "biologically active portion of a 16816 or 16839 polypeptide" can be prepared by isolating a portion of the nucleotide sequence of
20 SEQ ID NO:1 or SEQ ID NO:4, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, which encodes a polypeptide having a 16816 or 16839 biological activity (e.g., the biological activities of the 16816 or 16839 proteins as described herein), expressing the encoded portion of the 16816 or 16839 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the
25 encoded portion of the 16816 or 16839 protein. For example, a nucleic acid fragment encoding a biologically active portion of 16816 or 16839 includes a phospholipase C domain (e.g., about amino acid residues 290-435 of SEQ ID NO:2 or 155-299 of SEQ ID NO:5). A nucleic acid fragment encoding a biologically active portion of a 16816 or 16839 polypeptide, may comprise a nucleotide sequence which is greater than 300-1200
30 or more nucleotides in length.

In preferred embodiments, nucleic acids include a nucleotide sequence which is about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400 nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule

of SEQ ID NO:1, or SEQ ID NO:3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

16816 or 16839 Nucleic Acid Variants

5 The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:4, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. Such differences can be due to degeneracy of the genetic code (and result in a nucleic acid which encodes the same 16816 or 16839 proteins as those encoded by the nucleotide
10 sequence disclosed herein. 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 less than 5, 10, 20, 50, or 100 amino acid residues that shown in SEQ ID NO:2 or SEQ ID NO:5. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out
15 sequences from deletions or insertions, or mismatches, are considered differences.

 Nucleic acids of the inventor can be chosen for having codons, which are preferred, or non preferred, for a particular expression system. E.g., the nucleic acid can be one in which at least one codon, at preferably at least 10%, or 20% of the codons has been altered such that the sequence is optimized for expression in E. coli, yeast, human,
20 insect, or CHO cells.

 Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (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
25 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).

 In a preferred embodiment, the nucleic acid differs from that of SEQ ID NO:1 or
30 SEQ ID NO:4, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, e.g., as follows: by at least one but less than 10, 20, 30, or 40 nucleotides; at least one but less than 1%, 5%, 10% or 20% of the in the subject nucleic acid. If necessary for this analysis the sequences should be aligned for maximum

homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

Orthologs, homologs, and allelic variants can be identified using methods known in the art. These variants comprise a nucleotide sequence encoding a polypeptide that is 50%,
5 at least about 55%, typically at least about 70-75%, more typically at least about 80-85%,
and most typically at least about 90-95% or more identical to the amino acid sequence
shown in SEQ ID NO:2 or SEQ ID NO:5 or a fragment of this sequence. Such nucleic acid
molecules can readily be obtained as being able to hybridize under stringent conditions, to
the nucleotide sequence shown in SEQ ID NO:3 or a fragment of this sequence. Nucleic
10 acid molecules corresponding to orthologs, homologs, and allelic variants of the 16816 or
16839 cDNAs of the invention can further be isolated by mapping to the same
chromosome or locus as the 16816 or 16839 gene. Preferred variants include those that are
correlated with phospholipase C activity.

Allelic variants of 16816 or 16839, e.g., human 16816 or 16839, include both
15 functional and non-functional proteins. Functional allelic variants are naturally occurring
amino acid sequence variants of the 16816 or 16839 protein within a population that
maintain the ability to modulate the phosphorylation state of itself or another protein or
polypeptide. Functional allelic variants will typically contain only conservative
substitution of one or more amino acids of SEQ ID NO:2 or SEQ ID NO:5, or
20 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 16816 or 16839, e.g., human 16816 or 16839, protein within a population
that do not have the ability to mediate signal transduction pathways. Non-functional
allelic variants will typically contain a non-conservative substitution, a deletion, or
25 insertion, or premature truncation of the amino acid sequence of SEQ ID NO:2 or SEQ ID
NO:5, or a substitution, insertion, or deletion in critical residues or critical regions of the
protein.

Moreover, nucleic acid molecules encoding other 16816 or 16839 family members
and, thus, which have a nucleotide sequence which differs from the 16816 or 16839
30 sequences of SEQ ID NO:1 or SEQ ID NO:4, or the nucleotide sequence of the DNA
insert of the plasmid deposited with ATCC as Accession Number _____ are intended to
be within the scope of the invention.

Antisense Nucleic Acid Molecules, Ribozymes and Modified 16816 or 16839
Nucleic Acid Molecules

In another aspect, the invention features, an isolated nucleic acid molecule which is antisense to 16816 or 16839. An "antisense" nucleic acid can include a nucleotide
5 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. The antisense nucleic acid can be complementary to an entire 16816 or 16839 coding strand, or to only a portion thereof (e.g., the coding region of human 16816 or 16839 corresponding to SEQ ID NO:3 or SEQ ID NO:6). In
10 another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding 16816 or 16839 (e.g., the 5' and 3' untranslated regions).

An antisense nucleic acid can be designed such that it is complementary to the entire coding region of 16816 or 16839 mRNA, but more preferably is an oligonucleotide
15 which is antisense to only a portion of the coding or noncoding region of 16816 or 16839 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 16816 or 16839 mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65,
20 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
25 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. The antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed
30 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 (e.g., by direct injection at a tissue site), or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 16816 or

16839 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or
5 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
10 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*.
15 *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. A ribozyme having specificity for a 16816 or 16839-encoding nucleic acid can
20 include one or more sequences complementary to the nucleotide sequence of a 16816 or 16839 cDNA disclosed herein (i.e., SEQ ID NO:1, or SEQ ID NO:3), and a sequence having known catalytic sequence responsible for mRNA cleavage (see U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach, (1988) *Nature* 334:585-591). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide
25 sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 16816 or 16839-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, 16816 or 16839 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.

30 16816 or 16839 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 16816 or 16839 (e.g., the 16816 or 16839 promoter and/or enhancers) to form triple helical structures that prevent transcription of the 16816 or 16839 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) *Bioassays* 14(12):807-15. The potential sequences that can be targeted for triple helix formation can be increased by creating a so-called “switchback” nucleic acid molecule. Switchback molecules are synthesized in an alternating 5’-3’, 3’-5’ manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

The invention also provides detectably labeled oligonucleotide primer and probe molecules. Typically, such labels are chemiluminescent, fluorescent, radioactive, or colorimetric.

10 A 16816 or 16839 nucleic acid molecule 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. *et al.*, (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms
15 “peptide nucleic acid” or “PNA” refers to a nucleic acid mimic, e.g., a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can 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
20 synthesis protocols as described in Hyrup B. *et al.*, (1996) *supra*; Perry-O’Keefe *et al.*, *Proc. Natl. Acad. Sci.* 93: 14670-675.

PNAs of 16816 or 16839 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
25 transcription or translation arrest or inhibiting replication. PNAs of 16816 or 16839 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 B., (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.*, (1996)
30 *supra*; Perry-O’Keefe *supra*).

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. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, e.g., Krol *et al.*, (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, e.g., Zon, (1988) *Pharm. Res.* 5:539-549). To this end, the
5 oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

The invention also includes molecular beacon oligonucleotide primer and probe molecules having at least one region which is complementary to a 16816 or 16839 nucleic acid of the invention, two complementary regions one having a fluorophore and one a
10 quencher such that the molecular beacon is useful for quantitating the presence of the 16816 or 16839 nucleic acid of the invention in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi *et al.*, U.S. Patent No. 5,854,033; Nazarenko *et al.*, U.S. Patent No. 5,866,336, and Livak *et al.*, U.S. Patent 5,876,930.

15 Isolated 16816 or 16839 Polypeptides

In another aspect, the invention features, an isolated 16816 or 16839 protein, or fragment, e.g., a biologically active portion, for use as immunogens or antigens to raise or test (or more generally to bind) anti-16816 or 16839 antibodies. 16816 or 16839 protein can be isolated from cells or tissue sources using standard protein purification techniques.
20 16816 or 16839 protein or fragments thereof can be produced by recombinant DNA techniques or synthesized chemically.

Polypeptides of the invention include those which arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing events, and alternative translational and postranslational events. The polypeptide can be expressed in
25 systems, e.g., cultured cells, which result in substantially the same postranslational modifications present when expressed the polypeptide is expressed in a native cell, or in systems which result in the alteration or omission of postranslational modifications, e.g., glycosylation or cleavage, present when expressed in a native cell.

In a preferred embodiment, a 16816 or 16839 polypeptide has one or more of the
30 following characteristics:

- (i) it has the ability to promote the hydrolysis of a phosphatidylinositol;
- (ii) it has a molecular weight, e.g., a deduced molecular weight, amino acid composition or other physical characteristic of the polypeptide of SEQ ID NO:2 or SEQ ID NO:5;

(iii) it has an overall sequence similarity of at least 50%, preferably at least 60%, more preferably at least 70, 80, 90, or 95%, with a polypeptide of SEQ ID NO:2 or SEQ ID NO:5;

5 (iv) it has a PH domain which is preferably about 70%, 80%, 90% or 95% with amino acid residues 17-124 of SEQ ID NO:2;

(v) it has a PLC-X domain which is preferably about 70%, 80%, 90% or 95% with amino acid residues 290-435 of SEQ ID NO:2 or amino acid residues 155-299 of SEQ ID NO:5;

10 (vi) it has a PLC-Y domain which is preferably about 70%, 80%, 90% or 95% with amino acid residues 493-609 of SEQ ID NO:2 or amino acid residues 349-465 of SEQ ID NO:5;

(vii) it has a C2 domain which is preferably about 70%, 80%, 90% or 95% with amino acid residues 628-734 of SEQ ID NO:2 or amino acid residues 484-572 of SEQ ID NO:5;

15 (viii) it has an EF hand domain which is preferably about 70%, 80%, 90% or 95% with amino acid residues 138-166 or 174-202 of SEQ ID NO:2 or amino acid residues 39-67 of SEQ ID NO:5; or

(iv) it has at least 70%, preferably 80%, and most preferably 90% of the cysteines found amino acid sequence of the native protein.

20 In a preferred embodiment the 16816 or 16839 protein, or fragment thereof, differs from the corresponding sequence in SEQ ID NO:2 or SEQ ID NO:5. In one embodiment it differs by at least one but by less than 15, 10 or 5 amino acid residues. In another it differs from the corresponding sequence in SEQ ID NO:2 or SEQ ID NO:5 by at least one residue but less than 20%, 15%, 10% or 5% of the residues in it differ from
25 the corresponding sequence in SEQ ID NO:2 or SEQ ID NO:5. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) The differences are, preferably, differences or changes at a non-essential residue or a conservative substitution. In a preferred embodiment the differences are not in the
30 phospholipase C domain. In another preferred embodiment one or more differences are in non-active site residues, e.g. outside of the phospholipase C domain.

Other embodiments include a protein that contain one or more changes in amino acid sequence, e.g., a change in an amino acid residue which is not essential for activity.

Such 16816 or 16839 proteins differ in amino acid sequence from SEQ ID NO:2 or SEQ ID NO:5, yet retain biological activity.

In one embodiment, a biologically active portion of a 16816 or 16839 protein includes an phospholipase C domain. In another embodiment, a biologically active
5 portion of a 16816 or 16839 protein includes a protein kinase C phosphorylation domain. 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 16816 or 16839 protein.

In a preferred embodiment, the 16816 or 16839 protein has an amino acid
10 sequence shown in SEQ ID NO:2 or SEQ ID NO:5. In other embodiments, the 16816 or 16839 protein is substantially identical to SEQ ID NO:2 or SEQ ID NO:5. In yet another embodiment, the 16816 or 16839 protein is substantially identical to SEQ ID NO:2 or SEQ ID NO:5 and retains the functional activity of the protein of SEQ ID NO:2 or SEQ ID NO:5, as described in detail above. Accordingly, in another embodiment, the 16816
15 or 16839 protein is a protein which includes an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to SEQ ID NO:2 or SEQ ID NO:5.

16816 or 16839 Chimeric or Fusion Proteins

In another aspect, the invention provides 16816 or 16839 chimeric or fusion
20 proteins. As used herein, a 16816 or 16839 "chimeric protein" or "fusion protein" includes a 16816 or 16839 polypeptide linked to a non-16816 or 16839 polypeptide. A "non-16816 or 16839 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the 16816 or 16839
25 protein, e.g., a protein which is different from the 16816 or 16839 protein and which is derived from the same or a different organism. The 16816 or 16839 polypeptide of the fusion protein can correspond to all or a portion e.g., a fragment described herein of a 16816 or 16839 amino acid sequence. In a preferred embodiment, a 16816 or 16839 fusion protein includes at least one (or two) biologically active portion of a 16816 or
30 16839 protein. The non-16816 or 16839 polypeptide can be fused to the N-terminus or C-terminus of the 16816 or 16839 polypeptide.

The fusion protein can include a moiety which has a high affinity for a ligand. For example, the fusion protein can be a GST-16816 or 16839 fusion protein in which the 16816 or 16839 sequences are fused to the C-terminus of the GST sequences. Such

fusion proteins can facilitate the purification of recombinant 16816 or 16839.

Alternatively, the fusion protein can be a 16816 or 16839 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of 16816 or 16839 can be increased through use of a heterologous signal sequence.

Fusion proteins can include all or a part of a serum protein, e.g., an IgG constant region, or human serum albumin.

The 16816 or 16839 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The 16816 or 16839 fusion proteins can be used to affect the bioavailability of a 16816 or 16839 substrate. 16816 or 16839 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 16816 or 16839 protein; (ii) mis-regulation of the 16816 or 16839 gene; and (iii) aberrant post-translational modification of a 16816 or 16839 protein.

Moreover, the 16816 or 16839-fusion proteins of the invention can be used as immunogens to produce anti-16816 or 16839 antibodies in a subject, to purify 16816 or 16839 ligands and in screening assays to identify molecules which inhibit the interaction of 16816 or 16839 with a 16816 or 16839 substrate.

Expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A 16816 or 16839-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 16816 or 16839 protein.

Variants of 16816 or 16839 Proteins

In another aspect, the invention also features a variant of a 16816 or 16839 polypeptide, e.g., which functions as an agonist (mimetics) or as an antagonist. Variants of the 16816 or 16839 proteins can be generated by mutagenesis, e.g., discrete point mutation, the insertion or deletion of sequences or the truncation of a 16816 or 16839 protein. An agonist of the 16816 or 16839 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 16816 or 16839 protein. An antagonist of a 16816 or 16839 protein can inhibit one or more of the activities of the naturally occurring form of the 16816 or 16839 protein by, for example, competitively modulating a 16816 or 16839-mediated activity of a 16816 or 16839 protein. Thus, specific biological effects can be elicited by treatment with a variant of

limited function. Preferably, 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 16816 or 16839 protein.

5 Variants of a 16816 or 16839 protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a 16816 or 16839 protein for agonist or antagonist activity.

Libraries of fragments e.g., N terminal, C terminal, or internal fragments, of a 16816 or 16839 protein coding sequence can be used to generate a variegated population
10 of fragments for screening and subsequent selection of variants of a 16816 or 16839 protein.

Variants in which a cysteine residues is added or deleted or in which a residue which is glycosylated is added or deleted are particularly preferred.

Methods for screening gene products of combinatorial libraries made by point
15 mutations or truncation, and for screening cDNA libraries for gene products having a selected property. 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 16816 or 16839 variants (Arkin and Yourvan, (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.*, (1993) *Protein Engineering*
20 6(3):327-331).

Cell based assays can be exploited to analyze a variegated 16816 or 16839 library. For example, a library of expression vectors can be transfected into a cell line, e.g., a cell line, which ordinarily responds to 16816 or 16839 in a substrate-dependent manner. The transfected cells are then contacted with 16816 or 16839 and the effect of the expression
25 of the mutant on signaling by the 16816 or 16839 substrate can be detected, e.g., by measuring phospholipase C activity. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the 16816 or 16839 substrate, and the individual clones further characterized.

In another aspect, the invention features a method of making a 16816 or 16839
30 polypeptide, e.g., a peptide having a non-wild type activity, e.g., an antagonist, agonist, or super agonist of a naturally occurring 16816 or 16839 polypeptide, e.g., a naturally occurring 16816 or 16839 polypeptide. The method includes: altering the sequence of a 16816 or 16839 polypeptide, e.g., altering the sequence, e.g., by substitution or deletion

of one or more residues of a non-conserved region, a domain or residue disclosed herein, and testing the altered polypeptide for the desired activity.

In another aspect, the invention features a method of making a fragment or analog of a 16816 or 16839 polypeptide a biological activity of a naturally occurring 16816 or 16839 polypeptide. The method includes: altering the sequence, e.g., by substitution or deletion of one or more residues, of a 16816 or 16839 polypeptide, e.g., altering the sequence of a non-conserved region, or a domain or residue described herein, and testing the altered polypeptide for the desired activity.

10 Anti-16816 or 16839 Antibodies

In another aspect, the invention provides an anti-16816 or 16839 antibody. The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. 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 antibody can be a polyclonal, monoclonal, recombinant, e.g., a chimeric or humanized, fully human, non-human, e.g., murine, or single chain antibody. In a preferred embodiment it has effector function and can fix complement. The antibody can be coupled to a toxin or imaging agent.

A full-length 16816 or 16839 protein or, antigenic peptide fragment of 16816 or 16839 can be used as an immunogen or can be used to identify anti-16816 or 16839 antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. The antigenic peptide of 16816 or 16839 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 or SEQ ID NO:5 and encompasses an epitope of 16816 or 16839. Preferably, the antigenic peptide includes 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.

Fragments of 16816 or 16839 which include, e.g., residues 461-491 of SEQ ID NO:2 or 466-486 of SEQ ID NO:5 can be, e.g., used as immunogens, or used to characterize the specificity of an antibody or antibodies against what are believed to be hydrophilic regions of the 16816 or 16839 protein. Similarly, a fragment of 16816 or 16839 which includes, e.g., residues 686-706 of SEQ ID NO:2 or residues 486-501 of

SEQ ID NO:5 can be used to make an antibody against what is believed to be a hydrophobic region of the 16816 or 16839 protein; a fragment of 16816 or 16839 which includes residues 290-435 of SEQ ID NO:2 or residues 156-300 of SEQ ID NO:5 can be used to make an antibody against the phospholipase C region of the 16816 or 16839 protein.

Antibodies reactive with, or specific for, any of these regions, or other regions or domains described herein are provided.

In a preferred embodiment the antibody fails to bind an Fc receptor, e.g. it is a type which does not support Fc receptor binding or has been modified, e.g., by deletion or other mutation, such that it does not have a functional Fc receptor binding region.

Preferred epitopes encompassed by the antigenic peptide are regions of 16816 or 16839 are located on the surface of the protein, e.g., hydrophilic regions, as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human 16816 or 16839 protein sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the 16816 or 16839 protein and are thus likely to constitute surface residues useful for targeting antibody production.

In a preferred embodiment the antibody binds an epitope on any domain or region on 16816 or 16839 proteins described herein.

Chimeric, humanized, but most preferably, completely human antibodies are desirable for applications which include repeated administration, e.g., therapeutic treatment (and some diagnostic applications) of human patients.

The anti-16816 or 16839 antibody can be a single chain antibody. A single-chain antibody (scFV) may be engineered (see, for example, Colcher, D. *et al.*, *Ann. NY Acad. Sci.* 1999 Jun 30;880:263-80; and Reiter, Y., *Clin. Cancer Res.* 1996 Feb;2(2):245-52). The single chain antibody can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target 16816 or 16839 protein.

An anti-16816 or 16839 antibody (e.g., monoclonal antibody) can be used to isolate 16816 or 16839 by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an anti-16816 or 16839 antibody can be used to detect 16816 or 16839 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-16816 or 16839 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 (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Recombinant Expression Vectors, Host Cells and Genetically Engineered Cells

In another aspect, the invention includes, vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide described herein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors include, e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses.

A vector can include a 16816 or 16839 nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably the recombinant expression vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. 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. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or polypeptides, including fusion proteins or polypeptides, encoded by nucleic acids as

described herein (e.g., 16816 or 16839 proteins, mutant forms of 16816 or 16839 proteins, fusion proteins, and the like).

The recombinant expression vectors of the invention can be designed for expression of 16816 or 16839 proteins in prokaryotic or eukaryotic cells. For example, polypeptides of the invention can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

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 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, 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 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 used in 16816 or 16839 activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific for 16816 or 16839 proteins. In a preferred embodiment, a fusion protein expressed in a retroviral expression vector of the present invention can be used 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).

To maximize recombinant protein expression in *E. coli* is to express the protein in host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic

Press, San Diego, California (1990) 119-128). 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.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

The 16816 or 16839 expression vector can be a yeast expression vector, a vector for expression in insect cells, e.g., a baculovirus expression vector or a vector suitable for expression in mammalian cells.

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.

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). 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. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus. 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 the invention provides a host cell which includes a nucleic acid molecule described herein, e.g., a 16816 or 16839 nucleic acid molecule within a recombinant expression vector or a 16816 or 16839 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. Such terms refer not only to the particular subject cell but rather also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a 16816 or 16839 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 host 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

A host cell of the invention can be used to produce (i.e., express) a 16816 or 16839 protein. Accordingly, the invention further provides methods for producing a 16816 or 16839 protein using the host cells of the invention. In one embodiment, the method includes culturing the host cell of the invention (into which a recombinant expression vector encoding a 16816 or 16839 protein has been introduced) in a suitable medium such that a 16816 or 16839 protein is produced. In another embodiment, the method further includes isolating a 16816 or 16839 protein from the medium or the host cell.

In another aspect, the invention features, a cell or purified preparation of cells which include a 16816 or 16839 transgene, or which otherwise misexpress 16816 or 16839. The cell preparation can consist of human or non-human cells, e.g., rodent cells, e.g., mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a 16816 or 16839 transgene, e.g., a heterologous form of a 16816 or 16839,

e.g., a gene derived from humans (in the case of a non-human cell). The 16816 or 16839 transgene can be misexpressed, e.g., overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpress an endogenous 16816 or 16839, e.g., a gene the expression of which is disrupted, e.g., a knockout. Such cells can
5 serve as a model for studying disorders which are related to mutated or mis-expressed 16816 or 16839 alleles or for use in drug screening.

In another aspect, the invention features, a human cell, e.g., a hematopoietic stem cell, transformed with nucleic acid which encodes a subject 16816 or 16839 polypeptide.

Also provided are cells or a purified preparation thereof, e.g., human cells, in
10 which an endogenous 16816 or 16839 is under the control of a regulatory sequence that does not normally control the expression of the endogenous 16816 or 16839 gene. The expression characteristics of an endogenous gene within a cell, e.g., a cell line or microorganism, can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the
15 endogenous 16816 or 16839 gene. For example, an endogenous 16816 or 16839 gene, e.g., a gene which is "transcriptionally silent," e.g., not normally expressed, or expressed only at very low levels, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the
20 heterologous DNA as described in, e.g., Chappel, US 5,272,071; WO 91/06667, published on May 16, 1991.

Transgenic Animals

The invention provides non-human transgenic animals. Such animals are useful
25 for studying the function and/or activity of a 16816 or 16839 protein and for identifying and/or evaluating modulators of 16816 or 16839 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,
30 chickens, amphibians, and the like. A transgene is exogenous DNA or a rearrangement, e.g., a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic animal. A transgene can direct the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, other transgenes, e.g., a knockout, reduce expression. Thus, a

transgenic animal can be one in which an endogenous 16816 or 16839 gene has been altered by, e.g., 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.

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 transgene of the invention to direct expression of a 16816 or 16839 protein to particular cells. A transgenic founder animal can be identified based upon the presence of a 16816 or 16839 transgene in its genome
10 and/or expression of 16816 or 16839 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 16816 or 16839 protein can further be bred to other transgenic animals carrying other transgenes.

16816 or 16839 proteins or polypeptides can be expressed in transgenic animals or
15 plants, e.g., a nucleic acid encoding the protein or polypeptide can be introduced into the genome of an animal. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, e.g., a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Suitable animals are mice, pigs, cows, goats, and sheep.

20 The invention also includes a population of cells from a transgenic animal, as discussed herein.

Uses

The nucleic acid molecules, proteins, protein homologues, and antibodies
25 described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic). The isolated nucleic acid molecules of the invention can be used, for example, to express a 16816 or 16839 protein (e.g., via a recombinant expression vector
30 in a host cell in gene therapy applications), to detect a 16816 or 16839 mRNA (e.g., in a biological sample) or a genetic alteration in a 16816 or 16839 gene, and to modulate 16816 or 16839 activity, as described further below. The 16816 or 16839 proteins can be used to treat disorders characterized by insufficient or excessive production of a 16816 or 16839 substrate or production of 16816 or 16839 inhibitors. In addition, the 16816 or

16839 proteins can be used to screen for naturally occurring 16816 or 16839 substrates, to screen for drugs or compounds which modulate 16816 or 16839 activity, as well as to treat disorders characterized by insufficient or excessive production of 16816 or 16839 protein or production of 16816 or 16839 protein forms which have decreased, aberrant or unwanted activity compared to 16816 or 16839 wild-type protein. Such disorders include those characterized by aberrant signaling or aberrant, e.g., hyperproliferative, cell growth. Moreover, the anti-16816 or 16839 antibodies of the invention can be used to detect and isolate 16816 or 16839 proteins, regulate the bioavailability of 16816 or 16839 proteins, and modulate 16816 or 16839 activity.

A method of evaluating a compound for the ability to interact with, e.g., bind, a subject 16816 or 16839 polypeptide is provided. The method includes: contacting the compound with the subject 16816 or 16839 polypeptide; and evaluating ability of the compound to interact with, e.g., to bind or form a complex with the subject 16816 or 16839 polypeptide. This method can be performed *in vitro*, e.g., in a cell free system, or *in vivo*, e.g., in a two-hybrid interaction trap assay. This method can be used to identify naturally occurring molecules which interact with subject 16816 or 16839 polypeptide. It can also be used to find natural or synthetic inhibitors of subject 16816 or 16839 polypeptide. Screening methods are discussed in more detail below.

Screening Assays:

The invention provides methods (also referred to herein as "screening assays") for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to 16816 or 16839 proteins, have a stimulatory or inhibitory effect on, for example, 16816 or 16839 expression or 16816 or 16839 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 16816 or 16839 substrate. Compounds thus identified can be used to modulate the activity of target gene products (e.g., 16816 or 16839 genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions.

In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a 16816 or 16839 protein or polypeptide or a 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 16816 or 16839 protein or polypeptide or a biologically active portion thereof.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries [libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic
5 degradation but which nevertheless remain bioactive] (see, e.g., Zuckermann, R.N. *et al.*, *J. Med. Chem.* 1994, 37: 2678-85); 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 and peptoid library approaches are limited to peptide
10 libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

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

Libraries of compounds may be presented in solution (e.g., Houghten, (1992) *Biotechniques* 13:412-421), or on beads (Lam, (1991) *Nature* 354:82-84), chips (Fodor, (1993) *Nature* 364:555-556), bacteria or spores (Ladner, United States Patent No. 5,223,409), 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.* 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 16816 or 16839 protein or biologically active portion thereof is contacted with a test compound, and the ability of the test compound to modulate 16816 or 16839 activity is determined. Determining the ability of the test compound to modulate 16816 or 16839
30 activity can be accomplished by monitoring, for example, phospholipase C activity. The cell, for example, can be of mammalian origin, e.g., human. Cell homogenates, or fractions, preferably membrane containing fractions, can also be tested.

The ability of the test compound to modulate 16816 or 16839 binding to a compound, e.g., a 16816 or 16839 substrate, or to bind to 16816 or 16839 can also be

evaluated. This can be accomplished, for example, by coupling the compound, e.g., the substrate, with a radioisotope or enzymatic label such that binding of the compound, e.g., the substrate, to 16816 or 16839 can be determined by detecting the labeled compound, e.g., substrate, in a complex. Alternatively, 16816 or 16839 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate 16816 or 16839 binding to a 16816 or 16839 substrate in a complex. For example, compounds (e.g., 16816 or 16839 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 radioemmission 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.

The ability of a compound (e.g., a 16816 or 16839 substrate) to interact with 16816 or 16839 with or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with 16816 or 16839 without the labeling of either the compound or the 16816 or 16839. 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 16816 or 16839.

In yet another embodiment, a cell-free assay is provided in which a 16816 or 16839 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the 16816 or 16839 protein or biologically active portion thereof is evaluated. Preferred biologically active portions of the 16816 or 16839 proteins to be used in assays of the present invention include fragments which participate in interactions with non-16816 or 16839 molecules, e.g., fragments with high surface probability scores.

Soluble and/or membrane-bound forms of isolated proteins (e.g., 16816 or 16839 proteins or biologically active portions thereof) can be used in the cell-free assays of the invention. When membrane-bound forms of the protein are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic 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 N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

In one embodiment, assays are performed where the ability of an agent to block phospholipase C activity within a cell is evaluated.

The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FET) (see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos, *et al.*, U.S. Patent No. 4,868,103). A fluorophore label on the first, 'donor' molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

In another embodiment, determining the ability of the 16816 or 16839 protein to bind to a target molecule can be accomplished using real-time Biomolecular Interaction Analysis (BIA) (see, e.g., Sjolander, S. and Urbaniczky, C., (1991) *Anal. Chem.* **63**:2338-2345 and Szabo *et al.*, (1995) *Curr. Opin. Struct. Biol.* **5**:699-705). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the

optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid
5 phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

It may be desirable to immobilize either 16816 or 16839, an anti-16816 or 16839 antibody or its target molecule to facilitate separation of complexed from uncomplexed
10 forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a 16816 or 16839 protein, or interaction of a 16816 or 16839 protein with a 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.
15 In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/16816 or 16839 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test
20 compound or the test compound and either the non-adsorbed target protein or 16816 or 16839 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,
25 for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 16816 or 16839 binding or activity determined using standard techniques.

Other techniques for immobilizing either a 16816 or 16839 protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated
30 16816 or 16839 protein 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).

In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of
5 complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the
10 immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

In one embodiment, this assay is performed utilizing antibodies reactive with 16816 or 16839 protein or target molecules but which do not interfere with binding of the 16816 or 16839 protein to its target molecule. Such antibodies can be derivatized to the
15 wells of the plate, and unbound target or 16816 or 16839 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 16816 or 16839 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated
20 with the 16816 or 16839 protein or target molecule.

Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (see, for example, Rivas, G., and Minton, A.P., *Trends Biochem Sci* 1993 Aug; 18(8):284-
25 7); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel, F. *et al.*, eds. *Current Protocols in Molecular Biology* 1999, J. Wiley: New York.); and immunoprecipitation (see, for example, Ausubel, F. *et al.*, eds. *Current Protocols in Molecular Biology* 1999, J. Wiley: New York). Such resins and chromatographic techniques are known to one skilled in the art (see, e.g., Heegaard, N.H., *J Mol. Recognit.* 1998 Winter; 11(1-6):141-8; Hage, D.S., and Tweed, S.A., *J. Chromatogr. B Biomed. Sci. Appl.* 1997 Oct 10; 699(1-2):499-525). Further,
30 fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

In a preferred embodiment, the assay includes contacting the 16816 or 16839 protein or biologically active portion thereof with a known compound which binds 16816 or 16839 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 16816 or 16839 protein, wherein determining the ability of the test compound to interact with a 16816 or 16839 protein includes determining the ability of the test compound to preferentially bind to 16816 or 16839 or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

The target gene products of the invention can, *in vivo*, interact with one or more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as "binding partners." Compounds that disrupt such interactions can be useful in regulating the activity of the target gene product. Such compounds can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the 16816 or 16839 genes herein identified. In an alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of a 16816 or 16839 protein through modulation of the activity of a downstream effector of a 16816 or 16839 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.

To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), e.g., a substrate, a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within

reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the species to be anchored can be used to anchor the species to the solid surface.

In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

In yet another aspect, the 16816 or 16839 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 16816 or 16839 ("16816 or 16839-binding proteins" or "16816 or 16839-bp") and are involved in 16816 or 16839 activity. Such 16816 or 16839-bps can be activators or inhibitors of signals by the 16816 or 16839 proteins or 16816 or 16839 targets as, for example, downstream elements of a 16816 or 16839-mediated signaling pathway.

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 16816 or 16839 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.

(Alternatively the: 16816 or 16839 protein can be the fused to the activator domain.) If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a 16816 or 16839-

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
5 containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the 16816 or 16839 protein.

In another embodiment, modulators of 16816 or 16839 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of 16816 or 16839 mRNA or protein evaluated relative to the level of
10 expression of 16816 or 16839 mRNA or protein in the absence of the candidate compound. When expression of 16816 or 16839 mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 16816 or 16839 mRNA or protein expression. Alternatively,
15 less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 16816 or 16839 mRNA or protein expression. The level of 16816 or 16839 mRNA or protein expression can be determined by methods described herein for detecting 16816 or 16839 mRNA or protein.

In another aspect, the invention pertains to a combination of two or more of the
20 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 16816 or 16839 protein can be confirmed *in vivo*, e.g., in an animal.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an
25 agent identified as described herein (e.g., a 16816 or 16839 modulating agent, an antisense 16816 or 16839 nucleic acid molecule, a 16816 or 16839-specific antibody, or a 16816 or 16839-binding partner) in an appropriate animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be used
30 for treatments as described herein.

Detection Assays

Portions or fragments of the nucleic acid sequences identified herein can be used as polynucleotide reagents. For example, these sequences can be used to: (i) map their

respective genes on a chromosome e.g., to locate gene regions associated with genetic disease or to associate 16816 or 16839 with a 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.

5

Chromosome Mapping

The 16816 or 16839 nucleotide sequences or portions thereof can be used to map the location of the 16816 or 16839 genes on a chromosome. This process is called chromosome mapping. Chromosome mapping is useful in correlating the 16816 or 16839
10 sequences with genes associated with disease.

Briefly, 16816 or 16839 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the 16816 or 16839 nucleotide sequences. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene
15 corresponding to the 16816 or 16839 sequences will yield an amplified fragment.

A panel of somatic cell hybrids in which each cell line contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, can allow easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. *et al.*, (1983) *Science* 220:919-924).

20 Other mapping strategies e.g., 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 can be used to map 16816 or 16839 to a chromosomal location.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase
25 chromosomal spread can further be used to provide a precise chromosomal location in one step. 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
30 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 V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, 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 16816 or 16839 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.

Tissue Typing

16816 or 16839 sequences can be used to identify individuals from biological samples using, e.g., restriction fragment length polymorphism (RFLP). In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, the fragments separated, e.g., in a Southern blot, and probed to yield bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 16816 or 16839 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.

Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. 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 or SEQ ID NO:4 can 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 or SEQ ID NO:6 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from 16816 or 16839 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.

Use of Partial 16816 or 16839 Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. 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 or SEQ ID NO:4 (e.g., fragments derived from the noncoding regions of SEQ ID NO:1 or SEQ

ID NO:4 having a length of at least 20 bases, preferably at least 30 bases) are particularly appropriate for this use.

The 16816 or 16839 nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., a tissue containing phospholipase C activity. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 16816 or 16839 probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, e.g., 16816 or 16839 primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual.

Generally, the invention provides, a method of determining if a subject is at risk for a disorder related to a lesion in or the misexpression of a gene which encodes 16816 or 16839.

Such disorders include, e.g., a disorder associated with the misexpression of 16816 or 16839, or lipid metabolism related disorder.

The method includes one or more of the following:

detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the 16816 or 16839 gene, or detecting the presence or absence of a mutation in a region which controls the expression of the gene, e.g., a mutation in the 5' control region;

detecting, in a tissue of the subject, the presence or absence of a mutation which alters the structure of the 16816 or 16839 gene;

detecting, in a tissue of the subject, the misexpression of the 16816 or 16839 gene, at the mRNA level, e.g., detecting a non-wild type level of a mRNA;

detecting, in a tissue of the subject, the misexpression of the gene, at the protein level, e.g., detecting a non-wild type level of a 16816 or 16839 polypeptide.

In preferred embodiments the method includes: ascertaining the existence of at least one of: a deletion of one or more nucleotides from the 16816 or 16839 gene; an

insertion of one or more nucleotides into the gene, a point mutation, e.g., a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, e.g., a translocation, inversion, or deletion.

For example, detecting the genetic lesion can include: (i) providing a
5 probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence from SEQ ID NO:1 naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the 16816 or 16839 gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting, by hybridization, e.g., *in situ* hybridization, of the probe/primer to the nucleic acid, the
10 presence or absence of the genetic lesion.

In preferred embodiments detecting the misexpression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the 16816 or 16839 gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; or a non-wild type level of 16816 or 16839.

15 Methods of the invention can be used prenatally or to determine if a subject's offspring will be at risk for a disorder.

In preferred embodiments the method includes determining the structure of a 16816 or 16839 gene, an abnormal structure being indicative of risk for the disorder.

In preferred embodiments the method includes contacting a sample from the
20 subject with an antibody to the 16816 or 16839 protein or a nucleic acid, which hybridizes specifically with the gene. These and other embodiments are discussed below.

Diagnostic and Prognostic Assays

The presence, level, or absence of 16816 or 16839 protein or nucleic acid in a
25 biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 16816 or 16839 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes 16816 or 16839 protein such that the presence of 16816 or 16839 protein or nucleic acid is detected in the biological sample. The term "biological sample" includes tissues, cells
30 and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. A preferred biological sample is serum. The level of expression of the 16816 or 16839 gene can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the 16816 or 16839 genes; measuring the amount of

protein encoded by the 16816 or 16839 genes; or measuring the activity of the protein encoded by the 16816 or 16839 genes.

The level of mRNA corresponding to the 16816 or 16839 gene in a cell can be determined both by *in situ* and by *in vitro* formats.

5 The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe
10 can be, for example, a full-length 16816 or 16839 nucleic acid, such as the nucleic acid of SEQ ID NO:1, or the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 16816 or 16839 mRNA or genomic DNA. Other suitable probes
15 for use in the diagnostic assays are described herein.

 In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with the
20 probes, for example, in a two-dimensional gene chip array. A skilled artisan can adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the 16816 or 16839 genes.

 The level of mRNA in a sample that is encoded by one of 16816 or 16839 can be evaluated with nucleic acid amplification, e.g., by rtPCR (Mullis, 1987, U.S. Patent No.
25 4,683,202), ligase chain reaction (Barany, 1991, *Proc. Natl. Acad. Sci. USA* 88:189-193), self sustained sequence replication (Guatelli *et al.*, 1990, *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.*, 1989, *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.*, 1988, *Bio/Technology* 6:1197), rolling circle replication (Lizardi *et al.*, U.S. Patent No. 5,854,033) or any other nucleic
30 acid amplification method, followed by the detection of the amplified molecules using techniques known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a

region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

For *in situ* methods, a cell or tissue sample can be prepared/processed and
5 immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the 16816 or 16839 gene being analyzed.

In another embodiment, the methods further contacting a control sample with a compound or agent capable of detecting 16816 or 16839 mRNA, or genomic DNA, and comparing the presence of 16816 or 16839 mRNA or genomic DNA in the control
10 sample with the presence of 16816 or 16839 mRNA or genomic DNA in the test sample.

A variety of methods can be used to determine the level of protein encoded by 16816 or 16839. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample. In a preferred embodiment, the antibody bears a detectable label. Antibodies
15 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 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 a detectable substance.
20 Examples of detectable substances are provided herein.

The detection methods can be used to detect 16816 or 16839 protein in a biological sample *in vitro* as well as *in vivo*. *In vitro* techniques for detection of 16816 or 16839 protein include enzyme linked immunosorbent assays (ELISAs), immunoprecipitations, immunofluorescence, enzyme immunoassay (EIA),
25 radioimmunoassay (RIA), and Western blot analysis. *In vivo* techniques for detection of 16816 or 16839 protein include introducing into a subject a labeled anti-16816 or 16839 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.

In another embodiment, the methods further include contacting the control sample
30 with a compound or agent capable of detecting 16816 or 16839 protein, and comparing the presence of 16816 or 16839 protein in the control sample with the presence of 16816 or 16839 protein in the test sample.

The invention also includes kits for detecting the presence of 16816 or 16839 in a biological sample. For example, the kit can include a compound or agent capable of detecting 16816 or 16839 protein or mRNA in a biological sample; and 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 16816 or 16839 protein or nucleic acid.

For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also include a buffering agent, a preservative, or a protein-stabilizing agent. The kit can also include components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with misexpressed or aberrant or unwanted 16816 or 16839 expression or activity. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as pain or deregulated cell proliferation.

In one embodiment, a disease or disorder associated with aberrant or unwanted 16816 or 16839 expression or activity is identified. A test sample is obtained from a subject and 16816 or 16839 protein or nucleic acid (e.g., mRNA or genomic DNA) is evaluated, wherein the level, e.g., the presence or absence, of 16816 or 16839 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 16816 or 16839 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest, including a biological fluid (e.g., serum), cell sample, or tissue.

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 16816 or 16839 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a cellular growth related disorder.

The methods of the invention can also be used to detect genetic alterations in a 16816 or 16839 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in 16816 or 16839 protein activity or nucleic acid expression, such as a cellular growth related disorder. In preferred embodiments, the methods include detecting, in a sample 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 16816 or 16839-protein, or the mis-expression of the 16816 or 16839 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 16816 or 16839 gene; 2) an addition of one or more nucleotides to a 16816 or 16839 gene; 3) a substitution of one or more nucleotides of a 16816 or 16839 gene, 4) a chromosomal rearrangement of a 16816 or 16839 gene; 5) an alteration in the level of a messenger RNA transcript of a 16816 or 16839 gene, 6) aberrant modification of a 16816 or 16839 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 16816 or 16839 gene, 8) a non-wild type level of a 16816 or 16839-protein, 9) allelic loss of a 16816 or 16839 gene, and 10) inappropriate post-translational modification of a 16816 or 16839-protein.

An alteration can be detected without a probe/primer in a polymerase chain reaction, such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR), the latter of which can be particularly useful for detecting point mutations in the 16816 or 16839-gene. This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (e.g., genomic, mRNA or both) from the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a 16816 or 16839 gene under conditions such that hybridization and amplification of the 16816 or 16839-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. 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), transcriptional
5 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 other nucleic acid amplification methods, followed by the detection of the amplified molecules using techniques known to those of skill in the art.

In another embodiment, mutations in a 16816 or 16839 gene from a sample cell
10 can be identified by detecting 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, e.g., by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence
15 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.

In other embodiments, genetic mutations in 16816 or 16839 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, two-dimensional arrays, e.g., chip based arrays. Such arrays include a plurality of addresses, each of which
20 is positionally distinguishable from the other. A different probe is located at each address of the plurality. The arrays can have a high density of addresses, e.g., can contain hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.*, (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.*, (1996) *Nature Medicine* 2:753-759). For example, genetic mutations in 16816 or 16839 can be identified in two dimensional arrays
25 containing light-generated DNA probes as described in Cronin, M.T. *et al.* *supra*.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 16816 or 16839 gene and detect mutations by comparing the sequence of the sample 16816 or 16839 with the corresponding wild-type (control) sequence. Automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry.

Other methods for detecting mutations in the 16816 or 16839 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; Cotton *et al.*, (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba *et al.*, (1992) *Methods Enzymol.* 217:286-295).

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 16816 or 16839 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; U.S. Patent No. 5,459,039).

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 16816 or 16839 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 16816 or 16839 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 (Saiki *et al.*, (1986) *Nature* 324:163); Saiki *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:6230).

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. 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 16816 or 16839 gene.

Use of 16816 or 16839 Molecules as Surrogate Markers

The 16816 or 16839 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 16816 or 16839 molecules of the invention may be detected, and may be correlated with one or more biological states *in vivo*. For example, the 16816 or 16839 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 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 16816 or 16839 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
5 small amount of a drug may be sufficient to activate multiple rounds of marker (e.g., a 16816 or 16839 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-16816 or 16839 antibodies may be employed in an
10 immune-based detection system for a 16816 or 16839 protein marker, or 16816 or 16839-specific radiolabeled probes may be used to detect a 16816 or 16839 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.* US
15 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 16816 or 16839 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an
20 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
25 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., 16816 or 16839 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
30 present in the subject. Similarly, the presence or absence of a specific sequence mutation in 16816 or 16839 DNA may correlate 16816 or 16839 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.

Pharmaceutical Compositions

The nucleic acid and polypeptides, fragments thereof, as well as anti-16816 or 16839 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions. Such compositions typically include the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration.

Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in 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 syringability exists. It should 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 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. 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, e.g., gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

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

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be

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

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

10 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.
15 Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled
20 in the art, for example, as described in U.S. Patent No. 4,522,811.

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

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
30 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 ratio LD₅₀/ED₅₀. Compounds which exhibit high 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.

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 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 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 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 protein or polypeptide can be administered 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. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

For antibodies, the preferred dosage is 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is

described by Cruikshank *et al.*, ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules
5 include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), 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
10 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.

Exemplary doses include milligram or microgram amounts of the small molecule
15 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.
20 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
25 animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

An antibody (or fragment thereof) may be conjugated to a therapeutic moiety such
30 as a cytotoxin, a therapeutic agent or a radioactive metal ion. A 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, 5 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 10 (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 15 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 macrophase colony stimulating factor ("GM-CSF"), granulocyte colony 20 stimulating factor ("G-CSF"), or other growth factors.

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 25 example, intravenous injection, local administration (see U.S. Patent 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 the complete gene delivery 30 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.

Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted 16816 or 16839 expression or activity. With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. As used herein, the term "treatment" is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

"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 genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 16816 or 16839 molecules of the present invention or 16816 or 16839 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 treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted 16816 or 16839 expression or activity, by administering to the subject a 16816 or 16839 or an agent which modulates 16816 or 16839 expression or at least one 16816 or 16839 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted 16816 or 16839 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 16816 or 16839 aberrance, such that a disease or disorder is prevented or, alternatively, delayed in

its progression. Depending on the type of 16816 or 16839 aberrance, for example, a 16816 or 16839, 16816 or 16839 agonist or 16816 or 16839 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

5 It is possible that some 16816 or 16839 disorders can be caused, at least in part, by an abnormal level of gene product, or by the presence of a gene product exhibiting abnormal activity. As such, the reduction in the level and/or activity of such gene products would bring about the amelioration of disorder symptoms.

 As discussed, successful treatment of 16816 or 16839 disorders can be brought
10 about by techniques that serve to inhibit the expression or activity of target gene products. For example, compounds, e.g., an agent identified using an assays described above, that proves to exhibit negative modulatory activity, can be used in accordance with the invention to prevent and/or ameliorate symptoms of 16816 or 16839 disorders. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or
15 inorganic molecules, or antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and FAb, F(ab')₂ and FAb expression library fragments, scFV molecules, and epitope-binding fragments thereof).

 Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene
20 expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

 It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription
25 (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that
30 the target gene encodes an extracellular protein, it can be preferable to co-administer normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

 Another method by which nucleic acid molecules may be utilized in treating or preventing a disease characterized by 16816 or 16839 expression is through the use of

aptamer molecules specific for 16816 or 16839 protein. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to protein ligands (see, e.g., Osborne, *et al.*, *Curr. Opin. Chem. Biol.* 1997, 1(1): 5-9; and Patel, D.J., *Curr. Opin. Chem. Biol.* 1997 Jun;1(1):32-46). Since nucleic acid molecules may in many cases be more conveniently introduced into target cells than therapeutic protein molecules may be, aptamers offer a method by which 16816 or 16839 protein activity may be specifically decreased without the introduction of drugs or other molecules which may have pluripotent effects.

Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies may, therefore, be administered in instances whereby negative modulatory techniques are appropriate for the treatment of 16816 or 16839 disorders. For a description of antibodies, see the Antibody section above.

In circumstances wherein injection of an animal or a human subject with a 16816 or 16839 protein or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against 16816 or 16839 through the use of anti-idiotypic antibodies (see, for example, Herlyn, D., *Ann. Med.* 1999;31(1):66-78; and Bhattacharya-Chatterjee, M., and Foon, K.A., *Cancer Treat. Res.* 1998;94:51-68). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the 16816 or 16839 protein. Vaccines directed to a disease characterized by 16816 or 16839 expression may also be generated in this fashion.

In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (see e.g., Marasco *et al.*, (1993, *Proc. Natl. Acad. Sci. USA* 90:7889-7893).

The identified compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat

or ameliorate 16816 or 16839 disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for
5 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 ratio LD₅₀/ED₅₀. Compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system
10 that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

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
15 no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (i.e., the concentration of the test
20 compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

Another example of determination of effective dose for an individual is the ability
25 to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays may utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. The compound which is able to modulate 16816 or 16839 activity is used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents.
30 The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell, R. J. *et al.*, (1996) *Current Opinion in Biotechnology* 7:89-94 and in Shea, K.J., (1994) *Trends in Polymer Science* 2:166-173. Such "imprinted" affinity matrixes

are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, G. *et al.*, (1993) *Nature* 361:645-647.

Through the use of isotope-labeling, the “free” concentration of compound which
5 modulates the expression or activity of 16816 or 16839 can be readily monitored and used in calculations of IC₅₀.

Such “imprinted” affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using
10 appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC₅₀. A rudimentary example of such a “biosensor” is discussed in Kriz, D. *et al.*, (1995) *Analytical Chemistry* 67:2142-2144.

Another aspect of the invention pertains to methods of modulating 16816 or 16839 expression or activity for therapeutic purposes. Accordingly, in an exemplary
15 embodiment, the modulatory method of the invention involves contacting a cell with a 16816 or 16839 or agent that modulates one or more of the activities of 16816 or 16839 protein activity associated with the cell. An agent that modulates 16816 or 16839 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a 16816 or 16839 protein (e.g., a 16816 or 16839
20 substrate or receptor), a 16816 or 16839 antibody, a 16816 or 16839 agonist or antagonist, a peptidomimetic of a 16816 or 16839 agonist or antagonist, or other small molecule.

In one embodiment, the agent stimulates one or 16816 or 16839 activities. Examples of such stimulatory agents include active 16816 or 16839 protein and a nucleic
25 acid molecule encoding 16816 or 16839. In another embodiment, the agent inhibits one or more 16816 or 16839 activities. Examples of such inhibitory agents include antisense 16816 or 16839 nucleic acid molecules, anti-16816 or 16839 antibodies, and 16816 or 16839 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a
30 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 16816 or 16839 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or

downregulates) 16816 or 16839 expression or activity. In another embodiment, the method involves administering a 16816 or 16839 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted 16816 or 16839 expression or activity.

- 5 Stimulation of 16816 or 16839 activity is desirable in situations in which 16816 or 16839 is abnormally downregulated and/or in which increased 16816 or 16839 activity is likely to have a beneficial effect. For example, stimulation of 16816 or 16839 activity is desirable in situations in which a 16816 or 16839 is downregulated and/or in which increased 16816 or 16839 activity is likely to have a beneficial effect. Likewise,
- 10 inhibition of 16816 or 16839 activity is desirable in situations in which 16816 or 16839 is abnormally upregulated and/or in which decreased 16816 or 16839 activity is likely to have a beneficial effect.

- The 16816 or 16839 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more of cellular proliferative and/or differentiative disorders, cardiovascular disorders, as described above, as well as disorders associated with bone
- 15 metabolism, hematopoietic disorders, liver disorders, viral diseases, pain or metabolic disorders.

- Aberrant expression and/or activity of 16816 or 16839 molecules may mediate disorders associated with bone metabolism. "Bone metabolism" refers to direct or
- 20 indirect effects in the formation or degeneration of bone structures, e.g., bone formation, bone resorption, etc., which may ultimately affect the concentrations in serum of calcium and phosphate. This term also includes activities mediated by 16816 or 16839 molecules effects in bone cells, e.g. osteoclasts and osteoblasts, that may in turn result in bone formation and degeneration. For example, 16816 or 16839 molecules may support
- 25 different activities of bone resorbing osteoclasts such as the stimulation of differentiation of monocytes and mononuclear phagocytes into osteoclasts. Accordingly, 16816 or 16839 molecules that modulate the production of bone cells can influence bone formation and degeneration, and thus may be used to treat bone disorders. Examples of such disorders include, but are not limited to, osteoporosis, osteodystrophy, osteomalacia,
- 30 rickets, osteitis fibrosa cystica, renal osteodystrophy, osteosclerosis, anti-convulsant treatment, osteopenia, fibrogenesis-imperfecta ossium, secondary hyperparathyroidism, hypoparathyroidism, hyperparathyroidism, cirrhosis, obstructive jaundice, drug induced metabolism, medullary carcinoma, chronic renal disease, rickets, sarcoidosis,

glucocorticoid antagonism, malabsorption syndrome, steatorrhea, tropical sprue, idiopathic hypercalcemia and milk fever.

Examples of hematopoietic disorders include, but are not limited to, autoimmune diseases (including, for example, diabetes mellitus, arthritis (including rheumatoid arthritis, juvenile rheumatoid arthritis, osteoarthritis, psoriatic arthritis), multiple sclerosis, encephalomyelitis, myasthenia gravis, systemic lupus erythematosus, autoimmune thyroiditis, dermatitis (including atopic dermatitis and eczematous dermatitis), psoriasis, Sjögren's Syndrome, Crohn's disease, aphthous ulcer, iritis, conjunctivitis, keratoconjunctivitis, ulcerative colitis, asthma, allergic asthma, cutaneous lupus erythematosus, scleroderma, vaginitis, proctitis, drug eruptions, leprosy reversal reactions, erythema nodosum leprosum, autoimmune uveitis, allergic encephalomyelitis, acute necrotizing hemorrhagic encephalopathy, idiopathic bilateral progressive sensorineural hearing loss, aplastic anemia, pure red cell anemia, idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis, chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue, lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis, uveitis posterior, and interstitial lung fibrosis), graft-versus-host disease, cases of transplantation, and allergy such as, atopic allergy.

Disorders which may be treated or diagnosed by methods described herein include, but are not limited to, disorders associated with an accumulation in the liver of fibrous tissue, such as that resulting from an imbalance between production and degradation of the extracellular matrix accompanied by the collapse and condensation of preexisting fibers. The methods described herein can be used to diagnose or treat hepatocellular necrosis or injury induced by a wide variety of agents including processes which disturb homeostasis, such as an inflammatory process, tissue damage resulting from toxic injury or altered hepatic blood flow, and infections (e.g., bacterial, viral and parasitic). For example, the methods can be used for the early detection of hepatic injury, such as portal hypertension or hepatic fibrosis. In addition, the methods can be employed to detect liver fibrosis attributed to inborn errors of metabolism, for example, fibrosis resulting from a storage disorder such as Gaucher's disease (lipid abnormalities) or a glycogen storage disease, A1-antitrypsin deficiency; a disorder mediating the accumulation (e.g., storage) of an exogenous substance, for example, hemochromatosis (iron-overload syndrome) and copper storage diseases (Wilson's disease), disorders resulting in the accumulation of a toxic metabolite (e.g., tyrosinemia, fructosemia and galactosemia) and peroxisomal disorders (e.g., Zellweger syndrome). Additionally, the

methods described herein may be useful for the early detection and treatment of liver injury associated with the administration of various chemicals or drugs, such as for example, methotrexate, isoniazid, oxyphenisatin, methyl dopa, chlorpromazine, tolbutamide or alcohol, or which represents a hepatic manifestation of a vascular disorder such as obstruction of either the intrahepatic or extrahepatic bile flow or an alteration in hepatic circulation resulting, for example, from chronic heart failure, veno-occlusive disease, portal vein thrombosis or Budd-Chiari syndrome.

Additionally, 16816 or 16839 molecules may play an important role in the etiology of certain viral diseases, including but not limited to, Hepatitis B, Hepatitis C and Herpes Simplex Virus (HSV). Modulators of 16816 or 16839 activity could be used to control viral diseases. The modulators can be used in the treatment and/or diagnosis of viral infected tissue or virus-associated tissue fibrosis, especially liver and liver fibrosis. Also, 16816 or 16839 modulators can be used in the treatment and/or diagnosis of virus-associated carcinoma, especially hepatocellular cancer.

Additionally, 16816 or 16839 may play an important role in the regulation of metabolism or pain disorders. Diseases of metabolic imbalance include, but are not limited to, obesity, anorexia nervosa, cachexia, lipid disorders, and diabetes. Examples of pain disorders include, but are not limited to, pain response elicited during various forms of tissue injury, e.g., inflammation, infection, and ischemia, usually referred to as hyperalgesia (described in, for example, Fields, H.L., (1987) *Pain*, New York:McGraw-Hill); pain associated with musculoskeletal disorders, e.g., joint pain; tooth pain; headaches; pain associated with surgery; pain related to irritable bowel syndrome; or chest pain.

Pharmacogenomics

The 16816 or 16839 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 16816 or 16839 activity (e.g., 16816 or 16839 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) 16816 or 16839 associated disorders (e.g., cellular growth related disorders) associated with aberrant or unwanted 16816 or 16839 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's 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 whether to administer a 16816 or 16839 molecule or 16816 or 16839 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 16816 or 16839 molecule or 16816 or 16839 modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected 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 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 dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, 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
5 encodes a drug’s target is known (e.g., a 16816 or 16839 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.

Alternatively, a method termed the “gene expression profiling”, can be utilized to
10 identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a 16816 or 16839 molecule or 16816 or 16839 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
15 approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of 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 16816 or 16839 molecule or 16816 or 16839 modulator, such as a modulator identified by one of
20 the exemplary screening assays described herein.

The present invention further provides methods for identifying new agents, or combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the 16816 or 16839 genes of the present invention, wherein these products may be associated with resistance of the cells to
25 a therapeutic agent. Specifically, the activity of the proteins encoded by the 16816 or 16839 genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By blocking the activity of one or more of the resistance proteins, target cells, e.g., cancer cells, will become sensitive to treatment with an agent that the unmodified target cells were resistant to.

30 Monitoring the influence of agents (e.g., drugs) on the expression or activity of a 16816 or 16839 protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 16816 or 16839 gene expression, protein levels, or upregulate 16816 or 16839 activity, can be monitored in clinical trials of subjects exhibiting decreased 16816 or 16839 gene expression, protein

levels, or downregulated 16816 or 16839 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 16816 or 16839 gene expression, protein levels, or downregulate 16816 or 16839 activity, can be monitored in clinical trials of subjects exhibiting increased 16816 or 16839 gene expression, protein levels, or upregulated 16816 or 16839 activity. In such clinical trials, the expression or activity of a 16816 or 16839 gene, and preferably, other genes that have been implicated in, for example, a 16816 or 16839-associated disorder can be used as a “read out” or markers of the phenotype of a particular cell.

10 Other Embodiments

In another aspect, the invention features, a method of analyzing a plurality of capture probes. The method can be used, e.g., to analyze gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., a nucleic acid or peptide sequence; contacting the array with a 16816 or 16839, preferably purified, nucleic acid, preferably purified, polypeptide, preferably purified, or antibody, and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the 16816 or 16839 nucleic acid, polypeptide, or antibody.

The capture probes can be a set of nucleic acids from a selected sample, e.g., a sample of nucleic acids derived from a control or non-stimulated tissue or cell.

The method can include contacting the 16816 or 16839 nucleic acid, polypeptide, or antibody with a first array having a plurality of capture probes and a second array having a different plurality of capture probes. The results of each hybridization can be compared, e.g., to analyze differences in expression between a first and second sample. The first plurality of capture probes can be from a control sample, e.g., a wild type, normal, or non-diseased, non-stimulated, sample, e.g., a biological fluid, tissue, or cell sample. The second plurality of capture probes can be from an experimental sample, e.g., a mutant type, at risk, disease-state or disorder-state, or stimulated, sample, e.g., a biological fluid, tissue, or cell sample.

The plurality of capture probes can be a plurality of nucleic acid probes each of which specifically hybridizes, with an allele of 16816 or 16839. Such methods can be

used to diagnose a subject, e.g., to evaluate risk for a disease or disorder, to evaluate suitability of a selected treatment for a subject, to evaluate whether a subject has a disease or disorder. 16816 or 16839 is associated with phospholipase C activity, thus it is useful for disorders associated with abnormal lipid metabolism.

5 The method can be used to detect SNPs, as described above.

 In another aspect, the invention features, a method of analyzing a plurality of probes. The method is useful, e.g., for analyzing gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality
10 having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which express or mis express 16816 or 16839 or from a cell or subject in which a 16816 or 16839 mediated response has been elicited, e.g., by contact of the cell with 16816 or 16839 nucleic acid or protein, or administration to the cell or subject 16816 or 16839 nucleic acid or protein; contacting the array with one or more inquiry probe, wherein an
15 inquiry probe can be a nucleic acid, polypeptide, or antibody (which is preferably other than 16816 or 16839 nucleic acid, polypeptide, or antibody); providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject
20 which does not express 16816 or 16839 (or does not express as highly as in the case of the 16816 or 16839 positive plurality of capture probes) or from a cell or subject which in which a 16816 or 16839 mediated response has not been elicited (or has been elicited to a lesser extent than in the first sample); contacting the array with one or more inquiry probes (which is preferably other than a 16816 or 16839 nucleic acid, polypeptide, or
25 antibody), and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody.

 In another aspect, the invention features, a method of analyzing 16816 or 16839,
30 e.g., analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 16816 or 16839 nucleic acid or amino acid sequence; comparing the 16816 or 16839 sequence with one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database; to thereby analyze 16816 or 16839.

Preferred databases include GenBank™. The method can include evaluating the sequence identity between a 16816 or 16839 sequence and a database sequence. The method can be performed by accessing the database at a second site, e.g., over the internet.

5 In another aspect, the invention features, a set of oligonucleotides, useful, e.g., for identifying SNP's, or identifying specific alleles of 16816 or 16839. The set includes a plurality of oligonucleotides, each of which has a different nucleotide at an interrogation position, e.g., an SNP or the site of a mutation. In a preferred embodiment, the oligonucleotides of the plurality identical in sequence with one another (except for
10 differences in length). The oligonucleotides can be provided with different labels, such that an oligonucleotides which hybridizes to one allele provides a signal that is distinguishable from an oligonucleotides which hybridizes to a second allele.

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
15 applications cited throughout this application are incorporated herein by reference.

EXAMPLES

Example 1: Identification and Characterization of Human 16816 or 16839 cDNAs

20 The human 16816 or 16839 sequence (Figure 1A-C; SEQ ID NO:1 or Figure 3A-B; SEQ ID NO:4), which is approximately 2629 or 2171 nucleotides long, respectively, including untranslated regions, contains a predicted methionine-initiated coding sequence of about 2289 or 1827 nucleotides, respectively (nucleotides 257-2545 of SEQ ID NO:1; SEQ ID NO:3; or nucleotides 232-2058 of SEQ ID NO:4; SEQ ID NO:6). The coding
25 sequence encodes a 762 or 608 amino acid protein (SEQ ID NO:2 or SEQ ID NO:5), respectively.

Example 2: Tissue Distribution of 16816 or 16839 mRNA

Northern blot hybridizations with various RNA samples can be performed under
30 standard conditions and washed under stringent conditions, i.e., 0.2xSSC at 65°C. A DNA probe corresponding to all or a portion of the 16816 or 16839 cDNA (SEQ ID NO:1) or 16816 or 16839 cDNA (SEQ ID NO:4) can be used. The DNA was radioactively labeled with ³²P-dCTP using the Prime-It Kit (Stratagene, La Jolla, CA)

according to the instructions of the supplier. Filters containing mRNA from mouse hematopoietic and endocrine tissues, and cancer cell lines (Clontech, Palo Alto, CA) can be probed in ExpressHyb hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

5

Example 3: Gene Expression Analysis

Total RNA was prepared from various human tissues by a single step extraction method using RNA STAT-60 according to the manufacturer's instructions (TelTest, Inc). Each RNA preparation was treated with DNase I (Ambion) at 37°C for 1 hour. DNase I
10 treatment was determined to be complete if the sample required at least 38 PCR amplification cycles to reach a threshold level of fluorescence using β -2 microglobulin as an internal amplicon reference. The integrity of the RNA samples following DNase I treatment was confirmed by agarose gel electrophoresis and ethidium bromide staining. After phenol extraction cDNA was prepared from the sample using the SUPERSCRIPT™
15 Choice System following the manufacturer's instructions (GibcoBRL). A negative control of RNA without reverse transcriptase was mock reverse transcribed for each RNA sample.

Human 16816 or 16839 expression was measured by TaqMan® quantitative PCR (Perkin Elmer Applied Biosystems) in cDNA prepared from a variety of normal and
20 diseased (e.g., cancerous) human tissues or cell lines. Probes were designed by PrimerExpress software (PE Biosystems) based on the sequence of the human 16816 or 16839 gene. Each human 16816 or 16839 gene probe was labeled using FAM (6-carboxyfluorescein), and the β 2-microglobulin reference probe was labeled with a different fluorescent dye, VIC. The differential labeling of the target gene and internal
25 reference gene thus enabled measurement in same well. Forward and reverse primers and the probes for both β 2-microglobulin and target gene were added to the TaqMan® Universal PCR Master Mix (PE Applied Biosystems). Although the final concentration of primer and probe could vary, each was internally consistent within a given experiment. A typical experiment contained 200nM of forward and reverse primers plus 100nM probe
30 for β -2 microglobulin and 600 nM forward and reverse primers plus 200 nM probe for the target gene. TaqMan matrix experiments were carried out on an ABI PRISM 7700 Sequence Detection System (PE Applied Biosystems). The thermal cycler conditions were as follows: hold for 2 min at 50°C and 10 min at 95°C, followed by two-step PCR for 40 cycles of 95°C for 15 sec followed by 60°C for 1 min.

The following method was used to quantitatively calculate human 16816 or 16839 gene expression in the various tissues relative to β -2 microglobulin expression in the same tissue. The threshold cycle (Ct) value is defined as the cycle at which a statistically significant increase in fluorescence is detected. A lower Ct value is indicative of a higher mRNA concentration. The Ct value of the human 16816 or 16839 gene is normalized by subtracting the Ct value of the β -2 microglobulin gene to obtain a Δ Ct value using the following formula: Δ Ct = Ct_{human 59914 and 59921} - Ct _{β -2 microglobulin}. Expression is then calibrated against a cDNA sample showing a comparatively low level of expression of the human 16816 or 16839 gene. The Δ Ct value for the calibrator sample is then subtracted from Δ Ct for each tissue sample according to the following formula: $\Delta\Delta$ Ct = Δ Ct_{sample} - Δ Ct_{calibrator}. Relative expression is then calculated using the arithmetic formula given by $2^{-\Delta\Delta$ Ct}. Expression of the target human 16816 or 16839 gene in each of the tissues tested is then graphically represented as discussed in more detail below.

TaqMan real-time quantitative RT-PCR is used to detect the presence of RNA transcript corresponding to human 16816 relative to a no template control in a panel of human tissues or cells. It is found that the highest expression of 16816 orthologs are expressed in skeletal muscle tissue as shown in Table 1.

Table 1

	<u>Tissue Type</u>	<u>mean 16816</u>	<u>mean Beta 2</u>	<u>ΔCt</u>	<u>expression</u>
	Aorta / normal	38.87	23.105	15.765	0
5	Fetal heart/ normal	33.43	20.82	12.61	0.159960132
	Heart normal	34.62	18.985	15.635	0.019651514
	Heart/ CHF	35.64	20.47	15.17	0
	Vein/ Normal	33.81	19.15	14.66	0.038627826
	Spinal cord/ Normal	33.175	19.43	13.745	0.072835434
10	Brain cortex/ Normal	30.42	21.2	9.22	1.676885618
	Brain hypothalamus/ Normal	30.415	20.155	10.26	0.815515546
	Glial cells (Astrocytes)	33.495	21.58	11.915	0.258956968
	Brain/ Glioblastoma	29.77	18.41	11.36	0.380451455
	Breast / Normal	34.145	19.195	14.95	0.031593778
15	Breast tumor/ IDC	33.215	18.82	14.395	0.046416588
	OVARY/ Normal	35.47	21.23	14.24	0
	OVARY/ Tumor	37.815	19.785	18.03	0
	Pancreas	35.73	18.125	17.605	0
	Prostate/ Normal	30.895	19.255	11.64	0.313336401
20	Prostate/ Tumor	31.695	18.255	13.44	0.089982252
	Colon/ normal	31.035	17.45	13.585	0.081378093
	Colon/tumor	36.88	18.84	18.04	0
	Colon/IBD	35.185	18.445	16.74	0
	Kidney/ normal	30.255	21.005	9.25	1.642375811
25	Liver/ normal	36.48	19.18	17.3	0
	Liver fibrosis	35.16	20.015	15.145	0
	Fetal Liver/normal	37.02	22.14	14.88	0
	Lung / normal	33.93	18.02	15.91	0.016241
	Lung/ tumor	34.845	18.955	15.89	0.016467716
30	Lung/ COPD	33.345	18.13	15.215	0.026292302
	Spleen/ normal	39.23	20.19	19.04	0
	Tonsil/ normal	35.72	17.955	17.765	0
	Lymph node/ normal	34.55	18.485	16.065	0.014586568
	Thymus/ normal	33.475	19.675	13.8	0.070110984
35	Epithelial Cells (prostate)	31.825	20.57	11.255	0.409173406
	Endothelial Cells (aortic)	32.815	20.775	12.04	0.237464587
	Skeletal Muscle/ Normal	24.19	19.34	4.85	34.674046
	Fibroblasts (Dermal)	31.89	18.915	12.975	0.124204064
	Skin/ normal	34.805	20.945	13.86	0.067254951

	Adipose/ Normal	32.94	18.88	14.06	0.058548835
	Osteoblasts (primary)	32.595	20.175	12.42	0.182476715
	Osteoblasts (Undiff)	31.05	19.025	12.025	0.239946435
	Osteoblasts(Diff)	30.815	18.315	12.5	0.172633492
5	Osteoclasts	34.17	17.725	16.445	0.011208867
	Aortic SMC Early	34.705	20.23	14.475	0.043912768
	Aortic SMC Late	34.18	22.97	11.21	0.42213732
	shear HUVEC	31.315	20.955	10.36	0.76090291
	static HUVEC	31.465	20.875	10.59	0.64877237

10

TaqMan real-time quantitative RT-PCR is used to detect the presence of RNA transcript corresponding to human 16839 relative to a no template control in a panel of human tissues or cells. It is found that 16839 orthologs are expressed in teste as shown in the Table 2 and DRG as shown in Table 3.

15 Table 2

	<u>Tissue Type</u>	<u>16839.20</u>	<u>β2.803</u>	<u>ΔCt</u>	<u>Expression</u>
	Adrenal Gland	40.00	19.21	20.80	0.00
	Brain	40.00	21.01	18.99	0.00
	Heart	40.00	18.97	21.04	0.00
20	Kidney	39.90	18.98	20.92	0.00
	Liver	40.00	18.61	21.40	0.00
	Lung	40.00	17.12	22.88	0.00
	Mammary Gland	40.00	17.94	22.07	0.00
	Pancreas	40.00	20.50	19.50	0.00
25	Placenta	40.00	17.83	22.18	0.00
	Prostate	40.00	17.20	22.80	0.00
	Salivary Gland	40.00	18.48	21.52	0.00
	Muscle	40.00	20.56	19.45	0.00
	Sm. Intestine	40.00	18.82	21.19	0.00
30	Spleen	40.00	16.64	23.36	0.00
	Stomach	40.00	18.52	21.48	0.00
	Teste	26.97	20.05	6.92	1.69

	Thymus	40.00	17.66	22.35	0.00
	Trachea	40.00	18.93	21.07	0.00
	Uterus	40.00	18.80	21.20	0.00
	Spinal Cord	40.00	18.94	21.07	0.00
5	DRG	40.00	19.50	20.51	0.00
	Skin	39.10	18.44	20.66	0.00

Table 3

	<u>Tissue Type</u>	<u>Mean</u>	<u>β 2 Mean</u>	<u>ΔΔ Ct</u>	<u>Expression</u>
10	Artery normal	40	19.38	20.62	0
	Vein normal	39.38	18.47	20.91	0
	Aortic SMC EARLY	38.88	19.23	19.64	0
	Coronary SMC	39.92	20.55	19.38	0
	Static HUVEC	39.06	18.66	20.4	0
15	Heart normal	40	17.21	22.79	0
	Heart CHF	39.26	17.18	22.07	0
	Kidney	34.14	17.21	16.93	0.008
	Skeletal Muscle	40	20.26	19.74	0
20	Adipose normal	40	18.7	21.31	0
	Pancreas	39.41	19.23	20.18	0
	primary osteoblasts	37.65	17.54	20.11	0
	Osteoclasts (diff)	35.13	15.82	19.31	0
	Skin normal	39.13	19.16	19.98	0
25	Spinal cord normal	40	18.59	21.41	0
	Brain Cortex normal	38.95	18.59	20.36	0
	Brain Hypothalamus normal	40	19.43	20.57	0
	Nerve	40	22.34	17.66	0
	DRG (Dorsal Root Ganglion)	34.58	19.82	14.76	0.0362
30	Glial Cells (Astrocytes)	38.84	20.34	18.5	0
	Glioblastoma	38.91	16.52	22.39	0
	Breast normal	39.6	18.93	20.67	0
	Breast tumor	39.98	17.11	22.88	0
35	Ovary normal	38.79	18.03	20.76	0
	Ovary Tumor	40	18.23	21.77	0
	Prostate Normal	38.15	18.3	19.85	0

	Prostate Tumor	38.36	16.03	22.33	
	0				
	Epithelial Cells (Prostate)	39.45	19.29	20.16	0
	Colon normal	39.44	16.6	22.84	0
5	Colon Tumor	38.73	16.22	22.51	0
	Lung normal	38.99	16.12	22.87	0
	Lung tumor	37.2	16.58	20.63	0
	Lung COPD	38.73	16.72	22	0
	Colon IBD	38.3	15.77	22.54	0
10	Liver normal	39.88	17.98	21.91	0
	Liver fibrosis	37.62	19.28	18.34	0
	Dermal Cells- fibroblasts	37.84	17.66	20.18	0
	Spleen normal	37.22	18.06	19.16	0
	Tonsil normal	37.55	15.44	22.11	0
15	Lymph node	39.7	17.23	22.47	0
	Resting PBMC	40	18.8	21.2	
	0				
	Skin-Decubitus	36.45	18.82	17.64	
	0				
20	Synovium	37.98	17.66	20.32	0
	BM-MNC				
	(Bone marrow mononuclear cells)	38.51	14.83	23.68	0
	Activated PBMC	37.06	13.93	23.13	0
	Shear HUVEC	38	17.99	20.01	0.0009

25

Expression of 16839 was also detected in a panel of tissues and liver cell lines as shown in Table 4.

Table 4

30	<u>Tissue Type</u>	<u>Mean</u>	<u>18S Mean</u>	<u>ΔΔ Ct</u>	<u>Expression</u>
	PIT 278/ Heart	33.06	13.77	19.29	0.0016
	PIT 351/ Kidney	30.36	12.56	17.81	0.0044
	PIT 915/ Skeletal Muscle	38.28	11.87	26.41	0
	NDR 63/ Liver	33.23	11.69	21.54	0.0003
35	NDR 242/ Liver	34.26	12.63	21.63	0.0003
	PIT 260/ Liver	33.58	11.77	21.81	0.0003
	CHT 756/ Liver	31.77	11.65	20.13	0.0009
	MPI 155/ Liver	34.91	14.68	20.23	0.0008
	MPI 146/ Liver	30.91	11.95	18.97	0.002

	CHT 902/ Liver	35.72	12.49	23.22	0
	PIT 45/ Liver	33.97	12.14	21.84	0.0003
	PIT 292/ Liver	35.46	13.65	21.82	0
	CLN 784/ Liver	32.56	11.76	20.8	0.0005
5	NDR 752/ Liver	31.09	11.04	20.06	0.0009
	CHT 1679/ Liver	35.38	11.7	23.68	0
	CHT 1420/ Liver	33.06	11.15	21.91	0.0003
	CHT 339/ Liver	35.68	11.23	24.45	0
	CHT 1237/ Liver	33.41	11.02	22.4	0.0002

10

Expression of 16839 was also detected in a panel of breast tumor cell lines as shown in Table 5.

Table 5

15

	<u>Tissue Type</u>	<u>Mean 16839.4</u>	<u>β 2 Mean</u>	<u>ΔΔ Ct</u>	<u>Expression</u>
	MCF10MS	35.1	18.84	16.25	0
	MCF10A	33.88	18.81	15.07	0.03
	MCF10AT.cl1	34.2	18.88	15.32	0.02
20	MCF10AT.cl3	35.19	18.22	16.98	0
	MCF10AT1	36.74	19.19	17.56	0
	MCF10AT3B	36.23	18.93	17.3	0
	MCF10CA1a.cl1	36.94	16.34	20.59	0
	MCF10AT3B Agar	37.76	25.07	12.68	0
25	MCF10CA1a.cl1 Agar	37.93	23.13	14.8	0
	MCF10A.m25 Plastic	32.39	23.05	9.35	1.54
	MCF10CA Agar	33.63	20.71	12.93	0.13
	MCF10CA Plastic	35.98	20.56	15.43	0
	MCF3B Agar	37.34	21	16.34	0
30	MCF3B Plastic	32.67	20.54	12.14	0.22
	MCF10A EGF 0 hr	28.23	16.57	11.66	0.31
	MCF10A EGF 0.5 hr	28.84	16.61	12.23	0.21
	MCF10A EGF 1 hr	28.68	16.79	11.9	0.26
	MCF10A EGF 2 hr	29.54	16.88	12.66	0.15
35	MCF10A EGF 4 hr	31.21	16.97	14.24	0.05
	MCF10A EGF 8 hr	31.16	16.65	14.51	0.04
	MCF10A IGF1A 0 hr	30.54	20.41	10.13	0.90
	MCF10A IGF1A 0.5 hr	30.89	21.36	9.52	1.36
	MCF10A IGF1A 1 hr	29.7	20.84	8.86	2.15

	MCF10A IGF1A 3 hr	30.14	21.09	9.05	1.89
	MCF10A IGF1A 24 hr	28.53	20.43	8.1	3.66
	MCF10AT3B.cl5 Plastic	36.57	20.54	16.03	0
	MCF10AT3B.cl6 Plastic	38.18	20.7	17.48	0
5	MCF10AT3B.cl3 Plastic	39.13	20.8	18.34	0
	MCF10AT3B.cl1 Plastic	35.92	21.07	14.85	0
	MCF10AT3B.cl4 Plastic	38.1	20.95	17.15	0
	MCF10AT3B.cl2 Plastic	36.49	20.75	15.74	0
	MCF10AT3B.cl5 Agar	39.22	22.68	16.54	0
10	MCF10AT3B.cl6 Agar	35.7	23.13	12.57	0
	MCF-7	37.46	22.25	15.22	0
	ZR--75	37.54	21.25	16.3	0
	T47D	37.19	20.72	16.47	0
	MDA-231	33.55	19.7	13.85	0.07
15	MDA-435	34.73	19.39	15.35	0.02
	SkBr3	37.16	19.86	17.3	0
	Hs578Bst	36.52	18.93	17.59	0
	Hs578T	31.36	19	12.35	0.19

20 Expression of 16839 was also detected in an oncology phase panel as shown in Table 6 and shows highest relative expression in lung tumor (CHT 832 lung T-PDNCSCCL) and upregulation of 16839 was found in 3/7 lung tumor tissue or cell samples.

25 Table 6

	<u>Tissue Type</u>	<u>Mean 16839.4</u>	<u>β 2 Mean</u>	<u>$\Delta\Delta$ Ct</u>	<u>Expression</u>
	PIT 400 Breast N	35.4	19.18	16.23	0
	PIT 372 Breast N	35.38	19.66	15.72	0
	CHT 558 Breast N	36.38	19.02	17.36	0
30	CLN 168 Breast T: IDC	38.45	19.84	18.62	0
	MDA 304 Breast T: MD-IDC	37.04	17.84	19.2	0
	NDR 58 Breast T: IDC	38.83	17.84	20.99	0
	NDR 05 Breast T: IDC	35.22	20.57	14.65	0
	CHT 562 Breast T: IDC	34.08	18.71	15.37	0.02
35	NDR 12 Breast T	35.52	21.39	14.13	0
	PIT 208 Ovary N	37.06	18.56	18.5	0
	CHT 620 Ovary N	36.05	19.2	16.85	0
	CLN 03 Ovary T	40	19.83	20.17	0
	CLN 17 Ovary T	39.49	20	19.5	0

	MDA 25 Ovary T	38.42	21.65	16.78	0
	MDA 216 Ovary T	38.04	20.11	17.93	0
	CLN 012 Ovary T	39.27	21.3	17.98	0
	MDA 185 Lung N	40	19.43	20.57	0
5	CLN 930 Lung N	37.99	20.29	17.7	0
	MDA 183 Lung N	39.65	18.11	21.54	0
	MPI 215 Lung T--SmC	31.34	18.62	12.72	0.15
	MDA 259 Lung T-PDNSCCL	37.42	19.93	17.49	0
	CHT 832 Lung T-PDNSCCL	28.18	18.86	9.32	1.56
10	MDA 253 Lung T-PDNSCC	33.39	18.05	15.35	0.02
	MDA 262 Lung T-SCC	38.79	22.28	16.51	0
	CHT 211 Lung T-AC	39.1	19.29	19.82	0
	CHT 793 Lung T-ACA	35.74	18.55	17.19	0
	CHT 396 Colon N	39.82	17.7	22.13	0
15	CHT 523 Colon N	39.99	18.52	21.47	0
	CHT 452 Colon N	40	17.5	22.5	0
	CHT 382 Colon T: MD	38.42	17.97	20.45	0
	CHT 528 Colon T: MD	34.85	17.54	17.32	0.01
	CLN 609 Colon T	38.1	18.82	19.29	0
20	CHT 372 Colon T: MD-PD	37.83	19.24	18.59	0
	CHT 340 Colon-Liver Met	37.93	19.99	17.94	0
	NDR 100 Colon-Liver Met	36.52	18.26	18.27	0
	PIT 260 Liver N (female)	35.31	16.93	18.37	0
	ONC 102 Hemangioma	36.36	19.04	17.32	0
25	A24 HMVEC-Arr	35.31	18.75	16.55	0
	C48 HMVEC-Prol	35.12	19.18	15.94	0

Example 4: Recombinant Expression of 16816 or 16839 in Bacterial Cells

In this example, 16816 or 16839 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, 16816 or 16839 is fused to GST and this fusion polypeptide is expressed in *E. coli*, e.g., strain PEB199. Expression of the GST-16816 or 16839 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 electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

Example 5: Expression of Recombinant 16816 or 16839 Protein in COS Cells

To express the 16816 or 16839 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 16816 or 16839 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 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 16816 or 16839 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the 16816 or 16839 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 16816 or 16839 coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the 16816 or 16839 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, 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 16816 or 16839-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., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the 16816 or 16839 polypeptide is detected by radiolabelling (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 16816 or 16839 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 16816 or 16839 polypeptide is detected by radiolabelling and immunoprecipitation using a 16816 or 16839 specific monoclonal antibody.

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 invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed is:

1. An isolated 16816 or 16839 nucleic acid molecule selected from the group consisting of:
 - a) a nucleic acid molecule comprising a nucleotide sequence which is at least
5 60% identical to the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____;
 - b) a nucleic acid molecule comprising a fragment of at least 15 nucleotides of
10 the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____;
 - c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession
15 Number _____;
 - d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____, wherein the fragment comprises at least 15 contiguous amino
20 acids of SEQ ID NO:2, SEQ ID NO:5, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____;
 - e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the
25 ATCC as Accession Number _____, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, or a complement thereof, under stringent conditions;
 - f) a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, or the nucleotide sequence of the
30 DNA insert of the plasmid deposited with ATCC as Accession Number _____; and
 - g) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____.

2. The isolated nucleic acid molecule of claim 1, which is the nucleotide sequence SEQ ID NO:1 or SEQ ID NO:4.
3. A host cell which contains the nucleic acid molecule of claim 1.
4. An isolated 16816 or 16839 polypeptide selected from the group consisting of:
 - a) 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, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a complement thereof;
 - b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, or a complement thereof under stringent conditions;
 - c) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2 or SEQ ID NO:5; and
 - d) the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:5.
5. An antibody which selectively binds to a polypeptide of claim 4.
6. A method for producing a polypeptide selected from the group consisting of:
 - a) a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____;
 - b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____, wherein the fragment

comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:5, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____;

5 c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4 or SEQ ID NO:6; and

10 d) the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:5; comprising culturing the host cell of claim 3 under conditions in which the nucleic acid molecule is expressed.

7. A method for detecting the presence of a nucleic acid molecule of claim 1 or a polypeptide encoded by the nucleic acid molecule in a sample, comprising:

15 a) contacting the sample with a compound which selectively hybridizes to the nucleic acid molecule of claim 1 or binds to the polypeptide encoded by the nucleic acid molecule; and

b) determining whether the compound hybridizes to the nucleic acid or binds to the polypeptide in the sample.

20 8. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 or binds to a polypeptide encoded by the nucleic acid molecule and instructions for use.

9. A method for identifying a compound which binds to a polypeptide or modulates the activity of the polypeptide of claim 4 comprising the steps of:

25 a) contacting a polypeptide, or a cell expressing a polypeptide of claim 4 with a test compound; and

b) determining whether the polypeptide binds to the test compound or determining the effect of the test compound on the activity of the polypeptide.

30 10. A method for modulating the activity of a polypeptide of claim 4 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.

11. A method of identifying a nucleic acid molecule associated with a disorder comprising:

- 5 a) contacting a sample from a subject with or at risk of developing a disorder comprising nucleic acid molecules with a hybridization probe comprising at least 25 contiguous nucleotides of SEQ ID NO:1 or SEQ ID NO:4 defined in claim 2; and
- b) detecting the presence of a nucleic acid molecule in the sample that hybridizes to the probe, thereby identifying a nucleic acid molecule associated with a
- 10 disorder.

12. A method of identifying a nucleic acid associated with a disorder comprising:

- a) contacting a sample from a subject having a disorder or at risk of developing a disorder comprising nucleic acid molecules with a first and a second
- 15 amplification primer, the first primer comprising at least 25 contiguous nucleotides of SEQ ID NO:1 or SEQ ID NO:4 defined in claim 2 and the second primer comprising at least 25 contiguous nucleotides from the complement of SEQ ID NO:1 or SEQ ID NO:4;
- b) incubating the sample under conditions that allow nucleic acid amplification; and
- 20 c) detecting the presence of a nucleic acid molecule in the sample that is amplified, thereby identifying the nucleic acid molecule associated with a disorder.

13. A method of identifying a polypeptide associated with a disorder comprising:

- a) contacting a sample comprising polypeptides with a 16816 or 16839
- 25 binding partner of the 16816 or 16839 polypeptide defined in claim 4; and
- b) detecting the presence of a polypeptide in the sample that binds to the 16816 or 16839 binding partner, thereby identifying the polypeptide associated with a disorder.

14. A method of identifying a subject having a disorder or at risk for

30 developing a disorder comprising:

- a) contacting a sample obtained from the subject comprising nucleic acid molecules with a hybridization probe comprising at least 25 contiguous nucleotides of SEQ ID NO:1 or SEQ ID NO:4 defined in claim 2; and
- b) detecting the presence of a nucleic acid molecule in the sample that
5 hybridizes to the probe, thereby identifying a subject having a disorder or at risk for developing a disorder.

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

- a) contacting a sample obtained from the subject comprising nucleic acid
10 molecules with a first and a second amplification primer, the first primer comprising at least 25 contiguous nucleotides of SEQ ID NO:1 or SEQ ID NO:4 defined in claim 2 and the second primer comprising at least 25 contiguous nucleotides from the complement of SEQ ID NO:1 or SEQ ID NO:4;
- b) incubating the sample under conditions that allow nucleic acid
15 amplification; and
- c) detecting the presence of a nucleic acid molecule in the sample that is amplified, thereby identifying a subject having a disorder or at risk for developing a disorder.

20 16. A method of identifying a subject having a disorder or at risk for developing a disorder comprising:

- a) contacting a sample obtained from the subject comprising polypeptides with a 16816 or 16839 binding partner of the 16816 or 16839 polypeptide defined in claim 4; and
- b) detecting the presence of a polypeptide in the sample that binds to the
25 16816 or 16839 binding partner, thereby identifying a subject having a disorder or at risk for developing a disorder.

17. A method for identifying a compound capable of treating a disorder characterized by aberrant 16816 or 16839 nucleic acid expression or 16816 or 16839 polypeptide activity comprising assaying the ability of the compound to modulate 16816
30 or 16839 nucleic acid expression or 16816 or 16839 polypeptide activity, thereby

identifying a compound capable of treating a disorder characterized by aberrant 16816 or 16839 nucleic acid expression or 16816 or 16839 polypeptide activity.

18. A method for treating a subject having a disorder or at risk of developing a disorder comprising administering to the subject a 16816 or 16839 modulator of the nucleic acid molecule defined in claim 1 or the polypeptide encoded by the nucleic acid molecule or contacting a cell with a 16816 or 16839 modulator.

19. The method of claim 18, wherein the 16816 or 16839 modulator is
- a) a small molecule;
 - b) peptide;
 - 10 c) phosphopeptide;
 - d) anti-16816 or 16839 antibody;
 - e) a 16816 or 16839 polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, or a fragment thereof;
 - f) a 16816 or 16839 polypeptide comprising an amino acid sequence which
 - 15 is at least 90 percent identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:5, wherein the percent identity is calculated using the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; or
 - g) an isolated naturally occurring allelic variant of a polypeptide consisting of
 - 20 the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:5, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a complement of a nucleic acid molecule consisting of SEQ ID NO:1 or SEQ ID NO:4 at 6X SSC at 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C.

20. The method of claim 18, wherein the 16816 or 16839 modulator is
- 25 a) an antisense 16816 or 16839 nucleic acid molecule;
 - b) is a ribozyme;
 - c) the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:4, or a fragment thereof;
 - d) a nucleic acid molecule encoding a polypeptide comprising an amino acid
 - 30 sequence which is at least 90 percent identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:5, wherein the percent identity is calculated using the ALIGN

program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4;

- e) a nucleic acid molecule encoding a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:5,
5 wherein the nucleic acid molecule which hybridizes to a complement of a nucleic acid molecule consisting of SEQ ID NO:1 or SEQ ID NO:4 at 6X SSC at 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C; or
f) a gene therapy vector.

21. A method for evaluating the efficacy of a treatment of a disorder, in a
10 subject, comprising:

treating a subject with a protocol under evaluation;

assessing the expression level of a 16816 or 16839 nucleic acid molecule defined in claim 1 or 16816 or 16839 polypeptide encoded by the 16816 or 16839 nucleic acid molecule,

15 wherein a change in the expression level of 16816 or 16839 nucleic acid or 16816 or 16839 polypeptide after the treatment, relative to the level before the treatment, is indicative of the efficacy of the treatment of a disorder.

22. A method of diagnosing a disorder in a subject, comprising:
evaluating the expression or activity of a 16816 or 16839 nucleic acid molecule defined in
20 claim 1 or a 16816 or 16839 polypeptide encoded by the 16816 or 16839 nucleic acid molecule, such that a difference in the level of 16816 or 16839 nucleic acid or 16816 or 16839 polypeptide relative to a normal subject or a cohort of normal subjects is indicative of a disorder.

23. The method defined in claim 18, wherein the disorder is cancer or aberrant cellular proliferation and/or differentiation.

24. The method defined in claim 23, wherein the cancer or aberrant cellular proliferation and/or differentiation is lung cancer.

30

1/15

GCACGAGACACCTCTCCCCTTCTTACTGCTTCCCTCCGGCTATAACTTGCCAGTCACAGCAGCCAGCTGCTGTAGAAGA [SEQ ID NO:1]

GGGGAGGAAACAAGCCAGTGCAAGGGGAGCAAAGAGAAAAGGAGCCAGGCTGGGCTTCCTGATCCCACAGCATCGCAG

AGCTCGGGAGGCACAGCTCACAGACACAGGAAACACAGGACTGCTATTCTGCTCTCCTGCCACGGTGATCTGGTGCCA

M A S L L Q D Q L T T D Q D L [SEQ ID NO:2]
 GCTGGTGGAACAGTGGGTG ATG GCG TCC CTG CTG CAA GAC CAG CTG ACC ACT GAT CAG GAC TTG 45

↑SEQ ID NO:3→

L L M Q E G M P M R K V R S K S W K K L 35
 CTG CTG ATG CAG GAA GGC ATG CCG ATG CGC AAG GTG AGG TCC AAA AGC TGG AAG AAG CTA 105

R Y F R L Q N D G M T V W H A R Q A R G 55
 AGA TAC TTC AGA CTT CAG AAT GAC GGC ATG ACA GTC TGG CAT GCA CGG CAG GCC AGG GGC 165

S A K P S F S I S D V E T I R N G H D S 75
 AGT GCC AAG CCC AGC TTC TCA ATC TCT GAT GTG GAG ACA ATA CGT AAT GGC CAT GAT TCC 225

E L L R S L A E E L P L E Q G F T I V F 95
 GAG TTG CTG CGT AGC CTG GCA GAG GAG CTC CCC CTG GAG CAG GGC TTC ACC ATT GTC TTC 285

H G R R S N L D L M A N S V E E A Q I W 115
 CAT GGC CGC CGC TCC AAC CTG GAC CTG ATG GCC AAC AGT GTT GAG GAG GCC CAG ATA TGG 345

M R G L Q L L V D L V T S M D H Q E R L 135
 ATG CGA GGG CTC CAG CTG TTG GTG GAT CTT GTC ACC AGC ATG GAC CAT CAG GAG CGC CTG 405

D Q W L S D W F Q R G D K N Q D G K M S 155
 GAC CAA TGG CTG AGC GAT TGG TTT CAA CGT GGA GAC AAA AAT CAG GAT GGT AAG ATG AGT 465

F Q E V Q R L L H L M N V E M D Q E Y A 175
 TTC CAA GAA GTT CAG CGG TTA TTG CAC CTA ATG AAT GTG GAA ATG GAC CAA GAA TAT GCC 525

FIG. 1A

2/15

F S L F Q A A D T S Q S G T L E G E E F	195
TTC AGT CTT TTT CAG GCA GCA GAC ACG TCC CAG TCT GGA ACC CTG GAA GGA GAA GAA TTC	585
V Q F Y K A L T K R A E V Q E L F E S F	215
GTA CAG TTC TAT AAG GCA TTG ACT AAA CGT GCT GAG GTG CAG GAA CTG TTT GAA AGT TTT	645
S A D G Q K L T L L E F L D F L Q E E Q	235
TCA GCT GAT GGG CAG AAG CTG ACT CTG CTG GAA TTT TTG GAT TTC CTC CAA GAG GAG CAG	705
K E R D C T S E L A L E L I D R Y E R S	255
AAG GAG AGA GAC TGC ACC TCT GAG CTT GCT CTG GAA CTC ATT GAC CGC TAT GAA CCT TCA	765
D S G K L R H V L S M D G F L S Y L C S	275
GAC AGT GGC AAA CTG CGG CAT GTG CTG AGT ATG GAT GGC TTC CTC AGC TAC CTC TGC TCT	825
K D G D I F N P A C L P I Y Q D M T Q P	295
AAG GAT GGA GAC ATC TTC AAC CCA GCC TGC CTC CCC ATC TAT CAG GAT ATG ACT CAA CCC	885
L N H Y F I C S S H N T Y L V G D Q L C	315
CTG AAC CAC TAC TTC ATC TGC TCT TCT CAT AAC ACC TAC CTA GTG GGG GAC CAG CTT TGC	945
G Q S S V E G Y I R A L K R G C R C V E	335
GGC CAG AGC AGC GTC GAG GGA TAT ATA CGG GCC CTG AAG CGG GGG TGC CGC TGC GTG GAG	1005
V D V W D G P S G E P V V Y H G H T L T	355
GTG GAT GTA TGG GAT GGA CCT AGC GGG GAA CCT GTC GTT TAC CAC GGA CAC ACC CTG ACC	1065
S R I L F K D V V A T V A Q Y A F Q T S	375
TCC CGC ATC CTG TTC AAA GAT GTC GTG GCC ACA GTA GCA CAG TAT GCC TTC CAG ACA TCA	1125
D Y P V I L S L E T H C S W E Q Q Q T M	395
GAC TAC CCA GTC ATC TTG TCC CTG GAG ACC CAC TGC AGC TGG GAG CAG CAG CAG ACC ATG	1185

FIG. 1B

3/15

A R H L T E I L G E Q L L S T T L D G V	415
GCC CGT CAT CTG ACT GAG ATC CTG GGG GAG CAG CTG CTG AGC ACC ACC TTG GAT GGG GTG	1245
L P T Q L P S P E E L R R K I L V K G K	435
CTG CCC ACT CAG CTG CCC TCG CCT GAG GAG CTT CGG AGG AAG ATC CTG GTG AAG GGG AAG	1305
K L T L E E D L E Y E E E E A E P E L E	455
AAG TTA ACA CTT GAG GAA GAC CTG GAA TAT GAG GAA GAG GAA GCA GAA CCT GAG TTG GAA	1365
E S E L A L E S Q F E T E P E P Q E Q N	475
GAG TCA GAA TTG GCG CTG GAG TCC CAG TTT GAG ACT GAG CCT GAG CCC CAG GAG CAG AAC	1425
L Q N K D K K K K S K P I L C P A L S S	495
CTT CAG AAT AAG GAC AAA AAG AAG AAA TCC AAG CCC ATC TTG TGT CCA GCC CTC TCT TCC	1485
L V I Y L K S V S F R S F T H S K E H Y	515
CTG GTT ATC TAC TTG AAG TCT GTC TCA TTC CGC AGC TTC ACA CAT TCA AAG GAG CAC TAC	1545
H F Y E I S S F S E T K A K R L I K E A	535
CAC TTC TAC GAG ATA TCA TCT TTC TCT GAA ACC AAG GCC AAG CGC CTC ATC AAG GAG GCT	1605
G N E F V Q H N T W Q L S R V Y P S G L	555
GGC AAT GAG TTT GTG CAG CAC AAT ACT TGG CAG TTA AGC CGT GTG TAT CCC AGC GGC CTG	1665
R T D S S N Y N P Q E L W N A G C Q M V	575
AGG ACA GAC TCT TCC AAC TAC AAC CCC CAG GAA CTC TGG AAT GCA GGC TGC CAG ATG GTG	1725
A M N M Q T A G L E M D I C D G H F R Q	595
GCC ATG AAT ATG CAG ACT GCA GGG CTT GAA ATG GAC ATC TGT GAT GGG CAT TTC CGC CAG	1785
N G G C G Y V L K P D F L R D I Q S S F	615
AAT GGC GGC TGT GGC TAT GTG CTG AAG CCA GAC TTC CTG CGT GAT ATC CAG AGT TCT TTC	1845
H P E K P I S P F K A Q T L L I Q V I S	635
CAC CCT GAG AAG CCC ATC AGC CCT TTC AAA GCC CAG ACT CTC TTA ATC CAG GTG ATC AGC	1905

FIG. 1C

4/15

G Q Q L P K V D K T K E G S I V D P L V	655
GGT CAG CAA CTC CCC AAA GTG GAC AAG ACC AAA GAG GGG TCC ATT GTG GAT CCA CTG GTG	1965
K V Q I F G V R L D T A R Q E T N Y V E	675
AAA GTG CAG ATC TTT CGC GTT CGT CTA GAC ACA GCA CGG CAG GAG ACC AAC TAT GTG GAG	2025
N N G F N P Y W G Q T L C F R V L V P E	695
AAC AAT GGT TTT AAT CCA TAC TGG GGG CAG ACA CTA TGT TTC CGG GTG CTG GTG CCT GAA	2085
L A M L R F V V M D Y D W K S R N D F I	715
CTT GCC ATG CTG CGT TTT GTG GTA ATG GAT TAT GAC TGG AAA TCC CGA AAT GAC TTT ATT	2145
G Q Y T L P W T C M Q Q G Y R H I H L L	735
GGT CAG TAC ACC CTG CCT TGG ACC TGC ATG CAA CAA GGT TAC CGC CAC ATT CAC CTG CTG	2205
S K D G I S L R P A S I F V Y I C I Q E	755
TCC AAA GAT GGC ATC AGC CTC CGC CCA GCT TCC ATC TTT GTG TAT ATC TGC ATC CAG GAA	2265
G L E G D E S *	763
GGC CTG GAG GGG GAT GAG TCC TGA	2289
← SEQ ID NO:3↑	

GGTGGGCATTTACGGGAAGGGTTGGTGTGCTGGCTTTAGACGGGGAGAAACATCTGGAAGGATGCTCGAGGGGGGGCC

CGGGC

FIG. 1D

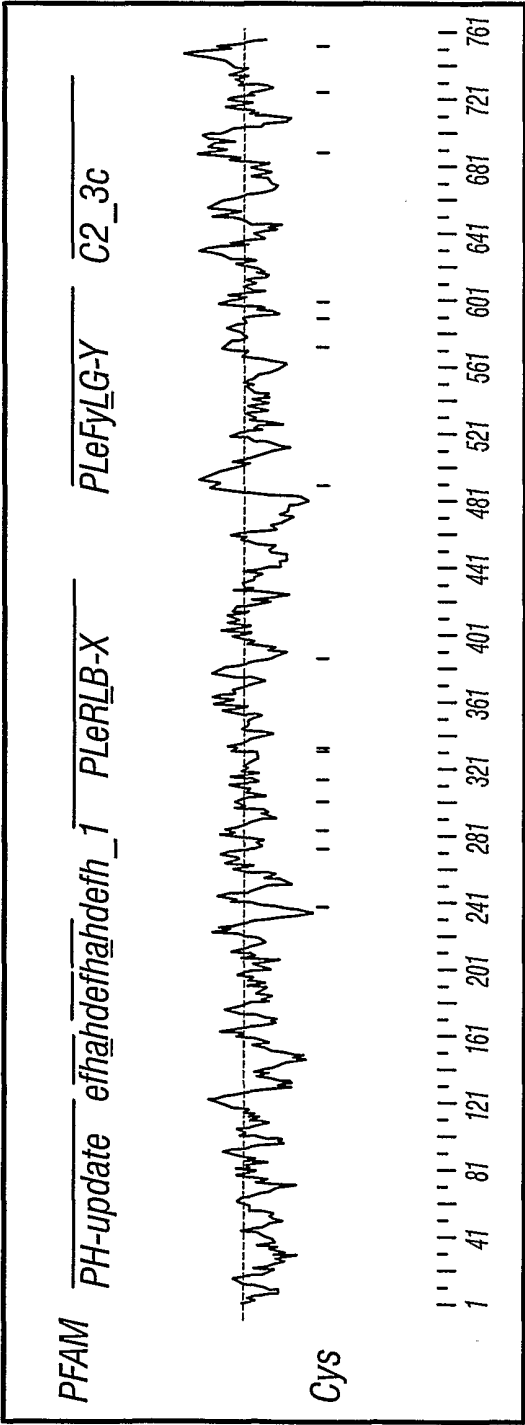


FIG. 2

6/15

```

      *->vikeGwLlkks...kswkkRyfvLfnnvLllykds..kkpkgsipL [SEQ ID NO:7]
      ++ eG ++k +++++ k Ryf L+n+ + ++ ++ + + k s+++
16816 17  LMQEGMPMRKVrskSWKKLRYFRLQNDGMTVWHARQarGSAKPSFSI 63

      sgc.qvek.pd.....kncFeirt.dr..tlllqaeseeerk
      s++++++ ++d++ ++ ++ + + +F i++++r++ l l+a+s ee++
16816 64 SDVeTIRNgHDSellrslaeelpEQGFTIVFhGRrsNLDLMANSVEEAQ 113

      eWvkaiqsair<-*
      W++ +q ++
16816' 114 IWMRGLQLLVD 124

```

FIG. 3

```

      *->elkeaFkefDkDgDGkIsfeEfkaalkkl<-* [SEQ ID NO:8]
      l ++F+ Dk++DGk sf+E +++l+ +
16816 138 WLSDFWQRGDKNQDGKMSFQEVQRLHLHM 166

```

FIG. 4A

```

      *->elkeaFkefDkDgDGkIsfeEfkaalkkl<-* [SEQ ID NO:9]
      + +F+ +D+ ++G+++ eEf+++ k+l
16816 174 YAFSLFQAADTSQSGTLEGEFVQFYKAL 202

```

FIG. 4B

7/15

```

*->dmsiPLsHYfisSshntYlTgkQlwGkssvesYrqgLdaGcRcvELD [SEQ ID NO:10]
    dm +PL+HYfi SshntYL g+Ql G+ssve+Y++ L++GcRcvE D
16816 291 DMTQPLNHYFICSSHNTYLVGDQLCGQSSVEGYIRALKRGCRCEVD 337

    cwdGkpddepiIyHGhtltleiklkdVleaIkdfafkPtSpyPvIlSlen
    +wdG p +ep++yHGhtlt+ i++kdV+ +++++af tS yPvIlSle
16816 338 VWDG-PSGEPVYHGHTLTSRILFKDVVATVAQYAFQ-TSDYPVILSLET 385

    HcnsddqQrkmakyfkeiFgdmLltkPtldslttepglpLPslkdlrgKI
    Hc+ + qQ +ma+++ ei+g++Ll t+ + +++ LPs+++lr+KI
16816 386 HCSWE-QQQTMARHLTEILGEQLLS--TTLDGVLP--TQLPSPEELRRKI 430

    LLknkk<-*
    L+k+kk
16816 431 LVKGKK 436

```

FIG. 5

```

*->ElsnLvnYiqsikFrsFelsGeekntsyeisSFsErkvkakkllkes [SEQ ID NO:11]
    ls+Lv+Y++s+ FrsF s +e +yeisSFsE+ kak+l ke+
16816 492 ALSSLVIYLKSVSFRSFTHS-KEHYHFYEISSFSET--KAKRLIKEA 535

    pvefVkyNkrqLsRvYPkGtRvDSSNfmPqvWnaGCQmVALNfQTsDlp
    ++efV++N qLsRvYP G+R+DSSN++Pq+ WnaGCQmVA N QT +l+
16816 536 GNEFVQHNTWQLSRVYPSGLRTDSSNYPQELWNAGCQMAMNMQTAGLE 585

    mqiNdGmFeyNgggPdGsfsGYLLKPeflR<-*
    m+i dG F Ngg +GY+LKP+flR
16816 586 MDICDGHFRQNGG-----CGYVLKPDFLR 609

```

FIG. 6

8/15

*->LtVtvieArnLpkmDk..vngrlsDPYVkvslldgdkkdlkkfkTkvv [SEQ ID NO:12]
 L + vi++++Lpk Dk++ ++ ++DP Vkv+++g++ d+ +++T v
 16816 629 LLIQVISGQQLPKVDKtkeGS-IVDPLVKVQIFGVRLDTARQETNYV 674
 kktNGLNPvWneEtFvFekvplpelasktLrfaVyDedrfssrdDfiGqvt
 ++ G+NP+W+ t++F+ v++pela Lrf+V+D+d +sr+DfiGq t
 16816 675 ENN-GFNPYWG-QTLCFR-VLPELAM--LRFVMDYDWKSRNDFIGQYT 719

FIG. 7

Query: 275 SKDGDIFNPACLPYQDMTQPLNHYFICSSHNTYLVGDQLC-GQSSVEGYIRALKRGCR 333
 S D ++F+P ++QDM QPL+HYFI SSHNTYL G+QL G+SSVE Y +AL +GCRC
 Sbjct: 2 SPDCNVFDPEHKQVHQDMNQPLSHYFINSSHNTYLTGNQLSSGESSVEMYRQALLKGCRC 61 [SEQ ID NO:13]
 Query: 334 VEVDVWDGPSG--EPVVYHGHTLTSRILFKDVVATVAQYAFQTSYPVILSLETHC-SWE 390
 +E+D WDG G EP++ HGHT+T+ I FKD + + ++AF TS+YPVILSLE HC S
 Sbjct: 62 IELDCWDGKGDPEPIITHGHTMTTEISFKDCLEAIKEHAFVTSEYPVILSLENHCDSTP 121
 Query: 391 QOQT-MARHLTEILGEQLLSTTLDG--VLP-TQLPSPEELRRKILVKGKK 436
 QOQ MA + E+ G+ L + L+ + P +LPSPEEL+RKIL+K KK
 Sbjct: 122 QQAKMAEYCKEVFGDMLFTEPLEESPLEPGKELPSPEELKRKILIKNKK 171

FIG. 8

9/15

Query: 1 MASXXXXXXXXXXXXMQEGMPMKVRSKSWKKLRYFRLQNDGMTVW---HARQARGSA 57
 MAS MQ+G ++KVR S WK RYF L +D T+W + S

Sbjct: 1 MASQIKKISSTNDCLQFMQKGSELKKVRSNSWKYNRYFTLDDMQTLWWEPHWFSKKDSE 60 [SEQ ID NO:14]

Query: 58 KPSFSISDVETIRNGHDSSELLRLAEELPLEQG---FTIVFHGR--RSNLDLMANSVEE 111
 KP F ISD++ IR G ++E R+ +E +++ F+I+F +LDL+ANS +

Sbjct: 61 KPKFDISDIKEIRMGKNTETFRNNGKEFQIQEPEDCCFSIIFGENYFHESLDLVANSADV 120

Query: 112 AQIWMRGLQLLDLVTSM--DHQERLDQWLSDFQRGDKNQDGKMSFQEVQRLHLMNVE 169
 A IW+ GL+ LVD M ++QE+LDQWL +WFQ+ D+N+D +MSF+E Q LL LMNV+

Sbjct: 121 ANIWWGLRYLVDYAKHMLDNYQEQLDQWLREWFQQADRNKDSRMSFREANLLKLMNVQ 180

Query: 170 MDQEYAFSLFQAADTSQSGTLE 191
 MD+EYAFS+F+ D SQS TL+

Sbjct: 181 MDEEYFASIFRECDFSQSNTLD 202

FIG. 9

Query: 491 PALSSSVIYLKSVSFRSFTHSKEHYHFYEISSFSETKAKRLI-KEAGNEFVQHNTWQLSR 549
 P LS LV Y + V F+ F ++E +Y +SSFSE KA+L+ KE EFV++N L R

Sbjct: 11 PELSDLVNYCQPVKFKGFEMAEKNTYYHMSSFSENKAEKLVNKEHPKEFVRYNQRLLR 70 [SEQ ID NO:15]

Query: 550 VYPSGLRTDSSNYPQELWNAGCOMVAMNMQTAGLEMDICDGHFRQNGCGYVLKPDFL 608
 VYP G R DSSNYP WN GCOMVA+N QT G M + G FR NGCGYVLKPDFL

Sbjct: 71 VYPKGTRIDSSNYPMVFWNHGCOMVALNFQTHGRSMWLNQGMFRANGCGYVLKPDFL 129

FIG. 10

Query: 722 WTCMQQGYRHIHLLSKDGLSLRPASIFVYICIQEGLEGDE 761
 WTCM+ GYRH+ LLSKDG SL PAS FVY C+QE L+ DE

Sbjct: 1 WTCMKHGYRHVSLLSKDGTSLHPASHFVYTCMQEDLDMDE 40 [SEQ ID NO:16]

FIG. 11

10/15

Query: 562 YNPQELWNAGCQMAMMOTAGLEMDICDGHFRQNGCGYVLKPDFLRDIQSS--FHP-E 618

Y PQ W G QM+A++ Q ++ + G FR NGCGYV KP+FL + SS F+P E

Sbjct: 9 YKPQRAWMHGAQMIALSRQDDKEKLWLMQGMFRANGCGYVKKPNFLLNAGSSGVFYPT E 68 [SEQ ID NO:17]

Query: 619 KPI 621

P+

Sbjct: 69 NPV 71

FIG. 12

CCACGCGTCCGCCCACGCGTCCGCGAGCCAACCAGCTATACCTCTTTTGAAGATTTTAAGAACTTAGCCTCCTGAACAGT [SEQ ID NO:4]

CTTCTTCGAAAGTGAAAAGTGGTAACAGCTGATGAGTATCAAGAAATTATTTCCGCAAAGGGGCAGAGTTAATTGTAT [SEQ ID NO:5]

TTGGAACCCATGACAGCACCTACTGGGGAAGACTTCTAAGTGAGGAGAAACGGCTCTACAGGTCATGAAACT ATG M 3
 ↑SEQ ID NO:6→

E M R W F L S K I Q D D F R G G K I N L 21
 GAA ATG AGA TGG TTT TTG TCA AAG ATT CAG GAT GAC TTC AGA GGT GGA AAA ATT AAC CTA 63

E K T Q R L L E K L D I R C S Y I H V K 41
 GAA AAA ACT CAG AGG TTA CTT GAA AAA TTA GAT ATT CGG TGC AGT TAT ATT CAT GTG AAA 123

Q I F K D N D R L K Q G R I T I E E F R 61
 CAG ATT TTT AAG GAC AAT GAC AGG CTG AAA CAA GGA AGA ATC ACC ATA GAA GAA TTT AGA 183

A I Y R I I T H R E E I I E I F N T Y S 81
 GCA ATT TAT CGA ATT ATC ACG CAC AGA GAA GAA ATT ATT GAG ATT TTC AAC ACA TAT TCT 243

E N R K I L L A S N L A Q F L T Q E Q Y 101
 GAA AAC CGG AAA ATT CTT TTA GCA AGT AAT CTG GCT CAA TTT CTG ACA CAA GAA CAA TAT 303

A A E M S K A I A F E I I Q K Y E P I E 121
 GCA GCT GAG ATG AGT AAA GCT ATT GCT TTT GAG ATC ATT CAG AAA TAC GAG CCT ATC GAA 363

FIG. 13A

11/15

E V R K A H Q M S L E G F T R Y M D S R	141
GAA GTT AGG AAA GCA CAC CAA ATG TCA TTA GAA GGT TTT ACA AGA TAC ATG GAT TCA CGT	423
E C L L F K N E C R K V Y Q D M T H P L	161
GAA TGT CTA CTG TTT AAA AAT GAA TGT AGA AAA GTT TAT CAA GAT ATG ACT CAT CCA TTA	483
N D Y F I S S S H N T Y L V S D Q L L G	181
AAT GAT TAT TTT ATT TCA TCT TCA CAT AAC ACA TAT TTG GTA TCT GAT CAA TTA TTG GGA	543
P S D L W G Y A S A L V K G C R C L E I	201
CCA AGT GAC CTT TGG GGA TAT GCA AGT GCC CTT GTG AAA GGA TGC CGT TGT TTG GAG ATT	603
D C W D G A Q N E P V V Y H G Y T L T S	221
GAC TGC TGG GAT GGA GCA CAA AAT GAA CCT GTT GTA TAT CAT GGC TAC ACA CTC ACA AGC	663
K L L F K T V I Q A I H K Y A F M T S D	241
AAA CTT CTG TTT AAA ACT GTT ATC CAA GCT ATA CAC AAG TAT GCA TTC ATG ACA TCT GAC	723
Y P V V L S L E N H C S T A Q Q E V M A	261
TAC CCA GTG GTG CTC TCT TTA GAA AAT CAC TGC TCC ACT GCC CAA CAA GAA GTA ATG GCA	783
D N L Q A T F G E S L L S D M L D D F P	281
GAC AAT TTG CAG GCT ACT TTT GGA GAG TCC TTG CTT TCT GAT ATG CTT GAT GAT TTT CCT	843
D T L P S P E A L K F K I L V K N K K I	301
GAT ACT CTA CCA TCA CCA GAG GCA CTA AAA TTC AAA ATA TTA GTT AAA AAT AAG AAA ATA	903
G T L K E T H E R K G S D K H G D N Q D	321
GGA ACC TTA AAG GAA ACC CAT GAA AGA AAA GGT TCT GAT AAG CAT GGA GAC AAT CAA GAC	963
K E T G V K K L P G V M L F K K K K T R	341
AAG GAA ACA GGG GTA AAA AAG TTA CCT GGA GTA ATG CTT TTC AAG AAA AAG AAG ACC AGG	1023
K L K I A L A L S D L V I Y T K A E K F	361
AAG CTA AAA ATT GCT CTG GCC TTA TCT GAT CTT GTC ATT TAT ACG AAA GCT GAG AAA TTC	1083
K S F Q H S R L Y Q Q F N E N N S I G E	381
AAA AGC TTT CAA CAT TCA AGA TTA TAT CAG CAA TTT AAT GAA ATT AAT TCT ATT GGG GAG	1143

FIG. 13B

12/15

T	Q	A	R	K	L	S	K	L	R	V	H	E	F	I	F	H	T	R	K	401
ACA	CAA	GCC	CGA	AAA	CTT	TCA	AAA	TTG	CGA	GTC	CAT	GAG	TTT	ATT	TTT	CAC	ACC	AGG	AAG	1203
F	I	T	R	I	Y	P	K	A	T	R	A	D	S	S	N	F	N	P	Q	421
TTC	ATT	ACC	AGA	ATA	TAT	CCC	AAA	GCA	ACA	AGA	GCA	GAC	TCT	TCT	AAT	TTT	AAT	CCC	CAA	1263
E	F	W	N	I	G	C	Q	M	V	A	L	N	F	Q	T	P	G	L	P	441
GAA	TTT	TGG	AAT	ATA	GGT	TGT	CAA	ATG	GTG	GCT	TTA	AAT	TTC	CAG	ACC	CCT	GGT	CTG	CCC	1323
M	D	L	Q	N	G	K	F	L	D	N	G	G	S	G	Y	I	L	K	P	461
ATG	GAT	CTG	CAA	AAT	GGG	AAA	TTT	TTG	GAT	AAT	GGT	GGT	TCT	GGA	TAT	ATT	TTG	AAA	CCA	1383
H	F	L	R	E	S	K	S	Y	F	N	P	S	N	I	K	E	G	M	P	481
CAT	TTC	TTA	AGA	GAG	AGT	AAA	TCA	TAC	TTT	AAC	CCA	AGT	AAC	ATA	AAA	GAG	GGT	ATG	CCA	1443
I	T	L	T	I	R	L	I	S	G	I	Q	L	P	L	T	H	S	S	S	501
ATT	ACA	CTT	ACA	ATA	AGG	CTC	ATC	AGT	GGT	ATC	CAG	TTG	CCT	CTT	ACT	CAT	TCA	TCA	TCT	1503
N	K	G	D	S	L	V	I	I	E	V	F	G	V	P	N	D	Q	M	K	521
AAC	AAA	GGT	GAT	TCA	TTA	GTA	ATT	ATA	GAA	GTT	TTT	GGT	GTT	CCA	AAT	GAT	CAA	ATG	AAG	1563
Q	Q	T	R	V	I	K	K	N	A	F	S	P	R	W	N	E	T	F	T	541
CAG	CAG	ACT	CGT	GTA	ATT	AAA	AAA	AAT	GCT	TTT	AGT	CCA	AGA	TGG	AAT	GAA	ACA	TTC	ACA	1623
F	I	I	H	V	P	E	L	A	L	I	R	F	V	V	E	G	Q	G	L	561
TTT	ATT	ATT	CAT	GTC	CCA	GAA	TTG	GCA	TTG	ATA	CGT	TTT	GTT	GTT	GAA	GGT	CAA	GGT	TTA	1683
I	A	G	N	E	F	L	G	Q	Y	T	L	P	L	L	C	M	N	K	G	581
ATA	GCA	GGA	AAT	GAA	TTT	CTT	GGG	CAA	TAT	ACT	TTG	CCA	CTT	CTA	TGC	ATG	AAC	AAA	GGT	1743
Y	R	R	I	P	L	F	S	R	M	G	E	S	L	E	P	A	S	L	F	601
TAT	CGT	CGT	ATT	CCT	CTG	TTT	TCC	AGA	ATG	GGT	GAG	AGC	CTT	GAG	CCT	GCT	TCA	CTG	TTT	1803
V	Y	V	W	Y	V	R	*													609
GTT	TAT	GTT	TGG	TAC	GTC	AGA	TAA													1827

←SEQ ID NO:6↑

CAGCTAATGATAAATGACATATCATTAGCTATGCATCGCAATAAAACAGCCAAAATGAAAAAAAAAAAAAAAAAAAAA
 AATTGGCGGCCGCAAGCTTATTCCTTTAGTAAG

FIG. 13C

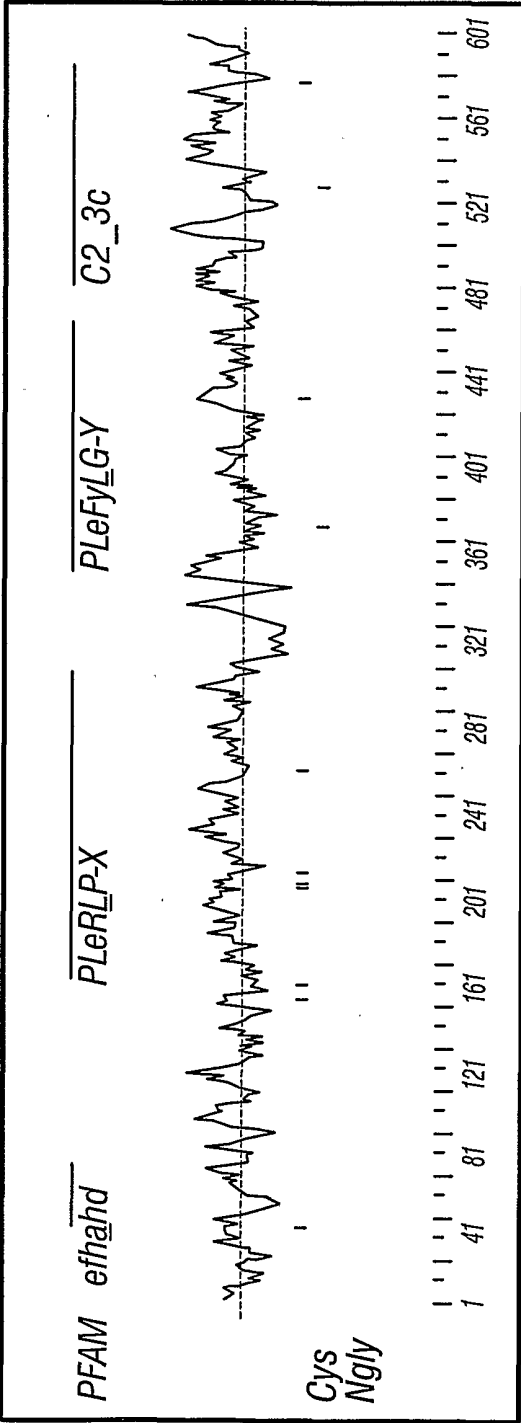


FIG. 14

14/15

efhand: domain 1 of 1, from 39 to 67: score 16.9, E = 0.48

```

      *->elkeaFkefDkDgDGkIsfeEfkaalkkl<-*      [SEQ ID NO:18]
      +k++Fk+ D+ + G+I++eEf+a+ + +<-*
16839   39   HVKQIFKDNDRLKQGRITIEEFRAIYRII   67

```

FIG. 15

```

      *->dmsiPLsHYfisSshntYlTgkQlwGkssvesYrqqLdaGcRcvELD      [SEQ ID NO:19]
      dm PL+ YfisSshntYL +Ql G+s +Y L +GcRc+E+D
16839  156   DMTHPLNDYFISSSHNTYLVSDQLLGPDDLWGYASALVKGCRCLEID  202

      cwdGkpddepiIyHGhtltleiklkdVleaIkdfafkPtSpyPvIlSlen
      cwdG + +ep++yHG+tl++ ++k+V++aI ++af ts yPv+lSlen
16839  203   CWDG-AQNEPVVYHGYTLTSKLLFKTVIQAIHKYAFM-TSDYPVVLLEN  250

      HcnsddqQrkmakyfkeiFgdmLltkPtldslttepglpLPslkdrlrgKI
      Hc+ + qQ ma+ + +Fg+ Ll + ld+ +++ Lps++ l+ KI
16839  251   HCSTA-QQEVMA DN LQATFGESLLSDM-LDDFPDT----LPSPEALKFKI  294

      LLknkk<-*
      L+knkk
16839  295   LVKNKK   300

```

FIG. 16

```

      *->ElsnLvnYiqsikFrsFelsGeekntsyeisSFsErkvkakkllkes      [SEQ ID NO:20]
      ls+Lv+Y + kF+sF s + + e S Et +a kl k +
16839  348   ALSDLVIYTKAEKFKSFQHS-RLYQQFNENNSIGET--QARKLSKLR  391

      pvefVkyNkrqLsRvYPkGtRvDSSNfmPqvFWnaGCQmVALNfQTSDlp
      + ef + ++ R+YPk tR+DSSNf+Pq+fWn+GCQmVALNfQT++lp
16839  392   VHEFIFHTRKFITRIYPKATRADSSNfNPQEFWNIGCQMVALNfQTPGLP  441

      mqINdGmFeyNggqPdGsfskGYLLKPeflR<-*
      m++ G+F++Ngg sGY+LKP flR
16893  442   MDLQNGKFLDNNG-----SGYILKPHFLR   465

```

FIG. 17

15/15

*->LtVtvieArnLpkmDk..vngrlsDPYVkvslldgkdklkkfkTkvv [SEQ ID NO:21]
 Lt++ i++ +Lp + +++++g D+ V ++++g+++d+ k++T+v+
 16839 484 LTIRLISGIQLPLTHSssNKG---DSLVIIEVFGVPNDQMKQOTRVI 527

kktNGLNPvWneEtFvFekvplpelasktLrfaVyDedrfdsrdDfiGqvt
 kk+ ++ P+Wn EtF+F +++pela +rf+V ++ + ++f+Gq t
 16839 528 KKN-AFSPRW-ETFTFI-IHVPELAL--IRFVVEGQGLIAGNEFLGQYT 572

FIG. 18

Query: 140 SRECLLFKNECRKVYQDMTHPLNDYFISSSHNTYLVSDQLL-GPSDLWGYASALVKGRC 198
 S +C +F E ++V+QDM PL+ YFI+SSHNTYL +QL G S + Y AL+KGRC
 Sbjct: 2 SPDCNVFDPEHKQVHQDMNQPLSHYFINSSHNTYLTGNQLSSGESSVEMYRQALLKGRC 61 [SEQ ID NO:22]

Query: 199 LEIDCWDGAQN--EPVVYHGYTLTSKLLFKTVIQAIHKYAFMTSDYPVVLSENHC-STA 225
 +E+DCWDG EP++ HG+T+T+++ FK ++AI ++AF+TS+YPV+LSENHC ST
 Sbjct: 62 IELDCWDGKGDPEPIITHGHTMTTEISFKDCLEAIKEHAFVTSEYPVILSENHCDSTP 121

Query: 256 QQEV-MADNLQATFGESILLSMDLDDFP----DTLPSPEALKFKILVKNKKIGTLKETHER 310
 QQ+ MA+ + FG+ L ++ L++ P LPSPE LK KIL+KNKK LKE E
 Sbjct: 122 QQQAKMAEYCKEVFGDMLFTEPLEESPLEPGKELPSPEELKRKILIKNKK---LKEHSEE 178

Query: 311 KGSDKHGDNQDKET 324
 K S++ +++ E+
 Sbjct: 179 KESEKKTDDETES 192

FIG. 19

Query: 350 SDLVIYTKAEKFSFQHSRLYQQFNENNSIGETQARKL-SKLRVHEFIFHTRKFITRIYP 408
 SDLV Y + KFK F+ + + +S E +A KL +K EF+ + ++ + R+YP
 Sbjct: 14 SDLVNYCQPVKFKGFEMAEKNYYHMSFSSENKAEKLVNKEHPKEFVRYNQRLRVYP 73 [SEQ ID NO:23]

Query: 409 KATRADSSNFPQEFWNIGCOMVALNFQTPGLPMDLQNGKFLDNGGSGYILKPHFL---R 465
 K TR DSSN+NP FWN GCOMVALNFQT G M L G F NGG GY+LKP FL R
 Sbjct: 74 KGTRIDSSNYPMVFWNHGCOMVALNFQTHGRSMWLNQGMFRANGCGYVLKPDFLLKAR 133

Query: 466 ESKSYFNP 473
 + F+P
 Sbjct: 134 PNDEVFDP 141

FIG. 20