The invention provides human nucleic acid-associated protein (NAAP) and polymolecules which identify and encode NAAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of NAAP.
NUCLEIC ACID-ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of nucleic acid-associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of nucleic acid-associated proteins.

BACKGROUND OF THE INVENTION

Multicellular organisms are comprised of diverse cell types that differ dramatically both in structure and function. The identity of a cell is determined by its characteristic pattern of gene expression, and different cell types express overlapping but distinctive sets of genes throughout development. Spatial and temporal regulation of gene expression is critical for the control of cell proliferation, cell differentiation, apoptosis, and other processes that contribute to organismal development. Furthermore, gene expression is regulated in response to extracellular signals that mediate cell-cell communication and coordinate the activities of different cell types. Appropriate gene regulation also ensures that cells function efficiently by expressing only those genes whose functions are required at a given time.

Transcription Factors

Transcriptional regulatory proteins are essential for the control of gene expression. Some of these proteins function as transcription factors that initiate, activate, repress, or terminate gene transcription. Transcription factors generally bind to the promoter, enhancer, and upstream regulatory regions of a gene in a sequence-specific manner, although some factors bind regulatory elements within or downstream of a gene coding region. Transcription factors may bind to a specific region of DNA singly or as a complex with other accessory factors. (Reviewed in Lewin, B. (1990) *Genes IV*, Oxford University Press, New York, NY, and Cell Press, Cambridge, MA, pp. 554-570.)

The double helix structure and repeated sequences of DNA create topological and chemical features which can be recognized by transcription factors. These features are hydrogen bond donor and acceptor groups, hydrophobic patches, major and minor grooves, and regular, repeated stretches of sequence which induce distinct bends in the helix. Typically, transcription factors recognize specific DNA sequence motifs of about 20 nucleotides in length. Multiple, adjacent transcription factor-binding motifs may be required for gene regulation.

Many transcription factors incorporate DNA-binding structural motifs which comprise either α helices or β sheets that bind to the major groove of DNA. Four well-characterized structural motifs
are helix-turn-helix, zinc finger, leucine zipper, and helix-loop-helix. Proteins containing these motifs may act alone as monomers, or they may form homo- or heterodimers that interact with DNA.

The helix-turn-helix motif consists of two α helices connected at a fixed angle by a short chain of amino acids. One of the helices binds to the major groove. Helix-turn-helix motifs are exemplified by the homeobox motif which is present in homeodomain proteins. These proteins are critical for specifying the anterior-posterior body axis during development and are conserved throughout the animal kingdom. The Antennapedia and Ultrabithorax proteins of Drosophila melanogaster are prototypical homeodomain proteins. (Pabo, C.O. and R.T. Sauer (1992) Annu. Rev. Biochem. 61:1053-1095.)

The zinc finger motif, which binds zinc ions, generally contains tandem repeats of about 30 amino acids consisting of periodically spaced cysteine and histidine residues. Examples of this sequence pattern, designated C2H2 and C3HC4 (“RING” finger), have been described. (Lewin, supra.) Zinc finger proteins each contain an α helix and an antiparallel β sheet whose proximity and conformation are maintained by the zinc ion. Contact with DNA is made by the arginine preceding the α helix and by the second, third, and sixth residues of the α helix. Variants of the zinc finger motif include poorly defined cysteine-rich motifs which bind zinc or other metal ions. These motifs may not contain histidine residues and are generally nonrepetitive. The zinc finger motif may be repeated in a tandem array within a protein, such that the α helix of each zinc finger in the protein makes contact with the major groove of the DNA double helix. This repeated contact between the protein and the DNA produces a strong and specific DNA-protein interaction. The strength and specificity of the interaction can be regulated by the number of zinc finger motifs within the protein. Though originally identified in DNA-binding proteins as regions that interact directly with DNA, zinc fingers occur in a variety of proteins that do not bind DNA (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York, NY, pp. 447-451). For example, Galcheva-Gargova, Z. et al. (1996) Science 272:1797-1802 have identified zinc finger proteins that interact with various cytokine receptors.

The C2H2-type zinc finger signature motif contains a 28 amino acid sequence, including 2 conserved Cys and 2 conserved His residues in a C-2-C-12-H-3-H type motif. The motif generally occurs in multiple tandem repeats. A cysteine-rich domain including the motif Asp-His-His-Cys (DHHC-CRD) has been identified as a distinct subgroup of zinc finger proteins. The DHHC-CRD region has been implicated in growth and development. One DHHC-CRD mutant shows defective function of Ras, a small membrane-associated GTP-binding protein that regulates cell growth and differentiation, while other DHHC-CRD proteins probably function in pathways not involving Ras (Bartels, D.J. et al. (1999) Mol. Cell Biol. 19:6775-6787).

Zinc-finger transcription factors are often accompanied by modular sequence motifs such as
the Kruppel-associated box (KRAB) and the SCAN domain. For example, the hypoalphalipoproteinemia susceptibility gene ZNF202 encodes a SCAN box and a KRAB domain followed by eight C2H2 zinc-finger motifs (Honer, C. et al. (2001) Biochim. Biophys. Acta 1517:441-448). The SCAN domain is a highly conserved, leucine-rich motif of approximately 60 amino acids found at the amino-terminal end of zinc finger transcription factors. SCAN domains are most often linked to C2H2 zinc finger motifs through their carboxyl-terminal end. Biochemical binding studies have established the SCAN domain as a selective hetero- and homotypic oligomerization domain. SCAN domain-mediated protein complexes may function to modulate the biological function of transcription factors (Schumacher, C. et al. (2000) J. Biol. Chem. 275:17173-17179).

The KRAB (Kruppel-associated box) domain is a conserved amino acid sequence spanning approximately 75 amino acids and is found in almost one-third of the 300 to 700 genes encoding C2H2 zinc fingers. The KRAB domain is found N-terminally with respect to the finger repeats. The KRAB domain is generally encoded by two exons; the KRAB-A region or box is encoded by one exon and the KRAB-B region or box is encoded by a second exon. The function of the KRAB domain is the repression of transcription. Transcription repression is accomplished by recruitment of either the KRAB-associated protein-1, a transcriptional corepressor, or the KRAB-A interacting protein. Proteins containing the KRAB domain are likely to play a regulatory role during development (Williams, A.J. et al. (1999) Mol. Cell Biol. 19:8526-8535). A subgroup of highly related human KRAB zinc finger proteins detectable in all human tissues is highly expressed in human T lymphoid cells (Bellefroid, E.J. et al. (1993) EMBO J. 12:1363-1374). The ZNF85 KRAB zinc finger gene, a member of the human ZNF91 family, is highly expressed in normal adult testis, in seminomas, and in the NT2/D1 teratocarcinoma cell line (Poncelet, D.A. et al. (1998) DNA Cell Biol.17:931-943).

The C4 motif is found in hormone-regulated proteins. The C4 motif generally includes only 2 repeats. A number of eukaryotic and viral proteins contain a conserved cysteine-rich domain of 40 to 60 residues (called C3HC4 zinc-finger or RING finger) that binds two atoms of zinc, and is probably involved in mediating protein-protein interactions. The 3D “cross-brace” structure of the zinc ligation system is unique to the RING domain. The spacing of the cysteines in such a domain is C-x(2)-C-x(9 to 39)-C-x(1 to 3)-H-x(2 to 3)-C-x(2)-C-x(4 to 48)-C-x(2)-C. The PHD finger is a C4HC3 zinc-finger-like motif found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation.

GATA-type transcription factors contain one or two zinc finger domains which bind specifically to a region of DNA that contains the consecutive nucleotide sequence GATA. NMR studies indicate that the zinc finger comprises two irregular anti-parallel β sheets and an α helix,
followed by a long loop to the C-terminal end of the finger (Ominchinski, J.G. (1993) Science 261:438-446). The helix and the loop connecting the two β-sheets contact the major groove of the DNA, while the C-terminal part, which determines the specificity of binding, wraps around into the minor groove.

The LIM motif consists of about 60 amino acid residues and contains seven conserved cysteine residues and a histidine within a consensus sequence (Schmeichel, K.L. and Beckerle, M.C. (1994) Cell 79:211-219). The LIM family includes transcription factors and cytoskeletal proteins which may be involved in development, differentiation, and cell growth. One example is actin-binding LIM protein, which may play roles in regulation of the cytoskeleton and cellular morphogenesis (Roof, D.J. et al. (1997) J. Cell Biol. 138:575-588). The N-terminal domain of actin-binding LIM protein has four double zinc finger motifs with the LIM consensus sequence. The C-terminal domain of actin-binding LIM protein shows sequence similarity to known actin-binding proteins such as dematin and villin. Actin-binding LIM protein binds to F-actin through its dematin-like C-terminal domain. The LIM domain may mediate protein-protein interactions with other LIM-binding proteins.

Myeloid cell development is controlled by tissue-specific transcription factors. Myeloid zinc finger proteins (MZF) include MZF-1 and MZF-2. MZF-1 functions in regulation of the development of neutrophilic granulocytes. A murine homolog MZF-2 is expressed in myeloid cells, particularly in the cells committed to the neutrophilic lineage. MZF-2 is down-regulated by G-CSF and appears to have a unique function in neutrophil development (Murai, K. et al. (1997) Genes Cells 2:581-591).

The leucine zipper motif comprises a stretch of amino acids rich in leucine which can form an amphipathic α helix. This structure provides the basis for dimerization of two leucine zipper proteins. The region adjacent to the leucine zipper is usually basic, and upon protein dimerization, is optimally positioned for binding to the major groove. Proteins containing such motifs are generally referred to as bZIP transcription factors. The leucine zipper motif is found in the proto-oncogenes Fos and Jun, which comprise the heterodimeric transcription factor AP1 involved in cell growth and the determination of cell lineage (Papavassiliou, A.G. (1995) N. Engl. J. Med. 332:45-47).

The helix-loop-helix motif (HLH) consists of a short α helix connected by a loop to a longer α helix. The loop is flexible and allows the two helices to fold back against each other and to bind to DNA. The transcription factor Myc contains a prototypical HLH motif.

The NF-kappa-B/Rel signature defines a family of eukaryotic transcription factors involved in oncogenesis, embryonic development, differentiation and immune response. Most transcription factors containing the Rel homology domain (RHD) bind as dimers to a consensus DNA sequence motif termed kappa-B. Members of the Rel family share a highly conserved 300 amino acid domain termed the Rel homology domain. The characteristic Rel C-terminal domain is involved in gene
activation and cytoplasmic anchoring functions. Proteins known to contain the RHD domain include vertebrate nuclear factor NF-kappa-B, which is a heterodimer of a DNA-binding subunit and the transcription factor p65, mammalian transcription factor RelB, and vertebrate proto-oncogene c-rel, a protein associated with differentiation and lymphopoiesis (Kabrun, N. and Enrietto, P.J. (1994) Semin. Cancer Biol. 5:103-112).

A DNA binding motif termed ARID (AT-rich interactive domain) distinguishes an evolutionarily conserved family of proteins. The approximately 100-residue ARID sequence is present in a series of proteins strongly implicated in the regulation of cell growth, development, and tissue-specific gene expression. ARID proteins include Bright (a regulator of B-cell-specific gene expression), dead ringer (involved in development), and MRF-2 (which represses expression from the cytomegalovirus enhancer) (Dallas, P.B. et al. (2000) Mol. Cell Biol. 20:3137-3146).

The ELM2 (Egl-27 and MTA1 homology 2) domain is found in metastasis-associated protein MTA1 and protein ER1. The Caenorhabditis elegans gene egl-27 is required for embryonic patterning MTA1, a human gene with elevated expression in metastatic carcinomas, is a component of a protein complex with histone deacetylase and nucleosome remodelling activities (Solari, F. et al. (1999) Development 126:2483-2494). The ELM2 domain is usually found to the N terminus of a myb-like DNA binding domain. ELM2 is also found associated with an ARID DNA.

Most transcription factors contain characteristic DNA binding motifs, and variations on the above motifs and new motifs have been and are currently being characterized. (Faisst, S. and S. Meyer (1992) Nucleic Acids Res. 20:3-26.)

**Chromatin Associated Proteins**

In the nucleus, DNA is packaged into chromatin, the compact organization of which limits the accessibility of DNA to transcription factors and plays a key role in gene regulation (Lewin, supra, pp. 409-410). The compact structure of chromatin is determined and influenced by chromatin-associated proteins such as the histones, the high mobility group (HMG) proteins, and the chromodomain proteins. There are five classes of histones, H1, H2A, H2B, H3, and H4, all of which are highly basic, low molecular weight proteins. The fundamental unit of chromatin, the nucleosome, consists of 200 base pairs of DNA associated with two copies each of H2A, H2B, H3, and H4. H1 links adjacent nucleosomes. HMG proteins are low molecular weight, non-histone proteins that may play a role in unwinding DNA and stabilizing single-stranded DNA. Chromodomain proteins play a key role in the formation of highly compacted heterochromatin, which is transcriptionally silent.

During mitosis, all DNA is compacted into heterochromatin and transcription ceases. Transcription in interphase begins with the activation of a region of chromatin. Active chromatin is decondensed. Decondensation appears to be accompanied by changes in the binding coefficient, phosphorylation, and acetylation states of chromatin histones. HMG proteins HMG13 and HMG17
selectively bind activated chromatin. Topoisomerases remove superhelical tension on DNA. The activated region decondenses, allowing gene regulatory proteins and transcription factors to assemble on the DNA.

Replication of eukaryotic chromosomes is integrated into the regulation of the cell division cycle by various proteins. For example, nucleosome assembly on replicating DNA molecules requires the nuclear chromatin assembly factor 1 (CAF-1), which mediates deposition of H3/H4 tetramers onto the DNA. Nucleosome assembly by CAF-1 depends upon reversible phosphorylation by G1/S phase-specific cyclin-dependent protein kinase 2 and type 1 protein phosphatase (PP1), thus coupling the cell cycle machinery to DNA replication (Keller, C. and Krude, T. (2000) J. Biol. Chem. 275:35512-35521). The regulatory subunit of PP1, sds22, is required for the completion of mitosis in yeast. The sds22 protein contains 11 leucine-rich repeats which are believed to mediate interactions with PP1 and other proteins (Ceulemans, H. et al. (1999) Eur. J. Biochem. 262:36-42).

The regulator of chromosome condensation (RCC1) is a eukaryotic protein that binds to chromatin and acts as a guanine nucleotide exchange factor for the nuclear GTP-binding protein Ran. The loss of RCC1 results in the premature entry of cells into mitosis, suggesting that RCC1 senses the status of the chromatin and couples it to cell cycle progression through Ran (Nishijima, H. et al. (2000) Prog. Cell Cycle Res. 4:145-156).

Patterns of chromatin structure can be stably inherited, producing heritable patterns of gene expression. In mammals, one of the two X chromosomes in each female cell is inactivated by condensation to heterochromatin during zygote development. The inactive state of this chromosome is inherited, so that adult females are mosaics of clusters of paternal-X and maternal-X clonal cell groups. The condensed X chromosome is reactivated in meiosis.

Chromatin is associated with disorders of protein expression such as thalassemia, a genetic anemia resulting from the removal of the locus control region (LCR) required for decondensation of the globin gene locus.

**Diseases and Disorders Related to Gene Regulation**

Many neoplastic disorders in humans can be attributed to inappropriate gene expression. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes. (Cleary, M.L. (1992) Cancer Surv. 15:89-104.)

The zinc finger-type transcriptional regulator WT1 is a tumor-suppressor protein that is inactivated in children with Wilm’s tumor. The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, A.G. (1995) N. Engl. J. Med. 332:45-47). Chromosomal translocations may also produce chimeric loci that fuse the coding sequence of one gene with the regulatory regions of a second unrelated gene. Such an arrangement likely results in inappropriate gene transcription, potentially contributing to malignancy. In Burkitt’s lymphoma, for
example, the transcription factor Myc is translocated to the immunoglobulin heavy chain locus, greatly enhancing Myc expression and resulting in rapid cell growth leading to leukemia (Latchman, D.S. (1996) N. Engl. J. Med. 334:28-33).

In addition, the immune system responds to infection or trauma by activating a cascade of events that coordinate the progressive selection, amplification, and mobilization of cellular defense mechanisms. A complex and balanced program of gene activation and repression is involved in this process. However, hyperactivity of the immune system as a result of improper or insufficient regulation of gene expression may result in considerable tissue or organ damage. This damage is well-documented in immunological responses associated with arthritis, allergens, heart attack, stroke, and infections. (Isselbacher et al. Harrison’s Principles of Internal Medicine, 13/e, McGraw Hill, Inc. and Teton Data Systems Software, 1996.) The causative gene for autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) was recently isolated and found to encode a protein with two PHD-type zinc finger motifs (Bjorres, P. et al. (1998) Hum. Mol. Genet. 7:1547-1553).

Furthermore, the generation of multicellular organisms is based upon the induction and coordination of cell differentiation at the appropriate stages of development. Central to this process is differential gene expression, which confers the distinct identities of cells and tissues throughout the body. Failure to regulate gene expression during development could result in developmental disorders. Human developmental disorders caused by mutations in zinc finger-type transcriptional regulators include: urogenital developmental abnormalities associated with WT1; Greig cephalopolysyndactyly, Pallister-Hall syndrome, and postaxial polydactyly type A (GLI3), and Townes-Brocks syndrome, characterized by anal, renal, limb, and ear abnormalities (SALL1) (Engelkamp, D. and V. van Heyningen (1996) Curr. Opin. Genet. Dev. 6:334-342; Kohlhase, J. et al. (1999) Am. J. Hum. Genet. 64:435-445).

SYNTHESIS OF NUCLEIC ACIDS

Polymerases

DNA and RNA replication are critical processes for cell replication and function. DNA and RNA replication are mediated by the enzymes DNA and RNA polymerase, respectively, by a “templating” process in which the nucleotide sequence of a DNA or RNA strand is copied by complementary base-pairing into a complementary nucleic acid sequence of either DNA or RNA.

However, there are fundamental differences between the two processes.

DNA polymerase catalyzes the stepwise addition of a deoxyribonucleotide to the 3'-OH end of a polynucleotide strand (the primer strand) that is paired to a second (template) strand. The new DNA strand therefore grows in the 5' to 3' direction (Alberts, B. et al. (1994) The Molecular Biology of the Cell, Garland Publishing Inc., New York, NY, pp 251-254). The substrates for the polymerization reaction are the corresponding deoxynucleotide triphosphates which must base-pair
with the correct nucleotide on the template strand in order to be recognized by the polymerase. Because DNA exists as a double-stranded helix, each of the two strands may serve as a template for the formation of a new complementary strand. Each of the two daughter cells of a dividing cell therefore inherits a new DNA double helix containing one old and one new strand. Thus, DNA is said to be replicated “semiconservatively” by DNA polymerase. In addition to the synthesis of new DNA, DNA polymerase is also involved in the repair of damaged DNA as discussed below under “Ligases.”

In contrast to DNA polymerase, RNA polymerase uses a DNA template strand to “transcribe” DNA into RNA using ribonucleotide triphosphates as substrates. Like DNA polymerization, RNA polymerization proceeds in a 5’ to 3’ direction by addition of a ribonucleoside monophosphate to the 3’-OH end of a growing RNA chain. DNA transcription generates messenger RNAs (mRNA) that carry information for protein synthesis, as well as the transfer, ribosomal, and other RNAs that have structural or catalytic functions. In eukaryotes, three discrete RNA polymerases synthesize the three different types of RNA (Alberts et al., supra, pp. 367-368). RNA polymerase I makes the large ribosomal RNAs, RNA polymerase II makes the mRNAs that will be translated into proteins, and RNA polymerase III makes a variety of small, stable RNAs, including 5S ribosomal RNA and the transfer RNAs (tRNA). In all cases, RNA synthesis is initiated by binding of the RNA polymerase to a promoter region on the DNA and synthesis begins at a start site within the promoter. Synthesis is completed at a stop (termination) signal in the DNA whereupon both the polymerase and the completed RNA chain are released.

**Ligases**

DNA repair is the process by which accidental base changes, such as those produced by oxidative damage, hydrolytic attack, or uncontrolled methylation of DNA, are corrected before replication or transcription of the DNA can occur. Because of the efficiency of the DNA repair process, fewer than one in a thousand accidental base changes causes a mutation (Alberts et al., supra, pp. 245-249). The three steps common to most types of DNA repair are (1) excision of the damaged or altered base or nucleotide by DNA nucleases, (2) insertion of the correct nucleotide in the gap left by the excised nucleotide by DNA polymerase using the complementary strand as the template and, (3) sealing the break left between the inserted nucleotide(s) and the existing DNA strand by DNA ligase. In the last reaction, DNA ligase uses the energy from ATP hydrolysis to activate the 5’ end of the broken phosphodiester bond before forming the new bond with the 3’-OH of the DNA strand. In Bloom’s syndrome, an inherited human disease, individuals are partially deficient in DNA ligation and consequently have an increased incidence of cancer (Alberts et al., supra p. 247).

**Nucleases**

Nucleases comprise enzymes that hydrolyze both DNA (DNase) and RNA (Rnase). They
serve different purposes in nucleic acid metabolism. Nucleases hydrolyze the phosphodiester bonds between adjacent nucleotides either at internal positions (endonucleases) or at the terminal 3' or 5' nucleotide positions (exonucleases). A DNA exonuclease activity in DNA polymerase, for example, serves to remove improperly paired nucleotides attached to the 3'-OH end of the growing DNA strand by the polymerase and thereby serves a "proofreading" function. As mentioned above, DNA endonuclease activity is involved in the excision step of the DNA repair process.

RNases also serve a variety of functions. For example, RNase P is a ribonucleoprotein enzyme which cleaves the 5' end of pre-tRNAs as part of their maturation process. RNase H digests the RNA strand of an RNA/DNA hybrid. Such hybrids occur in cells invaded by retroviruses, and RNase H is an important enzyme in the retroviral replication cycle. Pancreatic RNase secreted by the pancreas into the intestine hydrolyzes RNA present in ingested foods. RNase activity in serum and cell extracts is elevated in a variety of cancers and infectious diseases (Schein, C.H. (1997) Nat. Biotechnol. 15:529-536). Regulation of RNase activity is being investigated as a means to control tumor angiogenesis, allergic reactions, viral infection and replication, and fungal infections.

MODIFICATION OF NUCLEIC ACIDS

Methyllases

Methylation of specific nucleotides occurs in both DNA and RNA, and serves different functions in the two macromolecules. Methylation of cytosine residues to form 5-methyl cytosine in DNA occurs specifically in CG sequences which are base-paired with one another in the DNA double-helix. The pattern of methylation is passed from generation to generation during DNA replication by an enzyme called "maintenance methylase" that acts preferentially on those CG sequences that are base-paired with a CG sequence that is already methylated. Such methylation appears to distinguish active from inactive genes by preventing the binding of regulatory proteins that "turn on" the gene, but permitting the binding of proteins that inactivate the gene (Alberts et al. supra pp. 448-451). In RNA metabolism, "tRNA methylase" produces one of several nucleotide modifications in tRNA that affect the conformation and base-pairing of the molecule and facilitate the recognition of the appropriate mRNA codons by specific tRNAs. The primary methylation pattern is the dimethylation of guanine residues to form N,N-dimethyl guanine.

Helicases and Single-stranded Binding Proteins

Helicases are enzymes that destabilize and unwind double helix structures in both DNA and RNA. Since DNA replication occurs more or less simultaneously on both strands, the two strands must first separate to generate a replication "fork" for DNA polymerase to act on. Two types of replication proteins contribute to this process, DNA helicases and single-stranded binding proteins. DNA helicases hydrolyze ATP and use the energy of hydrolysis to separate the DNA strands. Single-stranded binding proteins (SSBs) then bind to the exposed DNA strands, without covering the bases,
thereby temporarily stabilizing them for templating by the DNA polymerase (Alberts et al. supra, pp. 255-256).

RNA helicases also alter and regulate RNA conformation and secondary structure. Like the DNA helicases, RNA helicases utilize energy derived from ATP hydrolysis to destabilize and unwind RNA duplexes. The most well-characterized and ubiquitous family of RNA helicases is the DEAD-box family, so named for the conserved B-type ATP-binding motif which is diagnostic of proteins in this family. Over 40 DEAD-box helicases have been identified in organisms as diverse as bacteria, insects, yeast, amphibians, mammals, and plants. DEAD-box helicases function in diverse processes such as translation initiation, splicing, ribosome assembly, and RNA editing, transport, and stability.

Examples of these RNA helicases include yeast Drs1 protein, which is involved in ribosomal RNA processing; yeast TIF1 and TIF2 and mammalian eIF-4A, which are essential to the initiation of RNA translation; and human p68 antigen, which regulates cell growth and division (Ripmaster, T.L. et al. (1992) Proc. Natl. Acad. Sci. USA 89:11131-11135; Chang, T.-H. et al. (1990) Proc. Natl. Acad. Sci. USA 87:1571-1575). These RNA helicases demonstrate strong sequence homology over a stretch of some 420 amino acids. Included among these conserved sequences are the consensus sequence for the A motif of an ATP binding protein; the “DEAD box” sequence, associated with ATPase activity; the sequence SAT, associated with the actual helicase unwinding region; and an octapeptide consensus sequence, required for RNA binding and ATP hydrolysis (Pause, A. et al. (1993) Mol. Cell Biol. 13:6789-6798). Differences outside of these conserved regions are believed to reflect differences in the functional roles of individual proteins (Chang, T.H. et al. (1990) Proc. Natl. Acad. Sci. USA 87:1571-1575).

Some DEAD-box helicases play tissue- and stage-specific roles in spermatogenesis and embryogenesis. Overexpression of the DEAD-box 1 protein (DDX1) may play a role in the progression of neuroblastoma (Nb) and retinoblastoma (Rb) tumors (Godbout, R. et al. (1998) J. Biol. Chem. 273:21161-21168). These observations suggest that DDX1 may promote or enhance tumor progression by altering the normal secondary structure and expression levels of RNA in cancer cells. Other DEAD-box helicases have been implicated either directly or indirectly in tumorigenesis. (Discussed in Godbout, supra.) For example, murine p68 is mutated in ultraviolet light-induced tumors, and human DDX6 is located at a chromosomal breakpoint associated with B-cell lymphoma.

Similarly, a chimeric protein comprised of DDX10 and NUP98, a nucleoporin protein, may be involved in the pathogenesis of certain myeloid malignancies.

**Topoisomerases**

Besides the need to separate DNA strands prior to replication, the two strands must be "unwound" from one another prior to their separation by DNA helicases. This function is performed by proteins known as DNA topoisomerases. DNA topoisomerase effectively acts as a reversible
nuclease that hydrolyzes a phosphodiesterase bond in a DNA strand, permits the two strands to rotate freely about one another to remove the strain of the helix, and then rejoins the original phosphodiester bond between the two strands. Topoisomerases are essential enzymes responsible for the topological rearrangement of DNA brought about by transcription, replication, chromatin formation, recombination, and chromosome segregation. Superhelical coils are introduced into DNA by the passage of processive enzymes such as RNA polymerase, or by the separation of DNA strands by a helicase prior to replication. Knotting and concatenation can occur in the process of DNA synthesis, storage, and repair. All topoisomerases work by breaking a phosphodiester bond in the ribose-phosphate backbone of DNA. A catalytic tyrosine residue on the enzyme makes a nucleophilic attack on the scissile phosphodiester bond, resulting in a reaction intermediate in which a covalent bond is formed between the enzyme and one end of the broken strand. A tyrosine-DNA phosphodiesterase functions in DNA repair by hydrolyzing this bond in occasional dead-end topoisomerase I-DNA intermediates (Pouliot, J.J. et al. (1999) Science 286:552-555).

Two types of DNA topoisomerase exist, types I and II. Type I topoisomerases work as monomers, making a break in a single strand of DNA while type II topoisomerases, working as homodimers, cleave both strands. DNA Topoisomerase I causes a single-strand break in a DNA helix to allow the rotation of the two strands of the helix about the remaining phosphodiester bond in the opposite strand. DNA topoisomerase II causes a transient break in both strands of a DNA helix where two double helices cross over one another. This type of topoisomerase can efficiently separate two interlocked DNA circles (Albers et al. supra, pp.260-262). Type II topoisomerases are largely confined to proliferating cells in eukaryotes, such as cancer cells. For this reason they are targets for anticancer drugs. Topoisomerase II has been implicated in multi-drug resistance (MDR) as it appears to aid in the repair of DNA damage inflicted by DNA binding agents such as doxorubicin and vincristine.

The topoisomerase I family includes topoisomerases I and III (topo I and topo III). The crystal structure of human topoisomerase I suggests that rotation about the intact DNA strand is partially controlled by the enzyme. In this “controlled rotation” model, protein-DNA interactions limit the rotation, which is driven by torsional strain in the DNA (Stewart, L. et al. (1998) Science 379:1534-1541). Structurally, topo I can be recognized by its catalytic tyrosine residue and a number of other conserved residues in the active site region. Topo I is thought to function during transcription. Two topo IIIIs are known in humans, and they are homologous to prokaryotic topoisomerase I, with a conserved tyrosine and active site signature specific to this family. Topo III has been suggested to play a role in meiotic recombination. A mouse topo III is highly expressed in testis tissue and its expression increases with the increase in the number of cells in pachytene (Seki, T. et al. (1998) J. Biol. Chem. 273:28553-28556).
The topoisomerase II family includes two isozymes (IIα and IIβ) encoded by different genes. Topo II cleaves double stranded DNA in a reproducible, nonrandom fashion, preferentially in an AT rich region, but the basis of cleavage site selectivity is not known. Structurally, topo II is made up of four domains, the first two of which are structurally similar and probably distantly homologous to similar domains in eukaryotic topo I. The second domain bears the catalytic tyrosine, as well as a highly conserved pentapeptide. The IIα isoform appears to be responsible for unlinking DNA during chromosome segregation. Cell lines expressing IIα but not IIβ suggest that IIβ is dispensable in cellular processes; however, IIβ knockout mice died perinatally due to a failure in neural development. That the major abnormalities occurred in predominantly late developmental events (neurogenesis) suggests that IIβ is needed not at mitosis, but rather during DNA repair (Yang, X. et al. (2000) Science 287:131-134).


**Recombinases**

Genetic recombination is the process of rearranging DNA sequences within an organism’s genome to provide genetic variation for the organism in response to changes in the environment.
DNA recombination allows variation in the particular combination of genes present in an individual’s genome, as well as the timing and level of expression of these genes. (See Alberts et al. supra pp. 263-273.) Two broad classes of genetic recombination are commonly recognized, general recombination and site-specific recombination. General recombination involves genetic exchange between any homologous pair of DNA sequences usually located on two copies of the same chromosome. The process is aided by enzymes, recombinases, that “nick” one strand of a DNA duplex more or less randomly and permit exchange with a complementary strand on another duplex. The process does not normally change the arrangement of genes in a chromosome. In site-specific recombination, the recombinase recognizes specific nucleotide sequences present in one or both of the recombining molecules. Base-pairing is not involved in this form of recombination and therefore it does not require DNA homology between the recombining molecules. Unlike general recombination, this form of recombination can alter the relative positions of nucleotide sequences in chromosomes.

RNA METABOLISM

Ribonucleic acid (RNA) is a linear single-stranded polymer of four nucleotides, ATP, CTP, UTP, and GTP. In most organisms, RNA is transcribed as a copy of deoxyribonucleic acid (DNA), the genetic material of the organism. In retroviruses RNA rather than DNA serves as the genetic material. RNA copies of the genetic material encode proteins or serve various structural, catalytic, or regulatory roles in organisms. RNA is classified according to its cellular localization and function.

Messenger RNAs (mRNAs) encode polypeptides. Ribosomal RNAs (rRNAs) are assembled, along with ribosomal proteins, into ribosomes, which are cytoplasmic particles that translate mRNA into polypeptides. Transfer RNAs (tRNAs) are cytosolic adaptor molecules that function in mRNA translation by recognizing both an mRNA codon and the amino acid that matches that codon. Heterogeneous nuclear RNAs (hnRNAs) include mRNA precursors and other nuclear RNAs of various sizes. Small nuclear RNAs (snRNAs) are a part of the nuclear spliceosome complex that removes intervening, non-coding sequences (introns) and rejoins exons in pre-mRNAs.

Proteins are associated with RNA during its transcription from DNA, RNA processing, and translation of mRNA into protein. Proteins are also associated with RNA as it is used for structural, catalytic, and regulatory purposes.

RNA Processing

Ribosomal RNAs (rRNAs) are assembled, along with ribosomal proteins, into ribosomes, which are cytoplasmic particles that translate messenger RNA (mRNA) into polypeptides. The eukaryotic ribosome is composed of a 60S (large) subunit and a 40S (small) subunit, which together form the 80S ribosome. In addition to the 18S, 28S, 5S, and 5.8S rRNAs, ribosomes contain from 50 to over 80 different ribosomal proteins, depending on the organism. Ribosomal proteins are classified
according to which subunit they belong (i.e., L, if associated with the large 60S large subunit or S if associated with the small 40S subunit). E. coli ribosomes have been the most thoroughly studied and contain 50 proteins, many of which are conserved in all life forms. The structures of nine ribosomal proteins have been solved to less than 3.0D resolution (i.e., S5, S6, S17, L1, L6, L9, L12, L14, L30), revealing common motifs, such as b-a-b protein folds in addition to acidic and basic RNA-binding motifs positioned between b-strands. Most ribosomal proteins are believed to contact rRNA directly (reviewed in Liljas, A. and Garber, M. (1995) Curr. Opin. Struct. Biol. 5:721–727; see also Woodson, S.A. and Leontis, N.B. (1998) Curr. Opin. Struct. Biol. 8:294-300; Ramakrishnan, V. and White, S.W. (1998) Trends Biochem. Sci. 23:208-212).

Ribosomal proteins may undergo post-translational modifications or interact with other ribosome-associated proteins to regulate translation. For example, the highly homologous 40S ribosomal protein S6 kinases (S6K1 and S6K2) play a key role in the regulation of cell growth by controlling the biosynthesis of translational components which make up the protein synthetic apparatus (including the ribosomal proteins). In the case of S6K1, at least eight phosphorylation sites are believed to mediate kinase activation in a hierarchical fashion (Dufner and Thomas (1999) Exp. Cell. Res. 253:100-109). Some of the ribosomal proteins, including L1, also function as translational repressors by binding to polycistronic mRNAs encoding ribosomal proteins (reviewed in Liljas, A. supra and Garber, M. supra).

Recent evidence suggests that a number of ribosomal proteins have secondary functions independent of their involvement in protein biosynthesis. These proteins function as regulators of cell proliferation and, in some instances, as inducers of cell death. For example, the expression of human ribosomal protein L13a has been shown to induce apoptosis by arresting cell growth in the G2/M phase of the cell cycle. Inhibition of expression of L13a induces apoptosis in target cells, which suggests that this protein is necessary, in the appropriate amount, for cell survival. Similar results have been obtained in yeast where inactivation of yeast homologues of L13a, rp22 and rp23, results in severe growth retardation and death. A closely related ribosomal protein, L7, arrests cells in G1 and also induces apoptosis. Thus, it appears that a subset of ribosomal proteins may function as cell cycle checkpoints and compose a new family of cell proliferation regulators.

Mapping of individual ribosomal proteins on the surface of intact ribosomes is accomplished using 3D immunocryoelectronmicroscopy, whereby antibodies raised against specific ribosomal proteins are visualized. Progress has been made toward the mapping of L1, L7, and L12 while the structure of the intact ribosome has been solved to only 20-25D resolution and inconsistencies exist among different crude structures (Frank, J. (1997) Curr. Opin. Struct. Biol. 7:266-272).

Three distinct sites have been identified on the ribosome. The aminoacyl-tRNA acceptor site (A site) receives charged tRNAs (with the exception of the initiator-tRNA). The peptidyl-tRNA site

Various proteins are necessary for processing of transcribed RNAs in the nucleus. Pre-mRNA processing steps include capping at the 5' end with methylguanosine, polyadenylating the 3' end, and splicing to remove introns. The primary RNA transcript from DNA is a faithful copy of the gene containing both exon and intron sequences, and the latter sequences must be cut out of the RNA transcript to produce a mRNA that codes for a protein. This “splicing” of the mRNA sequence takes place in the nucleus with the aid of a large, multicomponent ribonucleoprotein complex known as a spliceosome. The spliceosomal complex is comprised of five small nuclear ribonucleoprotein particles (snRNPs) designated U1, U2, U4, U5, and U6. Each snRNP contains a single species of snRNA and about ten proteins. The RNA components of some snRNPs recognize and base-pair with intron consensus sequences. The protein components mediate spliceosome assembly and the splicing reaction. Autoantibodies to snRNP proteins are found in the blood of patients with systemic lupus erythematosus (Stryer, L. (1995) *Biochemistry*, W.H. Freeman and Company, New York NY, p. 863).

Heterogeneous nuclear ribonucleoproteins (hnRNPs) have been identified that have roles in splicing, exporting of the mature RNAs to the cytoplasm, and mRNA translation (Biamonti, G. et al. (1998) Clin. Exp. Rheumatol. 16:317-326). Some examples of hnRNPs include the yeast proteins Hrp1p, involved in cleavage and polyadenylation at the 3' end of the RNA; Cbp80p, involved in capping the 5' end of the RNA; and Npl3p, a homolog of mammalian hnRNP A1, involved in export of mRNA from the nucleus (Shen, E.C. et al. (1998) Genes Dev. 12:679-691). HnRNPs have been shown to be important targets of the autoimmune response in rheumatic diseases (Biamonti, supra).

Many snRNP and hnRNP proteins are characterized by an RNA recognition motif (RRM). (Reviewed in Birney, E. et al. (1993) Nucleic Acids Res. 21:5803-5816.) The RRM is about 80 amino acids in length and forms four β-strands and two α-helices arranged in an α/β sandwich. The RRM contains a core RNP-1 octapeptide motif along with surrounding conserved sequences. In addition to snRNP proteins, examples of RNA-binding proteins which contain the above motifs include heteronuclear ribonucleoproteins which stabilize nascent RNA and factors which regulate alternative splicing. Alternative splicing factors include developmentally regulated proteins, specific examples of which have been identified in lower eukaryotes such as Drosophila melanogaster and Caenorhabditis elegans. These proteins play key roles in developmental processes such as pattern formation and sex determination, respectively. (See, for example, Hodgkin, J. et al. (1994)
Development 120:3681-3689.)

The 3′ ends of most eukaryote mRNAs are also posttranscriptionally modified by polyadenylation. Polyadenylation proceeds through two enzymatically distinct steps: (i) the endonucleolytic cleavage of nascent mRNAs at cis-acting polyadenylation signals in the 3′-untranslated (non-coding) region and (ii) the addition of a poly(A) tract to the 5′ mRNA fragment. The presence of cis-acting RNA sequences is necessary for both steps. These sequences include 5′-AAUAAA-3′ located 10-30 nucleotides upstream of the cleavage site and a less well-conserved GU- or U-rich sequence element located 10-30 nucleotides downstream of the cleavage site. Cleavage stimulation factor (CstF), cleavage factor I (CF I), and cleavage factor II (CF II) are involved in the cleavage reaction while cleavage and polyadenylation specificity factor (CPSF) and poly(A) polymerase (PAP) are necessary for both cleavage and polyadenylation. An additional enzyme, poly(A)-binding protein II (PAB II), promotes poly(A) tract elongation (Rüegsegger, U. et al. (1996) J. Biol. Chem. 271:6107-6113; and references within).

TRANSLATION

The translation of eukaryotic mRNA is a highly competitive and tightly regulated step in gene expression. Control of this step is most commonly exerted at the rate-limiting initiation phase. Ribosomal proteins involved in translation initiation have been known for some time and their biochemical activities were used to build the currently accepted model for cap-dependent initiation of translation (Merrick, W.C. et al. (1996) in Translational Control, Hershey, J.W.B. et al. Ed., Cold Spring Harbor Laboratory Press, pp. 31-69). According to this model, the 5′ cap structure (m⁷GpppN) attracts the eukaryotic initiation factor 4F (eIF4F) complex to the mRNA. eIF4F is a heteromultimeric complex composed of the cap-binding protein eIF4E, the RNA-dependent ATPase eIF4A, and the modular factor eIF4G. The small (40S) ribosomal subunit binds to the 5′ end of an mRNA as a 43S complex which is thought to unwind secondary structure in the 5′ UTR. The resulting 48S complex then advances through the initiation cycle. A later movement of the 43S complex along the mRNA, termed scanning, is the most plausible explanation for a faithful recognition of the (usually) first AUG triplet as the start codon. Codon-anticodon base-pairing with Met-tRNA⁴ triggers eukaryotic initiation factor 2 (eIF2)-bound GTP hydrolysis, catalysed by eukaryotic initiation factor 5 (eIF5). It has been thought that this causes dissociation of initiation factors and the large (60S) subunit joining to form the 80S ribosome.

Correct translation of the genetic code depends upon each amino acid forming a linkage with the appropriate transfer RNA (tRNA). The aminoacyl-tRNA synthetases (aaRSs) are essential proteins found in all living organisms. The aaRSs are responsible for the activation and correct attachment of an amino acid with its cognate tRNA, as the first step in protein biosynthesis.

Prokaryotic organisms have at least twenty different types of aaRSs, one for each different amino
acid, while eukaryotes usually have two aaRSs, a cytosolic form and a mitochondrial form, for each
different amino acid. The 20 aaRS enzymes can be divided into two structural classes. Class I
enzymes add amino acids to the 2' hydroxyl at the 3' end of tRNAs while Class II enzymes add amino
acids to the 3' hydroxyl at the 3' end of tRNAs. Each class is characterized by a distinctive topology
of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding
Rossman ‘fold’. In particular, a consensus tetrapeptide motif is highly conserved (Prosite Document
PD00161, Aminoacyl-transfer RNA synthetases class-I signature). Class I enzymes are specific
for arginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan,
and valine. Class II enzymes contain a central catalytic domain, which consists of a seven-stranded
antiparallel β-sheet domain, as well as N- and C- terminal regulatory domains. Class II enzymes are
separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the
latter group is further subdivided by the structure of the N- and C-terminal regulatory domains
alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and
threonine.

Certain aaRSs also have editing functions. IleRS, for example, can misactivate valine to form
Val-tRNA\textsuperscript{Val}, but this product is cleared by a hydrolytic activity that destroys the mischarged product.
This editing activity is located within a second catalytic site found in the connective polypeptide 1
region (CP1), a long insertion sequence within the Rossman fold domain of Class I enzymes
(Schimmel, P. et al. (1998) FASEB J. 12:1599-1609). AaRSs also play a role in tRNA processing. It
has been shown that mature tRNAs are charged with their respective amino acids in the nucleus
before export to the cytoplasm, and charging may serve as a quality control mechanism to insure the

Under optimal conditions, polypeptide synthesis proceeds at a rate of approximately 40
amino acid residues per second. The rate of misincorporation during translation in on the order of 10\textsuperscript{-4}
and is primarily the result of aminoacyl-t-RNAs being charged with the incorrect amino acid.
Incorrectly charged tRNA are toxic to cells as they result in the incorporation of incorrect amino acid
residues into an elongating polypeptide. The rate of translation is presumed to be a compromise
between the optimal rate of elongation and the need for translational fidelity. Mathematical
calculations predict that 10\textsuperscript{-4} is indeed the maximum acceptable error rate for protein synthesis in a
biological system (reviewed in Stryer, L. supra; and Watson, J. et al. (1987) The Benjamin/Cummings
Publishing Co., Inc. Menlo Park, CA). A particularly error prone aminoacyl-tRNA charging event is
the charging of tRNA\textsuperscript{Gln} with Gln. A mechanism exits for the correction of this mischarging event
which likely has its origins in evolution. Gln was among the last of the 20 naturally occurring amino
acids used in polypeptide synthesis to appear in nature. Gram positive eubacteria, cyanobacteria,
Archea, and eukaryotic organelles possess a noncanonical pathway for the synthesis of Gln-tRNA\textsuperscript{Gln} based on the transformation of Glu-tRNA\textsuperscript{Gln} (synthesized by Glu-tRNA synthetase, GluRS) using the enzyme Glu-tRNA\textsuperscript{Gln} amidotransferase (Glu-AdT). The reactions involved in the transamidation pathway are as follows (Curnow, A.W. et al. (1997) Nucleic Acids Symposium 36:2-4):

\[
\text{GluRS} \\
\text{tRNA}^{\text{Gln}} + \text{Glu} + \text{ATP} \rightarrow \text{Glu-tRNA}^{\text{Gln}} + \text{AMP} + \text{PP}_i
\]

\[
\text{Glu-AdT} \\
\text{Glu-tRNA}^{\text{Gln}} + \text{Gln} + \text{ATP} \rightarrow \text{Gln-tRNA}^{\text{Gln}} + \text{Glu} + \text{ADP} + \text{P}
\]

A similar enzyme, Asp-tRNA\textsuperscript{Asn} amidotransferase, exists in Archea, which transforms Asp-tRNA\textsuperscript{Asn} to Asn-tRNA\textsuperscript{Asn}. Formylase, the enzyme that transforms Met-tRNA\textsuperscript{Met} to fMet-tRNA\textsuperscript{Met} in eubacteria, is likely to be a related enzyme. A hydrolytic activity has also been identified that destroys mischarged Val-tRNA\textsuperscript{Val} (Schimmel, P. et al. (1998) FASEB J. 12:1599-1609). One likely scenario for the evolution of Glu-AdT in primitive life forms is the absence of a specific glutaminyl-tRNA synthetase (GlnRS), requiring an alternative pathway for the synthesis of Gln-tRNA\textsuperscript{Gln}. In fact, deletion of the Glu-AdT operon in Gram positive bacteria is lethal (Curnow, A.W. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11819-11826). The existence of GluRS activity in other organisms has been inferred by the high degree of conservation in translation machinery in nature; however, GluRS has not been identified in all organisms, including Homo sapiens. Such an enzyme would be responsible for ensuring translational fidelity and reducing the synthesis of defective polypeptides.

In addition to their function in protein synthesis, specific aminoacyl tRNA synthetases also play roles in cellular fidelity, RNA splicing, RNA trafficking, apoptosis, and transcriptional and translational regulation. For example, human tyrosyl-tRNA synthetase can be proteolytically cleaved into two fragments with distinct cytokine activities. The carboxy-terminal domain exhibits monocyte and leukocyte chemotaxis activity as well as stimulating production of myeloperoxidase, tumor necrosis factor-\(\alpha\), and tissue factor. The N-terminal domain binds to the interleukin-8 type A receptor and functions as an interleukin-8-like cytokine. Human tyrosyl-tRNA synthetase is secreted from apoptotic tumor cells and may accelerate apoptosis (Wakasugi, K., and Schimmel, P. (1999) Science 284:147-151). Mitochondrial Neurospora crassa TyrRS and S. cerevisiae LeuRS are essential factors for certain group I intron splicing activities, and human mitochondrial LeuRS can substitute for the yeast LeuRS in a yeast null strain. Certain bacterial aaRSs are involved in regulating their own transcription or translation (Martinis, supra). Several aaRSs are able to synthesize diadenosine oligophosphates, a class of signalling molecules with roles in cell proliferation, differentiation, and apoptosis (Kisselev, L.L et al. (1998) FEBS Lett. 427:157-163; Vartanian, A. et al. (1999) FEBS Lett.
Autoantibodies against aminoacyl-tRNAs are generated by patients with autoimmune diseases such as rheumatic arthritis, dermatomyositis and polymyositis, and correlate strongly with complicating interstitial lung disease (ILD) (Freist, W. et al. (1999) Biol. Chem. 380:623-646; Freist, W. et al. (1996) Biol. Chem. Hoppe Seyler 377:343-356). These antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

Comparison of aaRS structures between humans and pathogens has been useful in the design of novel antibiotics (Schimmel, supra). Genetically engineered aaRSs have been utilized to allow site-specific incorporation of unnatural amino acids into proteins in vivo (Liu, D.R. et al. (1997) Proc. Natl. Acad. Sci. USA 94:10092-10097).

**tRNA Modifications**

The modified ribonucleoside, pseudouridine (ψ), is present ubiquitously in the anticodon regions of transfer RNAs (tRNAs), large and small ribosomal RNAs (rRNAs), and small nuclear RNAs (snRNAs). ψ is the most common of the modified nucleosides (i.e., other than G, A, U, and C) present in tRNAs. Only a few yeast tRNAs that are not involved in protein synthesis do not contain ψ (Cortese, R. et al. (1974) J. Biol. Chem. 249:1103-1108). The enzyme responsible for the conversion of uridine to ψ, pseudouridine synthase (pseudouridylate synthase), was first isolated from *Salmonella typhimurium* (Arena, F. et al. (1978) Nucleic Acids Res. 5:4523-4536). The enzyme has since been isolated from a number of mammals, including steer and mice (Green, C.J. et al. (1982) J. Biol. Chem. 257:3045-52; and Chen, J. and Patton, J.R. (1999) RNA 5:409-419). tRNA pseudouridine synthases have been the most extensively studied members of the family. They require a thiol donor (e.g., cysteine) and a monovalent cation (e.g., ammonia or potassium) for optimal activity. Additional cofactors or high energy molecules (e.g., ATP or GTP) are not required (Green, supra). Other eukaryotic pseudouridine synthases have been identified that appear to be specific for rRNA (reviewed in Smith, C.M. and Steitz, J.A. (1997) Cell 89:669-672) and a dual-specificity enzyme has been identified that uses both tRNA and rRNA substrates (Wrzesinski, J. et al. (1995) RNA 1: 437-448). The absence of ψ in the anticodon loop of tRNAs results in reduced growth in both bacteria (Singer, C.E. et al. (1972) Nature New Biol. 238:72-74) and yeast (Lecointe, F. (1998) J. Biol. Chem. 273:1316-1323), although the genetic defect is not lethal.

Another ribonucleoside modification that occurs primarily in eukaryotic cells is the conversion of guanosine to \( \text{N}^2,\text{N}^2\)-dimethylguanosine (\( \text{m}^2\text{G} \)) at position 26 or 10 at the base of the D-stem of cytosolic and mitochondrial tRNAs. This post-transcriptional modification is believed to stabilize tRNA structure by preventing the formation of alternative tRNA secondary and tertiary structures. Yeast tRNA\(^{\text{m}2\text{G}}\) is unusual in that it does not contain this modification. The modification
does not occur in eubacteria, presumably because the structure of tRNAs in these cells and organelles is sequence constrained and does not require posttranscriptional modification to prevent the formation of alternative structures (Steinberg, S. and Cedergren, R. (1995) RNA 1:886-891, and references within). The enzyme responsible for the conversion of guanosine to m^2,G is a 63 kDa S-
adenosylmethionine (SAM)-dependent tRNA N^2,N^2-dimethyl-guanosine methyltransferase (also referred to as the TRM1 gene product and herein referred to as TRM) (Edqvist, J. (1995) Biochimie 77:54-61). The enzyme localizes to both the nucleus and the mitochondria (Li, J-M. et al. (1989) J. Cell Biol. 109:1411-1419). Based on studies with TRM from Xenopus laevis, there appears to be a requirement for base pairing at positions C11-G24 and G10-C25 immediately preceding the G26 to be modified, with other structural features of the tRNA also being required for the proper presentation of the G26 substrate (Edqvist. J. et al. (1992) Nucleic Acids Res. 20:6575-6581). Studies in yeast suggest that cells carrying a weak ochre tRNA suppressor (sup3-i) are unable to suppress translation termination in the absence of TRM activity, suggesting a role for TRM in modifying the frequency of suppression in eukaryotic cells (Niederberger, C. et al. (1999) FEBS Lett. 464:67-70), in addition to the more general function of ensuring the proper three-dimensional structures for tRNA.

**Translation Initiation**

Initiation of translation can be divided into three stages. The first stage brings an initiator transfer RNA (Met-tRNA\textsubscript{i}) together with the 40S ribosomal subunit to form the 43S preinitiation complex. The second stage binds the 43S preinitiation complex to the mRNA, followed by migration of the complex to the correct AUG initiation codon. The third stage brings the 60S ribosomal subunit to the 40S subunit to generate an 80S ribosome at the initiation codon. Regulation of translation primarily involves the first and second stage in the initiation process (V.M. Pain (1996) Eur. J. Biochem. 236:747-771).

Several initiation factors, many of which contain multiple subunits, are involved in bringing an initiator tRNA and the 40S ribosomal subunit together. eIF2, a guanine nucleotide binding protein, recruits the initiator tRNA to the 40S ribosomal subunit. Only when eIF2 is bound to GTP does it associate with the initiator tRNA. eIF2B, a guanine nucleotide exchange protein, is responsible for converting eIF2 from the GDP-bound inactive form to the GTP-bound active form. Two other factors, eIF1A and eIF3 bind and stabilize the 40S subunit by interacting with the 18S ribosomal RNA and specific ribosomal structural proteins. eIF3 is also involved in association of the 40S ribosomal subunit with mRNA. The Met-tRNA\textsubscript{i}, eIF1A, eIF3, and 40S ribosomal subunit together make up the 43S preinitiation complex (Pain, supra).

strains of *S. cerevisiae* which lack the gene which encodes yeast IF2 can be used to demonstrate this evolutionary conservation with respect to IF2 activity. Protein biosynthetic activity of translation extracts prepared from such mutant strains can be restored by addition of recombinant yIF2 as described in Choi et al. (supra). Evidence that the biologic activity of these same translation extracts can be restored by addition of either human or archael IF2 (Lee et al. supra), supports the idea of universal conservation of IF2 function throughout evolution.

Additional factors are required for binding of the 43S preinitiation complex to an mRNA molecule, and the process is regulated at several levels. eIF4F is a complex consisting of three proteins: eIF4E, eIF4A, and eIF4G. eIF4E recognizes and binds to the mRNA 5' terminal m7GTP cap, eIF4A is a bidirectional RNA-dependent helicase, and eIF4G is a scaffolding polypeptide. eIF4G has three binding domains. The N-terminal third of eIF4G interacts with eIF4E, the central third interacts with eIF4A, and the C-terminal third interacts with eIF3 bound to the 43S preinitiation complex. Thus, eIF4G acts as a bridge between the 40S ribosomal subunit and the mRNA (M.W. Hentze (1997) Science 275:500-501).

The ability of eIF4F to initiate binding of the 43S preinitiation complex is regulated by structural features of the mRNA. The mRNA molecule has an untranslated region (UTR) between the 5' cap and the AUG start codon. In some mRNAs this region forms secondary structures that impede binding of the 43S preinitiation complex. The helicase activity of eIF4A is thought to function in removing this secondary structure to facilitate binding of the 43S preinitiation complex (Pain, supra).

Overexpression of eIF4E results in rapid cell or tissue proliferation and malignant transformation. eIF4E facilitates the synthesis of two powerful tumor angiogenic factors (VEGF and FGF-2) by selectively enhancing their translation. eIF4E is overexpressed not only in all head and neck squamous cell cancers but also in some dysplastic margins. Tumorigenesis in the head and neck is proposed to be a multistep process preceded by clinically evident precancerous lesions (Nathan et al. (1999) Laryngoscope 109:1253-1258; De Benedetti and Harris (1999) Int. J. Biochem. Cell Biol. 31:59-72).

**Translation Elongation**

Elongation is the process whereby additional amino acids are joined to the initiator methionine to form the complete polypeptide chain. The elongation factors EF1 α, EF1 β γ, and EF2 are involved in elongating the polypeptide chain following initiation. EF1 α is a GTP-binding protein. In EF1 α's GTP-bound form, it brings an aminoacyl-tRNA to the ribosome's A site. The amino acid attached to the newly arrived aminoacyl-tRNA forms a peptide bond with the initiator methionine. The GTP on EF1 α is hydrolyzed to GDP, and EF1 α-GDP dissociates from the ribosome. EF1 β γ binds EF1 α-GDP and induces the dissociation of GDP from EF1 α, allowing
EFI α to bind GTP and a new cycle to begin.

As subsequent aminoacyl-tRNAs are brought to the ribosome, EF-G, another GTP-binding protein, catalyzes the translocation of tRNAs from the A site to the P site and finally to the E site of the ribosome. This allows the ribosome and the mRNA to remain attached during translation. Elongation factor 2 (eEF-2) is a 100-kDa protein that catalyzes the ribosomal translocation reaction, resulting in the movement of ribosomes along mRNA. eEF-2 is the target for a very specific Ca²⁺/calmodulin-dependent eEF-2 kinase. Phosphorylation of eEF-2 makes it inactive in translation, which suggests that protein synthesis can be regulated by Ca²⁺ through eEF-2 phosphorylation. eEF-2 phosphorylation therefore regulates the cell-cycle and other processes where changes of intracellular Ca²⁺ concentration induce a new physiological state of a cell. The main role of eEF-2 phosphorylation in these processes is temporary inhibition of overall translation in response to transient elevation of the Ca²⁺ concentrations in the cytoplasm. Temporary inhibition of translation may trigger the transition of a cell from one physiologic state into another because of the disappearance of short-lived repressors and thus the activation of expression of new genes (Ryazanov and Spirin (1990) New Biol. 2:843-850).

Other ribosomal proteins which modulate translation of mRNA include the retinoblastoma protein (Rb1), HIV-1 TAR RNA binding protein (TARBP-b), v-fos transformation effector protein (Fte-1), the colin carcinoma laminin-binding protein, the Wilm’s tumor-related protein (QM), the ribosomal phosphoproteins P0, P1, and P2, ubiquitin, and the Epstein-Barr virus small RNAs-associated protein (EAP).

**Translation Termination**

The release factor eRF carries out termination of translation. eRF recognizes stop codons in the mRNA, leading to the release of the polypeptide chain from the ribosome.

The discovery of new nucleic acid-associated proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of nucleic acid-associated proteins.

**SUMMARY OF THE INVENTION**

The invention features purified polypeptides, nucleic acid-associated proteins, referred to collectively as “NAAP” and individually as “NAAP-1,” “NAAP-2,” “NAAP-3,” “NAAP-4,” “NAAP-5,” “NAAP-6,” “NAAP-7,” “NAAP-8,” “NAAP-9,” and “NAAP-10.” In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an
amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide
comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid
sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of
a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10,
and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the
group consisting of SEQ ID NO:1-10. In one alternative, the invention provides an isolated
polypeptide comprising the amino acid sequence of SEQ ID NO:1-10.

The invention further provides an isolated polynucleotide encoding a polypeptide selected
from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the
group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid
sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ
ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence
selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a
polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.
In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of
SEQ ID NO:1-10. In another alternative, the polynucleotide is selected from the group consisting of
SEQ ID NO:11-20.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter
sequence operably linked to a polynucleotide encoding a polypeptide selected from the group
consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting
of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least
90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a
biologically active fragment of a polypeptide having an amino acid sequence selected from the group
consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino
acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the
invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the
invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group
consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting
of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least
90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a
biologically active fragment of a polypeptide having an amino acid sequence selected from the group
consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino
acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a)
culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is
transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at
least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional NAAP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional NAAP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid
sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional NAAP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, the method
comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

**BRIEF DESCRIPTION OF THE TABLES**

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for
analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to “a host cell” includes a plurality of such host cells, and a reference to “an antibody” is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

“NAAP” refers to the amino acid sequences of substantially purified NAAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term “agonist” refers to a molecule which intensifies or mimics the biological activity of
NAAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NAAP either by directly interacting with NAAP or by acting on components of the biological pathway in which NAAP participates.

An "allelic variant" is an alternative form of the gene encoding NAAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding NAAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as NAAP or a polypeptide with at least one functional characteristic of NAAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding NAAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding NAAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent NAAP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of NAAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine.

Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity
of NAAP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NAAP either by directly interacting with NAAP or by acting on components of the biological pathway in which NAAP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')2, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind NAAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules.

The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH2), which may improve a desired property, e.g., resistance to nuclease or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system.

Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-
handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic NAAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding NAAP or fragments of NAAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5’ and/or the 3’ direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison...
WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

“Conservative amino acid substitutions” are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

<table>
<thead>
<tr>
<th>Original Residue</th>
<th>Conservative Substitution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>Gly, Ser</td>
</tr>
<tr>
<td>Arg</td>
<td>His, Lys</td>
</tr>
<tr>
<td>Asn</td>
<td>Asp, Gln, His</td>
</tr>
<tr>
<td>Asp</td>
<td>Asn, Glu</td>
</tr>
<tr>
<td>Cys</td>
<td>Ala, Ser</td>
</tr>
<tr>
<td>Gln</td>
<td>Asn, Glu, His</td>
</tr>
<tr>
<td>Glu</td>
<td>Asp, Gln, His</td>
</tr>
<tr>
<td>Gly</td>
<td>Ala</td>
</tr>
<tr>
<td>His</td>
<td>Asn, Arg, Gln, Glu</td>
</tr>
<tr>
<td>Ile</td>
<td>Leu, Val</td>
</tr>
<tr>
<td>Leu</td>
<td>Ile, Val</td>
</tr>
<tr>
<td>Lys</td>
<td>Arg, Gln, Glu</td>
</tr>
<tr>
<td>Met</td>
<td>Leu, Ile</td>
</tr>
<tr>
<td>Phe</td>
<td>His, Met, Leu, Trp, Tyr</td>
</tr>
<tr>
<td>Ser</td>
<td>Cys, Thr</td>
</tr>
<tr>
<td>Thr</td>
<td>Ser, Val</td>
</tr>
<tr>
<td>Trp</td>
<td>Phe, Tyr</td>
</tr>
<tr>
<td>Tyr</td>
<td>His, Phe, Trp</td>
</tr>
<tr>
<td>Val</td>
<td>Ile, Leu, Thr</td>
</tr>
</tbody>
</table>

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A “deletion” refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term “derivative” refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A “detectable label” refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.
“Differential expression” refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

“Exon shuffling” refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassertion of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A “fragment” is a unique portion of NAAP or the polynucleotide encoding NAAP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/aminio acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:11-20 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:11-20, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:11-20 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:11-20 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:11-20 and the region of SEQ ID NO:11-20 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-10 is encoded by a fragment of SEQ ID NO:11-20. A fragment of SEQ ID NO:1-10 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-10. For example, a fragment of SEQ ID NO:1-10 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-10. The precise length of a fragment of SEQ ID NO:1-10 and the region of SEQ ID NO:1-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A “full length” polynucleotide sequence is one containing at least a translation initiation
codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A “full length” polynucleotide sequence encodes a “full length” polypeptide sequence.

“Homology” refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktupel=2, gap penalty=5, window=4, and “diagonals saved”=4. The “weighted” residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including “blastn,” that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called “BLAST 2 Sequences” that is used for direct pairwise comparison of two nucleotide sequences. “BLAST 2 Sequences” can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The “BLAST 2 Sequences” tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

- **Matrix:** BLOSUM62
- **Reward for match:** 1
- **Penalty for mismatch:** -2
- **Open Gap:** 5 and **Extension Gap:** 2 penalties


Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases “percent identity” and “% identity,” as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and “diagonals saved”=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50
Expect: 10
Word Size: 3
Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.

Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press,
High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term “hybridization complex” refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., Cq,t or Rq,t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words “insertion” and “addition” refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

“Immune response” can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An “immunogenic fragment” is a polypeptide or oligopeptide fragment of NAAP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term “immunogenic fragment” also includes any polypeptide or oligopeptide fragment of NAAP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term “microarray” refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms “element” and “array element” refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term “modulate” refers to a change in the activity of NAAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other
biological, functional, or immunological properties of NAAP.

The phrases “nucleic acid” and “nucleic acid sequence” refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

“Peptide nucleic acid” (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

“Post-translational modification” of an NAAP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of NAAP.

“Probe” refers to nucleic acid sequences encoding NAAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. “Primers” are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter.
sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A “regulatory element” refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

“Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An “RNA equivalent,” in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term “sample” is used in its broadest sense. A sample suspected of containing NAAP, nucleic acids encoding NAAP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms “specific binding” and “specifically binding” refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope “A,” the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term “substantially purified” refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A “substitution” refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

“Substrate” refers to any suitable rigid or semi-rigid support including membranes, filters,
chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A “transcript image” or “expression profile” refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

“Transformation” describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroproporation, heat shock, lipofection, and particle bombardment. The term “transformed cells” includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A “transgenic organism,” as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A “variant” of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an “allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have
significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass “single nucleotide polymorphisms” (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A “variant” of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human nucleic acid-associated proteins (NAAP), the polynucleotides encoding NAAP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the
nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are nucleic acid-associated proteins. For example, SEQ ID NO:1 is 72% identical, from residue W321 to residue T721, to chicken DNA topoisomerase I (GenBank ID g1786132) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.3e-225, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a eukaryotic DNA topoisomerase I domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a DNA topoisomerase. In an alternative example, SEQ ID NO:2 is 64% identical, from residue G187 to residue P610, to human topoisomerase-related function protein (GenBank ID g5565687) as determined by BLAST, with a probability score of 7.9e-145. (See Table 2.) SEQ ID NO:2 also contains a nucleotide/transferase domain as determined by searching for statistically significant matches in the HMM-based PFAM database. (See Table 3.) Data from MOTIFS and BLAST analyses provide further corroborative evidence that SEQ ID NO:2 is a topoisomerase. In an alternative example, SEQ ID NO:3 is 71% identical, from residue W159 to residue F718, to chicken DNA topoisomerase I (GenBank ID g1786132) as determined by BLAST, with a probability score of 1.0e-231. SEQ ID NO:3 is also 71% identical, from residue W159 to residue F718, to human DNA topoisomerase I (GenBank IDs g339804 and g339806) as determined by BLAST, with probability scores of 9.1e-231 and 9.1e-231, respectively. SEQ ID NO:3 is also 70% and 69% identical, from residue W159 to residue F718, to Chinese hamster and mouse DNA.
topoisomerase I (GenBank IDs g297079 and g220618), respectively, as determined by BLAST analysis. The BLAST probability scores are 5.0e-230 and 4.5e-229, respectively. (See Table 2.)

SEQ ID NO:3 also contains a eukaryotic DNA topoisomerase I domain as determined by searching for statistically significant matches in the HMM-based PFAM database. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:3 is a eukaryotic DNA topoisomerase. In an alternative example, SEQ ID NO:5 is 70% identical, from residue M1 to residue M531, to human RCC1-like G exchanging factor RLG (GenBank ID g3789799) as determined by BLAST, with a probability score of 4.2e-211. (See Table 2.) SEQ ID NO:5 also contains a regulator of chromosome condensation (RCC1) domain and a BTB/POZ domain as determined by searching for statistically significant matches in the HMM-based PFAM database. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:5 is an RCC1 family protein. In an alternative example, SEQ ID NO:7 is 99% identical, from residue M48 to residue A860, and 94% identical over the entire length of the sequence, to Oreitogagus cuniculus translation initiation factor eIF2C (GenBank ID g3253159) as determined by BLAST, with a probability score of 0.0. (See Table 2.) SEQ ID NO:6 is 59% identical, from residue T2 to residue E340, to Arabidopsis thaliana putative translation initiation factor eIF-2B alpha subunit (GenBank ID g4006818) as determined by BLAST with a probability score of 9.93-93. SEQ ID NO:6 also contains an initiation factor 2 subunit family domain as determined by searching for statistically significant matches in the HMM-based PFAM database. (See Table 3.) Data from BLIMPS and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:6 and SEQ ID NO:7 are protein translation initiation factors.

In an alternative example, SEQ ID NO:8 is 38% identical, from residue K129 to residue K494, and 51% identical, from residue R13 to residue N168, to Arabidopsis thaliana ATP-dependent RNA helicase (GenBank ID g4895231) as determined by BLAST, with a probability score of 2.6e-91. (See Table 2.) SEQ ID NO:8 also contains a helicase conserved C-terminal domain and a DEAD/DEAH box helicase domain as determined by searching for statistically significant matches in the HMM-based PFAM database. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:8 is a DEAD-box subfamily ATP-dependent helicase. In an alternative example, SEQ ID NO:10 is 100% identical, from residue M172 to residue R212, and 95% identical from residue L285 to residue E865, to human topoisomerase I (GenBank ID g15919359) as determined by BLAST, with a probability score of 0.0. (See Table 2.) SEQ ID NO:10 also has homology to proteins that are localized to the nucleus, and are topoisomerases, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:10 also contains a eukaryotic DNA topoisomerase I catalytic core domain, as well as a eukaryotic DNA topoisomerase I DNA-binding domain as determined by searching for statistically significant matches in the HMM-based
PFAM database. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:10 is a topoisomerase. SEQ ID NO:4 and SEQ ID NO:9 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-10 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO.), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:11-20 or that distinguish between SEQ ID NO:11-20 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (i.e., those sequences including the designation “ENST”). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (i.e., those sequences including the designation “NM” or “NT”) or the NCBI RefSeq Protein Sequence Records (i.e., those sequences including the designation “NP”). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon stitching” algorithm. For example, a polynucleotide sequence identified as FL_XXXXX_N1_N2_YYYY_N3_N4 represents a “stitched” sequence in which XYYYY is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N1,2,3,…, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an “exon-stretching” algorithm. For example, a polynucleotide sequence identified as FLXXXXX_gAAAAA_gBBBBB_1_N is a “stretched” sequence, with XYYYY being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the “exon-stretching” algorithm was applied, gBBBBB being the
GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and \( N \) referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (i.e., gBBBAAA).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Type of analysis and/or examples of programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>GNN, GFG, ENST</td>
<td>Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).</td>
</tr>
<tr>
<td>GBI</td>
<td>Hand-edited analysis of genomic sequences.</td>
</tr>
<tr>
<td>FL</td>
<td>Stitched or stretched genomic sequences (see Example V).</td>
</tr>
<tr>
<td>INCY</td>
<td>Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.</td>
</tr>
</tbody>
</table>

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses NAAP variants. A preferred NAAP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the NAAP amino acid sequence, and which contains at least one functional or structural characteristic of NAAP.

The invention also encompasses polynucleotides which encode NAAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20, which encodes NAAP. The polynucleotide sequences of SEQ ID NO:11-20, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the
sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding NAAP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding NAAP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:11-20. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of NAAP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding NAAP. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding NAAP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding NAAP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding NAAP. For example, a polynucleotide comprising a sequence of SEQ ID NO:11, a polynucleotide comprising a sequence of SEQ ID NO:13, and a polynucleotide comprising a sequence of SEQ ID NO:20, are all splice variants of each other. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of NAAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding NAAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring NAAP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode NAAP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring NAAP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding NAAP or
its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding NAAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode NAAP and NAAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding NAAP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:11-20 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in “Definitions.”

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klencow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding NAAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed,
restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Appl. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Appl. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as Oligo 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5’ regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5’ non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polymers of nucleotide sequences or fragments thereof which encode NAAP may be cloned in recombinant DNA molecules that direct expression of NAAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent
degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express NAAP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter NAAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of NAAP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding NAAP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, NAAP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of NAAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or
a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active NAAP, the nucleotide sequences encoding NAAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5’ and 3’ untranslated regions in the vector and in polynucleotide sequences encoding NAAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding NAAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding NAAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding NAAP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding NAAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Hecke, G. and S.M. Schuster

The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding NAAP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding NAAP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding NAAP into the vector’s multiple cloning site disrupts the lacZ gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of NAAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of NAAP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of NAAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of NAAP. Transcription of sequences encoding NAAP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al.

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding NAAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses NAAP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of NAAP in cell lines is preferred. For example, sequences encoding NAAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in tk and apr cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, dhfr confers resistance to methotrexate; neo confers resistance to the aminoglycosides neomycin and G-418; and als and pat confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., trpB and hisD, which
alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.) Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding NAAP is inserted within a marker gene sequence, transformed cells containing sequences encoding NAAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding NAAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding NAAP and that express NAAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of NAAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on NAAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding NAAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding NAAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase.

54
such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding NAAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode NAAP may be designed to contain signal sequences which direct secretion of NAAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding NAAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric NAAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of NAAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, c-myc, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, c-myc, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the NAAP encoding sequence and the heterologous protein sequence, so that NAAP may be cleaved away from the heterologous moiety following purification.
Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled NAAP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, $^{35}$S-methionine.

NAAP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to NAAP. At least one and up to a plurality of test compounds may be screened for specific binding to NAAP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of NAAP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which NAAP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express NAAP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing NAAP or cell membrane fractions which contain NAAP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either NAAP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with NAAP, either in solution or affixed to a solid support, and detecting the binding of NAAP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

NAAP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of NAAP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for NAAP activity, wherein NAAP is combined with at least one test compound, and the activity of NAAP in the
presence of a test compound is compared with the activity of NAAP in the absence of the test compound. A change in the activity of NAAP in the presence of the test compound is indicative of a compound that modulates the activity of NAAP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising NAAP under conditions suitable for NAAP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of NAAP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding NAAP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding NAAP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding NAAP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding NAAP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress NAAP, e.g., by secreting NAAP in its milk, may also

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of NAAP and nucleic acid-associated proteins. In addition, examples of tissues expressing NAAP are brain tissue, brain tumor tissue, and normal eosinophils, and also can be found in Table 6. Therefore, NAAP appears to play a role in cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections. In the treatment of disorders associated with increased NAAP expression or activity, it is desirable to decrease the expression or activity of NAAP. In the treatment of disorders associated with decreased NAAP expression or activity, it is desirable to increase the expression or activity of NAAP.

Therefore, in one embodiment, NAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocytopenia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer’s disease, Pick’s disease, Huntington’s disease, dementia, Parkinson’s disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorder of the central nervous system, cerebral palsy, a neuroskeletal disorder, an autonomic nervous system disorder, a cranial nerve disorder, a spinal cord disease, muscular dystrophy and other neuromuscular disorder, a peripheral nervous system disorder, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathy, myasthenia gravis, periodic paralysis, a mental disorder including mood, anxiety, and schizophrenic disorder, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive
dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette’s disorder; a
developmental disorder such as renal tubular acidosis, anemia, Cushing’s syndrome, achondroplastic
dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR
syndrome (Wilms’ tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-
Magenis syndrome, myelodysplastic syndrome, hereditary mucopolithelial dysplasia, hereditary
keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis,
hypothyroidism, hydrocephalus, seizure disorders such as Sydenham’s chorea and cerebral palsy,
spina bifida, anencephaly, craniorachischis, congenital glaucoma, cataract, and sensorineural
hearing loss; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome
(AIDS), Addison’s disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis,
amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis,
autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis,
cholecystitis, contact dermatitis, Crohn’s disease, atopic dermatitis, dermatomyositis, diabetes
mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis,
erthema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture’s syndrome, gout, Graves’
disease, Hashimoto’s thyroiditis, hyperesinophilia, irritable bowel syndrome, multiple sclerosis,
myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis,
polymyositis, psoriasis, Reiter’s syndrome, rheumatoid arthritis, scleroderma, Sjögren’s syndrome,
 systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura,
ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and
extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and
trauma; and an infection, such as those caused by a viral agent classified as adenovirus, arenavirus,
bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus,
parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, or
togavirus; an infection caused by a bacterial agent classified as pneumococcus, staphylococcus,
streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria,
moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including
shigella, salmonella, or campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia,
bartonella, norcardium, actinomyces, mycobacterium, spirochaetae, rickettsia, chlamydia, or
mycoplasma; an infection caused by a fungal agent classified as aspergillus, blastomyces,
dermatophytes, cryptococcus, coccidioides, malasezzia, histoplasma, or other mycosis-causing fungal
agent; and an infection caused by a parasite classified as plasmodium or malaria-causing, parasitic
entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as
giardia, trichomonas, tissue nematode such as trichinella, intestinal nematode such as ascaris,
lymphatic filarial nematode, trematode such as schistosoma, and cestode such as tapeworm.
In another embodiment, a vector capable of expressing NAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified NAAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of NAAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of NAAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NAAP. Examples of such disorders include, but are not limited to, those cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections, described above. In one aspect, an antibody which specifically binds NAAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express NAAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding NAAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NAAP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of NAAP may be produced using methods which are generally known in the art. In particular, purified NAAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind NAAP. Antibodies to NAAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muylerdemans, S. (2001) J.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with NAAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lyssolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to NAAP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of NAAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.


Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for NAAP may also be generated.
For example, such fragments include, but are not limited to, F(ab')$_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.) Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between NAAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering NAAP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for NAAP. Affinity is expressed as an association constant, $K_a$, which is defined as the molar concentration of NAAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The $K_a$, determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple NAAP epitopes, represents the average affinity, or avidity, of the antibodies for NAAP. The $K_a$ determined for a preparation of monoclonal antibodies, which are monospecific for a particular NAAP epitope, represents a true measure of affinity. High-affinity antibody preparations with $K_a$ ranging from about $10^8$ to $10^{12}$ L/mole are preferred for use in immunoassays in which the NAAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with $K_a$ ranging from about $10^5$ to $10^7$ L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of NAAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of NAAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding NAAP, or any
fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding NAAP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding NAAP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)


case where a genetic deficiency in NAAP expression or regulation causes disease, the expression of NAAP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.


Expression vectors that may be effective for the expression of NAAP include, but are not limited to, the PCDNA 3.1, EPTTAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). NAAP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gosser, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossea, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PINZ; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding NAAP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to NAAP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding NAAP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding NAAP to cells which have one or more genetic abnormalities with respect to the expression of NAAP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinolzzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding NAAP to target cells which have one or more genetic abnormalities with respect to the expression of NAAP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing NAAP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S.
Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding NAAP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for NAAP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of NAAP-coding RNAs and the synthesis of high levels of NAAP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of NAAP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using
triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding NAAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis.

Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding NAAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2'-O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding NAAP.

Compounds which may be effective in altering expression of a specific polynucleotide may include,
but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased NAAP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding NAAP may be therapeutically useful, and in the treatment of disorders associated with decreased NAAP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding NAAP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding NAAP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding NAAP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding NAAP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable
for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of NAAP, antibodies to NAAP, and mimetics, agonists, antagonists, or inhibitors of NAAP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising NAAP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, NAAP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example NAAP or fragments thereof, antibodies of NAAP, and agonists, antagonists or inhibitors of NAAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED$_{50}$ (the dose therapeutically effective in 50% of the population) or LD$_{50}$ (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD$_{50}$/ED$_{50}$ ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED$_{50}$ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 $\mu$g to 100,000 $\mu$g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

**DIAGNOSTICS**

In another embodiment, antibodies which specifically bind NAAP may be used for the diagnosis of disorders characterized by expression of NAAP, or in assays to monitor patients being treated with NAAP or agonists, antagonists, or inhibitors of NAAP. Antibodies useful for diagnostic
purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays
for NAAP include methods which utilize the antibody and a label to detect NAAP in human body
fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and
may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of
reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring NAAP, including ELISAs, RIAs, and FACS, are known
in the art and provide a basis for diagnosing altered or abnormal levels of NAAP expression. Normal
or standard values for NAAP expression are established by combining body fluids or cell extracts
taken from normal mammalian subjects, for example, human subjects, with antibodies to NAAP
under conditions suitable for complex formation. The amount of standard complex formation may be
quantitated by various methods, such as photometric means. Quantities of NAAP expressed in
subject, control, and disease samples from biopsied tissues are compared with the standard values.
Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding NAAP may be used
for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences,
complementary RNA and DNA molecules, and PNA. The polynucleotides may be used to detect
and quantify gene expression in biopsied tissues in which expression of NAAP may be correlated
with disease. The diagnostic assay may be used to determine absence, presence, and excess
expression of NAAP, and to monitor regulation of NAAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide
sequences, including genomic sequences, encoding NAAP or closely related molecules may be used
to identify nucleic acid sequences which encode NAAP. The specificity of the probe, whether it is
made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a
conserved motif, and the stringency of the hybridization or amplification will determine whether the
probe identifies only naturally occurring sequences encoding NAAP, allelic variants, or related
sequences.

Probes may also be used for the detection of related sequences, and may have at least 50%
sequence identity to any of the NAAP encoding sequences. The hybridization probes of the subject
invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:11-20 or from
genomic sequences including promoters, enhancers, and introns of the NAAP gene.

Means for producing specific hybridization probes for DNAs encoding NAAP include the
cloning of polynucleotide sequences encoding NAAP or NAAP derivatives into vectors for the
production of mRNA probes. Such vectors are known in the art, are commercially available, and may
be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA
polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a
variety of reporter groups, for example, by radionuclides such as $^{32}$P or $^{35}$S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding NAAP may be used for the diagnosis of disorders associated with expression of NAAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocytopenia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebellarretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorder of the central nervous system, cerebral palsy, a neuroskeletal disorder, an autonomic nervous system disorder, a cranial nerve disorder, a spinal cord disease, muscular dystrophy and other neuromuscular disorder, a peripheral nervous system disorder, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathy, myasthenia gravis, periodic paralysis, a mental disorder including mood, anxiety, and schizophrenic disorder, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratoderma, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an autoimmune/inflammatory disorder such as acquired
immunodeficiency syndrome (AIDS), Addison’s disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn’s disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture’s syndrome, gout, Graves’ disease, Hashimoto’s thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter’s syndrome, rheumatoid arthritis, scleroderma, Sjögren’s syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and an infection, such as those caused by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, or togavirus; an infection caused by a bacterial agent classified as pneumococcus, staphylococcus, streptococcus, bacillus, Corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, or campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, or mycoplasma; an infection caused by a fungal agent classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malasezia, histoplasma, or other mycosis-causing fungal agent; and an infection caused by a parasite classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematode such as trichinella, intestinal nematode such as ascaris, lymphatic filarial nematode, trematode such as schistosoma, and cestode such as tapeworm. The polynucleotide sequences encoding NAAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered NAAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding NAAP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding NAAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a
suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding NAAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of NAAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding NAAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding NAAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding NAAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding NAAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding NAAP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are
substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic
disease in humans. Methods of SNP detection include, but are not limited to, single-stranded
conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP,
oligonucleotide primers derived from the polynucleotide sequences encoding NAAP are used to
amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example,
from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause
differences in the secondary and tertiary structures of PCR products in single-stranded form, and
these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the
oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-
throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis
methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the
sequence of individual overlapping DNA fragments which assemble into a common consensus
sequence. These computer-based methods filter out sequence variations due to laboratory preparation
of DNA and sequencing errors using statistical models and automated analyses of DNA sequence
chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry
using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16
common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also
useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis,
sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding
lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic
fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that
influence a patient’s response to a drug, such as life-threatening toxicity. For example, a variation in
N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the
anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in
diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase
pathway. Analysis of the distribution of SNPs in different populations is useful for investigating
genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations

Methods which may also be used to quantify the expression of NAAP include radiolabeling
or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from
et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be
accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of
interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, NAAP, fragments of NAAP, or antibodies specific for NAAP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., “Comparative Gene Transcript Analysis,” U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed
molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell’s proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical
density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for NAAP to quantify the levels of NAAP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoza, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are
incubated with antibodies specific to the polypeptides of the present invention. The amount of
protein recognized by the antibodies is quantified. The amount of protein in the treated biological
sample is compared with the amount in an untreated biological sample. A difference in the amount of
protein between the two samples is indicative of a toxic response to the test compound in the treated
sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g.,
USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al.
2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are
well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed.

In another embodiment of the invention, nucleic acid sequences encoding NAAP may be used
to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either
coding or noncoding sequences may be used, and in some instances, noncoding sequences may be
preferable over coding sequences. For example, conservation of a coding sequence among members
of a multi-gene family may potentially cause undesired cross hybridization during chromosomal
mapping. The sequences may be mapped to a particular chromosome, to a specific region of a
chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs),
yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial PI
constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat.
7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop
genetic linkage maps, for example, which correlate the inheritance of a disease state with the
inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP).
(See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic
map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic
map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man
(OMIM) World Wide Web site. Correlation between the location of the gene encoding NAAP on a
physical map and a specific disorder, or a predisposition to a specific disorder, may help define the
region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as
linkage analysis using established chromosomal markers, may be used for extending genetic maps.
Often the placement of a gene on the chromosome of another mammalian species, such as mouse,
may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, NAAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between NAAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with NAAP, or fragments thereof, and washed. Bound NAAP is then detected by methods well known in the art. Purified NAAP can also be coated directly onto plates for use in the aforementioned drug screening techniques.

Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding NAAP specifically compete with a test compound for binding NAAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with NAAP.

In additional embodiments, the nucleotide sequences which encode NAAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/270,858, U.S. Ser. No. 60/274,071, U.S. Ser. No. 60/283,496, U.S. Ser.
EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones
Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae.
Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); hidden
Markov model (HMM)-based protein family databases such as PFAM; and HMM-based protein
domain databases such as SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864;
Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which
Oin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST,
FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length
polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences,
stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to
extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on
Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using
programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were
translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide
of the invention may begin at any of the methionine residues of the full length translated polypeptide.

Full length polypeptide sequences were subsequently analyzed by querying against databases such as
the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS,
PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases
such as PFAM; and HMM-based protein domain databases such as SMART. Full length
polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software
Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide
and polypeptide sequence alignments are generated using default parameters specified by the
CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program
(DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of
Incyte cDNA and full length sequences and provides applicable descriptions, references, and
threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used,
the second column provides brief descriptions thereof, the third column presents appropriate
references, all of which are incorporated by reference herein in their entirety, and the fourth column
presents, where applicable, the scores, probability values, and other parameters used to evaluate the
strength of a match between two sequences (the higher the score or the lower the probability value,
the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide
and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ
ID NO:11-20. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization
and amplification technologies are described in Table 4, column 2.
IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative nucleic acid-associated proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode nucleic acid-associated proteins, the encoded polypeptides were analyzed by querying against PFAM models for nucleic acid-associated proteins. Potential nucleic acid-associated proteins were also identified by homology to Incyte cDNA sequences that had been annotated as nucleic acid-associated proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated
but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then “stitched” together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpr public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore “stretched” or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of NAAP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:11-20 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:11-20 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome’s p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between
chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Génethon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:15 was mapped to the X chromosome within the interval from 138.8 to 198.1 centiMorgans, and SEQ ID NO:17 was mapped to chromosome 8 within the interval from 152.50 centiMorgans to the q terminus.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESeq (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

\[
\text{BLAST Score} \times \text{Percent Identity} = 5 \times \text{minimum (length(Seq. 1), length(Seq. 2))}
\]

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the
other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding NAAP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding NAAP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of NAAP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5'extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLG0 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 pmol of each primer, reaction buffer containing Mg^2+, (NH_4)_2SO_4, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the
alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1% agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviII cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in NAAP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were
identified in SEQ ID NO:11-20 using the LIFESSEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:11-20 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of $[^{32}P]$ adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing $10^7$ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases:

Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).
The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)^+ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)^+ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse
transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

**Microarray Preparation**

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water.

Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

**Hybridization**

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample
mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

**Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-83SH analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and
measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore’s emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XII. Complementary Polynucleotides

Sequences complementary to the NAAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring NAAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of NAAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5’ sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the NAAP-encoding transcript.

XIII. Expression of NAAP

Expression and purification of NAAP is achieved using bacterial or virus-based expression systems. For expression of NAAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the trp-lac (tac) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the lac operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express NAAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of NAAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding NAAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, NAAP is synthesized as a fusion protein with, e.g., glutathione
S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from NAAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified NAAP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

XIV. Functional Assays

NAAP function is assessed by expressing the sequences encoding NAAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of NAAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding NAAP and either CD64 or CD64-GFP. CD64 and
CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art.

Expression of mRNA encoding NAAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of NAAP Specific Antibodies

NAAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the NAAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-NAAP activity by, for example, binding the peptide or NAAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring NAAP Using Specific Antibodies

Naturally occurring or recombinant NAAP is substantially purified by immunoaffinity chromatography using antibodies specific for NAAP. An immunoaffinity column is constructed by covalently coupling anti-NAAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing NAAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of NAAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/NAAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and NAAP is collected.

XVII. Identification of Molecules Which Interact with NAAP
NAAP, or biologically active fragments thereof, are labeled with $^{125}$I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled NAAP, washed, and any wells with labeled NAAP complex are assayed. Data obtained using different concentrations of NAAP are used to calculate values for the number, affinity, and association of NAAP with the candidate molecules.

Alternatively, molecules interacting with NAAP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

NAAP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

**XVIII. Demonstration of NAAP Activity**

NAAP activity is measured by its ability to stimulate transcription of a reporter gene (Liu, H.Y. et al. (1997) EMBO J. 16:5289-5298). The assay entails the use of a well characterized reporter gene construct, LexAop-LacZ, that consists of LexA DNA transcriptional control elements (LexAop) fused to sequences encoding the E. coli LacZ enzyme. The methods for constructing and expressing fusion genes, introducing them into cells, and measuring LacZ enzyme activity, are well known to those skilled in the art. Sequences encoding NAAP are cloned into a plasmid that directs the synthesis of a fusion protein, LexA-NAAP, consisting of NAAP and a DNA binding domain derived from the LexA transcription factor. The resulting plasmid, encoding a LexA-NAAP fusion protein, is introduced into yeast cells along with a plasmid containing the LexAop-LacZ reporter gene. The amount of LacZ enzyme activity associated with LexA-NAAP transfected cells, relative to control cells, is proportional to the amount of transcription stimulated by the NAAP.

Alternatively, NAAP activity is measured by its ability to bind zinc. A 5-10 μM sample solution in 2.5 mM ammonium acetate solution at pH 7.4 is combined with 0.05 M zinc sulfate solution (Aldrich, Milwaukee WI) in the presence of 100 μM dithiothreitol with 10% methanol added. The sample and zinc sulfate solutions are allowed to incubate for 20 minutes. The reaction solution is passed through a Vydac column (Grace Vydac, Hesperia, CA) with approximately 300 Angstrom bore size and 5 μM particle size to isolate zinc-sample complex from the solution, and into a mass spectrometer (PE Sciex, Ontario, Canada). Zinc bound to sample is quantified using the functional atomic mass of 63.5 Da observed by Whittal, R. M. et al. ((2000) Biochemistry 39:8406-8417).

In the alternative, a method to determine nucleic acid binding activity of NAAP involves a
polyacrylamide gel mobility-shift assay. In preparation for this assay, NAAP is expressed by
transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector
containing NAAP cDNA. The cells are incubated for 48-72 hours after transformation under
conditions appropriate for the cell line to allow expression and accumulation of NAAP. Extracts
containing solubilized proteins can be prepared from cells expressing NAAP by methods well known
in the art. Portions of the extract containing NAAP are added to [32P]-labeled RNA or DNA.
Radioactive nucleic acid can be synthesized in vitro by techniques well known in the art. The
mixtures are incubated at 25°C in the presence of RNase- and DNase-inhibitors under buffered
conditions for 5-10 minutes. After incubation, the samples are analyzed by polyacrylamide gel
electrophoresis followed by autoradiography. The presence of a band on the autoradiogram indicates
the formation of a complex between NAAP and the radioactive transcript. A band of similar mobility
will not be present in samples prepared using control extracts prepared from untransformed cells.

In the alternative, a method to determine methylase activity of NAAP measures transfer of
radiolabeled methyl groups between a donor substrate and an acceptor substrate. Reaction mixtures
(50 μl final volume) contain 15 mM HEPES, pH 7.9, 1.5 mM MgCl2, 10 mM dithiothreitol, 3%
polyvinylalcohol, 1.5 μCi [methyl-3H]AdoMet (0.375 μM AdoMet) (DuPont-NEN), 0.6 μg NAAP,
and acceptor substrate (e.g., 0.4 μg [35S]RNA, or 6-mercaptopurine (6-MP) to 1 mM final
concentration). Reaction mixtures are incubated at 30°C for 30 minutes, then 65°C for 5 minutes.

Analysis of [methyl-3H]RNA is as follows: (1) 50 μl of 2 x loading buffer (20 mM Tris-HCl,
pH 7.6, 1 M LiCl, 1 mM EDTA, 1% sodium dodecyl sulphate (SDS)) and 50 μl oligo d(T)-cellulose
(10 mg/ml in 1 x loading buffer) are added to the reaction mixture, and incubated at ambient
temperature with shaking for 30 minutes. (2) Reaction mixtures are transferred to a 96-well filtration
plate attached to a vacuum apparatus. (3) Each sample is washed sequentially with three 2.4 ml
aliquots of 1 x oligo d(T) loading buffer containing 0.5% SDS, 0.1% SDS, or no SDS. (4) RNA is
eluted with 300 μl of water into a 96-well collection plate, transferred to scintillation vials containing
liquid scintillant, and radioactivity determined.

Analysis of [methyl-3H]6-MP is as follows: (1) 500 μl 0.5 M borate buffer, pH 10.0, and then
2.5 ml of 20% (v/v) isoamyl alcohol in toluene are added to the reaction mixtures. (2) The samples
are mixed by vigorous vortexing for ten seconds. (3) After centrifugation at 700g for 10 minutes, 1.5
ml of the organic phase is transferred to scintillation vials containing 0.5 ml absolute ethanol and
liquid scintillant, and radioactivity determined. (4) Results are corrected for the extraction of 6-MP
into the organic phase (approximately 41%).

In the alternative, type I topoisomerase activity of NAAP can be assayed based on the
relaxation of a supercoiled DNA substrate. NAAP is incubated with its substrate in a buffer lacking
Mg2+ and ATP, the reaction is terminated, and the products are loaded on an agarose gel. Altered
topoisomers can be distinguished from supercoiled substrate electrophoretically. This assay is specific for type I topoisomerase activity because Mg$^{2+}$ and ATP are necessary cofactors for type II topoisomerases.

In the alternative, Type II topoisomerase activity of NAAP can be assayed based on the decatenation of a kinetoplast DNA (KDNA) substrate. NAAP is incubated with KDNA, the reaction is terminated, and the products are loaded on an agarose gel. Monomeric circular KDNA can be distinguished from catenated KDNA electrophoretically. Kits for measuring type I and type II topoisomerase activities are available commercially from Topogen (Columbus OH).

In the alternative, ATP-dependent RNA helicase unwinding activity of NAAP can be measured by the method described by Zhang and Grosse (1994; Biochemistry 33:3906-3912). The substrate for RNA unwinding consists of $^{32}$P-labeled RNA composed of two RNA strands of 194 and 130 nucleotides in length containing a duplex region of 17 base-pairs. The RNA substrate is incubated together with ATP, Mg$^{2+}$, and varying amounts of NAAP in a Tris-HCl buffer, pH 7.5, at 37°C for 30 minutes. The single-stranded RNA product is then separated from the double-stranded RNA substrate by electrophoresis through a 10% SDS-polyacrylamide gel, and quantitated by autoradiography. The amount of single-stranded RNA recovered is proportional to the amount of NAAP in the preparation.

In the alternative, NAAP function is assessed by expressing the sequences encoding NAAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen Corporation, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected.

Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; CLONTECH), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties.

FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as
measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow

Cytoometry, Oxford, New York NY.

The influence of NAAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding NAAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Inc., Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding NAAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

Pseudouridine synthase activity of NAAP is assayed using a tritium (^3H) release assay

modified from Nurse et al. (1995) RNA 1:102-112), which measures the release of ^3H from the C5 position of the pyrimidine component of uridylic (U) when ^3H-radiolabeled U in RNA is isomerized to pseudouridine (Ψ). A typical 500 μl assay mixture contains 50 mM HEPES buffer (pH 7.5), 100 mM ammonium acetate, 5 mM dithiothreitol, 1 mM EDTA, 30 units RNase inhibitor, and 0.1-4.2 μM [5-^3H]RNA (approximately 1 μCi/nmol tRNA). The reaction is initiated by the addition of <5 μl of a concentrated solution of NAAP (or sample containing NAAP) and incubated for 5 min at 37 °C.

Portions of the reaction mixture are removed at various times (up to 30 min) following the addition of NAAP and quenched by dilution into 1 ml 0.1 M HCl containing Norit-SA3 (12% w/v). The quenched reaction mixtures are centrifuged for 5 min at maximum speed in a microcentrifuge, and the supernatants are filtered through a plug of glass wool. The pellet is washed twice by resuspension in 1 ml 0.1 M HCl, followed by centrifugation. The supernatants from the washes are separately passed through the glass wool plug and combined with the original filtrate. A portion of the combined filtrate is mixed with scintillation fluid (up to 10 ml) and counted using a scintillation counter. The amount of ^3H released from the RNA and present in the soluble filtrate is proportional to the amount of pseudouridine synthase activity in the sample (Ramamurthy, V. (1999) J. Biol. Chem. 274:22225-22230).

In the alternative, pseudouridine synthase activity of NAAP is assayed at 30 °C to 37 °C in a mixture containing 100 mM Tris-HCl (pH 8.0), 100 mM ammonium acetate, 5 mM MgCl₂, 2 mM dithiothreitol, 0.1 mM EDTA, and 1-2 fmol of [³²P]-radiolabeled runoff transcripts (generated in vitro by an appropriate RNA polymerase, i.e., T7 or SP6) as substrates. NAAP is added to initiate the reaction or omitted from the reaction in control samples. Following incubation, the RNA is extracted
with phenol-chloroform, precipitated in ethanol, and hydrolyzed completely to 3-nucleotide
monophosphates using RNase T2. The hydrolysates are analyzed by two-dimensional thin layer
chromatography, and the amount of 32P radiolabel present in the ψMP and UMP spots are evaluated
after exposing the thin layer chromatography plates to film or a PhosphorImager screen. Taking into
account the relative number of uridylic residues in the substrate RNA, the relative amount ψMP and
UMP are determined and used to calculate the relative amount of ψ per tRNA molecule (expressed in
mol ψ/mol of tRNA or mol ψ/min of tRNA/minute), which corresponds to the amount of
273:1316-1323).

N2,N2-dimethylguanosine transferase ((m2, G)methyltransferase) activity of NAAP is
measured in a 160 μl reaction mixture containing 100 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 10 mM
MgCl2, 20 mM NH4Cl, 1mM dithiothreitol, 6.2 μM S-adenosyl-L-[methyl-3H]methionine (30-70
Ci/mM), 8 μg m2,G-deficient tRNA or wild type tRNA from yeast, and approximately 100 μg of
purified NAAP or a sample comprising NAAP. The reactions are incubated at 30 °C for 90 min and
chilled on ice. A portion of each reaction is diluted to 1 ml in water containing 100 μg BSA. 1 ml of
2 M HCl is added to each sample and the acid insoluble products are allowed to precipitate on ice for
20 min before being collected by filtration through glass fiber filters. The collected material is
washed several times with HCl and quantitated using a liquid scintillation counter. The amount of 3H
incorporated into the m2,G-deficient, acid-insoluble tRNAs is proportional to the amount of
N2,N2-dimethylguanosine transferase activity in the NAAP sample. Reactions comprising no
substrate tRNAs, or wild-type tRNAs that have already been modified, serve as control reactions
which should not yield acid-insoluble 3H-labeled products.

Polyadenylation activity of NAAP is measured using an in vitro polyadenylation reaction.
The reaction mixture is assembled on ice and comprises 10 μl of 5 mM dithiothreitol, 0.025% (v/v)
NONIDET P-40, 50 mM creatine phosphate, 6.5% (w/v) polyvinyl alcohol, 0.5 unit/μl RNAGUARD
(Pharmacia), 0.025 μg/μl creatine kinase, 1.25 mM cordycepin 5'-triphosphate, and 3.75 mM MgCl2,
in a total volume of 25 μl. 60 fmol of CstF, 50 fmol of CPSF, 240 fmol of PAP, 4 μl of crude or
partially purified CF II and various amounts of amounts CF I are then added to the reaction mix. The
volume is adjusted to 23.5 μl with a buffer containing 50 mM TrisHCl, pH 7.9, 10% (v/v) glycerol,
and 0.1 mM Na-EDTA. The final ammonium sulfate concentration should be below 20 mM. The
reaction is initiated (on ice) by the addition of 15 fmol of 32P-labeled pre-mRNA template, along with
2.5 μg of unlabeled tRNA, in 1.5 μl of water. Reactions are then incubated at 30 °C for 75-90 min
and stopped by the addition of 75 μl (approximately two-volumes) of proteinase K mix (0.2 M Tris-
HCl, pH 7.9, 300 mM NaCl, 25 mM Na-EDTA, 2% (w/v) SDS), 1 μl of 10 mg/ml proteinase K, 0.25
μl of 20 mg/ml glycogen, and 23.75 μl of water). Following incubation, the RNA is precipitated with
ethanol and analyzed on a 6% (w/v) polyacrylamide, 8.3 M urea sequencing gel. The dried gel is developed by autoradiography or using a phosphoimager. Cleavage activity is determined by comparing the amount of cleavage product to the amount of pre-mRNA template. The omission of any of the polypeptide components of the reaction and substitution of NAAP is useful for identifying the specific biological function of NAAP in pre-mRNA polyadenylation (Riegsegger, U. et al. (1996) J. Biol. Chem. 271:6107-6113; and references within).

tRNA synthetase activity is measured as the aminoaacylation of a substrate tRNA in the presence of [14C]-labeled amino acid. NAAP is incubated with [14C]-labeled amino acid and the appropriate cognate tRNA (for example, [14C]alanine and tRNA_{Ala}) in a buffered solution. 14C-labeled product is separated from free [14C]amino acid by chromatography, and the incorporated 14C is quantified by scintillation counter. The amount of 14C-labeled product detected is proportional to the activity of NAAP in this assay.

In the alternative, NAAP activity is measured by incubating a sample containing NAAP in a solution containing 1 mM ATP, 5 mM Hepes-KOH (pH 7.0), 2.5 mM KCl, 1.5 mM magnesium chloride, and 0.5 mM DTT along with misacylated [14C]-Glu-tRNA_{Gln} (e.g., 1 μM) and a similar concentration of unlabeled L-glutamine. Following the quenching of the reaction with 3 M sodium acetate (pH 5.0), the mixture is extracted with an equal volume of water-saturated phenol, and the aqueous and organic phases are separated by centrifugation at 15,000 × g at room temperature for 1 min. The aqueous phase is removed and precipitated with 3 volumes of ethanol at -70°C for 15 min. The precipitated aminoaacyl-tRNAs are recovered by centrifugation at 15,000 × g at 4°C for 15 min. The pellet is resuspended in of 25 mM KOH, deacylated at 65°C for 10 min., neutralized with 0.1 M HCl (to final pH 6-7), and dried under vacuum. The dried pellet is resuspended in water and spotted onto a cellulose TLC plate. The plate is developed in either isopropanol/formic acid/water or ammonia/water/chloroform/methanol. The image is subjected to densitometric analysis and the relative amounts of Glu and Gln are calculated based on the Rf values and relative intensities of the spots. NAAP activity is calculated based on the amount of Gln resulting from the transformation of Glu while acylated as Glu-tRNA_{Gln} (adapted from Curnow, A.W. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11819-11826).

In the alternative, NAAP activity can be demonstrated by the use of in vitro translation assays which utilize mutant strains of S. cerevisiae lacking the FUN12 gene which encodes yeast translation initiation factor 2 (IF2). These strains exhibit a slow growth phenotype which can be rescued (made to grow at a normal rate) by the addition of IF2 from any source, including IF2 which is produced by recombinant methods or IF2 which is isolated from another organism. Briefly, the fun12Δ strain J133 is transformed with either the low copy-number FUN12 plasmid pC479, an expression plasmid carrying NAAP, or the vector only. The control strains and the test strains are streaked on synthetic
minimal medium containing 10% galactose plus the required nutrient supplements, and the plates are incubated at 30°C for 5 days. In vitro translation extracts are prepared from the fun 12Δ strain J133. Extracts are incubated with 200 ng of luciferase mRNA and increasing amounts of the control strains or the test strains containing a source of IF2. Luminescence of the samples is plotted as a function of the amount of test protein added to the translation reaction (Lee, J.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:4342-4347).

In the alternative, chromatin molecule activity of NAAP is demonstrated by measuring sensitivity to DNase I (Dawson, B.A. et al. (1989) J. Biol. Chem. 264:12830-12837). Samples are treated with DNase I, followed by insertion of a cleavable biotinylated nucleotide analog, 5-[(N-biotinamido)hexanoamido-ethyl]-3-thioproionyl-3-aminoallyl]-2'-deoxyuridine 5'-triphosphate using nick-repair techniques well known to those skilled in the art. Following purification and digestion with EcoRI restriction endonuclease, biotinylated sequences are affinity isolated by sequential binding to streptavidin and biotin cellulose.

XIX. Identification of NAAP Agonists and Antagonists

Agonists or antagonists of NAAP activation or inhibition may be tested using the assays described in section XVII. Agonists cause an increase in NAAP activity and antagonists cause a decrease in NAAP activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.
Table 1

<table>
<thead>
<tr>
<th>Incyte Project ID</th>
<th>Polypeptide SEQ ID NO:</th>
<th>Incyte Polypeptide ID</th>
<th>Polynucleotide SEQ ID NO:</th>
<th>Incyte Polynucleotide ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>3804141</td>
<td>1</td>
<td>3804141CD1</td>
<td>11</td>
<td>3804141CB1</td>
</tr>
<tr>
<td>7947732</td>
<td>2</td>
<td>7947732CD1</td>
<td>12</td>
<td>7947732CB1</td>
</tr>
<tr>
<td>2620845</td>
<td>3</td>
<td>2620845CD1</td>
<td>13</td>
<td>2620845CB1</td>
</tr>
<tr>
<td>7473330</td>
<td>4</td>
<td>7473330CD1</td>
<td>14</td>
<td>7473330CB1</td>
</tr>
<tr>
<td>4029969</td>
<td>5</td>
<td>4029969CD1</td>
<td>15</td>
<td>4029969CB1</td>
</tr>
<tr>
<td>2676571</td>
<td>6</td>
<td>2676571CD1</td>
<td>16</td>
<td>2676571CB1</td>
</tr>
<tr>
<td>1725129</td>
<td>7</td>
<td>1725129CD1</td>
<td>17</td>
<td>1725129CB1</td>
</tr>
<tr>
<td>2626405</td>
<td>8</td>
<td>2626405CD1</td>
<td>18</td>
<td>2626405CB1</td>
</tr>
<tr>
<td>429930</td>
<td>9</td>
<td>429930CD1</td>
<td>19</td>
<td>429930CB1</td>
</tr>
<tr>
<td>7504129</td>
<td>10</td>
<td>7504129CD1</td>
<td>20</td>
<td>7504129CB1</td>
</tr>
<tr>
<td>Polypeptide SEQ ID NO:</td>
<td>Incyte Polypeptide ID</td>
<td>GenBank ID NO:</td>
<td>Probability Score</td>
<td>GenBank Homolog</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------</td>
<td>---------------</td>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>1</td>
<td>3804141CD1</td>
<td>g1786132</td>
<td>4.3E-225</td>
<td>[Gallus gallus] DNA topoisomerase I</td>
</tr>
<tr>
<td>3</td>
<td>2620845CD1</td>
<td>g1786132</td>
<td>1.0E-231</td>
<td>[Gallus gallus] DNA topoisomerase I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>g297079</td>
<td>5.0E-230</td>
<td>[Cricetulus griseus] DNA topoisomerase I (Tanizawa, A. et al. (993) J. Biol. Chem. 268:25463-25468)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>g220618</td>
<td>4.5E-229</td>
<td>[Mus musculus] DNA topoisomerase I (Koivari, O. et al. (1993) Gene 125:211-216)</td>
</tr>
<tr>
<td>5</td>
<td>4029969CD1</td>
<td>g3789799</td>
<td>4.2E-211</td>
<td>[Homo sapiens] RCC1-like G exchanging factor RLG (Devilder, M.C. et al. (1998) Genomics 54:99-106)</td>
</tr>
<tr>
<td>7</td>
<td>1725129CD1</td>
<td>g3253159</td>
<td>0.0</td>
<td>[Oryctolagus cuniculus] translation initiation factor eIF2C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>g6002623</td>
<td>0.0</td>
<td>[Homo sapiens] putative RNA-binding protein Q99 (Koesters, R. et al. (1999) Genomics 61:210-218)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>g16566550</td>
<td>0.0</td>
<td>DEAD/DEEH helicase DDX31 [Homo sapiens]</td>
</tr>
<tr>
<td>Polypeptide SEQ ID NO:</td>
<td>Incyte Polypeptide ID</td>
<td>GenBank ID NO: or PROTEOM ID NO:</td>
<td>Probability Score</td>
<td>GenBank Homolog</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------</td>
<td>----------------------------------</td>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>9</td>
<td>429930CD1</td>
<td>g190848</td>
<td>2.6E-71</td>
<td>[Homo sapiens] ribonuclease/angiogenin inhibitor (Schneider, R. et al. (1988) EMBO J. 7:4151-4156)</td>
</tr>
<tr>
<td>10</td>
<td>7504129CD1</td>
<td>g15919359</td>
<td>0.0</td>
<td>[Homo sapiens] topoisomerase I (Zhang, H. et al. (2001) Proc. Natl. Acad. Sci. USA 98:10608-10613)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>623762</td>
<td>TOP1</td>
<td>5.5E-227</td>
</tr>
<tr>
<td></td>
<td></td>
<td>724265</td>
<td>la36_A</td>
<td>7.4E-225</td>
</tr>
<tr>
<td></td>
<td></td>
<td>326672</td>
<td>Topl</td>
<td>1.9E-224</td>
</tr>
<tr>
<td></td>
<td></td>
<td>721997</td>
<td>lef9_A</td>
<td>5.2E-224</td>
</tr>
<tr>
<td></td>
<td></td>
<td>720602</td>
<td>la35_A</td>
<td>3.2E-209</td>
</tr>
<tr>
<td>SEQ ID Polypeptide ID</td>
<td>Amino Acid Residues</td>
<td>Potential Phosphorylation Sites</td>
<td>Potential Glycosylation Sites</td>
<td>Signature Sequences, Domains and Motifs</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>1 3804141CD1</td>
<td>1568</td>
<td>S43 S73 S80 S98 S134 S202 S231 S435 S538 S621 S802 S833 S877 S946 S1468 S1501 S1507 S1564 T64 T129 T205 T230 T304 T493 T561 T700 T712 T901 T913 T942 T1464 Y183 Y255 Y663 Y939 Y948</td>
<td>N274 N500 N708</td>
<td>Signal peptide: ML-A66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eukaryotic DNA topoisomerase I active site: E905-I923</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eukaryotic DNA topoisomerase I: R246-G950</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Transmembrane domains: T277-R296, H1313-W1330; N-terminus is cytosolic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eukaryotic DNA topoisomerase I proteins BL00176: D902-T942, D459-P508, B612-V627, V640-L666, D578-L717</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eukaryotic DNA topoisomerase I active site: L886-E945</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eukaryotic DNA topoisomerase I signature PR00416: L475-K484, Q589-K608, A613-V627, V640-V656, G698-T712, G912-I923</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Topoisomerase I, DNA-binding PD000422: E318-R682, Q828-K915, P1117-R1133, E93-K138</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Transmembrane domain: L362-L390</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N-terminus is non-cytosolic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LAK1 PD127162: M451-P610</td>
</tr>
<tr>
<td>Seq ID</td>
<td>Incyte Polypeptide ID</td>
<td>Amino Acid Residues</td>
<td>Potential Phosphorylation Sites</td>
<td>Potential Glycosylation Sites</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------</td>
<td>---------------------</td>
<td>--------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEQ NO</td>
<td>Incyte Polypeptide ID</td>
<td>Amino Acid Residues</td>
<td>Potential Phosphorylation Sites</td>
<td>Potential Glycosylation Sites</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------</td>
<td>---------------------</td>
<td>--------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>SEQ NO.</td>
<td>Incyte Polypeptide ID</td>
<td>Amino Acid Residues</td>
<td>Potential Phosphorylation Sites</td>
<td>Potential Glycosylation Sites</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------</td>
<td>---------------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1725129CD1</td>
<td>860</td>
<td>S35 S132</td>
<td>N284 N360</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S149 S172</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S206 S243</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S254 S372</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S470 S479</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S611 S657</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S776 T38</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>T184 T286</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>T358 T369</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>T410 T445</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>T629 T719</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>T735 T745</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>T760 T853</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2626405CD1</td>
<td>565</td>
<td>S3 S62 S87</td>
<td>N363 N469</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S150 S158</td>
<td></td>
</tr>
</tbody>
</table>

I: Insertions
D: Deletions
-: Not determined

<table>
<thead>
<tr>
<th>Seq ID</th>
<th>Incyte Polypeptide ID</th>
<th>Amino Acid Residues</th>
<th>Potential Phosphorylation Sites</th>
<th>Potential Glycosylation Sites</th>
<th>Signature Sequences, Domains and Motifs</th>
<th>Analytical Methods and Databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ ID NO:</td>
<td>Incyte Polypeptide ID</td>
<td>Amino Acid Residues</td>
<td>Potential Phosphorylation Sites</td>
<td>Potential Glycosylation Sites</td>
<td>Signature Sequences, Domains and Motifs</td>
<td>Analytical Methods and Databases</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>--------------------------------</td>
<td>-------------------------------</td>
<td>----------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RECEPTOR ANGIOTENSIN/VASOPRESSIN AII/AVP VASOPRESSIN PD156095: N2-Q104</td>
<td>BLAST-PRODOM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>signal cleavage: M1-A66</td>
<td>SPCAN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eukaryotic DNA topoisomerase BL00176: D459-P508, E612-V627, V640-L666, D678-L717, D822-T862</td>
<td>BLIMPS-BLOCKS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eukaryotic DNA topoisomerase I active site: L806-E865</td>
<td>PROFILESCAN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eukaryotic DNA topoisomerase PRO00416: L475-K484, Q589-K608, A613-V627, V640-V656, G698-T712, Q832-I843</td>
<td>BLIMPS-PRINTS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eukaryotic DNA topoisomerase I active site: E825-I843</td>
<td>MOTIFS</td>
</tr>
<tr>
<td>Polynucleotide Seq ID NO.: Encyte ID/ Sequence Length</td>
<td>Sequence Fragments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>--------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polynucleotide</td>
<td>Sequence Fragments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEQ ID NO:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incyte ID/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence Length</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 4 (cont.)

<table>
<thead>
<tr>
<th>Polynucleotide</th>
<th>SEQ ID NO:/</th>
<th>Sequence Fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynucleotide Index</td>
<td>Sequence Fragments</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------</td>
<td></td>
</tr>
<tr>
<td>Polynucleotide Seq ID No.:</td>
<td>Incyte ID/Sequence Length</td>
<td>Sequence Fragments</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Polynucleotide SEQ ID No.</td>
<td>Incyte Project ID</td>
<td>Representative Library</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>11</td>
<td>3804141CBL</td>
<td>ITRSNOT01</td>
</tr>
<tr>
<td>12</td>
<td>7947732CBL</td>
<td>ENDCNOT01</td>
</tr>
<tr>
<td>13</td>
<td>2620845CBL</td>
<td>BLATUT05</td>
</tr>
<tr>
<td>14</td>
<td>7473330CBL</td>
<td>NOSETUE01</td>
</tr>
<tr>
<td>15</td>
<td>4029969CBL</td>
<td>SINITMR01</td>
</tr>
<tr>
<td>16</td>
<td>2676571CBL</td>
<td>BRAIFER05</td>
</tr>
<tr>
<td>17</td>
<td>1725129CBL</td>
<td>FIBPEN06</td>
</tr>
<tr>
<td>18</td>
<td>2626405CBL</td>
<td>BRSTNOT03</td>
</tr>
<tr>
<td>19</td>
<td>429930CBL</td>
<td>NEUTGTM01</td>
</tr>
<tr>
<td>20</td>
<td>7504129CBL</td>
<td>SYMONOT01</td>
</tr>
<tr>
<td>Library</td>
<td>Vector</td>
<td>Library Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>BLADTUT05</td>
<td>pINCY</td>
<td>Library was constructed using RNA isolated from bladder tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma on the anterior wall of the bladder. Patient history included lung neoplasm and tobacco abuse in remission. Family history included malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease, and lung cancer.</td>
</tr>
<tr>
<td>BRAIFFER05</td>
<td>pINCY</td>
<td>Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.</td>
</tr>
<tr>
<td>BRSTNOT03</td>
<td>PSORT1</td>
<td>Library was constructed using RNA isolated from diseased breast tissue removed from a 54-year-old Caucasian female during a bilateral radical mastectomy. Pathology for the associated tumor tissue indicated residual invasive grade 3 mammary ductal adenocarcinoma. Patient history included kidney infection and condyloma acuminatum. Family history included benign hypertension, hyperlipidemia and a malignant neoplasm of the colon.</td>
</tr>
<tr>
<td>ENDCN0101</td>
<td>pINCY</td>
<td>Library was constructed using RNA isolated from endothelial cells removed from the coronary artery of a 58-year-old Hispanic male.</td>
</tr>
<tr>
<td>FIBPPE06</td>
<td>pINCY</td>
<td>The normalized prostate stromal fibroblast tissue libraries were constructed from 1.56 million independent clones from a prostate fibroblast library. Starting RNA was made from fibroblasts of prostate stroma removed from a male fetus, who died after 26 weeks' gestation. The libraries were normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48-hours/round) renanelling hybridization was used. The library was then linearized and recircularized to select for insert containing clones as follows: plasmid DNA was prepped from approximately 1 million clones from the normalized prostate stromal fibroblast tissue libraries following soft agar transformation.</td>
</tr>
<tr>
<td>NEUWMT01</td>
<td>PSORT1</td>
<td>Library was constructed using RNA isolated from peripheral blood granulocytes collected by density gradient centrifugation through Ficoll-Hypaque. The cells were isolated from buffy coat units obtained from 20 unrelated male and female donors. Cells were cultured in 10 nM GM-CSF for 1 hour before washing and harvesting for total RNA preparation.</td>
</tr>
<tr>
<td>Library</td>
<td>Vector</td>
<td>Library Description</td>
</tr>
<tr>
<td>---------------</td>
<td>--------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>NOSETUE01</td>
<td>FCDNA2.1</td>
<td>This 5' biased random primed library was constructed using RNA isolated from nasal and cribriform tumor tissue removed from a 45-year-old Caucasian male during total face ostectomy with reconstruction, rhinotomy and craniotomy. Pathology indicated olfactory neuroblastoma in the nasal cavity and cribriform region. The patient presented with cancer of the head, face and neck, and epistaxis. Patient history included extrinsic asthma, cancer of the head, face and neck, and epistaxis. Previous surgeries included total face ostectomy with reconstruction. Patient medications included Biaxin, Atessalon, and Valium. The patient received radiation treatments. Family history included chronic sinusitis in the mother and type II diabetes in the father.</td>
</tr>
<tr>
<td>SINITMR01</td>
<td>FCDNA2.1</td>
<td>This random primed library was constructed using RNA isolated from ileum tissue removed from a 70-year-old Caucasian female during right hemicolectomy, open liver biopsy, flexible sigmoidoscopy, colonoscopy, and permanent colostomy. Pathology for the matched tumor tissue indicated invasive grade 2 adenocarcinoma forming an ulcerated mass, situated 2 cm distal to the ileocecal valve. Patient history included a malignant breast neoplasm, type II diabetes, hyperlipidemia, viral hepatitis, an unspecified thyroid disorder, osteoarthritis, a malignant skin neoplasm, deficiency anemia, and normal delivery. Family history included breast cancer, atherosclerotic coronary artery disease, benign hypertension, cerebrovascular disease, ovarian cancer, and hyperlipidemia.</td>
</tr>
<tr>
<td>SYNONOT01</td>
<td>pINCY</td>
<td>Library was constructed using RNA isolated from synovial tissue removed from a 75-year-old Caucasian male.</td>
</tr>
<tr>
<td>UTRSNOT01</td>
<td>PSPORT1</td>
<td>Library was constructed using RNA isolated from the uterine tissue of a 59-year-old female who died of a myocardial infarction. Patient history included cardiomyopathy, coronary artery disease, previous myocardial infarctions, hypercholesterolemia, hypotension, and arthritis.</td>
</tr>
<tr>
<td>Program</td>
<td>Description</td>
<td>Reference</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ABI FACTURA</td>
<td>A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.</td>
<td>Applied Biosystems, Foster City, CA.</td>
</tr>
<tr>
<td>ABI/PARACEL PDF</td>
<td>A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.</td>
<td>Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.</td>
</tr>
<tr>
<td>ABI AutoAssembler</td>
<td>A program that assembles nucleic acid sequences.</td>
<td>Applied Biosystems, Foster City, CA.</td>
</tr>
<tr>
<td>BLAST</td>
<td>A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastx, and tblastx.</td>
<td>Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.</td>
</tr>
<tr>
<td>Program</td>
<td>Description</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Phred</td>
<td>A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.</td>
<td>Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.</td>
</tr>
</tbody>
</table>
What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
   a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10,
   b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 and SEQ ID NO:9-10,
   c) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to the amino acid sequence of SEQ ID NO:7,
   d) a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to the amino acid sequence of SEQ ID NO:8,
   e) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and
   f) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.


9. A method of producing a polypeptide of claim 1, the method comprising:
a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-16 and SEQ ID NO:19-20,

c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 99% identical to the polynucleotide sequence of SEQ ID NO:17,

d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 98% identical to the polynucleotide sequence of SEQ ID NO:18,

e) a polynucleotide complementary to a polynucleotide of a),

f) a polynucleotide complementary to a polynucleotide of b),

g) a polynucleotide complementary to a polynucleotide of c),

h) a polynucleotide complementary to a polynucleotide of d), and

i) an RNA equivalent of a)-h).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under
conditions whereby a hybridization complex is formed between said probe and said
target polynucleotide or fragments thereof, and
b) detecting the presence or absence of said hybridization complex, and, optionally, if
present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide
having a sequence of a polynucleotide of claim 12, the method comprising:
a) amplifying said target polynucleotide or fragment thereof using polymerase chain
reaction amplification, and
b) detecting the presence or absence of said amplified target polynucleotide or fragment
thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable
excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence
selected from the group consisting of SEQ ID NO:1-10.

19. A method for treating a disease or condition associated with decreased expression of
functional NAAP, comprising administering to a patient in need of such treatment the composition of
claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of
claim 1, the method comprising:
a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a
pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of
functional NAAP, comprising administering to a patient in need of such treatment a composition of
claim 21.
23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:
   a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
   b) detecting antagonist activity in the sample.


25. A method for treating a disease or condition associated with overexpression of functional NAAP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:
   a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
   b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:
   a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
   b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
   c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:
   a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:
a) treating a biological sample containing nucleic acids with the test compound,
b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
c) quantifying the amount of hybridization complex, and
d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of NAAP in a biological sample, the method comprising:
a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:
a) a chimeric antibody,
b) a single chain antibody,
c) a Fab fragment,
d) a F(ab')$_2$ fragment, or
e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of NAAP
in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of NAAP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:
   a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
   b) isolating antibodies from said animal, and
   c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.


38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:
   a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
   b) isolating antibody producing cells from the animal,
   c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
   d) culturing the hybridoma cells, and
   e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 in a sample, the method comprising:
   a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
   b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 from a sample, the method comprising:
   a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
   b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:
   a) labeling the polynucleotides of the sample,
   b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.
58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:11.

67. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:12.

68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.

69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.

70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
INCYTE GENOMICS, INC.
THORNTON, Michael
HAPALLA, April J.A.
LU, Dyung Aina M.
ARVIZU, Chandra
SWARNAKAR, Anita
LU, Yan
WARREN, Bridget A.
BAUGHN, Mariah R.
TANG, Y. Tom
LEE, Ernestine A.
YAO, Monique G.
RAMKUMAR, Jayalaxmi
KHAN, Farrah A.
GANDHI, Ameena R.
DING, Li
YUE, Henry
GIETZEN, Kimberly J.
WALIA, Narinder K.
THANGAVELU, Kavitha
ELLIOTT, Vicki S.
MARQUIS, Joseph P.

NUCLEIC ACID-ASSOCIATED PROTEINS

PI-0368 PCT

To Be Assigned

Herewith

60/268,118; 60/270,963; 60/270,858; 60/271,194; 60/274,071;
60/283,496; 60/344,650
2001-02-09; 2001-02-21; 2001-02-22; 2001-02-23; 2001-03-07;
2001-04-12; 2001-11-09

20

PERL Program

1
1568
PRT
Homo sapiens

misc_feature

Incyte ID No: 3804141CD1

1
Met Gly Val Thr Ser Ala Ala Ser Gly Leu Val Gly Ser Ala Pro
1  5  10  15
Gln Cys Val Ala Leu Pro Ser Glu Gly Trp Thr Leu Ala Val Trp
20  25  30
Pro Val Ala Ala Cys Thr Cys Ser Gly Val Gly Ser Ser Pro Lys
35  40  45
Leu Thr Pro Gly Ser Phe Val His Cys Pro Trp Phe Leu Leu Leu
50  55  60
Thr Glu Ala Thr Arg Ala Glu Ile Lys Arg Pro Phe Ser Lys Ala
65  70  75
Glu Leu Lys Ala Ser Val Arg Pro Met Glu Gly Ser His Cys Trp
80  85  90
Gly Gly Glu Ala Arg Arg Thr Ser Gln Gly His Thr Glu Gln Lys
95 100 105
Gly Glu Leu Arg Ala Arg Arg Gln Ala Gln Asn Glu Asp Glu Asp
110 115 120
Val Lys Glu Val Trp Val Gly Lys Thr Lys Lys Glu Glu Ser Asp
125 130 135
Lys Leu Gly Cys Gln Gln Gly Ala Trp Val Pro Pro Arg Val Pro Ser
140 145 150
Trp Ile Gly Arg Phe Glu Leu Trp Val Val Gly Tyr Trp Glu Gln
155 160 165
Pro Ala Val Ser Trp Gln Met Val Arg Val Leu Leu Arg Leu
170 175 180
Arg Ala Ala Leu Thr Leu Leu Gly Glu Val Pro Arg Arg Pro Ala
185 190 195
Ser Arg Gly Val Pro Gly Ser Arg Thr Gln Lys Gly Ser Gly
200 205 210
Ala Arg Thr Asp Ser Ala Trp Arg Ala Ala Leu Thr Val Ile Ser
215 220 225
Thr Ser Pro Gly Thr Ser Arg Met Asp Pro Val Ala Leu Val Ala
230 235 240
Val Gly Gly Pro Arg Arg Phe Pro Gly Gly His Thr Leu Glu Arg
245 250 255
Leu Pro Val Ala Leu Arg Thr Leu Ile Pro Ala Asp Gln Ala His
260 265 270
Gln Ala Pro Asn Ser Thr Thr Trp Leu Gly Ser Arg Ala Gly Leu
275 280 285
Leu Ala Leu Ala Ala Gly Leu Gly Ile Arg Asp Ser Ala Ala
290 295 300
His Gly Pro Leu Gln Val Gly Gly Met Gly Tyr Thr Gly Met
305 310 315
Gly Ser Glu Val Arg Trp Glu Lys Glu Lys His Gly Asp Gly Val
320 325 330
Lys Trp Arg Gln Leu Glu His Lys Gly Pro Tyr Phe Ala Pro Pro
335 340 345
Tyr Glu Pro Leu Pro Asp Gly Val Arg Phe Phe Tyr Glu Gly Arg
350 355 360
Pro Val Arg Leu Ser Val Ala Ala Glu Val Ala Thr Phe Tyr
365 370 375
Gly Arg Met Leu Asp His Gly Tyr Thr Thr Lys Glu Val Phe Arg
380 385 390
Lys Asn Phe Phe Asp Trp Arg Lys Asp Met Ala Val Glu Glu
395 400 405
Arg Glu Val Ile Lys Ser Leu Asp Lys Cys Asp Phe Thr Glu Ile
410 415 420
His Arg Tyr Phe Val Asp Lys Ala Ala Ala Arg Val Leu Ser
425 430 435
Arg Glu Glu Lys Gln Lys Leu Lys Glu Glu Ala Glu Lys Leu Glu
440 445 450
Gln Glu Phe Gly Tyr Cys Ile Leu Asp Gly His Gln Glu Lys Ile
455 460 465
Gly Asn Phe Lys Ile Glu Pro Pro Gly Leu Phe Arg Gly Arg Gly
470 475 480
Asp His Pro Lys Met Gly Met Leu Lys Arg Arg Ile Thr Pro Glu
485 490 495
Asp Val Val Ile Asn Cys Ser Arg Asp Ser Lys Ile Pro Glu Pro
500 505 510
Pro Ala Gly His Gln Trp Lys Glu Val Arg Ser Asp Asn Thr Val
515 520 525
Thr Trp Leu Ala Ala Trp Thr Glu Ser Val Gln Asn Ser Ile Lys
530 535 540
Tyr Thr Met Leu Asn Pro Cys Ser Lys Leu Lys Gly Glu Thr Ala
545 550 555
Trp Gln Lys Phe Glu Thr Ala Arg Arg Leu Arg Gly Phe Val Asp
560 565 570
Glu Ile Arg Ser Gln Tyr Arg Ala Asp Trp Lys Ser Arg Glu Met
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>575</td>
<td>Lys Thr Arg Glu Arg Ala Val Ala Leu Tyr Phe Ile Asp Lys Leu</td>
<td>580</td>
</tr>
<tr>
<td>590</td>
<td>Ala Leu Arg Ala Gly Asn Glu Lys Glu Asp Gly Glu Ala Ala Asp</td>
<td>595</td>
</tr>
<tr>
<td>605</td>
<td>Thr Val Gly Cys Cys Ser Leu Arg Val Glu His Val Glu Leu His</td>
<td>610</td>
</tr>
<tr>
<td>620</td>
<td>Pro Glu Ala Asp Gly Cys Glu His Val Glu Phe Asp Phe Leu</td>
<td>625</td>
</tr>
<tr>
<td>635</td>
<td>Gly Lys Asp Cys Ile Arg Tyr Tyr Asn Arg Val Pro Val Glu Lys</td>
<td>640</td>
</tr>
<tr>
<td>650</td>
<td>Pro Val Tyr Lys Asn Leu Glu Leu Phe Met Glu Asn Lys Asp Pro</td>
<td>655</td>
</tr>
<tr>
<td>665</td>
<td>Arg Asp Asp Leu Phe Asp Arg Leu Thr Met Thr Ser Leu Asn Lys</td>
<td>670</td>
</tr>
<tr>
<td>680</td>
<td>His Leu Glu Glu Leu Met Asp Gly Leu Thr Ala Lys Val Phe Arg</td>
<td>685</td>
</tr>
<tr>
<td>695</td>
<td>Thr Tyr Asn Ala Ser Ile Thr Leu Glu Glu Glu Leu Arg Ala Leu</td>
<td>700</td>
</tr>
<tr>
<td>710</td>
<td>Thr Arg Gly Leu Pro Gly Gly Gly Ala Glu Ala Gly Cys Ser Gly</td>
<td>715</td>
</tr>
<tr>
<td>725</td>
<td>Ala Leu Thr Leu Gly Met Cys Phe Ala Cys Thr Arg Gly Pro Gin</td>
<td>730</td>
</tr>
<tr>
<td>740</td>
<td>Arg Pro Val Leu Cys Pro Gly Asp Asn Leu Gly Ala Gly Ala</td>
<td>745</td>
</tr>
<tr>
<td>755</td>
<td>Ala Leu Gly Leu Trp Asp Ala Glu Glu Leu Ala Gly Ile Glu Val</td>
<td>760</td>
</tr>
<tr>
<td>770</td>
<td>Arg Ala Pro Gly Glu Glu Thr Pro Pro Pro Leu Arg Glu Ala Glu</td>
<td>775</td>
</tr>
<tr>
<td>785</td>
<td>Asp Ser Gly Pro Cys Leu Ser Ala Glu Asp Ser Ile Ala Ala Lys</td>
<td>790</td>
</tr>
<tr>
<td>800</td>
<td>Ile Leu Ser Tyr Asn Arg Ala Asn Arg Val Val Ala Ala Leu Cys</td>
<td>805</td>
</tr>
<tr>
<td>815</td>
<td>Asn His Glu Arg Ala Thr Pro Ser Thr Phe Glu Lys Ser Met Gin</td>
<td>820</td>
</tr>
<tr>
<td>830</td>
<td>Asn Leu Glu Thr Lys Ile Glu Ala Lys Lys Glu Glu Val Ala Glu</td>
<td>835</td>
</tr>
<tr>
<td>845</td>
<td>Ala Arg Ala Glu Leu Arg Arg Ala Arg Ala Glu His Lys Ala Gin</td>
<td>850</td>
</tr>
<tr>
<td>860</td>
<td>Gly Asp Gly Lys Ser Arg Ser Val Leu Glu Lys Lys Arg Arg Leu</td>
<td>865</td>
</tr>
<tr>
<td>875</td>
<td>Leu Glu Lys Leu Gin Glu Gin Leu Ala Gin Leu Ser Val Gin Ala</td>
<td>880</td>
</tr>
<tr>
<td>890</td>
<td>Thr Asp Lys Glu Asn Lys Gin Val Ala Leu Gly Thr Ser Lys</td>
<td>895</td>
</tr>
<tr>
<td>905</td>
<td>Leu Asn Tyr Leu Asp Pro Arg Ile Ser Ile Ala Trp Cys Lys Arg</td>
<td>910</td>
</tr>
<tr>
<td>920</td>
<td>Phe Arg Val Pro Val Glu Lys Ile Tyr Ser Lys Thr Gin Arg Glu</td>
<td>925</td>
</tr>
<tr>
<td>935</td>
<td>Ser Ser Tyr Asp Gly Leu Leu Pro Glu Leu Pro Asn Pro Val Ser</td>
<td>940</td>
</tr>
<tr>
<td>950</td>
<td>Ser Gin Ser Cys Leu Ser Ala Ala Leu Cys His Ala Gly Ser Ser</td>
<td>955</td>
</tr>
<tr>
<td>965</td>
<td>Ala Ala Leu Ser Ala Thr Arg Asp Pro Leu Pro Pro Ser Ser Leu</td>
<td>970</td>
</tr>
<tr>
<td>980</td>
<td>Pro Arg Gly Thr Leu Cys Arg Pro Leu Leu Cys His Ala Gly Pro</td>
<td>985</td>
</tr>
<tr>
<td>995</td>
<td>Ser Ala Ala Leu Phe Ser Ala Thr Arg Asp Pro Leu Pro Pro Ser</td>
<td>1000</td>
</tr>
<tr>
<td>1010</td>
<td>Leu Pro Arg Thr Leu Cys Arg Pro Leu Leu Cys His Thr Gly Gly</td>
<td>1015</td>
</tr>
<tr>
<td>1025</td>
<td>Pro Ser Ala Ala Leu Leu Cys His Ala Gly Pro Ser Ala Ala Leu</td>
<td>1030</td>
</tr>
<tr>
<td>1040</td>
<td>1045</td>
<td>1050</td>
</tr>
</tbody>
</table>
Phe Ser Ala Met Arg Asp Pro Leu Pro Pro Ser Ser Leu Pro Arg
1055 1060 1065
Gly Thr Leu Cys Pro Leu Leu Cys His Val Gly Pro Ser Ala
1070 1075 1080
Ala Leu Ser Ala Thr Arg Pro Leu Pro Pro Ser Pro Leu Pro
1085 1090 1095
Arg Gly Ile Cys Pro Leu Leu Cys His Val Gly Ser Ser
1100 1105 1110
Ala Met Arg Asp Pro Leu Pro Pro Ser Pro Leu Pro Pro Ser Pro
1115 1120 1125
Leu Pro Pro Ser Leu Pro Arg Gly Ile Leu Cys Arg Pro Leu
1130 1135 1140
Leu Cys His Ala Gly Thr Ser Ala Ala Leu Ser Ser Ala Thr Arg
1145 1150 1155
Asp Pro Leu Pro Pro Ser Pro Leu Pro Arg Gly Thr Leu Cys Arg
1160 1165 1170
Pro Leu Leu Cys His Ala Gly Pro Ser Ala Ala Leu Ser Ala Thr
1175 1180 1185
Gln Asp Pro Leu Pro Pro Ser Ser Leu Pro Arg Gly Thr Leu Cys
1190 1195 1200
Cys Pro Leu Leu Cys Arg Pro Leu Leu Cys Arg Ala Gly Pro Ser
1205 1210 1215
Ala Ala Leu Ser Ser Ala Thr Arg Asp Pro Leu Pro Pro Ser Pro
1220 1225 1230
Leu Pro Arg Gly Ile Leu Cys His Pro Leu Cys His Ala Gly Pro
1235 1240 1245
Ser Ala Ala Leu Phe Ser Ala Thr Arg Asp Pro Leu Leu Pro Ser
1250 1255 1260
Pro Leu Pro Arg Gly Thr Leu Cys Cys Pro Leu Cys His Ala Gly
1265 1270 1275
Pro Ser Ala Ala Leu Ser Ser Ala Thr Arg Asp Pro Leu Pro Pro
1280 1285 1290
Ser Ala Met Arg Asp Pro Leu Leu Pro Ser Leu Leu Pro Arg Gly
1295 1300 1305
Thr Leu Cys Arg Pro Leu Cys His Ala Gly Pro Ser Ala Ala Leu
1310 1315 1320
Phe Ser Ala Thr Leu Phe Ser Ala Thr Trp Asp Pro Leu Leu Pro
1325 1330 1335
Phe Pro Leu Leu Arg Ala Leu Cys Arg Pro Leu Cys His Thr
1340 1345 1350
Gly Pro Ser Ala Ala Leu Ser Ser Ala Thr Trp Asp Pro Leu Leu
1355 1360 1365
Pro Ser Ser Leu Pro Arg Gly Ile Leu Cys His Ala Gly Pro Ser
1370 1375 1380
Ala Ala Leu Phe Ser Ala Thr Trp Asp Pro Leu Leu Pro Ser Ala
1385 1390 1395
Thr Arg Asp Pro Leu Pro Pro Ser Pro Leu Arg Gly Ile Leu
1400 1405 1410
Cys Cys Pro Leu Leu Cys His Val Gly Ser Ser Ala Ile Cys Ser
1415 1420 1425
Ala Leu Cys Phe Arg Gly Cys Leu Cys Gly Cys Leu Pro His Thr
1430 1435 1440
Gly Ala Pro Ala Asn Pro Met Arg Gln Pro Leu Cys Lys Ala Pro
1445 1450 1455
Glu Lys Ser Trp Val Val Pro Ser Thr Gly Lys Ser Ser Gly Lys
1460 1465 1470
Ala Ser Cys Val Cys Trp Cys Glu Met Arg Phe His Asp Gly Gly
1475 1480 1485
Cys Arg Ala Leu Ser Leu Ser Cys Tyr Pro Val Ala Cys Glu Pro
1490 1495 1500
Ser Leu Lys Cys Thr Lys Ser Cys Lys Gly Leu Arg Ala Thr Ser
1505 1510 1515
Phe Thr Ala Ala Gln Ala Glu Lys Gly Ala Gln Ala Leu Gln Leu
Glu Leu Gly Asn Arg Pro Gly Gly Asp Ser His Trp Ile Pro Leu 1520 1525 1530
Ser His Asp His Lys Asn Arg Ala Thr Ala Ala Arg Val Ser Arg 1535 1540 1545
Ala Thr Asp Ser Asn Arg Glu Pro 1550 1555 1560 1565

<210> 2
<211> 713
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 79477732CD1

<400> 2
Met Asp Pro Arg Ile Ala Trp Phe Gln Pro Glu Gln Leu Gly Pro 1 5 10 15
Ser Asn Ser Leu Trp Met Gln Ile Trp Glu Thr Thr Gln Gly Leu 20 25 30
Arg Asn Leu Tyr Phe Asn His His Cys His Ser Ser Gly Gly Ala 35 40 45
Ser Gly Gly Gly Gly Ser Ser Ser Ser Ser Ser Thr Ala Thr Gly 50 55 60 65 70 75
Gly Ser Gly Ser Thr Gly Ser Pro Gly Gly Ala Ala Ser Ala 80 85 90
Pro Ala Pro Ala Pro Ala Gly Met Tyr Arg Ser Gly Glu Arg Leu 95 100 105
Leu Gly Ser His Ala Leu Pro Ala Glu Gln Arg Asp Phe Leu Pro 110 115 120
Leu Glu Thr Thr Asn Asn Asn Asn His His Gln Pro Gly Ala 125 130 135
Trp Ala Arg Arg Ala Gly Ser Ser Ala Ser Ser Pro Pro Ser Ala 140 145 150
Ser Ser Ser Pro His Pro Ser Ala Ala Val Pro Ala Ala Asp Pro 155 160 165
Ala Asp Ser Ala Ser Gly Ser Ser Asn Lys Arg Lys Arg Asp Asn 170 175 180
Lys Ala Ser Thr Tyr Gly Leu Asn Tyr Ser Leu Leu Gln Pro Ser 185 190 195
Gly Gly Arg Ala Ala Gly Gly Gly Arg Ala Asp Gly Gly Gly Val 200 205 210
Val Tyr Ser Gly Thr Pro Trp Lys Arg Arg Asn Tyr Asn Gln Gin 215 220 225
Val Val Gly Leu His Glu Glu Ile Ser Asp Phe Tyr Glu Tyr Met 230 235 240
Ser Pro Arg Pro Glu Glu Glu Lys Met Arg Met Glu Val Val Asn 245 250
Arg Ile Glu Ser Val Ile Lys Glu Leu Trp Pro Ser Ala Asp Val 260 265 270
Gln Ile Phe Gly Ser Phe Lys Thr Gly Leu Tyr Leu Pro Thr Ser 275 280 285
Asp Ile Asp Leu Val Val Phe Gly Lys Trp Glu Asn Leu Pro Leu 290 295 300
Trp Thr Leu Glu Ala Leu Arg Lys His Lys Val Ala Asp Glu 305 310 315
Asp Ser Val Lys Val Leu Asp Lys Ala Thr Val Pro Ile Ile Lys 320 325 330
Leu Thr Asp Ser Phe Thr Glu Val Lys Val Asp Ile Ser Phe Asn
Val Glu Asn Gly Val Arg Ala Ala Ala Asp Leu Ile Lys Asp Phe Thr

5/30
<table>
<thead>
<tr>
<th></th>
<th>335</th>
<th>340</th>
<th>345</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys</td>
<td>Lys</td>
<td>Tyr</td>
<td>Pro</td>
</tr>
<tr>
<td>Val</td>
<td>Leu</td>
<td>Pro</td>
<td>Tyr</td>
</tr>
<tr>
<td>Leu</td>
<td>Val</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td>Val</td>
<td>Val</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td>Gly</td>
<td>Ser</td>
<td>Tyr</td>
<td>Ser</td>
</tr>
<tr>
<td>Leu</td>
<td>Phe</td>
<td>Leu</td>
<td>Met</td>
</tr>
<tr>
<td>Ala</td>
<td>Val</td>
<td>Ser</td>
<td>Phe</td>
</tr>
<tr>
<td>Leu</td>
<td>Gln</td>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Leu</td>
<td>Gln</td>
<td>Arg</td>
</tr>
<tr>
<td>Asp</td>
<td>Leu</td>
<td>Asn</td>
<td>Glu</td>
</tr>
<tr>
<td>Gly</td>
<td>Val</td>
<td>Phe</td>
<td>Thr</td>
</tr>
<tr>
<td>Val</td>
<td>Gly</td>
<td>Gly</td>
<td>Ile</td>
</tr>
<tr>
<td>His</td>
<td>Pro</td>
<td>Arg</td>
<td>Glu</td>
</tr>
<tr>
<td>Ala</td>
<td>Cys</td>
<td>Ile</td>
<td>Pro</td>
</tr>
<tr>
<td>Thr</td>
<td>Asn</td>
<td>Tyr</td>
<td>Gly</td>
</tr>
<tr>
<td>Val</td>
<td>395</td>
<td>400</td>
<td>405</td>
</tr>
<tr>
<td>Leu</td>
<td>Ile</td>
<td>Glu</td>
<td>Phe</td>
</tr>
<tr>
<td>Phe</td>
<td>Glu</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td>Gly</td>
<td>Arg</td>
<td>His</td>
<td>Phe</td>
</tr>
<tr>
<td>Asn</td>
<td>Tyr</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>Leu</td>
<td>Lys</td>
<td>Thr</td>
<td>Gly</td>
</tr>
<tr>
<td>Ile</td>
<td>Arg</td>
<td>Ile</td>
<td>Lys</td>
</tr>
<tr>
<td>Asp</td>
<td>Gly</td>
<td>Gly</td>
<td>Ser</td>
</tr>
<tr>
<td>Tyr</td>
<td>Val</td>
<td>Ala</td>
<td>425</td>
</tr>
<tr>
<td>Lys</td>
<td>Asp</td>
<td>Glu</td>
<td>Val</td>
</tr>
<tr>
<td>Gln</td>
<td>Lys</td>
<td>Asn</td>
<td>Met</td>
</tr>
<tr>
<td>Leu</td>
<td>Asp</td>
<td>Gln</td>
<td>Arg</td>
</tr>
<tr>
<td>Phe</td>
<td>Tyr</td>
<td>Arg</td>
<td>Pro</td>
</tr>
<tr>
<td>Ser</td>
<td>440</td>
<td>445</td>
<td>450</td>
</tr>
<tr>
<td>Met</td>
<td>Leu</td>
<td>Tyr</td>
<td>Ile</td>
</tr>
<tr>
<td>Glu</td>
<td>Asp</td>
<td>Pro</td>
<td>Leu</td>
</tr>
<tr>
<td>Gln</td>
<td>Pro</td>
<td>Gly</td>
<td>Asn</td>
</tr>
<tr>
<td>Asp</td>
<td>Val</td>
<td>Gly</td>
<td>455</td>
</tr>
<tr>
<td>Arg</td>
<td>Ser</td>
<td>Ser</td>
<td>Tyr</td>
</tr>
<tr>
<td>Gly</td>
<td>Ala</td>
<td>Met</td>
<td>Gln</td>
</tr>
<tr>
<td>Val</td>
<td>Lys</td>
<td>Gln</td>
<td>Ala</td>
</tr>
<tr>
<td>Phe</td>
<td>Asp</td>
<td>Tyr</td>
<td>470</td>
</tr>
<tr>
<td>Ala</td>
<td>Tyr</td>
<td>Val</td>
<td>Val</td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>His</td>
<td>Ala</td>
</tr>
<tr>
<td>Val</td>
<td>Ser</td>
<td>Pro</td>
<td>Ile</td>
</tr>
<tr>
<td>Ala</td>
<td>Lys</td>
<td>Tyr</td>
<td>485</td>
</tr>
<tr>
<td>Tyr</td>
<td>Pro</td>
<td>Asn</td>
<td>Asn</td>
</tr>
<tr>
<td>Glu</td>
<td>Thr</td>
<td>Glu</td>
<td>Ser</td>
</tr>
<tr>
<td>Ile</td>
<td>Leu</td>
<td>Gly</td>
<td>Arg</td>
</tr>
<tr>
<td>Ile</td>
<td>Ile</td>
<td>Arg</td>
<td>500</td>
</tr>
<tr>
<td>Val</td>
<td>Thr</td>
<td>Asp</td>
<td>Glu</td>
</tr>
<tr>
<td>Val</td>
<td>Ala</td>
<td>Thr</td>
<td>Tyr</td>
</tr>
<tr>
<td>Arg</td>
<td>Asp</td>
<td>Trp</td>
<td>Ile</td>
</tr>
<tr>
<td>Ser</td>
<td>Lys</td>
<td>Gln</td>
<td>515</td>
</tr>
<tr>
<td>Trp</td>
<td>Gly</td>
<td>Leu</td>
<td>Lys</td>
</tr>
<tr>
<td>Asn</td>
<td>Arg</td>
<td>Pro</td>
<td>Glu</td>
</tr>
<tr>
<td>Glu</td>
<td>Pro</td>
<td>Ser</td>
<td>Cys</td>
</tr>
<tr>
<td>Asn</td>
<td>Gly</td>
<td>Asn</td>
<td>Gly</td>
</tr>
<tr>
<td>Val</td>
<td>Thr</td>
<td>Val</td>
<td>Ile</td>
</tr>
<tr>
<td>Val</td>
<td>Asp</td>
<td>Thr</td>
<td>Gln</td>
</tr>
<tr>
<td>Gln</td>
<td>Gln</td>
<td>Leu</td>
<td>Asp</td>
</tr>
<tr>
<td>Lys</td>
<td>Cys</td>
<td>Asn</td>
<td>545</td>
</tr>
<tr>
<td>Asn</td>
<td>Leu</td>
<td>Ser</td>
<td>Glu</td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Gly</td>
<td>Lys</td>
</tr>
<tr>
<td>Cys</td>
<td>Arg</td>
<td>Ser</td>
<td>Lys</td>
</tr>
<tr>
<td>Thr</td>
<td>Ser</td>
<td>Glu</td>
<td>Ser</td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>Lys</td>
<td>Ser</td>
</tr>
<tr>
<td>His</td>
<td>Ser</td>
<td>Asn</td>
<td>Ser</td>
</tr>
<tr>
<td>Ser</td>
<td>Ser</td>
<td>Ser</td>
<td>575</td>
</tr>
<tr>
<td>Pro</td>
<td>Val</td>
<td>Ser</td>
<td>Ser</td>
</tr>
<tr>
<td>Ser</td>
<td>Ala</td>
<td>Thr</td>
<td>Gln</td>
</tr>
<tr>
<td>Ser</td>
<td>Ser</td>
<td>Ser</td>
<td>590</td>
</tr>
<tr>
<td>Asp</td>
<td>Ser</td>
<td>Asp</td>
<td>Ala</td>
</tr>
<tr>
<td>Thr</td>
<td>Pro</td>
<td>Cys</td>
<td>Lys</td>
</tr>
<tr>
<td>Thr</td>
<td>Pro</td>
<td>Lys</td>
<td>Gln</td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Cys</td>
<td>605</td>
</tr>
<tr>
<td>Arg</td>
<td>Pro</td>
<td>Ser</td>
<td>Thr</td>
</tr>
<tr>
<td>Gly</td>
<td>Asn</td>
<td>Arg</td>
<td>Val</td>
</tr>
<tr>
<td>Gly</td>
<td>Ser</td>
<td>Gln</td>
<td>Asp</td>
</tr>
<tr>
<td>Val</td>
<td>Ser</td>
<td>Val</td>
<td>620</td>
</tr>
<tr>
<td>Glu</td>
<td>Ser</td>
<td>Ser</td>
<td>Gln</td>
</tr>
<tr>
<td>Ala</td>
<td>Val</td>
<td>Gly</td>
<td>Met</td>
</tr>
<tr>
<td>Gln</td>
<td>Ser</td>
<td>Thr</td>
<td>Gln</td>
</tr>
<tr>
<td>Thr</td>
<td>Thr</td>
<td>Thr</td>
<td>635</td>
</tr>
<tr>
<td>Asn</td>
<td>Thr</td>
<td>Ser</td>
<td>Asn</td>
</tr>
<tr>
<td>Ser</td>
<td>Thr</td>
<td>Asn</td>
<td>Ser</td>
</tr>
<tr>
<td>Lys</td>
<td>Ser</td>
<td>Gln</td>
<td>His</td>
</tr>
<tr>
<td>Gly</td>
<td>Ser</td>
<td>Ala</td>
<td>Arg</td>
</tr>
<tr>
<td>Leu</td>
<td>Phe</td>
<td>Arg</td>
<td>Ser</td>
</tr>
<tr>
<td>Ser</td>
<td>Ser</td>
<td>Lys</td>
<td>Gly</td>
</tr>
<tr>
<td>Phe</td>
<td>Gln</td>
<td>Gly</td>
<td>Thr</td>
</tr>
<tr>
<td>Thr</td>
<td>Gln</td>
<td>Thr</td>
<td>665</td>
</tr>
<tr>
<td>Ser</td>
<td>His</td>
<td>Gly</td>
<td>Ser</td>
</tr>
<tr>
<td>Leu</td>
<td>Met</td>
<td>Thr</td>
<td>Asn</td>
</tr>
<tr>
<td>Lys</td>
<td>Gln</td>
<td>His</td>
<td>Gln</td>
</tr>
<tr>
<td>Gly</td>
<td>Lys</td>
<td>Ser</td>
<td>680</td>
</tr>
<tr>
<td>Asn</td>
<td>Asn</td>
<td>Gln</td>
<td>Tyr</td>
</tr>
<tr>
<td>Leu</td>
<td>His</td>
<td>Gly</td>
<td>Lys</td>
</tr>
<tr>
<td>Lys</td>
<td>Arg</td>
<td>Lys</td>
<td>695</td>
</tr>
<tr>
<td>Ala</td>
<td>Pro</td>
<td>Leu</td>
<td>Ser</td>
</tr>
<tr>
<td>Asp</td>
<td>Leu</td>
<td>Cys</td>
<td>Arg</td>
</tr>
</tbody>
</table>

<210> 3
<211> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2620845CD1

<400> 3
Met Arg Val Val Arg Leu Leu Arg Leu Arg Ala Ala Leu Thr Cys
<table>
<thead>
<tr>
<th>Residue</th>
<th>Sequence Position</th>
<th>Residue</th>
<th>Sequence Position</th>
<th>Residue</th>
<th>Sequence Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ser Gly Arg Ser Pro</td>
<td>5</td>
<td>Pro Pro Ala Ser Arg Asp Val Arg Ser Arg</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Arg Thr Gln Lys Gly Ser Gly Ala Arg Thr Asp Ser Ala Trp Arg</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>55</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Arg Ala Leu Thr Val Ile Ser Thr Ser Pro Gly Thr Ser Arg Met</td>
<td>65</td>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>85</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Pro Ser Glu Thr Ser Cys Cys Ser Glu Asp Thr His Pro Ser Arg</td>
<td>95</td>
<td>100</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td></td>
<td>110</td>
<td>115</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Ala Gly Leu Leu Ala Leu Ala Ala Gly Leu Gly Gly Ile Arg Asp</td>
<td>125</td>
<td>130</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>145</td>
<td>150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Thr Gly Met Ser Gly Val Arg Gly Val Arg Gly Met Gly Gly Tyr</td>
<td>155</td>
<td>160</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td></td>
<td>170</td>
<td>175</td>
<td>180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Asp Gly Val Lys Trp Arg Gln Leu Gly His Lys Gly Pro Tyr Phe</td>
<td>185</td>
<td>190</td>
<td>195</td>
<td></td>
</tr>
<tr>
<td></td>
<td>205</td>
<td>210</td>
<td>215</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Ala Pro Pro Tyr Gly Pro Leu Pro Asp Gly Val Arg Phe Phe Tyr</td>
<td>215</td>
<td>220</td>
<td>225</td>
<td></td>
</tr>
<tr>
<td></td>
<td>230</td>
<td>235</td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Thr Phe Tyr Gly Arg Met Leu Asp Gly Met Tyr Thr Thr Lys Glu</td>
<td>245</td>
<td>250</td>
<td>255</td>
<td></td>
</tr>
<tr>
<td></td>
<td>260</td>
<td>265</td>
<td>270</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Val Leu Ser Arg Gly Val Ile Lys Ser Leu Asp Lys Cys Asp Phe</td>
<td>275</td>
<td>280</td>
<td>285</td>
<td></td>
</tr>
<tr>
<td></td>
<td>285</td>
<td>290</td>
<td>295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Lys Leu Gln Gln Gly Met Leu Lys Arg Arg Arg Arg</td>
<td>300</td>
<td>305</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td></td>
<td>315</td>
<td>320</td>
<td>325</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Gly Arg Gly Asp His Pro Gly Met Leu Lys Arg Arg Arg Arg</td>
<td>330</td>
<td>335</td>
<td>340</td>
<td></td>
</tr>
<tr>
<td></td>
<td>345</td>
<td>350</td>
<td>355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Thr Pro Gly Val Val Ile Asp Ser Arg Asp Ser Lys Ile</td>
<td>360</td>
<td>365</td>
<td>370</td>
<td></td>
</tr>
<tr>
<td></td>
<td>375</td>
<td>380</td>
<td>385</td>
<td>390</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Ser Ile Tyr Val Trp Ala Ala Trp Thr Gly Ser Ile Gln Asn</td>
<td>395</td>
<td>400</td>
<td>405</td>
<td></td>
</tr>
<tr>
<td></td>
<td>410</td>
<td>415</td>
<td>420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Arg Glu Met Lys Thr Arg Glu Arg Ala Val Ala Leu Tyr Phe Ile</td>
<td>425</td>
<td>430</td>
<td>435</td>
<td></td>
</tr>
<tr>
<td></td>
<td>440</td>
<td>445</td>
<td>450</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Asp Lys Leu Ala Arg Ala Gly Asp Lys Glu Asp Gly Glu</td>
<td>455</td>
<td>460</td>
<td>465</td>
<td></td>
</tr>
<tr>
<td></td>
<td>470</td>
<td>475</td>
<td>480</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Asp Phe Leu Gly Lys Asp Cys Ile Arg Tyr Tyr Asn Arg Arg Val Pro
485 490 495
Val Glu Lys Pro Val Tyr Lys Asn Leu Gln Leu Phe Met Glu Asn
500 505 510
Lys Asp Pro Arg Asp Leu Phe Asp Arg Leu Thr Thr Thr Ser
515 520 525
Leu Asn Arg Pro Leu Gln Gln Leu Met Asp Gly Leu Thr Ala Lys
530 535 540
Val Pro Arg Thr Tyr Asn Ala Ser Ile Thr Leu Gln Glu Gln Leu
545 550 555
Arg Ala Leu Thr Arg Ala Glu Asp Ser Ile Ala Ala Lys Ile Leu
560 565 570
Ser Tyr Asn Arg Ala Asn Arg Val Val Ala Ile Leu Cys Asn His
575 580 585
Gln Arg Ala Thr Pro Ser Thr Phe Glu Lys Ser Met Gln Asn Leu
590 595 600
Gln Thr Lys Ile Gln Ala Lys Lys Gln Val Ala Glu Ala Arg
605 610 615
Ala Glu Leu Arg Arg Ala Arg Ala Glu His Lys Ala Glu Gly Asp
620 625 630
Gly Lys Ser Arg Ser Val Leu Glu Lys Lys Arg Arg Leu Leu Glu
635 640 645
Lys Leu Gln Glu Gln Leu Ala Gln Leu Ser Val Gln Ala Thr Asp
650 655 660
Lys Glu Glu Asn Lys Gln Val Ala Leu Gly Thr Ser Lys Leu Asn
665 670 675
Tyr Leu Asp Pro Arg Ile Ser Ile Ala Trp Cys Lys Arg Phe Arg
680 685 690
Val Pro Val Glu Lys Ile Tyr Ser Lys Thr Glu Arg Glu Arg Phe
695 700 705
Ala Trp Ala Leu Ala Met Ala Gly Glu Asp Phe Glu Phe
710 715

<210> 4
<211> 611
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473330CD1

<400> 4
Met His Pro Glu Pro Ser Glu Pro Ala Thr Gly Gly Ala Ala Glu
1  5  10  15
Leu Asp Cys Ala Gln Glu Pro Gly Val Glu Glu Ser Ala Gly Asp
20  25  30
His Gly Ser Ala Gly Arg Gly Gly Cys Lys Glu Glu Ile Asn Asp
35  40  45
Pro Lys Glu Ile Cys Val Gly Ser Ser Asp Thr Ser Tyr His Ser
50  55  60
Gln Gln Lys Gln Ser Gly Asn Gly Ser Gly Gly His Phe Ala
65  70  75
His Pro Arg Glu Asp Arg Glu Asp Arg Gly Pro Arg Met Thr Lys
80  85  90
Ser Ser Leu Gln Leu Cys Lys Gln His Lys Leu Tyr Ile Thr
95 100 105
Pro Ala Leu Asp Thr Leu Tyr Leu His Phe Lys Gly Phe Asp
110 115 120
Arg Ile Glu Asn Leu Glu Gly Tyr Thr Gly Leu Arg Cys Leu Trp
125 130 135
Leu Gln Ser Asn Gly Ile Gln Lys Ile Glu Asn Leu Glu Ala Gln
140 145 150

8/30
<table>
<thead>
<tr>
<th>Residue</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thr</td>
<td>Glu Leu Arg Cys Leu Phe Leu Gln Met Asn Leu Leu Arg Lys</td>
</tr>
<tr>
<td></td>
<td>155</td>
</tr>
<tr>
<td>Ile</td>
<td>Glu Asn Leu Glu Pro Leu Gln Lys Leu Asp Ala Leu Asn Leu</td>
</tr>
<tr>
<td></td>
<td>170</td>
</tr>
<tr>
<td>Ser</td>
<td>Asn Tyr Ile Lys Thr Ile Glu Asn Leu Phe Leu Asn Thr</td>
</tr>
<tr>
<td></td>
<td>185</td>
</tr>
<tr>
<td>Leu</td>
<td>Glu Met Ala His Asn His Leu Glu Thr Val Glu Asp Ile Gln</td>
</tr>
<tr>
<td></td>
<td>200</td>
</tr>
<tr>
<td>His</td>
<td>Leu Glu Glu Cys Leu Arg Leu Cys Val Leu Asp Leu Ser His</td>
</tr>
<tr>
<td></td>
<td>215</td>
</tr>
<tr>
<td>Asn</td>
<td>Leu Ser Asp Pro Glu Ile Leu Ser Ile Leu Glu Ser Met</td>
</tr>
<tr>
<td></td>
<td>230</td>
</tr>
<tr>
<td>Pro</td>
<td>Asp Leu Arg Val Leu Asn Leu Met Gly Asn Pro Val Ile Arg</td>
</tr>
<tr>
<td></td>
<td>245</td>
</tr>
<tr>
<td>Gln</td>
<td>Ile Pro Asn Tyr Arg Thr Val Thr Val Arg Leu Lys His</td>
</tr>
<tr>
<td></td>
<td>260</td>
</tr>
<tr>
<td>Leu</td>
<td>Thr Tyr Leu Asp Asp Arg Pro Val Phe Pro Lys Asp Arg Ala</td>
</tr>
<tr>
<td></td>
<td>275</td>
</tr>
<tr>
<td>Cys</td>
<td>Ala Glu Ala Ala Trp Ala Arg Gly Gly Tyr Ala Ala Glu Lys Glu</td>
</tr>
<tr>
<td></td>
<td>290</td>
</tr>
<tr>
<td>Glu</td>
<td>Arg Gln Gln Trp Glu Ser Arg Glu Arg Lys Lys Ile Thr Asp</td>
</tr>
<tr>
<td></td>
<td>305</td>
</tr>
<tr>
<td>Ser</td>
<td>Ile Ala Leu Ala Met Ile Lys Gln Arg Ala Glu Arg</td>
</tr>
<tr>
<td></td>
<td>310</td>
</tr>
<tr>
<td>Lys</td>
<td>Arg Glu Arg Glu Ser Gln Glu Arg Gly Glu Met Thr Ser Ser</td>
</tr>
<tr>
<td></td>
<td>335</td>
</tr>
<tr>
<td>Asp</td>
<td>Asp Gly Glu Asn Val Pro Ala Ser Ala Glu Gly Lys Glu Glu</td>
</tr>
<tr>
<td></td>
<td>350</td>
</tr>
<tr>
<td>Pro</td>
<td>Pro Gly Asp Arg Glu Thr Arg Glu Lys Met Glu Leu Phe Val</td>
</tr>
<tr>
<td></td>
<td>365</td>
</tr>
<tr>
<td>Lys</td>
<td>Glu Ser Phe Glu Ala Lys Asp Glu Leu Cys Pro Glu Lys Pro</td>
</tr>
<tr>
<td></td>
<td>370</td>
</tr>
<tr>
<td>Ser</td>
<td>Gly Glu Glu Glu Pro Pro Val Glu Ala Lys Arg Glu Asp Gly Gly</td>
</tr>
<tr>
<td></td>
<td>380</td>
</tr>
<tr>
<td>Pro</td>
<td>Glu Pro Glu Gly Thr Leu Pro Ala Glu Thr Leu Leu Leu Ser</td>
</tr>
<tr>
<td></td>
<td>395</td>
</tr>
<tr>
<td>Ser</td>
<td>Pro Val Glu Val Lys Gly Glu Asp Gly Asp Gly Glu Pro Glu</td>
</tr>
<tr>
<td></td>
<td>410</td>
</tr>
<tr>
<td>Gly</td>
<td>Thr Leu Pro Ala Ala Glu Ala Pro Gly Asp Gly Ala Glu Pro</td>
</tr>
<tr>
<td></td>
<td>425</td>
</tr>
<tr>
<td>Val</td>
<td>Lys Gly Glu Asp Gly Asp Glu Glu Pro Glu Gly Thr Leu Pro</td>
</tr>
<tr>
<td></td>
<td>440</td>
</tr>
<tr>
<td>Ala</td>
<td>Glu Thr Leu Leu Ser Ser Pro Pro Val Lys Val Lys Gly Glu</td>
</tr>
<tr>
<td></td>
<td>455</td>
</tr>
<tr>
<td>Asp</td>
<td>Gly Asp Arg Glu Pro Glu Gly Thr Leu Pro Ala Glu Ala Pro</td>
</tr>
<tr>
<td></td>
<td>470</td>
</tr>
<tr>
<td>Pro</td>
<td>Pro Pro Pro Leu Gly Ala Ala Arg Glu Glu Pro Thr Pro Glu</td>
</tr>
<tr>
<td></td>
<td>485</td>
</tr>
<tr>
<td>Ala</td>
<td>Val Ala Thr Glu Gly Val Phe Val Thr Glu Leu Asp Gly Thr</td>
</tr>
<tr>
<td></td>
<td>495</td>
</tr>
<tr>
<td>Arg</td>
<td>Thr Glu Asp Leu Glu Thr Ile Arg Leu Glu Thr Lys Glu Thr</td>
</tr>
<tr>
<td></td>
<td>510</td>
</tr>
<tr>
<td>Phe</td>
<td>Cys Ile Asp Gly Leu Pro Asp Leu Glu Asp Asp Glu Thr</td>
</tr>
<tr>
<td></td>
<td>520</td>
</tr>
<tr>
<td>Gly</td>
<td>Lys Ser Leu Glu Asp Gln Asn Met Cys Phe Pro Lys Val Glu</td>
</tr>
<tr>
<td></td>
<td>530</td>
</tr>
<tr>
<td>Val</td>
<td>Ile Ser Ser Leu Ser Asp Ser Asp Pro Glu Leu Asp Tyr</td>
</tr>
<tr>
<td></td>
<td>540</td>
</tr>
<tr>
<td>Thr</td>
<td>Ser Leu Pro Val Leu Glu Asn Leu Pro Thr Asp Thr Leu Ser</td>
</tr>
<tr>
<td></td>
<td>550</td>
</tr>
<tr>
<td>Asn</td>
<td>Ile Phe Ala Val Ser Lys Asp Pro Pro Ser</td>
</tr>
<tr>
<td></td>
<td>560</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Met</td>
<td>Val</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Gln</td>
<td>Glu</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Ala</td>
<td>Ser</td>
</tr>
<tr>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Phe</td>
<td>Gly</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Ser</td>
<td>Thr</td>
</tr>
<tr>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>Ile</td>
<td>Lys</td>
</tr>
<tr>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Thr</td>
<td>Gly</td>
</tr>
<tr>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>Gln</td>
<td>Leu</td>
</tr>
<tr>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td>Val</td>
<td>Cys</td>
</tr>
<tr>
<td>125</td>
<td>125</td>
</tr>
<tr>
<td>Gly</td>
<td>Ser</td>
</tr>
<tr>
<td>140</td>
<td>140</td>
</tr>
<tr>
<td>Ala</td>
<td>Trp</td>
</tr>
<tr>
<td>155</td>
<td>155</td>
</tr>
<tr>
<td>Ala</td>
<td>Asn</td>
</tr>
<tr>
<td>170</td>
<td>170</td>
</tr>
<tr>
<td>Lys</td>
<td>Arg</td>
</tr>
<tr>
<td>185</td>
<td>185</td>
</tr>
<tr>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Gly</td>
<td>Gln</td>
</tr>
<tr>
<td>215</td>
<td>215</td>
</tr>
<tr>
<td>Arg</td>
<td>Val</td>
</tr>
<tr>
<td>230</td>
<td>230</td>
</tr>
<tr>
<td>Gly</td>
<td>Tyr</td>
</tr>
<tr>
<td>245</td>
<td>245</td>
</tr>
<tr>
<td>Ala</td>
<td>Trp</td>
</tr>
<tr>
<td>260</td>
<td>260</td>
</tr>
<tr>
<td>Asn</td>
<td>Asn</td>
</tr>
<tr>
<td>275</td>
<td>275</td>
</tr>
<tr>
<td>Val</td>
<td>Val</td>
</tr>
<tr>
<td>290</td>
<td>290</td>
</tr>
<tr>
<td>Lys</td>
<td>Thr</td>
</tr>
<tr>
<td>305</td>
<td>305</td>
</tr>
<tr>
<td>Gln</td>
<td>Ser</td>
</tr>
<tr>
<td>320</td>
<td>320</td>
</tr>
<tr>
<td>Asp</td>
<td>Val</td>
</tr>
<tr>
<td>335</td>
<td>335</td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
</tr>
<tr>
<td>350</td>
<td>350</td>
</tr>
<tr>
<td>Lys</td>
<td>Lys</td>
</tr>
<tr>
<td>365</td>
<td>365</td>
</tr>
<tr>
<td>Ile</td>
<td>Asp</td>
</tr>
<tr>
<td>380</td>
<td>380</td>
</tr>
<tr>
<td>Arg</td>
<td>Cys</td>
</tr>
</tbody>
</table>
Asp Met Lys Glu Val Ile Glu Ile Asp Gln Phe Ser Tyr Pro Val
Tyr Arg Ala Phe Leu Gln Tyr Leu Tyr Thr Asp Thr Val Asp Leu
410
415
420
425
430
435
Pro Pro Glu Asp Ala Ile Gly Leu Leu Asp Leu Ala Thr Ser Tyr
440
445
450
Cys Glu Asn Arg Leu Lys Lys Leu Cys Gln His Ile Ile Lys Arg
455
460
465
Gly Ile Thr Val Glu Asn Ala Phe Ser Leu Phe Ser Ala Ala Val
470
475
480
Arg Tyr Asp Ala Glu Leu Glu Glu Phe Cys Phe Lys Phe Cys
485
490
495
Ile Asn His Leu Thr Glu Val Thr Gln Thr Ala Ala Phe Trp Gln
500
505
510
Met Asp Gly Pro Leu Leu Lys Glu Phe Ile Ala Lys Ala Ser Lys
515
520
525
Cys Gly Ala Phe Lys Asn
530

<210> 6
<211> 369
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2676571CD1

<400> 6
Met Thr Leu Glu Ala Ile Arg Tyr Ser Arg Gly Ser Leu Gln Ile
1
5
10
15
Leu Asp Gln Leu Leu Leu Pro Lys Gln Ser Arg Tyr Glu Ala Val
20
25
30
Gly Ser Val His Gln Ala Trp Glu Ala Ile Arg Ala Met Lys Val
35
40
45
Arg Gly Ala Pro Ala Ile Ala Leu Val Gly Cys Leu Ser Leu Ala
50
55
60
Val Glu Leu Gln Ala Gly Ala Gly Gly Pro Gly Leu Ala Ala Leu
65
70
75
Val Ala Phe Val Arg Asp Lys Leu Ser Phe Leu Val Thr Ala Arg
80
85
90
Pro Thr Ala Val Met Arg Met Ala Arg Asp Leu Ala Asp
95
100
105
Val Ala Ala Arg Glu Ala Glu Arg Glu Ala Thr Glu Glu Ala
110
115
120
Val Arg Glu Arg Val Ile Cys Cys Thr Glu Asp Met Leu Glu Lys
125
130
135
Asp Leu Arg Asp Arg Arg Ser Ile Gly Asp Leu Gly Ala Arg His
140
145
150
Leu Leu Glu Arg Val Ala Pro Ser Gly Gly Lys Val Thr Val Leu
155
160
165
Thr His Cys Asn Thr Gly Ala Leu Ala Thr Ala Gly Tyr Thr
170
175
180
Ala Leu Gly Val Ile Arg Ser Leu His Ser Leu Gly Arg Leu Glu
185
190
195
His Ala Phe Cys Thr Glu Thr Arg Pro Tyr Asn Gln Gly Ala Arg
200
205
210
Leu Thr Ala Phe Glu Leu Val Tyr Glu Ile Pro Ala Thr Leu
215
220
225
Ile Thr Asp Ser Met Val Ala Ala Ala Met Ala His Arg Gly Val
230
235
240
Ser Ala Val Val Val Gly Ala Asp Arg Val Val Ala Asn Gly Asp

11/30
Thr Ala Asn Lys Val Gly Thr Tyr Gln Leu Ala Ile Val Ala Lys
245 250 255
Thr Ala Asn Lys Val Gly Thr Tyr Gln Leu Ala Ile Val Ala Lys
260 265 270
His His Gly Ile Pro Phe Tyr Val Ala Ala Pro Ser Ser Ser Cys
275 280 285
Asp Leu Arg Leu Glu Thr Gly Lys Glu Ile Ile Ile Glu Glu Arg
290 295 300
Pro Gly Gln Glu Leu Thr Asp Val Asn Gly Val Arg Ile Ala Ala
305 310 315
Pro Gly Ile Gly Val Trp Asn Pro Ala Phe Asp Val Thr Pro His
320 325 330
Asp Leu Ile Thr Gly Gly Ile Ile Thr Glu Leu Gly Val Phe Ala
335 340 345
Pro Glu Glu Leu Arg Thr Ala Leu Thr Thr Thr Ile Ser Ser Arg
350 355 360
Asp Gly Thr Leu Asp Gly Pro Gln Met
365

<210> 7
<211> 860
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1725129CD1

<400> 7
Met Gly Val Leu Ser Ala Ile Pro Ala Leu Ala Pro Pro Ala Pro
1  5  10  15
Pro Pro Pro Ile Gln Gly Tyr Ala Phe Lys Pro Pro Pro Arg Pro
20 25 30
Asp Phe Gly Thr Ser Gly Arg Thr Ile Lys Leu Gln Ala Asn Phe
35 40 45
Phe Glu Met Asp Ile Pro Lys Ile Asp Ile Tyr His Tyr Glu Leu
50 55 60
Asp Ile Lys Pro Glu Lys Cys Pro Arg Arg Val Asn Arg Glu Ile
65 70 75
Val Glu His Met Val Gln His Phe Lys Thr Gln Ile Phe Gly Asp
80 85 90
Arg Lys Pro Val Phe Asp Gly Arg Lys Asn Leu Tyr Thr Ala Met
95 100 105
Pro Leu Pro Ile Gly Arg Asp Lys Val Glu Leu Val Thr Leu
110 115 120
Pro Gly Glu Gly Lys Asp Arg Ile Phe Lys Val Ser Ile Lys Trp
125 130 135
Val Ser Cys Val Ser Leu Gln Ala Leu His Asp Ala Leu Ser Gly
140 145 150
Arg Leu Pro Ser Val Pro Phe Glu Thr Ile Gln Ala Leu Asp Val
155 160 165
Val Met Arg His Leu Pro Ser Met Arg Tyr Thr Pro Val Gly Arg
170 175 180
Ser Phe Phe Thr Ala Ser Glu Gly Cys Ser Asn Pro Leu Gly Glu
185 190 195
Gly Arg Glu Val Trp Phe Gly Phe His Glu Ser Val Arg Pro Ser
200 205 210
Leu Trp Lys Met Met Leu Asn Ile Asp Val Ser Ala Thr Ala Phe
215 220 225
Tyr Lys Ala Gln Pro Val Ile Glu Phe Val Cys Glu Val Leu Asp
230 235 240
Phe Lys Ser Ile Glu Gln Gln Lys Pro Leu Thr Asp Ser Gln
245 250 255
Arg Val Lys Phe Thr Lys Glu Ile Lys Gly Leu Lys Val Glu Ile

12/30
<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>260</td>
<td>270</td>
<td>Thr His Cys Gly Gln Met Lys Arg Lys Tyr Val Cys Asc Val</td>
</tr>
<tr>
<td>275</td>
<td>285</td>
<td>Arg Arg Arg Pro Ala Ser His Gln Thr Phe Pro Leu Gln Gln Glu</td>
</tr>
<tr>
<td>290</td>
<td>300</td>
<td>Ser Gly Gln Thr Val Glu Cys Thr Val Ala Gln Tyr Phe Lys Asp</td>
</tr>
<tr>
<td>305</td>
<td>315</td>
<td>Arg His Lys Leu Val Leu Arg Tyr Pro His Leu Pro Cys Leu Gln</td>
</tr>
<tr>
<td>320</td>
<td>330</td>
<td>Val Gly Gln Gln Gln Lys His Thr Tyr Leu Pro Leu Glu Val Cys</td>
</tr>
<tr>
<td>335</td>
<td>345</td>
<td>Asn Ile Val Ala Gly Gln Arg Cys Ile Lys Leu Thr Asp Asc</td>
</tr>
<tr>
<td>350</td>
<td>360</td>
<td>Gln Thr Ser Thr Met Ile Arg Ala Thr Ala Arg Ser Ala Pro Asp</td>
</tr>
<tr>
<td>365</td>
<td>375</td>
<td>Arg Gln Glu Glu Ile Ser Lys Leu Met Arg Ser Ala Ser Phe Asc</td>
</tr>
<tr>
<td>380</td>
<td>390</td>
<td>Thr Asp Pro Tyr Val Arg Glu Phe Gly Ile Met Val Lys Asp Glu</td>
</tr>
<tr>
<td>395</td>
<td>405</td>
<td>Met Thr Asp Val Thr Gly Arg Val Leu Gln Pro Pro Ser Ile Leu</td>
</tr>
<tr>
<td>410</td>
<td>420</td>
<td>Tyr Gly Arg Asn Lys Ala Lys Ala Thr Pro Val Glu Glu Val</td>
</tr>
<tr>
<td>425</td>
<td>435</td>
<td>Trp Asp Met Arg Asn Lys Gln Phe His Thr Gly Ile Glu Ile Lys</td>
</tr>
<tr>
<td>440</td>
<td>450</td>
<td>Val Trp Ala Ile Ala Cys Phe Ala Pro Gln Arg Glu Cys Thr Glu</td>
</tr>
<tr>
<td>455</td>
<td>465</td>
<td>Val His Leu Lys Ser Phe Thr Glu Gln Leu Arg Lys Ile Ser Arg</td>
</tr>
<tr>
<td>470</td>
<td>480</td>
<td>Asp Ala Gly Met Pro Ile Gln Gly Glu Pro Cys Phe Cys Lys Tyr</td>
</tr>
<tr>
<td>485</td>
<td>495</td>
<td>Ala Gln Gly Ala Asp Ser Val Glu Pro Met Phe Arg His Leu Lys</td>
</tr>
<tr>
<td>500</td>
<td>510</td>
<td>Asn Thr Tyr Ala Gly Leu Gln Leu Val Val Ile Leu Pro Gly</td>
</tr>
<tr>
<td>515</td>
<td>525</td>
<td>Lys Thr Pro Val Tyr Ala Glu Val Lys Arg Val Gly Asp Thr Val</td>
</tr>
<tr>
<td>530</td>
<td>540</td>
<td>Leu Gly Met Ala Thr Gln Cys Val Gln Met Lys Asn Val Gln Arg</td>
</tr>
<tr>
<td>545</td>
<td>555</td>
<td>Thr Thr Pro Gln Thr Leu Ser Asc Leu Cys Leu Lys Ile Asn Val</td>
</tr>
<tr>
<td>560</td>
<td>570</td>
<td>Lys Leu Gly Glu Val Asn Asn Ile Leu Leu Pro Gln Gly Arg Pro</td>
</tr>
<tr>
<td>575</td>
<td>585</td>
<td>Pro Val Phe Gln Gln Pro Val Ile Phe Leu Gly Ala Asp Val Thr</td>
</tr>
<tr>
<td>590</td>
<td>600</td>
<td>His Pro Pro Ala Gly Asp Gly Lys Lys Pro Ser Ile Ala Ala Val</td>
</tr>
<tr>
<td>605</td>
<td>615</td>
<td>Val Gly Ser Met Asp Ala His Pro Asc Arg Tyr Cys Ala Thr Val</td>
</tr>
<tr>
<td>620</td>
<td>630</td>
<td>Arg Val Gln Gln His Arg Glu Glu Ile Glu Asp Leu Ala Ala</td>
</tr>
<tr>
<td>635</td>
<td>645</td>
<td>Met Val Arg Glu Leu Leu Ile Gln Phe Tyr Lys Ser Thr Arg Phe</td>
</tr>
<tr>
<td>650</td>
<td>660</td>
<td>Lys Pro Thr Arg Ile Phe Tyr Arg Asp Gly Val Ser Glu Gly</td>
</tr>
<tr>
<td>665</td>
<td>675</td>
<td>Gln Phe Gln Gln Val Leu His His Glu Leu Leu Ala Ile Arg Glu</td>
</tr>
<tr>
<td>680</td>
<td>690</td>
<td>Ala Cys Ile Lys Leu Glu Lys Asp Tyr Gln Pro Gly Ile Thr Phe</td>
</tr>
<tr>
<td>695</td>
<td>705</td>
<td>Ile Val Val Gln Lys Arg His Thr Arg Leu Phe Cys Thr Asp</td>
</tr>
<tr>
<td>710</td>
<td>720</td>
<td>Lys Asn Glu Arg Val Gly Lys Ser Gly Asc Ile Pro Ala Gly Thr</td>
</tr>
<tr>
<td>725</td>
<td>735</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Val</td>
<td>Asp</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>740</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Cys</td>
<td>Ser</td>
</tr>
<tr>
<td>755</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>His</td>
<td>Val</td>
</tr>
<tr>
<td>770</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Ile</td>
<td>Leu</td>
</tr>
<tr>
<td>785</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Ser</td>
<td>Val</td>
</tr>
<tr>
<td>800</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Phe</td>
<td>Arg</td>
</tr>
<tr>
<td>815</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Glu</td>
<td>Gly</td>
</tr>
<tr>
<td>830</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td>845</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Met</td>
<td>Tyr</td>
</tr>
<tr>
<td>860</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Atala Ser Gly Gln Leu Pro Ser Ala Ser Met Arg Leu Lys Phe Leu
260 265 270
Arg Leu His Gly Gly Met Glu Gln Glu Glu Arg Thr Ala Val Phe
275 280 285
Gln Glu Phe Ser His Ser Arg Arg Gly Val Leu Leu Cys Thr Asp
290 295 300
Val Ala Ala Arg Gly Leu Asp Leu Pro Glu Val Thr Trp Ile Val
305 310 315
Gln Tyr Asn Ala Pro Ser Ser Pro Ala Gly Tyr Ile His Arg Ile
320 325 330
Gly Arg Thr Ala Arg Ile Gly Cys His Gly Ser Ser Leu Leu Ile
335 340 345
Leu Ala Pro Ser Glu Ala Glu Tyr Val Asn Ser Leu Ala Ser His
350 355 360
Lys Ile Asn Val Ser Glu Ile Lys Met Glu Asp Ile Leu Cys Val
365 370 375
Leu Thr Arg Asp Asp Cys Phe Lys Gly Lys Asp Arg Trp Gly Ala Gln
380 385 390
Lys Ser His Ala Val Gly Pro Gln Glu Ile Arg Glu Arg Ala Thr
395 400 405
Val Leu Gln Thr Val Phe Glu Asp Tyr Val His Ser Ser Glu Arg
410 415 420
Arg Val Ser Trp Ala Lys Ala Leu Gln Ser Phe Ile Gln Ala
425 430 435
Tyr Ala Thr Tyr Pro Arg Glu Leu Lys His Ile Phe His Val Arg
440 445 450
Ser Leu His Leu Gly His Val Ala Lys Ser Phe Gly Leu Arg Asp
455 460 465
Ala Pro Arg Asn Leu Ser Ala Leu Thr Arg Lys Lys Arg Lys Ala
470 475 480
Hist Val Lys Arg Pro Asp Leu His Lys Thr Gln Ser Lys His
485 490 495
Ser Leu Ala Glu Ile Leu Arg Ser Glu Tyr Ser Ser Gly Met Glu
500 505 510
Ala Asp Val Ala Lys Val Lys Lys Glu Asn Ala Pro Gly Glu Pro
515 520 525
Gly Gly Arg Pro Leu Gln His Ser Leu Gln Pro Thr Pro Cys Phe
530 535 540
Gly Arg Gly Lys Thr Leu Lys Trp Arg Lys Thr Gln Lys Gly Val
545 550 555
Gln Arg Asp Ser Lys Thr Ser Gln Lys Val
560 565

<210> 9
<211> 565
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 429930CD1

<400> 9
Met Asn Ile Phe Gln Lys Asp Ile Asn Cys Glu Arg Tyr Tyr Ser
1 10 15
Phe Ile His Leu Ser Phe Gln Glu Phe Phe Ala Ala Met Tyr Tyr
20 25 30
Ile Leu Asp Glu Gly Glu Gly Gly Ala Gly Pro Asp Glu Asp Val
35 40 45
Thr Arg Leu Leu Thr Glu Tyr Ala Phe Ser Glu Arg Ser Phe Leu
50 55 60
Ala Leu Thr Ser Arg Phe Leu Phe Gly Leu Leu Asn Glu Glu Thr
65 70 75

15/30
Arg Ser His Leu Glu Lys Ser Leu Cys Trp Lys Val Ser Pro His
  80  85  90
Ile Lys Met Asp Leu Leu Glu Trp Ile Gln Ser Lys Ala Gln Ser
  95 100 105
Asp Gly Ser Thr Leu Gln Gly Ser Leu Glu Phe Phe Ser Cys
 110 115 120
Leu Tyr Glu Ile Glu Glu Glu Phe Ile Gln Gln Ala Leu Ser
 125 130 135
His Phe Gln Val Ile Val Val Ser Asn Ile Ala Ser Lys Met Glu
 140 145 150
His Met Val Ser Ser Phe Cys Leu Lys Arg Cys Arg Ser Ala Gln
 155 160 165
Val Leu His Leu Tyr Gly Ala Thr Tyr Ser Ala Asp Gly Glu Asp
 170 175 180
Arg Ala Arg Cys Ser Ala Gly Ala His Thr Leu Leu Val Gln Leu
 185 190 195
Pro Glu Arg Thr Val Leu Leu Asp Ala Tyr Ser Glu His Leu Ala
 200 205 210
Ala Ala Leu Cys Thr Asn Pro Asn Leu Ile Glu Leu Ser Leu Tyr
 215 220 225
Arg Asn Ala Leu Gly Ser Arg Gly Val Lys Leu Leu Cys Gin Gly
 230 235 240
Leu Arg His Pro Asn Cys Lys Leu Gin Asn Leu Arg Leu Lys Arg
 245 250 255
Cys Arg Ile Ser Ser Ser Ala Cys Glu Asp Leu Ser Ala Ala Leu
 260 265 270
Ile Ala Asn Lys Asn Leu Thr Arg Met Asp Leu Ser Gly Asn Gly
 275 280 285
Val Gly Phe Pro Gly Met Met Leu Leu Cys Glu Gly Leu Arg His
 290 295 300
Pro Gln Cys Arg Leu Gln Met Ile Gln Leu Arg Lys Cys Gin Leu
 305 310 315
Glu Ser Gly Ala Cys Gin Glu Met Ala Ser Val Leu Gly Thr Asn
 320 325 330
Pro His Leu Val Glu Leu Asp Leu Thr Gly Asn Ala Leu Glu Asp
 335 340 345
Leu Gly Leu Arg Leu Leu Cys Gin Gly Leu Arg His Pro Val Cys
 350 355 360
Arg Leu Arg Thr Leu Trp Leu Lys Ile Cys Arg Leu Thr Ala Ala
 370 375
Ala Cys Asp Glu Leu Ala Ser Thr Leu Ser Val Asn Gin Ser Leu
 380 385 390
Arg Glu Leu Asp Leu Ser Leu Asn Leu Glu Gly Asp Leu Gly Val
 395 400 405
Leu Leu Leu Cys Glu Gly Leu Arg His Pro Thr Cys Lys Leu Gln
 410 415 420
Thr Leu Arg Leu Gly Ile Cys Arg Leu Gly Ser Ala Ala Cys Glu
 425 430 435
Gly Leu Ser Val Val Leu Gln Ala Asn His Asn Leu Arg Glu Leu
 440 445 450
Asp Leu Ser Phe Asn Asp Leu Gly Asp Trp Gly Leu Trp Leu Leu
 455 460 465
Ala Glu Gly Leu Gin His Pro Ala Cys Arg Leu Gln Lys Leu Trp
 470 475 480
Leu Asp Ser Cys Gly Leu Thr Ala Lys Ala Cys Gin Lys Gin Leu Tyr
 485 490 495
Phe Thr Leu Gly Ile Asn Gln Thr Leu Thr Asp Leu Tyr Leu Thr
 500 505 510
Asn Asn Ala Leu Gly Asp Thr Gly Val Arg Leu Leu Cys Gin Arg
 515 520 525
Leu Ser His Pro Gly Cys Lys Leu Arg Val Leu Trp Leu Phe Gly
 530 535 540
Met Asp Leu Asn Lys Met Thr His Ser Arg Leu Ala Ala Leu Arg

16/30
Val Thr Lys Pro Tyr Leu Asp Ile Gly Cys
545 550 555

<210> 10
<211> 1488
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7504129CD1

<400> 10
Met Gly Val Thr Ser Ala Ala Ser Gly Leu Val Gly Ser Ala Pro 1
5 10 15
Gln Cys Val Ala Leu Pro Ser Glu Gly Trp Thr Leu Ala Val Trp 20
25 30
Pro Val Ala Ala Cys Thr Cys Ser Gly Val Gly Ser Ser Pro Lys 35
40 45
Leu Thr Pro Gly Ser Phe Val His Cys Pro Trp Phe Leu Leu Leu 50
55 60
Thr Glu Ala Thr Arg Ala Gly Ile Lys Arg Pro Phe Ser Lys Ala 65
70 75
Glu Leu Lys Ala Ser Val Arg Pro Met Glu Gly Ser His Cys Trp 80
85 90
Gly Gly Glu Ala Arg Arg Thr Ser Glu Gly His Thr Glu Gln Lys 95
100 105
Gly Glu Leu Arg Ala Arg Glu Ala Gln Asn Glu Asp Glu Asp 110
115 120
Val Lys Glu Val Trp Val Gly Lys Thr Lys Glu Glu Ser Asp 125
130 135
Lys Leu Gly Cys Glu Gly Ala Trp Val Pro Ser Arg Val Val Pro Ser 140
145 150
Trp Ile Gly Arg Phe Glu Leu Trp Val Val Gly Tyr Trp Glu Gln 155
160 165
Pro Ala Val Ser Trp Glu Met Gln Val Arg Leu Leu Arg Leu 170
175 180
Arg Ala Ala Leu Thr Leu Leu Gly Val Pro Arg Arg Pro Ala 185
190 195
Ser Arg Gly Val Pro Glu Ser Arg Thr Glu Gly Leu Lys Ser Gly Gln 200
205 210
Ala Arg Thr Asp Ser Ala Trp Arg Ala Leu Thr Val Ile Ser 215
220 225
Thr Ser Pro Gly Thr Ser Arg Met Asp Arg Val Ala Leu Val Ala 230
235 240
Val Gly Gly Pro Arg Arg Glu Pro Gly Gly His Thr Leu Gln Arg 245
250 255
Leu Pro Val Ala Leu Arg Thr Leu Ile Pro Ala Asp Gln Ala His 260
265 270
Gln Ala Pro Asn Ser Thr Thr Trp Leu Gly Ser Arg Ala Gly Leu 275
280 285
Leu Ala Leu Ala Ala Gly Leu Gly Ile Arg Asp Ser Ala Ala 290
295 300
His Gly Pro Leu Gln Val Gly Gly Met Gly Gly Tyr Thr Gly Met 305
310 315
Gly Ser Glu Val Arg Trp Glu Lys Glu His Glu Asp Gly Val 320
325 330
Lys Trp Arg Glu Leu Gly His Lys Gly Pro Tyr Phe Ala Pro Pro 335
340 345
Tyr Glu Pro Leu Pro Asp Gly Val Arg Phe Phe Tyr Glu Gly Arg 350
355 360
Pro Val Arg Leu Ser Val Ala Ala Glu Val Val Ala Thr Phe Tyr
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>365</td>
<td>Gly Arg Met Leu Asp His Glu Tyr Thr Thr Lys Glu Val Phe Arg</td>
</tr>
<tr>
<td>380</td>
<td>Lys Asn Phe Phe Asn Asp Trp Arg Lys Glu Met Ala Val Glu Glu</td>
</tr>
<tr>
<td>395</td>
<td>Arg Glu Val Ile Lys Ser Leu Asp Lys Cys Asp Phe Thr Glu Ile</td>
</tr>
<tr>
<td>410</td>
<td>His Arg Tyr Phe Val Asp Lys Ala Ala Ala Arg Lys Val Leu Ser</td>
</tr>
<tr>
<td>425</td>
<td>Arg Glu Glu Lys Gln Lys Leu Lys Glu Glu Ala Glu Lys Leu Gln</td>
</tr>
<tr>
<td>440</td>
<td>Gln Glu Phe Gly Tyr Cys Ile Leu Asp Gly His Gln Glu Lys Ile</td>
</tr>
<tr>
<td>455</td>
<td>Gly Asn Phe Lys Ile Glu Pro Pro Gly Leu Phe Arg Gly Arg Gly</td>
</tr>
<tr>
<td>470</td>
<td>Asp His Pro Lys Met Gly Met Leu Lys Arg Arg Ile Thr Pro Glu</td>
</tr>
<tr>
<td>485</td>
<td>Asp Val Val Ile Asn Cys Ser Arg Asp Ser Lys Ile Pro Glu Pro</td>
</tr>
<tr>
<td>500</td>
<td>Pro Ala Gly His Gln Trp Lys Glu Val Arg Ser Asp Asn Thr Val</td>
</tr>
<tr>
<td>515</td>
<td>Thr Trp Leu Ala Ala Trp Thr Glu Ser Val Glu Asn Ser Ile Lys</td>
</tr>
<tr>
<td>530</td>
<td>Tyr Ile Met Leu Asn Pro Cys Ser Lys Leu Lys Gly Glu Thr Ala</td>
</tr>
<tr>
<td>545</td>
<td>Trp Gln Lys Phe Glu Thr Ala Arg Arg Leu Arg Gly Phe Val Asp</td>
</tr>
<tr>
<td>560</td>
<td>Glu Ile Arg Ser Gln Tyr Arg Ala Asp Trp Lys Ser Arg Glu Met</td>
</tr>
<tr>
<td>575</td>
<td>Lys Thr Arg Gln Arg Ala Val Ala Leu Tyr Phe Ile Asp Lys Leu</td>
</tr>
<tr>
<td>590</td>
<td>Ala Leu Arg Ala Gly Asn Glu Lys Glu Asp Gly Glu Ala Ala Asp</td>
</tr>
<tr>
<td>605</td>
<td>Thr Val Gly Cys Cys Ser Leu Arg Val Glu His Val Glu Leu His</td>
</tr>
<tr>
<td>620</td>
<td>Pro Glu Ala Asp Gly Cys Gln His Val Glu Phe Asp Phe Leu</td>
</tr>
<tr>
<td>635</td>
<td>Gly Lys Asp Cys Ile Arg Tyr Tyr Asn Arg Val Pro Val Glu Lys</td>
</tr>
<tr>
<td>650</td>
<td>Pro Val Tyr Lys Asn Leu Gln Leu Phe Met Glu Asn Lys Asp Pro</td>
</tr>
<tr>
<td>665</td>
<td>Arg Asp Arg Leu Phe Asp Arg Leu Thr Thr Thr Ser Leu Asn Lys</td>
</tr>
<tr>
<td>680</td>
<td>His Leu Gln Glu Leu Met Asp Gly Leu Thr Ala Lys Val Phe Arg</td>
</tr>
<tr>
<td>695</td>
<td>Thr Tyr Asn Ala Ser Ile Thr Leu Gln Glu Glu Leu Arg Ala Leu</td>
</tr>
<tr>
<td>710</td>
<td>Thr Arg Ala Glu Asp Ser Ile Ala Ala Lys Ile Leu Ser Tyr Asn</td>
</tr>
<tr>
<td>725</td>
<td>Arg Ala Asn Arg Val Val Ala Ile Leu Cys Asn His Gln Arg Ala</td>
</tr>
<tr>
<td>740</td>
<td>Thr Pro Ser Thr Phe Glu Lys Ser Met Glu Asn Leu Glu Thr Lys</td>
</tr>
<tr>
<td>755</td>
<td>Ile Gln Ala Lys Lys Glu Gln Val Ala Glu Ala Arg Ala Glu Leu</td>
</tr>
<tr>
<td>770</td>
<td>Arg Arg Ala Arg Ala Glu His Lys Ala Glu Gly Asp Gly Lys Ser</td>
</tr>
<tr>
<td>785</td>
<td>Arg Ser Val Leu Glu Lys Arg Arg Leu Leu Glu Lys Leu Gln</td>
</tr>
<tr>
<td>800</td>
<td>Glu Gln Leu Ala Gln Leu Ser Val Gln Ala Thr Asp Lys Glu Glu</td>
</tr>
<tr>
<td>815</td>
<td>Asn Lys Gln Val Ala Leu Gly Thr Ser Lys Leu Asn Tyr Leu Asp</td>
</tr>
</tbody>
</table>
Pro Arg Ile Ser Ile Ala Trp Cys Lys Arg Phe Arg Val Val
845 850 855
Glu Lys Ile Tyr Ser Lys Thr Gln Arg Glu Ser Ser Tyr Asp Gly
860 865 870
Leu Leu Pro Glu Leu Pro Asn Pro Val Ser Ser Gln Ser Cys Leu
875 880 885
Ser Ala Ala Leu Cys His Ala Gly Ser Ser Ala Ala Leu Ser Ala
890 895 900
Thr Arg Asp Pro Leu Pro Pro Ser Ser Leu Pro Arg Gly Thr Leu
905 910 915
Cys Arg Pro Leu Cys His Ala Gly Pro Ser Ala Ala Leu Phe
920 925 930
Ser Ala Thr Arg Asp Pro Leu Pro Pro Ser Leu Pro Arg Thr
935 940 945
Leu Cys Arg Pro Leu Cys His Thr Gly Pro Ser Ala Ala Leu
950 955 960
Leu Cys His Ala Gly Pro Ser Ala Ala Leu Phe Ser Ala Met Arg
965 970 975
Asp Pro Leu Pro Pro Ser Ser Leu Pro Arg Gly Thr Leu Cys Cys
980 985 990
Pro Leu Leu Cys His Val Gly Pro Ser Ala Ala Leu Ser Ala Thr
995 1000 1005
Arg Asp Pro Leu Pro Pro Ser Pro Leu Pro Arg Gly Ile Leu Cys
1010 1015 1020
Cys Pro Leu Leu Cys His Val Gly Ser Ser Ala Met Arg Asp Pro
1030 1035 1040
Leu Pro Pro Ser Pro Leu Pro Pro Ser Leu Pro Pro Ser Ser
1040 1045 1050
Leu Pro Arg Gly Ile Leu Cys Arg Pro Leu Leu Cys His Ala Gly
1055 1060 1065
Thr Ser Ala Ala Leu Ser Ser Ala Thr Arg Asp Pro Leu Pro Pro
1070 1075 1080
Ser Pro Leu Pro Arg Gly Thr Leu Cys Arg Pro Leu Leu Cys His
1085 1090 1095
Ala Gly Pro Ser Ala Ala Leu Ser Ala Thr Gln Asp Pro Leu Pro
1100 1105 1110
Pro Ser Ser Leu Pro Arg Gly Thr Leu Cys Cys Pro Leu Leu Cys
1115 1120 1125
Arg Pro Leu Leu Cys Arg Ala Gly Pro Ser Ala Ala Leu Ser Ser
1130 1135 1140
Ala Thr Arg Asp Pro Leu Pro Pro Ser Pro Leu Pro Arg Gly Ile
1145 1150 1155
Leu Cys His Pro Leu Cys His Ala Gly Pro Ser Ala Ala Leu Phe
1160 1165 1170
Ser Ala Thr Arg Asp Pro Leu Leu Pro Ser Pro Leu Pro Arg Gly
1175 1180 1185
Thr Leu Cys Cys Pro Leu Cys His Ala Gly Pro Ser Ala Ala Leu
1190 1195 1200
Ser Ser Ala Thr Arg Asp Pro Leu Pro Pro Ser Ala Met Arg Asp
1205 1210 1215
Pro Leu Leu Pro Ser Leu Pro Arg Gly Thr Leu Cys Arg Pro
1220 1225 1230
Leu Cys His Ala Gly Pro Ser Ala Ala Leu Phe Ser Ala Thr Leu
1235 1240 1245
Phe Ser Ala Thr Trp Asp Pro Leu Leu Pro Phe Pro Leu Leu Arg
1250 1255 1260
Arg Ala Leu Cys Arg Pro Leu Cys His Thr Gly Pro Ser Ala Ala
1265 1270 1275
Leu Ser Ser Ala Thr Trp Asp Pro Leu Leu Pro Ser Ser Leu Pro
1280 1285 1290
Arg Gly Ile Leu Cys His Ala Gly Pro Ser Ala Ala Leu Phe Ser
1295 1300 1305
Ala Thr Trp Asp Pro Leu Leu Pro Ser Ala Thr Arg Asp Pro Leu
1310 1315 1320
Pro Pro Ser Pro Leu Leu Arg Gly Ile Leu Cys Cys Pro Leu Leu
1325 1330 1335
Cys His Val Gly Ser Ser Ala Ile Cys Ser Ala Leu Cys Phe Arg
1340 1345 1350
Gly Cys Leu Cys Gly Cys Leu Pro His Thr Gly Ala Pro Ala Asn
1355 1360 1365
Pro Met Arg Glu Pro Leu Cys Lys Ala Pro Glu Lys Ser Trp Val
1370 1375 1380
Val Pro Ser Thr Gly Lys Ser Ser Gly Lys Ala Ser Cys Val Cys
1385 1390 1395
Trp Cys Glu Met Arg Phe His Asp Gly Gly Cys Arg Ala Leu Ser
1400 1405 1410
Leu Ser Cys Tyr Pro Val Ala Cys Glu Pro Ser Leu Lys Cys Thr
1415 1420 1425
Lys Ser Ser Lys Gly Leu Arg Ala Thr Ser Phe Thr Ala Ala Gln
1430 1435 1440
Ala Glu Lys Gln Ala Leu Gln Leu Glu Leu Gly Asn Arg
1445 1450 1455
Pro Gly Gly Asp Ser His Trp Ile Pro Leu Ser His Asp His Lys
1460 1465 1470
Asn Arg Ala Thr Ala Arg Val Ser Arg Ala Thr Asp Ser Asn
1475 1480 1485
Arg Glu Pro

<210> 11
<211> 4874
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3804141CB1

<400> 11
atggggtgtga cctcggtgctgc cagtggcgcgt gtgggctcttg gcccctcagcgt cg ttgctccttg 60
ccagtgaggg ggtggaaggct gcagcataggct cctggtcggcc cttgcctactct gcctgcattctgt 120
gggagcagcgc ccagacctgac cctcgctcct gcctgtcatct gcccctcgacttc tctctgctcttc 180
acagggccca caagggctgct gataacaggc cccatttctca aagctgtggtgatt aaccacattc 240
gttggcgccca tggagggcagc ccattgctgagc gggtgagggcg caacggagagc gagccagggc 300
cacacctgagc agaagaccacc gcctcgagcg ggagtggggtc cgagggcatgc cggagctggagc 360
gtaaaagagag tgttggtgtgg ggagaaaccacc aagagacaaga aagaagatgg tctgagctcag 420
gagacccctcg tgcctccccg gattccgctcg ctctgtcgagt gttgtgctgtgc tctctgtgcgt 480
ggtactggag gcgtcgcccc gcctgtagcag cagtagcgcc tggcgccggc ggccttcgcc gggcttcgcgg 540
coggggcggccc tgcctcgcttc cgggggctcc gcggtcccttg ggtgcttgccg gcctgggcggcc 600
gtgccgagca ggcggcgccgg gcgtggtgccg ggcgtggaagc attcctcttc gcctggggcg ccctgggcaag 660
tccactctgcc tgcctgtccct gcctggtcgcgt gctgcgtgctg gatgcggcggct gacgattgctc 720
gttgtgggtct ccaggggctg ctctgagggc cagccctcctg cagttcctcg cagttcctcg tcctcctcg 780
agagacactc ccacgctccagg ccacagcagg ccacctgccc ggtccccac gacgcatctc ccctctctctc 840
tcaagtcggct gtcctgctcc ctctctctcc ggctgctgcc gatgcgggtct gccgctgagc 900
cagggggggcc tgcctggtcc ggcgtggcgg gcgtgggcttc cagcagcagc ggcgtggtgcc ggcgtggtgcc 960
tggggagagag aagagacagg ccctccagcag ccctccagcag ccctccagcag ccctccagcag gcgtgggcttc 1020
tactgccgct cccctccagc ggcctccagc gcctgccggt ggcctccagc gcctgccggt gcctgccggt 1080
ccctgtgaccgt gcctgctgccgc ggcgtggcgg gcgtgggcttc tttggtgctgcc gcctgccggt gcctgccggt 1140
catgaggtgc cagcagcagg gcgtgggcttc tttggtgctgcc gcctgccggt gcctgccggt gcctgccggt 1200
atgaggtgc cagcagcagg gcgtgggcttc tttggtgctgcc gcctgccggt gcctgccggt gcctgccggt 1260
cacagaggtgc cagcagcagg gcgtgggcttc tttggtgctgcc gcctgccggt gcctgccggt gcctgccggt 1320
agacatgcag ggcgtggtgc ggcgtggtgc ggcgtggtgc ggcgtggtgc ggcgtggtgc ggcgtggtgc 1380
cacccgaagaa cccggcggcag ggcgctcttc tgcctccagg gcgtgggtgc cgtggggcgc ggcgtggtgc 1440
gccctccagc gcgtgggcttc tttggtgctgcc gcctgccggt gcctgccggt gcctgccggt gcctgccggt 1500
tggacgagagg acctgagaggt cccggagcgc ccccggcggc gcggcgccgag gcgtgggcttc gcctgccggt 1560
tccgataccgc gcgtcgccagt gcgtgggcttc gcgtgggctgc gcctgccggt gcctgccggt gcctgccggt 1620

20/30
WO 02/072630
PCT/US02/03844

tacaccatgc tgaaccccttg ctctgaagcttg aagggggaga caagcttgagca gaagtttggaa 1680
aaccgctccag gctctgggagg attcgtggac ggatagctggt cccacgtacc cgcggctcgg gaagttgcttg 1740
aagctccgggg aataagcgac gagaacacgc cggctgggacc ggtctttcttg cgatagttcat 1800
gacattgag aaggaaatga gaaggggagac ggtgagctgg gcgcaccacg ctgggtctggct 1860
tcccctcggcg ttggagacat cgctgaccgc cccggagcgc atggcggctc aacgtagctgg 1920
gaatttgact tctgggagaa ggacagcttc cctgcagcata acagagttgcc ggtgggagaa 1980
ccggtgtcata agaactactta cgtctttcttg gagaacaaag aacccgggaga gcactctcctg 2040
gacacgctgtc ccattgacac gctgaaacag gacacctcaag agctctgatgg ggggctgacgc 2100
gccaagtgctt tggccacctc caagcccttc ataatcctctc agagcagact gcggccctgc 2160
acgcgggggc ctcctgaggg aggccgctag gcctggtgctg cacggtgctgc ggttgccagcc 2220
aattgtcttc cgtccccgac cgggcggcag cctcggctgg gatcggcgag aagctgtgctg 2280
ctgggagctg ggcggcagct ggcctggtgg cagctggcag actgtagcctg cctcgaggtg 2340
agagccctgc ggaggagagc gcggccccca ctgggggagc cagaaaggtc tcggccgctgc 2400
tctcagccgg aggccagcat agagcagtca aagatatctt ccaacccgac ccagccggct 2460
gtcggacttc ttcgaccaac caacgctgag ggtttgcagaa tctggagcag tctgcagtgc 2520
aatccctcga cgcagagatcc gcagcaagaa gacgagggtg tgagcggcag gggagagctg 2580
agagggcgga gggctgagca ccaagccgga ggggatgcgc agttccagaggg tgcctcgag 2640
aagaagatgc ggcgcctgtca gaagctgactg gacagccttg cggagccagct cttgcagggc 2700
acggaaacag aggagaaacg cgcagctggc cttgaggccgt ccaagctctca ctacgctgac 2760
cccaagatct gcgtccgctgt gcctgctcgg ctcctgccgt ccctctgccg ctcctgtagc 2820
aaaccaagcc ggggccagtc ttacagctgc ctctctgcgg cttcctgctg agctctccgt 2880
tctcagagt gcgtgctgtgc gcggcccttg ccctctcgcct ccctgctcgc ctcctgcctg 2940
acggccggac ccctcgccgc ctctctctcc cccctgccgg ccctctctct ccctctctct 3000
tgcagagcag ggcacccctc gggccccctt cccctcgcctt ctcctgctcgc ctcctgctcgc 3060
tcctcagca ggcctgctcc ctctgccgct tgcgcgcgcc gcgcgcgccg cgcgcgcgccg 3120
tcctcctcgac ccctccgctc ctctgccgct tgcgcgcgcc gcgcgcgccg cgcgcgcgccg 3180
tcctcctctcg gtcctgcttc ccctctctct ccctctctct ccctctctct ccctctctct 3240
gcgcctcctc gcacggcggc ccctctctct ccctctctct ccctctctct ccctctctct 3300
ttcctcctctcg ccctccctct ccctctctct ccctctctct ccctctctct ccctctctct 3360
tccctcctctcg ccctccctct ccctctctct ccctctctct ccctctctct ccctctctct 3420
tcctcctctcg ccctccctct ccctctctct ccctctctct ccctctctct ccctctctct 3480
tctcctcctctcg ccctccctct ccctctctct ccctctctct ccctctctct ccctctctct 3540
gcgcctcctc gcacggcggc ccctctctct ccctctctct ccctctctct ccctctctct 3600
tgcctcctcctcg ccctccctct ccctctctct ccctctctct ccctctctct ccctctctct 3660
gcccacgggg aatccacctcg gcctccctct gcctccctct gcctccctct gcctccctct 3720
tgcagagcag gcaccccttc gcgcctccct gcgcctccct gcgcctccct gcgcctccct 3780
tcctcctgca gcgagacacc cggctgctcc gcgcctccct gcgcctccct gcgcctccct 3840
tcctcctgcc gccgacccct gcgcctgctc gcgcctccct gcgcctccct gcgcctccct 3900
tcctcctgcc gccgacccct gcgcctgctc gcgcctccct gcgcctccct gcgcctccct 3960
tcctcctgcc gccgacccct gcgcctgctc gcgcctccct gcgcctccct gcgcctccct 4020
agagcgggtc gccacggcgg gcgcctgctc gcgcctccct gcgcctccct gcgcctccct 4080
tgcgcgctcct gcctgcgtct gcctgcgtct gcctgcgtct gcctgcgtct gcctgcgtct 4140
gccgccccctc tctctcggcct gcgtggaccc ccctgtgctt gcgtggagcc gcgtggagcc 4200
gcgcctccctgc ccctctctct gcctgcgtct gcctgcgtct gcctgcgtct gcctgcgtct 4260
tgcctccctgc ccctctctct gcctgcgtct gcctgcgtct gcctgcgtct gcctgcgtct 4320
agggcccccc gcggacagcc gcgcctctct gcgcctccct gcgcctccct gcgcctccct 4380
tgcagagcag gcaccccttc gcgcctccct gcgcctccct gcgcctccct gcgcctccct 4440
ttcctcctctcg ccctctctct ccctctctct ccctctctct ccctctctct ccctctctct 4500
agagcgggtc gccacggcgg gcgcctgctc gcgcctccct gcgcctccct gcgcctccct 4560
gcgcctccctgc ccctctctct gcctgcgtct gcctgcgtct gcctgcgtct gcctgcgtct 4620
acagctcagac ctgcacgctg ctgcacgctg ctgcacgctg ctgcacgctg ctgcacgctg 4680
gcgcctccctgc ccctctctct gcctgcgtct gcctgcgtct gcctgcgtct gcctgcgtct 4740
cggagacgcc gcgccttggg ctggctttta ctgactctga ccctccctgc ccctccctgc 4800
cggagacgcc gcggcttggg cctgattttta ctcggccctcg ccctccctgc ccctccctgc 4860
<210> 12
<211> Homo sapiens
<220> misc_feature
<221> DNA
21/30
agattgacgcg tgccacgcgg gaagagtcgcg actattttattg ggaggtagtgt agatcagtaa 660
tacaaacaac aagagtttct gcggaggaac ttctctcaagt aactgctgcag gcgaagttcg 720
tggaaagagga gggaagctcat caaagcgctg gcaagctgtg actttcagcga gatcacaagc 780	tagctctttgg cacaagccgc gacccgctgg ctgccgcaga gcggagagaa gcgaagacta 840
aaaagaagag ccaaacatcg tcagcacagc ttggctcactgt tattttttaga tgggaccacaa 900
gaaaattaat gcaacttctca aatggagcccct gctggcttgg ccggtgaccct tcggagatcc 960
cccagagattt gggtcgttga agaagagtcct acggcagggct gacgagctca actctgttgaag 1020
aggagtctgga agatcctgagc ggcgcggcgg gcggagacag tgagagggga ggctgcagct 1080
aacacctgctca cttgggctgcct acgagtgac gacaacagcgc ggcgtggtctt cttggagctc 1140
atgctgtaacc cccagttcgaag gcgtggtggag ggcgaatgcct gcggagctgc gcgtggcttc 1200
gcagccctgc gggaggatttg gcagcgagtc gcgtcctgag cgcggcggcgc gcccggggtgg 1260
agggagataa cgagagagaca gggcggggtg gcgcctgcag tccctcgcgt ctggaggcttt 1320
agacgcagaa atgagaaagag gcagcggttag gcgggccgac gctggtgcttg cttgctccct 1380
aggaggagcc agctgagcctg gcacccgagc gcgcctgctg cggctgagac gcggccacggg 1440
gttcctctgg ggaagagctg cattgctctgc tcaacagag tggccggtgga gaagcccggtg 1500
tacaagactc acctacgtcttt tattggcaac aaggaccccgc gggagagacctt cttggcaggg 1560
ctttgacacag gccacgctgc cagcagcact caggtgctgac ttcggcccgg ctcgcgcagc 1620
gccagagacact cagcagactc ccttcacagg actcggtgccg ccgcgcagcg ctggacgggc 1680
attcctctggc gcatactagcc gagcagagcg ctgagatgtg gcagcagatgc gcgaggatct 1740
cagcgagaa cttggcgagaa gagaaggac gttgctgctgg gcagcagcag gcggctgtac 1800
gcgagggctgc ctgacagcagc ccaagggagt gacagactgg gcggcgtgcc gcgaggagtg 1860
agggcggcctc ttggaggaga gctagcagcg cttggagagtc cggcgcggcgc gcgcggcagc 1920
agaaggagaga ccagacagag gcccgaagct gcggggctgt gcggacagct ctcgaggtgc 1980
atccgctctgg cccagttgctaa gggcgggtcggt gctggctgtct ggagatgagc gcgcacgaca 2040
ccaggggcggc gcggctgggttg ccacccctgc tattgggagtgc ttcgagccttt cccagcagca 2100
gcggcggctgg gcggcggcgg ctaacgcttt gttcttctct tcggctcttc gtcagcctgg 2160
ccggcggcggc gcggcggcgc gcggcggcgg gcggcggcgg gcggcggcgg gcggcggcgg 2220
tactacatagc aactagcagttt cccactcgcag tcggctctct cggagagcag gcgggaggag 2280
<210> Homo sapiens
<220> Homo sapiens
<221> misc_feature
<222> Incyte Feature
<223> Incyte Feature
<400> 14
gggggaagcgt tggtgtcttga aaagactagg gcgcagccgcgc gcggcgaagag gaaaggtggc 60
acctctctggc tggtgtcgggg ggctgtagcag gcggcgctgg aacctgggccc ccccaacgcct 120
gcgggctcgt gcgtgtgcagag gcggcgcgcc gcgtgtgtcttg gttatgcaccc gcgtggtcac 180
gtgggtcagg gcagctgtgcag gcgcgggtgg gcggagagtcag gcgcgggtgg gcggagagtc 240
agggggagcg gcggcggcag gcggggagtcag gcggggagtttg gcggggagtttg gcggggagtttg 300
tgggtctctct gcgcacagcgc gcgcacgctc gcgcacagcg gcgcacagcg gcgcacagcg 360
tgagagtctcct gcgtcttgggt gcgtcttgggt gcgtcttgggt gcgtcttgggt gcgtcttgggt 420
ctccccaggacc cccagcagcgc cccagcagcgc cccagcagcgc cccagcagcgc cccagcagcgc 480
tgggtctcttc gacagagcagc gcggcgggctgc gcggcgggctgc gcggcgggctgc gcggcgggctgc 540
tgggtcttgggcagc gcggcgctgtgc gcggcgctgtgc gcggcgctgtgc gcggcgctgtgc gcggcgctgtgc 600
tgggtctctgg gcgtgtcttgc ctaggtcctgc gcgtgtcttgc gcgtgtcttgc gcgtgtcttgc 660
ctgccgagctg aacagctgttg ctggcccgac ctggcccgac ctggcccgac ctggcccgac 720
ctcctcgggtgc gcgtggtgctc gcggcgtggtgc gcgtggtgctc gcggcgtggtgc gcgtggtgctc 780
acacccagcagc gctgggcttc gcggcgtggtgc gcgtggtgctc gcggcgtggtgc gcgtggtgctc 840
tggggaggcg gcggcgggctgc gcggcgggctgc gcggcgggctgc gcggcgggctgc gcggcgggctgc 900
cttcgagagctt gcgtgtctgtgc gcgtgtctgtgc gcgtgtctgtgc gcgtgtctgtgc gcgtgtctgtgc 960
tggggagagc tcaggtgcac ggcgtgccgc gcgtggtgctc gcgtggtgctc gcgtggtgctc 1020
tggggagagc tcaggtgcac ggcgtgccgc gcgtggtgctc gcgtggtgctc gcgtggtgctc 1080
tggggagagc tcaggtgcac ggcgtgccgc gcgtggtgctc gcgtggtgctc gcgtggtgctc 1140
<400> 14
aggggaagcgt tggtgtcttga aaagactagg gcgcagccgcgc gcggcgaagag gaaaggtggc 60
acctctctggc tggtgtcgggg ggctgtagcag gcggcgctgg aacctgggccc ccccaacgcct 120
gcgggctcgt gcgtgtgcagag gcggcgcgcc gcgtgtgtcttg gttatgcaccc gcgtggtcac 180
gtgggtcagg gcagctgtgcag gcgcgggtgg gcggagagtcag gcgcgggtgg gcggagagtc 240
agggggagcg gcggcggcag gcggggagtcag gcggggagtttg gcggggagtttg gcggggagtttg 300
tgggtctctct gcgcacagcgc gcgcacgctc gcgcacagcg gcgcacagcg gcgcacagcg 360
tgagagtctcct gcgtcttgggt gcgtcttgggt gcgtcttgggt gcgtcttgggt gcgtcttgggt 420
ctccccaggacc cccagcagcgc cccagcagcgc cccagcagcgc cccagcagcgc cccagcagcgc 480
tgggtctcttc gacagagcagc gcggcgggctgc gcggcgggctgc gcggcgggctgc gcggcgggctgc 540
tgggtcttgggcagc gcggcgctgtgc gcggcgctgtgc gcggcgctgtgc gcggcgctgtgc gcggcgctgtgc 600
tgggtctctgg gcgtgtcttgc ctaggtcctgc gcgtgtcttgc gcgtgtcttgc gcgtgtcttgc 660
ctgccgagctg aacagctgttg ctggcccgac ctggcccgac ctggcccgac ctggcccgac 720
ctcctcgggtgc gcgtggtgctc gcggcgtggtgc gcgtggtgctc gcggcgtggtgc gcgtggtgctc 780
acacccagcagc gctgggcttc gcggcgtggtgc gcgtggtgctc gcggcgtggtgc gcgtggtgctc 840
tggggaggcg gcggcgggctgc gcggcgggctgc gcggcgggctgc gcggcgggctgc gcggcgggctgc 900
cttcgagagctt gcgtgtctgtgc gcgtggtgctc gcggcgtggtgc gcgtggtgctc gcgtggtgctc 960
WO 02/072630

PCT/US02/03844

tgccgcgccag tgcggagagcc aagggagacgc ctcocgggga ctcagagaaca agggcagaga 1260
tggagctatt tggagaaggg cagcaagcgac gcagcgcgcgc ctcctgcccag gagaaacccaa 1320
tgtggagagg aacggcctcgg gaggctaaag gagaggattg aggtcagagc ccagacgggga 1380
cctccgaccag tgcgaacctg tctacgtgct cacctgcag gcgttaaaga gaggacggag 1440
atggagaccc agacagagcg tgcagccccg cgcgcgccac ccttggcttgag ccttggacag 1500
ttaaaggagga ggtggcagat caagacgcag aaggacccct cccctgctgag acctgcctac 1560
tggctccgacc tgcgaaggtt aagggagagg atggagactg agagccagag ggcgcctccc 1620
caggctaggg cccacccacag cgcgcctcgg gcgtgcctcg ggaaagaccc actcccccag 1680
ccttgcccccag tgaggggtga ttgcctcattag aaccttgatgc aacagaaagcc gaaagatgtag 1740
aacacattcag aagagagacag ctgctgattg gttgctgacag ggcgtcctcg gactcctgaag 1800
atgatgatgaa aacagcgaata cgcttcgaga caggaaatgct tgtccctcgc agggttgtagg 1860
tcttcgctcg cttgagtggt cagagctgcc agtggcagtt gcagcctgca ctcctctgctc 1920
tggaaacacag ccacagccct gctccttcag atatacttgc atatctctct tccctcctcct 1980
gctgagagcc tgcctcagtc cccagctctg aaggctctaa gagaagacag ggaagacgtcg 2040
ggcgcctcaga tgtgaactaa atctccctgca aacagccccg ccagcogcaca ctctcttcat 2100
acccagcagct ctcacccttc gccttcctaa acatcttctg tctattaagc tcttttgatcag 2160
ctggaagaaaa ggg 2173

<210> 15
<211> 4037
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4029969CB1

<400> 15
cctcctctcgc tgcgtgcctgg gcaccgcgcc tgcgtgcctac tctggtgccgc attggagacct 60
cgccagccccct gcggctcctcg cgaggagccgg cggagcagag tacacgcgcag gcagcagatct 120
tgcgctcgagg tgcagacctg gttgaacacc gcgctcgcgc gcgctcgacc gcagctcaccc 180
ccctggctctca cagatcagct ggctatgcgg tgcgcagcgcgc ctcctggctca gcagtggtagg 240
ctctccctcg ggctcctgcc ggtcatggct tcagagaggt ctgtccctctaa atatactctct 300
cggtctgcgtt gcctgcctgc ggtgcctctct ggtgcctctct gcgcgctgcgc gcgcgcgcgc 360
ctcttccttc gcctgcgtct gcctcctgcgt gcctgcgtct gcgcgctgcgc gcgcgcgcgc 420
ctcttccttc gcctgcgtct gcctcctgcgt gcctgcgtct gcgcgctgcgc gcgcgcgcgc 480
ctcttccttc gcctgcgtct gcctcctgcgt gcctgcgtct gcgcgctgcgc gcgcgcgcgc 540
ctcttccttc gcctgcgtct gcctcctgcgt gcctgcgtct gcgcgctgcgc gcgcgcgcgc 600
ctcttccttc gcctgcgtct gcctcctgcgt gcctgcgtct gcgcgctgcgc gcgcgcgcgc 660
cctgcgtggtt gccttccttc gcctgcgtct gcctgcgtct gcgcgctgcgc gcgcgcgcgc 720
cctgcgtggtt gccttccttc gcctgcgtct gcctgcgtct gcgcgctgcgc gcgcgcgcgc 780
ctcttccttc gcctgcgtct gcctcctgcgt gcctgcgtct gcgcgctgcgc gcgcgcgcgc 840
ctcttccttc gcctgcgtct gcctcctgcgt gcctgcgtct gcgcgctgcgc gcgcgcgcgc 900
ctcttccttc gcctgcgtct gcctcctgcgt gcctgcgtct gcgcgctgcgc gcgcgcgcgc 960
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1020
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1080
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1140
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1200
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1260
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1320
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1380
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1440
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1500
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1560
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1620
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1680
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1740
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1800
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1860
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1920
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 1980
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 2040
acaccttctt gcacatcagct agaagacgtt ggtgtgctcc gcggagcagc gcaacatagct 2100

24/30
gttttggctt ttttattttc tttcagtc ggagatatct caagtttag catcactcact tttaggtgaa cgggagtt 2160
gccgaccctct cttgctggcct attatatcttt atcctgatgc cagggagagg cgttcataagt 2220
ggcagagaggt gatagatatt ggtggtgag cagggagagt cggggagt 2280
tgttaagtgt tcgctgtagg tctgtaatt gtaaatcatt tctggctggc 2340
gttcagaccc ggtctttggtg ggtcctcttg tgaattgaa cgggcatt 2400
gttagatatct ttagtttagt tagcttattg gatgatatag tctagtttag 2460
cctagactatt tctaggttct aattggttcttag tttagatt cactagtcg 2520
tggcgttggag gtttagaattta taagttcag ctgctggttttt 2580
tcttgactact ttcgctttta ttagaatattt gtagataatt 2640
aactagact gacactacttta cagatccttgc attttaaattta 2700
actatataattt catgcttaca aatagcctag cggccgggta 2760
tgctttaaatg attatgata tggagagagag atttgagaga 2820
tactttagttg atacagactt ctatattgct cagatcactt 2880
tgctcgtcct ttcgctcttct tattctatctt cactatattt 2940
ccgctctctg cggctgctatt attatttctt cagatcactt 3000
tgctcgtcct ttcgctcttct tattctatctt cactatattt 3060
ctgctcgtcct ttcgctcttct tattctatctt cactatattt 3120
tttagttgag gttttcattg ctagttttatt cagagagagag 3180
gagacccagac ggtcttattttt atcagagagag 3240
taagagatatt ttcggtattt cagagagagag 3300
tgctcgtcct ttcgctcttct tattctatctt cactatattt 3360
tgctcgtcct ttcgctcttct tattctatctt cactatattt 3420
tgctcgtcct ttcgctcttct tattctatctt cactatattt 3480
tgctcgtcct ttcgctcttct tattctatctt cactatattt 3540
tgctcgtcct ttcgctcttct tattctatctt cactatattt 3600
tgctcgtcct ttcgctcttct tattctatctt cactatattt 3660
tgctcgtcct ttcgctcttct tattctatctt cactatattt 3720
tgctcgtcct ttcgctcttct tattctatctt cactatattt 3780
tgctcgtcct ttcgctcttct tattctatctt cactatattt 3840
nttcctcact tattttttag tacatccttt gacacacatcg ggtcttttgtag 3900
tgctcgtcct ttcgctcttct tattctatctt cactatattt 3960
tgctcgtcct ttcgctcttct tattctatctt cactatattt 4020
nttcctcact tattttttag tacatccttt cagagatattt ttcgctcttct tattctatctt cactatattt 4080
<210> 16
<211> 1818
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2676571CB1

<400> 16
cctacccgcata gactgcgcgct gacccctact gcttcctctg gttcgctctg 60
gccatgctagc cagcgcgcgct gacccctact gcttcctctg gttcgctctg 120
tgctgctgct cagcgcgcgct gacccctact gcttcctctg gttcgctctg 180
cctcaczac gacccctact gcttcctctg gttcgctctg 240
tgcgctgctgc gacccctact gcttcctctg gttcgctctg 300
tgcgctgctgc gacccctact gcttcctctg gttcgctctg 360
cttcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 420
cttcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 480
ccgcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 540
cttcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 600
ccgcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 660
ccgcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 720
ccgcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 780
ccgcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 840
ccgcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 900
ccgcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 960
ccgcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 1020
ccgcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 1080
ccgcaccccg ccgctgctgc gacccctact gcttcctctg gttcgctctg 1140
<210> 18
<211> 2278
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2626405CB1

<400> 18
atgcgctcat caagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 60
tatgcgctctg tgtggtaata aagcgagatg ctaagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 120
aaactgctaa acgccttcat ctaagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 180
aatgctctac ccagcttcat ctaagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 240
tttgctttaa caagctcttc ataacacttc ctaagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 300
ttcttcctag tttttt 600
<400> 18
atgcgctcat caagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 60
tatgcgctctg tgtggtaata aagcgagatg ctaagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 120
aaactgctaa acgccttcat ctaagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 180
aatgctctac ccagcttcat ctaagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 240
tttgctttaa caagctcttc ataacacttc ctaagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 300
ttcttcctag tttttt 600
<400> 18
atgcgctcat caagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 60
tatgcgctctg tgtggtaata aagcgagatg ctaagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 120
aaactgctaa acgccttcat ctaagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 180
aatgctctac ccagcttcat ctaagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 240
tttgctttaa caagctcttc ataacacttc ctaagctctaa tccagctctg acctcgcggg gtcagccgac gtataagcc 300
ttcttcctag tttttt 600
<210> 20
<211> 4927
<212> DNA
<213> Homo sapiens
<220>  
<221>  misc_feature  
<223>  Incyto ID No: 7504129CB1  

<400> 20  
acagtgaaaa acctctctgct gaagtgcagc ccagacgggtgt gggagttagat cccgagaggg 60  
ccggttgtcgt aagtgcacct cccgaccccg ccgtccgacca ggacgttgtg 120  
gggccagcct ccgctctctg tctctctctgt aagtgttcac ccgctctctg tctgactcgc ggacgaccag 180  
agacccctcc gcggctccag ctgctgctcc gtgtgctctg cctgcctgcag ctgctgctctg ctggtgctctg 240  
ccacccaggg gtcgctccag ccgctctctg ccgtccgaccc ggcctcctcc gggtggtggt 300  
tccgttgtct gccgtgtggt gttgctctgc atcctctctg gggctctctg cctgtgctctg 360  
aggggtgcag ctggcttgtg ctggcttgtg gggcctccta gctgtggtcc ctgctgctctg ccgtccgaccc 420  
ctggtgtcct gccgctcctgt ggtggtggt gttggtggt ggtggtggt ggtggtggt 480  
gggtggtggt ggtggtggt ggtggtggt ggtggtggt ccgctctctg cccgctcctc ggcctcctcc 540  
ctggctctgc agggttgttg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 600  
agcagagaaa agaaggtagc gcggagacgg ggcgctccag gggaggagtgg ggtggtggt ggtggtggt 660  
aaggtgggtt ggaggaaaac cccacggcag gacatgcagc aattagagttg ccgaggtggtg 720  
gtggtggtgg ccctccctcag tctcctctct ggtggtggt ggtggtggt ggtggtggt 780  
ggggctccgg gtcgctctgt ccgctctctg ccgctctctg ccgctctctg ccgctctctg 840  
ctggctctgc agggttgttg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 900  
tcggaggcgc ccacggcagc cccgctcctc cggcctcctc cggcctcctc cggcctcctc 960  
tggtggctgg agcctccctc tgcgccctcg tgcgccctcg tgcgccctcg tgcgccctcg 1020  
actggtaggc cccggcctcc ccgctctctg ccgctctctg ccgctctctg ccgctctctg 1080  
tcgggttgtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 1140  
pattggaggg gcgggtgtgt gggcctcctg ctcgctctgc atcctctctg gcgggctctct 1200  
atcgggggtt gcggggtgtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 1260  
tccgtctctg ccccagctag ccccagctag ccccagctag ccccagctag ccccagctag 1320  
cggcctcctc ccgctctctg ccgctctctg ccgctctctg ccgctctctg ccgctctctg 1380  
agtggagagg gcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg 1440  
cacccacaaa gggggtgtgt gcgcgcgcgc cgccgcgcgc gcgcgcgcgc gcgcgcgcgc 1500  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 1560  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 1620  
agaagagagc aggctccctc aagagagagc cgggtgtgtg cgggtgtgtg cgggtgtgtg 1680  
ttcgtctctg gcgtctctct gcgtctctct gcgtctctct gcgtctctct gcgtctctct 1740  
agaagagagc aggctccctc aagagagagc cgggtgtgtg cgggtgtgtg cgggtgtgtg 1800  
cacccacaaa gggggtgtgt gcgcgcgcgc cgccgcgcgc gcgcgcgcgc gcgcgcgcgc 1860  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 1920  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 1980  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 2040  
ttcgtctctg gcgtctctct gcgtctctct gcgtctctct gcgtctctct gcgtctctct 2100  
ccggagcctg cggccgggcg ccggagcctg cggccgggcg ccggagcctg cggccgggcg 2160  
actgctgctgc gcggtgtggt gcggtgtgtg gcggtgtgtg gcggtgtgtg gcggtgtgtg 2220  
tccgtctctg gcgtctctct gcgtctctct gcgtctctct gcgtctctct gcgtctctct 2280  
acaagactat aggccaggtt aagggcgtt cggcgagcag cggcgagcag cggcgagcag 2340  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 2400  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 2460  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 2520  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 2580  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 2640  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 2700  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 2760  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 2820  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 2880  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 2940  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 3000  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 3060  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 3120  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 3180  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 3240  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 3300  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 3360  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 3420  
tcgggtgtgt gggggtggtg gggggtggtg gggggtggtg gggggtggtg gggggtggtg 3480
acgccccag cccttgccgct ccctctctct cgaccagcc ggctctctcc 3540
tgcacaaggg cacctctccg cgcctctccc tctgccacgc ggacccctct gcgcccccct 3600
cggcccaag ggacccctct cgccctctct cctgctgcag ccggacccct tgcgtcctctgc 3660
tctctcgccag cctctctctct gcgcctgccc gacccctctgc gccctctccct tctgctgcagc 3720
gggatctcttg gggcctccct cctctgccac gcgggtaccct tgtccacccct cctgcgtcagc 3780
caggccctct cggcgccctcc ctctgtgcca cacggcaccct tctgctgcagc cttctctctg 3840
cacgccccag cctctctcctgc cctctctgccac acgggcaccct cctgctgcagc cttctctctg 3900
cacgcctctcgccctctctccc cctctctctgc cctctctctgc acgggcaccct cctgctgcagc cttctctctg 4020
cacccctctct ctctgctgcag cggccctccc ctctctctgtc acgcctgtccct cctgcgtcagc 4080
tctgctgccc cctctctctgc tctctctccct cctctctctgc acgggcaccct cctgctgcagc 4140
tctctctctgc cctctctctgc cctctctctgc cctctctctgc acgggcaccct cctgctgcagc 4200
tctctctctgc cctctctctgc cctctctctgc cctctctctgc acgggcaccct cctgctgcagc 4260
tctctctctgc cctctctctgc cctctctctgc cctctctctgc acgggcaccct cctgctgcagc 4320
tctctctctgc cctctctctgc cctctctctgc cctctctctgc acgggcaccct cctgctgcagc 4380
cggccaccc ccctgctgctgctgccctgt ctgtcctgta aaagactcct gtaagctgcct gttgccctct 4440
cgcacccag ccctgctgctgctgccctgt ctgtcctgta aaagactcct gtaagctgcct gttgccctct 4500
agtgcaccag ccctgctgctgctgccctgt ctgtcctgta aaagactcct gtaagctgcct gttgccctct 4560
agtgcaccag ccctgctgctgctgccctgt ctgtcctgta aaagactcct gtaagctgcct gttgccctct 4620
ccgccccag ccctgctgctgctgccctgt ctgtcctgta aaagactcct gtaagctgcct gttgccctct 4680
tctctccctgg ccacagagagt gcgctctctgc ggctctctctt ctgctgcagc cctgctgcagc 4740
acacccctcg ccctgctgctgctgccctgt ctgtcctgta aaagactcct gtaagctgcct gttgccctct 4800
gccgcgcggatat ttgctgctgctgccctgt ctgctgcagc cctgctgcagc cctgctgcagc cctgctgcagc 4860
ccacgggttgt ttgactgtgct aacaacctctg atctgacccctgccctgctgcagc cctgctgcagc 4920
ccgcccctactgctgctgctgccctgt ctgctgcagc cctgctgcagc cctgctgcagc 4927