(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 March 2003 (13.03.2003)

PCT

(10) International Publication Number
WO 03/020947 A2

(51) International Patent Classification⁷: C12Q

(21) International Application Number: PCT/US02/26930

(22) International Filing Date: 23 August 2002 (23.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/316,710 31 August 2001 (31.08.2001) US

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

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **WILLIAMSON, Mark** [US/US]; 15 Stonecrest Drive, Saugus, MA 01906 (US).

(74) Agent: BOSSONE, Steven, A.; Millennium Pharmaceuticals, Inc., 75 Sidney Street, Cambridge, MA 02139 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

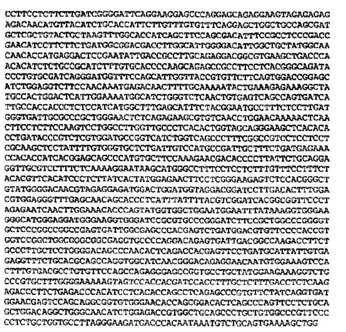
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI,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

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

[Continued on next page]

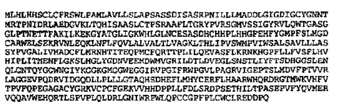
(54) Title: MID 9002, A HUMAN SULFATASE FAMILY MEMBER AND USES THEREFOR

cDNA Sequence of human MID 9002 (SEQ ID NO:1)



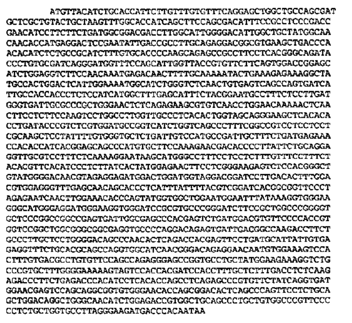
a

Predicted Amino Acid Sequence of MID 9002 (SEQ ID NO:2)



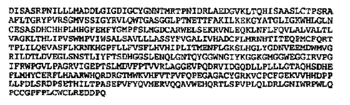
b

Coding Sequence of MID 9002 (SEQ ID NO:3)



c

Amino Acid Sequence of Mature Form of MID 9002 (SEQ ID NO:4)



d

(57) Abstract: The invention provides isolated nucleic acids molecules, designated MID 9002 nucleic acid molecules, which encode novel sulfatase family members. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing MID 9002 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a MID 9002 gene has been introduced or disrupted. The invention still further provides isolated Mill 9002 proteins, fusion proteins, antigenic pep tides and anti-Mill 9002 antibodies. Diagnostic and therapeutic methods utilizing compositions of the invention are also provided.

WO 03/020947 A2

MID 9002, A HUMAN SULFATASE FAMILY MEMBER AND USES THEREFOR

Related Applications

[0001] This application claims the benefit of prior-filed provisional application Serial No. 60/316,710 filed August 31, 2001, entitled "Mid 9002, A Human Sulfatase Family Member and Uses Therefor" (pending). The entire content of the above-referenced application is incorporated herein by this reference.

Background of the Invention

[0002] The sulfatases are a family of enzymes that catalyze the hydrolysis of sulfate ester bonds (*O*-sulfatase activity) present in a wide variety of substrates, *e.g.*, glycosaminoglycans (*e.g.*, heparan sulfate, chondroitin sulfate, and dermatan sulfate), 3β-hydroxysteroid sulfates, and sulfolipids (Parenti *et al.* "The sulfatase gene family" *Curr. Opin. Gen. Dev.* (1997) 7:386-391; Ballabio *et al.* Steroid sulfatase deficiency and X-linked ichthyosis. In *The Metabolic and Molecular Bases of Inherited Disease*, Scriver *et al.*, eds. (New York: McGraw-Hill), 2999-3022, 1995). Nine different human sulfatases have been described and characterized (Ballabio *et al.* Steroid sulfatase deficiency and X-linked ichthyosis. In *The Metabolic and Molecular Bases of Inherited Disease*, Scriver *et al.*, eds. (New York: McGraw-Hill), 2999-3022, 1995; Kolodny *et al.* Metachromatic leukodystrophy and multiple sulfatase deficiency: sulfatide lipidosis. In *The Metabolic and Molecular Bases of Inherited Diseases*. Scriver *et al.*, eds. (New York: McGraw-Hill), 2693-2741, 1995; Neufeld *et al.* The mucopolysaccharidoses. In *The Metabolic and Molecular Bases of Inherited Disease*, Scriver *et al.*, eds. (New York: McGraw-Hill), 2645-2494, 1995) The deficiency of sulfatases has been implicated in at least seven distinct inherited disorders (Kolodny *et al.* Metachromatic leukodystrophy and multiple sulfatase deficiency: sulfatide lipidosis. In *The Metabolic and Molecular Bases of Inherited Diseases*. Scriver *et al.*, eds. (New York: McGraw-Hill), 2693-2741, 1995; Neufeld *et al.* The mucopolysaccharidoses. In *The Metabolic and Molecular Bases of Inherited Disease*, Scriver *et al.*, eds. (New York: McGraw-Hill), 2645-2494, 1995). There exists a high degree of homology along the entire length of the known sulfatase proteins, particularly in the N-terminal region (Wilson *et al.* "Hunter Syndrome:

isolation of an iduronate-2-sulfatase cDNA clone and analysis of patient DNA" *Proc. Natl. Acad. Sci. USA* 87:8531-8535, 1990).

[0003] Three of the previously known human sulfatases, arylsulfatases A, B, and C (ARSA, ARSB, and ARSC), are known to hydrolyze sulfated artificial substrates containing a phenolic ring, such as *p*-nitrocatechol sulfate or 4-methylumbelliferyl sulfate (4-MU sulfate). ARSC is more specifically known as steroid sulfatase (STS) because of its ability to hydrolyze steroid sulfates. While ARSA, ARSB, and ARSC each act on a distinct natural substrate, they all have a "promiscuous" activity against artificial substrates, *i.e.*, arylsulfatase activity. ARSA and ARSB are lysosomal in their subcellular localization and have acidic pH optima, while ARSC is microsomal and has a neutral to alkaline pH optimum. Unlike the lysosomal sulfatases, STS is involved in the synthetic pathway of steroid hormones, rather than in a degradative pathway (Ballabio *et al.* Steroid sulfatase deficiency and X-linked ichthyosis. In *The Metabolic and Molecular Bases of Inherited Disease*, Scriver *et al.*, eds. (New York: McGraw-Hill), 2999-3022, 1995).

[0004] Three newly identified members of the sulfatase family, ARSD, ARSE, and ARSF, also appear to be non-lysosomal, as they are localized in the endoplasmic reticulum (ARSD and ARSF) and in the Golgi compartment (ARSE). It has been show that ARSE and ARSF are active on the fluorogenic artificial substrate 4-MU sulfate (Franco *et al. Cell* 81:15-25, 1995; Puca *et al. Genomics* 1997, in press).

[0005] Arylsulfatase E (ARSE), has been found to be involved in X-linked recessive chondrodysplasia punctata, a disorder of cartilage and bone development. The term "chondrodysplasia punctata" refers to a group of skeletal dysplasias characterized by abnormal calcium deposition in regions of endochondral bone formation (Franco *et al.* "A Cluster of Sulfatase Genes on Xp22.3: Mutations in Chondrodysplasia Punctata (CDPX) and Implications for Warfarin Embryopathy" *Cell* 81:15-25, April 7, 1995). This abnormality results in a peculiar radiological finding, commonly referred to as "stippling" of epiphyses or "paint-spattered" calcifications. These calcifications tend to disappear within the first few years of life as bone development progresses. Affected individuals display facial dysmorphism characterized by severe nasal hypoplasia, depressed nasal bridge, short nasal septum, and a deep groove between the nasal alae and the tip. They also have short stature and distal phalangeal hypoplasia.

[0006] Much attention has been paid in recent years to the enzymatic mechanism involved in the hydrolysis of sulfate ester bonds by sulfatases. One unique aspect of members of the sulfatase family is that they all seem to undergo a common and unique co- and post-translational modification, probably occurring in the endoplasmic reticulum, in which a cysteine residue is converted into a 2-amino-3-oxopropionic acid or serine semialdehyde (Schmidt *et al. Cell* 82:271-278, 1995). The machinery implicated in this post-translational modification is thought to recognize the sulfatases by a conserved linear amino acid sequence containing this cysteine residue. The conversion of the cysteine to serine semialdehyde is required for catalytic activity, and a defect in this conversion is thought to result in multiple sulfatase deficiency (MSD), a rare disease in which all sulfatase activities are simultaneously deficient (Kolodny *et al.* Metachromatic leukodystrophy and multiple sulfatase deficiency: sulfatide lipidosis. In *The Metabolic and Molecular Bases of Inherited Diseases.* Scriver *et al.*, eds. (New York: McGraw-Hill), 2693-2741, 1995). A study of the crystal structure of ARSB has shown that the modified cysteine residue and a metal ion are located at the base of a substrate-binding pocket (Bond *et al. Structure* 5:277-289, 1997).

Summary of the Invention

[0007] The present invention is based, in part, on the discovery of a novel sulfatase family member, referred to herein as "MID 9002". The nucleotide sequence of a cDNA encoding MID 9002 is shown in SEQ ID NO:1, and the amino acid sequence of a MID 9002 polypeptide is shown in SEQ ID NO:2. In addition, the nucleotide sequence of the coding region is depicted in SEQ ID NO:3.

[0008] Accordingly, in one aspect, the invention features a nucleic acid molecule which encodes a MID 9002 protein or polypeptide, *e.g.*, a biologically active portion of the MID 9002 protein. In a preferred embodiment, the isolated nucleic acid molecule encodes a polypeptide having the amino acid sequence of SEQ ID NO:2. In other embodiments, the invention provides isolated MID 9002 nucleic acid molecules having the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3 or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC Accession Number ____. In still other embodiments, the invention provides nucleic acid molecules that are substantially identical (*e.g.*, naturally occurring allelic variants) to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3 or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC Accession Number ____. In other embodiments, the invention provides a nucleic acid molecule which hybridizes under a stringent hybridization condition as described herein to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3 or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC Accession Number ____, wherein the nucleic acid encodes a full length MID 9002 protein or an active fragment thereof.

[0009] In a related aspect, the invention further provides nucleic acid constructs which include a MID 9002 nucleic acid molecule described herein. In certain embodiments, the nucleic acid molecules of the invention are operatively linked to native or heterologous regulatory sequences. Also included are vectors and host cells containing the MID 9002 nucleic acid molecules of the invention *e.g.*, vectors and host cells suitable for producing polypeptides.

[0010] In another related aspect, the invention provides nucleic acid fragments suitable as primers or hybridization probes for the detection of MID 9002-encoding nucleic acids.

[0011] In still another related aspect, isolated nucleic acid molecules that are antisense to a MID 9002 encoding nucleic acid molecule are provided.

[0012] In another aspect, the invention features MID 9002 polypeptides, and biologically active or antigenic fragments thereof that are useful, *e.g.*, as reagents or targets in assays applicable to treatment and diagnosis of sulfatase-associated or other MID 9002-associated disorders. In another embodiment, the invention provides MID 9002 polypeptides having a MID 9002 activity. Preferred polypeptides are MID 9002 proteins including at least one sulfatase domain, and, preferably, having a MID 9002 activity, *e.g.*, a MID 9002 activity as described herein.

[0013] In other embodiments, the invention provides MID 9002 polypeptides, *e.g.*, a MID 9002 polypeptide having the amino acid sequence shown in SEQ ID NO:2 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC Accession Number ____; an amino acid sequence that is substantially identical to the amino acid sequence shown in SEQ ID NO:2 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC Accession Number ____; or an amino acid sequence encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under a stringent hybridization condition as described herein to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3 or the nucleotide sequence of the insert of the plasmid deposited with ATCC Accession Number ____, wherein the nucleic acid encodes a full length MID 9002 protein or an active fragment thereof.

[0014] In a related aspect, the invention further provides nucleic acid constructs which include a MID 9002 nucleic acid molecule described herein.

[0015] In a related aspect, the invention provides MID 9002 polypeptides or fragments operatively linked to non-MID 9002 polypeptides to form fusion proteins.

[0016] In another aspect, the invention features antibodies and antigen-binding fragments thereof, that react with, or more preferably specifically or selectively bind MID 9002 polypeptides.

[0017] In another aspect, the invention provides methods of screening for compounds that modulate the expression or activity of the MID 9002 polypeptides or nucleic acids.

[0018] In still another aspect, the invention provides a process for modulating MID 9002 polypeptide or nucleic acid expression or activity, *e.g.*, using the compounds identified in the screens described herein. In certain embodiments, the methods involve treatment of conditions related to aberrant activity or expression of the MID 9002 polypeptides or nucleic acids, such as conditions or disorders involving aberrant or deficient sulfatase function or expression. Examples of such disorders include, but are not limited to, cellular proliferative and/or differentiative disorders, disorders associated with bone or cartilage metabolism, or disorders associated with the extracellular matrix.

[0019] The invention also provides assays for determining the activity of or the presence or absence of MID 9002 polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

[0020] In a further aspect, the invention provides assays for determining the presence or absence of a genetic alteration in a MID 9002 polypeptide or nucleic acid molecule, including for disease diagnosis.

[0021] In another aspect, the invention features a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, *e.g.*, a nucleic acid or peptide sequence. At least one address of the plurality has a capture probe that recognizes a MID 9002 molecule. In one embodiment, the capture probe is a nucleic acid, *e.g.*, a probe complementary to a MID 9002 nucleic acid sequence. In another embodiment, the capture probe is a polypeptide, *e.g.*, an antibody specific for MID 9002 polypeptides. Also featured is a method of analyzing a sample by contacting the sample to the aforementioned array and detecting binding of the sample to the array.

[0022] Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings

[0023] *Figure 1a-d* depicts a cDNA sequence (SEQ ID NO:1) and predicted amino acid sequence (SEQ ID NO:2) of human MID 9002. The methionine-initiated open reading frame of human MID 9002 (without the 5' and 3' untranslated regions of SEQ ID NO:1) is shown also as the coding sequence, SEQ ID NO:3. The mature form of the protein with the signal sequence removed is shown as SEQ ID NO:4.

[0024] *Figure 2* depicts a hydropathy plot of human MID 9002. Relatively hydrophobic residues are shown above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The numbers corresponding to the amino acid sequence of human MID 9002 are indicated. Polypeptides of the invention include fragments which include: all or part of a hydrophobic sequence, *e.g.*, a sequence above the dashed line, *e.g.*, the sequence from about amino acid 1 to 22, from about 32 to 45, from about 185 to 248, from about 265 to 275, from about 282 to 300, from about 428 to 438, and from about 566 to 575 of SEQ ID NO:2; all or part of a hydrophilic sequence, *e.g.*, a sequence below the dashed line, *e.g.*, the sequence from about amino acid 50 to 70, from about 145 to 193, from about 250 to 265, from about 275 to 285, from about 315 to 325, from about 345 to 385, from about 435 to 450, from about 455 to 470, from about 505 to 515, and from about 525 to 545 of SEQ ID NO:2; a sequence which includes a Cys, or a glycosylation site.

[0025] *Figure 3* depicts an alignment of the sulfatase domain of human MID 9002 with a consensus amino acid sequence derived from a hidden Markov model (HMM) from PFAM. The upper sequence is the consensus amino acid sequence (SEQ ID NO:X), while the lower amino acid sequence corresponds to amino acids 38 to 520 of SEQ ID NO:2.

[0026] *Figure 4a-f* depicts a BLAST alignment of the human MID 9002 arylsulfatase domain with a consensus amino acid sequence of domains derived from the ProDomain database ("Arylsulfatase hydrolase", No. PD001700, PD006857, PD208511, PD002587, PD345058, PD004614, PD335606, PD105130, PD013467, PD013468, PD016856; ProDomain Release 2001.1; http://www.toulouse.inra.fr/prodom.html). The lower sequences are amino acid residues of the consensus sequences from the ProDomain database, while the upper amino acid sequence corresponds to the arylsulfatase domain of human MID 9002 or portions thereof. The BLAST algorithm identifies multiple local alignments between the consensus amino acid sequences and human MID 9002.

[0027] *Figure 5* depicts a GAP alignment of human MID 9002 with human arylsulfatase E precursor (ARSE; Accessionnumber in Genbank P51690). The upper sequence in the figure is amino acids 68 to 1834 of human MID 9002 (SEQ ID NO:2) while the lower sequence is amino acids 1 to 589 of ARSE (P51690) (SEQ ID NO:X). GAP alignments use a matrix made by matblas from blosum62.iij.

Detailed Description of the Invention

[0028] The human MID 9002 sequence (Figure 1a; SEQ ID NO:1), which is approximately 1858 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 1770 nucleotides, including the termination codon (nucleotides indicated as coding of SEQ ID NO:1 in Fig. 1a; SEQ ID NO:3, Fig. 1c). The coding sequence encodes a 589 amino acid protein (SEQ ID NO:2). The human MID 9002 protein of SEQ ID NO:2, Figure 1d, and Figure 2 includes an amino-terminal hydrophobic amino acid sequence, consistent with a signal sequence, of about 31 amino acids (from amino acid 1 to about amino acid 31 of SEQ ID NO:2, PSORT, Nakai and Kanehisa (1992) *Genomics* 14:897-911), which upon cleavage results in the production of a mature protein form.)). This mature protein form (shown as SEQ ID NO:4 in Figure 1d) is approximately 558 amino acid residues in length (from about amino acid 32 to amino acid 589 of SEQ ID NO:2).

[0029] Human MID 9002 contains the following regions or other structural features (for general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer *et al.* (1997) *Protein* 28:405-420 and http://www.psc.edu/general/software/packages/pfam/pfam.html):

[0030] a sulfatase domain (PFAM Accession Number PF00884) located at about amino acid residues 38 to 520 of SEQ ID NO:2;

[0031] a arylsulfatase domain (ProDom No. PD001700, PD006857, PD208511, PD002587, PD105119, PD345058, PD004614, PD335606, PD105130, PD013467, PD013468, PD016856) located at about amino acid residues 37 to 589 of SEQ ID NO:2;

[0032] three transmembrane domains (predicted by MEMSAT, Jones *et al.* (1994) *Biochemistry* 33:3038-3049) at about amino acids 14 to 30, 227 to 249, and 291 to 307 of SEQ ID NO:2;

[0033] a sulfatase signature sequence (Prosite PS00149) located at about amino acids 137 to 147 (GYATGLIGKWH) of SEQ ID NO:2;

[0034] another sulfatase signature sequence (Prosite PS00523) located at about amino acids 84 to 96 (SLCTPSRAAFLTG) of SEQ ID NO:2;

[0035] three protein kinase C phosphorylation sites (Prosite PS00005) located at about amino acids 60 to 62 (TMR), 95 to 97 (TGR), and 187 to 189 (SEK) of SEQ ID NO:2;

[0036] four casein kinase II phosphorylation sites (Prosite PS00006) located at about amino acids 154 to 157 (SASD), 260 to 263 (TITE), 410 to 413 (SLMD), and 448 to 451 (SDHE)of SEQ ID NO:2;

[0037] one tyrosine kinase phosphorylation site (Prosite PS00007) located at about amino acids 131 to 138 (KILKEKGY) of SEQ ID NO:2;

[0038] eleven N-myristoylation site (Prosite PS00008) located at about amino acids 54 to 59 (GCYGNN), 73 to 78 (GVKLTQ), 103 to 108 (GMVSSI), 117 to 122 (GASGGL), 172 to 177 (GMPFSL), 341 to 346 (GLSNST), 356 to 361 (GSLENQ), 368 to 373 (GGWNGI), 381 to 386 (GGWEGG), 397 to 402 (GVLPAG), and 488 to 493 (GAGACY) of SEQ ID NO:2;

[0039] one amidation site (Prosite PS00009) located at about amino acids 493 to 496 (YGRK) of SEQ ID NO:2; and

[0040] four N-glycosylation sites (Prosite PS00001) located at about amino acids 58 to 61 (NNTM), 125 to 128 (NETT), 258 to 261 (NHTI), and 344 to 347 (NSTL) of SEQ ID NO:2.

[0041] A plasmid containing the nucleotide sequence encoding human MID 9002 was deposited with American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Number

______. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

[0042] The MID 9002 protein contains a significant number of structural characteristics in common with members of the sulfatase family. The term "family" when referring to the protein and nucleic acid molecules of the invention means two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well as other distinct proteins of human origin, or alternatively, can contain homologs of non-human origin, *e.g.*, rat or mouse proteins. Members of a family also can have common functional characteristics.

[0043] As used herein, the term "sulfatase" includes a protein or polypeptide which is capable of catalyzing the hydrolysis of sulfate esters. The substrates of sulfatases range from complex molecules, such as glycosaminoglycans and sulfolipids, to 3β-hydroxysteroid. Sulfatases are thought to be important in the degradation of the extracellular matrix. The gene encoding arylsulfatase E has been shown to be involved in X-linked recessive chondrodisplasia punctata, a disorder of cartilage and bone development. The sulfatase enzyme have also been implicated in at least six other inherited disorders. Previously described human sulfatases, arylsulfatases A, B, and C (ARSA, ARSB, ARSC) are able to hydrolyze sulfated artificial substrates containing a phenolic ring, such as *p*-nitrocatechol sulfate or 4-methylumbelliferyl sulfate (4-MU sulfate). ARSC is more specifically known as steroid sulfatase (STS) because of its ability to hydrolyze steroid sulfates. ARSA and ARSB are lysosomal in their subcellular localization and have acidic pH optima, while ARSC is microsomal and has a neutral to alkaline pH optima.

[0044] Members of a sulfatase family of proteins are typically associated with the lysosomes, microsomes, endoplasmic reticulum, or Golgi apparatus of the cell. The active site of sulfatases typically has a cysteine residue that has been converted into a serine semialdehyde as well as a metal (*e.g.*, Ca^{+2}) binding site. The modified cysteine residue (Cys86 of MID 9002 shown in SEQ ID NO:2) in the active site is able to covalently bind the sulfate of the substrate, thus forming an intermediate enzyme-sulfate complex. Members of the sulfatase family show a high degree of amino acid similarity along the entire length of the protein, particularly in the amino-terminal region.

[0045] An alignment of the MID 9002 protein with human arylsulfatase E precursor is shown in Figure 5 and demonstrates about 100% sequence identity between the two sequences (as calculated in matblas from the blosum62.iij matrix). An alignment of the MID 9002 protein with human arylsulfatase family members, arysulfatase D and F, also demonstrates significant homology (67-74% sequence identity) between the various sulfatase family members.

[0046] A MID 9002 polypeptide can include a "sulfatase domain" or regions homologous with a "sulfatase domain". A MID 9002 polypeptide can further include a "catalytic domain" responsible for the enzymatic activity of the protein or regions homologous with a "catalytic domain.".

[0047] As used herein, the term "sulfatase domain" includes an amino acid sequence of about 38 to 520 amino acid residues in length and having a bit score for the alignment of the sequence to the sulfatase domain (HMM) of at least 500, 600, or preferably 700. Preferably a sulfatase domain catalyzes the hydrolysis of sulfate esters. Preferably, a sulfatase domain includes at least about 100 to 700 amino acids, more preferably about 200 to 600 amino acid residues, or about 400 to 600 amino acids and has a bit score for the alignment of the sequence to the sulfatase domain (HMM) of at least 500, 600, 700, 750, 775, 790, or greater. A "sulfatase domain" may also include a sulfatase signature sequence 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the at least one of the signature sequences SLCTPSRAAFLTG, [SAP]-[LIVMST]-[CS]-[STAC]-P-[STA]-R-x(2)-[LIVMFW](2)-[TAR]-G, GYATGLIGKWH, or G-[YV]-x-[ST]-x(2)-[IVAS]-G-K-x(0,1)-[FYWMK]-[HL], described herein. A "sulfatase domain" may also contain a cysteine residue or a cysteine residue that has been converted to a serine semialdehyde (2-amino-3-oxopropionic acid). The "sulfatase domain" may also contain a metal ion (Ca^{+2}) binding site. The "sulfatase domain" may have the ability to bind and/or catalyze the hydrolysis of a sulfate substrate, preferably an aryl sulfate substrate.

[0048] The sulfatase domain can include a Prosite sulfatase signature sequence PS00149, G-[YV]-x-[ST]-x(2)-[IVAS]-G-K-x(0,1)-[FYWMK]-[HL], or sequences homologous thereto. The sulfatase domain also can include another Prosite sulfatase signature sequence PS00523, [SAP]-[LIVMST]-[CS]-[STAC]-P-[STA]-R-x(2)-[LIVMFW](2)-[TAR]-G, or sequences homologous thereto. In the above conserved signature sequence, and other motifs or signature sequences described herein, the standard IUPAC one-letter code for the amino acids is used. Each element in the pattern is separated by a dash (-); square brackets ([ ]) indicate the particular residues that are accepted at that position; x indicates that any residue is accepted at that position; and numbers in parentheses (()) indicate the number of residues represented by the accompanying amino acid. The Prosite PS00149 sulfatase signature sequence is located in the sulfatase domain of human MID 9002 polypeptide and corresponds to about amino acids 137 to 147 (GYATGLIGKWH) of SEQ ID NO:2. The Prosite PS00523 sulfatase signature sequence is also located in the sulfatase domain of human MID 9002 polypeptide and corresponds to about amino acids 84 to 96 (SLCTPSRAAFLTG) of SEQ ID NO:2. The sulfatase domain (HMM) has been assigned the PFAM Accession Number PF00884 (http;//genome.wustl.edu/Pfam/.html). As used herein, the "sulfatase domain" is a portion of the human MID 9002 protein which is homologous, *e.g.*, at least about 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, or 84% identical to either or both of the following ProDom family "arylsulfatase" domain (ProDomain Release 2001.1; http://www.toulouse.inra.fr/prodom.html, Figure 4a-g). An alignment of the sulfatase domain (amino acids 37 to 181 of SEQ ID NO:2) of human MID 9002 with PD001700, SEQ ID NO:X, derived from a BLAST search model shows 73% identity (as calculated in ProDomain from the blosum62 matrix, Figure 4a-g). An alignment of the sulfatase domain (amino acids 38 to 520 of SEQ ID NO:2) of human MID 9002 with the Pfam Sulfatase consensus amino acid sequence (SEQ ID NO:X) derived from a hidden Markov model is depicted in Figure 3.

[0049] In a preferred embodiment, a MID 9002 polypeptide or protein has a "sulfatase domain" or a region which includes at least about 100 to 700 more preferably about 300 to 600 or 400 to 600 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a "sulfatase domain," *e.g.*, the sulfatase domain of human MID 9002 (*e.g.*, residues 38 to 520 of SEQ ID NO:2).

[0050] To identify the presence of a "sulfatase" domain in a MID 9002 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against the

Pfam database of HMMs (*e.g.*, the Pfam database, release 2.1) using the default parameters (http://www.sanger.ac.uk/Software/Pfam/HMM_search). For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for MILPAT0063 and a score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (*e.g.*, to 8 bits). A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28:405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.* (1990) *Meth. Enzymol.*183:146-159; Gribskov *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.* (1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.* (1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference. A search was performed against the HMM database resulting in the identification of a "sulfatase domain" domain in the amino acid sequence of human MID 9002 at about residues 38 to 520 of SEQ ID NO:2 (see Figure 1b).

[0051] To identify the presence of a "sulfatase" domain in a MID 9002 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against a database of domains, *e.g.*, the ProDom database (Corpet *et al.* (1999), *Nucl. Acids Res.* 27:263-267). The ProDom protein domain database consists of an automatic compilation of homologous domains. Current versions of ProDom are built using recursive PSI-BLAST searches (Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402; Gouzy *et al.* (1999) *Computers and Chemistry* 23:333-340) of the SWISS-PROT 38 and TREMBL protein databases. The database automatically generates a consensus sequence for each domain. A BLAST search was performed against the HMM database resulting in the identification of a "sulfatase" domain in the amino acid sequence of human MID 9002 at about residues 37 to 589 of SEQ ID NO:2 (see Figure 1b).

[0052] A MID 9002 polypeptide can include at least one, preferably two "transmembrane domains" or regions homologous with a "transmembrane domain". As used herein, the term "transmembrane domain" includes an amino acid sequence of about 10 to 40 amino acid residues in length and spans the plasma membrane. Transmembrane domains are rich in hydrophobic residues, *e.g.*, at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids of a transmembrane domain are hydrophobic, *e.g.*, leucines, isoleucines, tyrosines, or tryptophans. Transmembrane domains typically have alpha-helical structures and are described in, for example, Zagotta *et al.*, (1996) *Annual Rev. Neurosci.* 19:235-263, the contents of which are incorporated herein by reference. The transmembrane domains of human MID 9002 are located at about residues 227 to 249 and 291 to 307 of SEQ ID NO:2.

[0053] In a preferred embodiment, a MID 9002 polypeptide or protein has at least one, preferably two "transmembrane domains" or regions which includes at least about 12 to 35 more preferably about 14 to 30 or 15 to 25 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a "transmembrane domain," *e.g.*, the transmembrane domains of human MID 9002 (*e.g.*, residues 227 to 249 and 291 to 307 of SEQ ID NO:2). The transmembrane domain of human MID 9002 is visualized in the hydropathy plot (Figure 2) as regions of about 15 to 25 amino acids where the hydropathy trace is mostly above the horizontal line.

[0054] To identify the presence of a "transmembrane" domain in a MID 9002 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be analyzed by a transmembrane prediction method that predicts the secondary structure and topology of integral membrane proteins based on the recognition of topological models (MEMSAT, Jones *et al.*, (1994) *Biochemistry* 33:3038-3049).

[0055] A MID 9002 polypeptide can include at least one, two, preferably three "non-transmembrane regions." As used herein, the term "non-transmembrane region" includes an amino acid sequence not identified as a transmembrane domain. The non-transmembrane regions in MID 9002 are located at about amino acids 1 to 227, 249 to 291, and 307 to 589 of SEQ ID NO:2. The non-transmembrane regions may be extracellular, cytoplasmic, or lumenal. A three-domain model has been proposed for members of the sulfatase family. Two luminally oriented domains are separated by a hydrophobic domain which spans the membrane twice in opposite directions. The luminally oriented domains may be glycosylated.

[0056] The non-transmembrane regions of MID 9002 include at least one, preferably two luminal regions. When located at the N-terminus, the luminal region is referred to herein as the "N-terminal luminal domain." As used herein, an "N-terminal luminal domain" includes an amino acid sequence having about 1 to 300, preferably about 1 to 250, more preferably about 1 to 225, or even more preferably about 1 to 200 amino acid residues in length, is located inside of lumen of a cellular organelle such as the endoplasmic reticulum or Golgi. The C-terminal amino acid residue of an "N-terminal luminal domain" is adjacent to an N-terminal amino acid residue of a transmembrane domain in a MID 9002 protein. For example, an N-terminal luminal domain is located at about amino acid residues 1 to 227 of SEQ ID NO:2.

[0057] In a preferred embodiment, a MID 9002 polypeptide or protein has an N-terminal luminal domain or a region which includes about 1 to 300, preferably about 1 to 250, and more preferably about 1 to 200 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with an "N-terminal luminal domain," *e.g.*, the N-terminal luminal domain of human MID 9002 (*e.g.*, residues 1 to 227 of SEQ ID NO:2).

[0058] In another embodiment, a MID 9002 cytoplasmic region includes at least one loops. As used herein, the term "loop" includes an amino acid sequence which is not included within a phospholipid membrane, having a length of at least about 4, preferably about 5 to 50, more preferably about 6 to 45 amino acid residues, and has an amino acid sequence that connects two transmembrane domains within a protein or polypeptide. Accordingly, the N-terminal amino acid of a loop is adjacent to a C-terminal amino acid of a transmembrane domain in a MID 9002 molecule, and the C-terminal amino acid of a loop is adjacent to an N-terminal amino acid of a transmembrane domain in a MID 9002 molecule. As used herein, a "loop" includes a loop located outside the lumen of an organelle of a cell or within the cytoplasm of a cell. For example, a "loop" can be found at about amino acid residues 249 to 291 of SEQ ID NO:2.

[0059] In a preferred embodiment, a MID 9002 polypeptide or protein has a loop or a region which includes at least about 4, preferably about 5 to 50, and more preferably about 6 to 45 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a loop," *e.g.*, a loop of human MID 9002 (*e.g.*, residues 249 to 291 of SEQ ID NO:2).

[0060] In another embodiment, a luminal region of a MID 9002 protein can include the C-terminus and can be a "C-terminal luminal domain," also referred to herein as a "C-terminal luminal tail." As used herein, a "C-terminal luminal domain" includes an amino acid sequence having a length of at least about 100, preferably about 200 to 400, more preferably about 250 to 300 amino acid residues, is located inside the lumen of an organelle of a cell.. The N-terminal amino acid residue of a

"C-terminal luminal domain" is adjacent to a C-terminal amino acid residue of a transmembrane domain in a MID 9002 protein. For example, a C-terminal luminal domain is located at about amino acid residues 307 to 589 of SEQ ID NO:2.

[0061] In a preferred embodiment, a MID 9002 polypeptide or protein has a C-terminal luminal domain or a region which includes at least about 100, preferably about 200 to 400, and more preferably about 250 to 300 amino acid residues and has at least about 60%, 70% 80%, 90%, 95%, 99%, or 100% homology with a "C-terminal luminal domain," *e.g.*, the C-terminal luminal domain of human MID 9002 (*e.g.*, residues 307 to 589 of SEQ ID NO:2).

[0062] A human MID 9002 protein can further include a sulfatase signature sequence such as Prosite PS00149 and PS00523. For example, a sulfatase signature sequence is located at about amino acid residues 137 to 147 and 84 to 96 of SEQ ID NO:2. In a preferred embodiment, a MID 9002 polypeptide or protein has a sulfatase signature sequence which include at least about 5, preferably about 5 to 20, and more preferably about 8 to 15 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a sulfatase signature sequence, *e.g.*, the sulfatase signature sequences of human MID 9002 (*e.g.*, residues 137 to 147 and 84 to 96 of SEQ ID NO:2).

[0063] A MID 9002 family member can include at least one sulfatase domain; and at least one, preferably two, transmembrane domains. A MID 9002 family member can include at least one sulfatase signature sequence (Prosite PS00149 and PS00523). Furthermore, a MID 9002 family member can include at least one, two, preferably three protein kinase C phosphorylation sites (Prosite PS00005); at least one, two, three, preferably four casein kinase II phosphorylation sites (Prosite PS00006); at least one, two three, preferably four N-glycosylation sites (Prosite PS00001); at least one tyrosine kinase phosphorylation site (Prosite PS00007); at least one amidation site (Prosite PS 00009); and at least one, two, three, four, five, six, seven, eight, nine, ten, and preferably eleven N-myristoylation sites (Prosite PS00008).

[0064] As the MID 9002 polypeptides of the invention can modulate MID 9002-mediated activities, they can be useful for developing novel diagnostic and therapeutic agents for sulfatase-associated or other MID 9002-associated disorders, as described below.

[0065] As used herein, a "sulfatase-associated activity" includes an activity which involves catalysis of the hydrolysis of a sulfate ester. The "sulfatase-associated activity" may be associated with the degradation of components of the extracellular matrix (ECM). Members of the sulfatase family can play a role in X-linked recessive chondrodysplasia punctata, osteochondrodysplasias, and brachytelephalangic dwarfism disease (Malou *et al. Arch. Pediatr.* 8(2):176-180, 2001 February; Savarirayan *Pediatr. Radiol.* 29(5):322, 1999 May; Sheffield *et al. J. Med. Genet.* 35(12):1004-1008, 1998 December; Daniele *et al. Am. J. Hum. Genet.* 62(3):562-572, 1998 March; Parenti *et al. Am. J. Med. Genet.* 73(2):139-143, 1997 December 1997; Franco *et al. Cell* 81(1):15-25, 1995 April 7; Baitner *et al. J. Pediatr. Orthop.* 20(5):594-605, 2000 September-October; Horton *Eur. J. Hum. Genet.* 3(6):357-373, 1995; Seidel *et al. Clin. Genet.* 59(2):115-121, 2001 February)

[0066] As used herein, a "MID 9002 activity", "biological activity of MID 9002" or "functional activity of MID 9002", refers to an activity exerted by a MID 9002 protein, polypeptide or nucleic acid molecule on *e.g.*, a MID 9002-responsive cell or on a MID 9002 substrate, *e.g.*, a protein, biomolecule, small molecule, carbohydrate substrate, as determined *in vivo* or *in vitro*. In one embodiment, a MID 9002 activity is a direct activity, such as an association with a MID 9002 target molecule. A "target molecule" or "binding partner" is a molecule with which a MID 9002 protein binds or interacts in nature. In an exemplary embodiment, MID 9002 is a sulfatase, *e.g.*, an arylsulfatase family member and thus it binds to and interacts in nature with a substrate preferably a biomolecule containing a sulfate ester moiety and catalyzes the hydrolysis of the sulfate ester linkage.

[0067] A MID 9002 activity can also be an indirect activity, *e.g.*, a cellular signaling activity mediated by interaction of the MID 9002 protein with a MID 9002 receptor. Based on the above-described sequence structures and similarities to molecules of known function, the MID 9002 molecules of the present invention can have similar biological activities as sulfatase family members. For example, the MID 9002 proteins of the present invention can have one or more of the following activities: (1) the ability to catalyze the hydrolysis of a sulfate ester (*e.g.*, a aryl sulfate ester); (2) the ability to degrade extracellular matrix; (3) the ability to bind a chemical compound comprising a sulfate ester; (4) the ability to affect bone and/or cartilage metabolism/formation/destruction; and (5) the ability to affect cancer progression (*e.g.*, invasion, metastasis).

[0068] The MID 9002 molecules of the invention can modulate the activities of cells in tissues where they are expressed. For example, MID 9002 mRNA is expressed in kidney, pancreas, dorsal root ganglion, colon, and liver. Accordingly, the MID 9002 molecules of the invention can act as therapeutic or diagnostic agents for renal, pancreatic, hepatic, neurological, gastrointestinal, or colonic disorders.

[0069] The MID 9002 molecules can be used to treat proliferative and/or differentiative disorders of the colon, gastrointestinal tract, and kidneys in part because the MID 9002 mRNA is expressed in the colon, liver, kidney, and pancreas. Examples of cellular proliferative and/or differentiative disorders include cancer, *e.g.*, carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic disorders, *e.g.*, leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of prostate, colon, lung, breast, and liver origin.

[0070] As used herein, the term "cancer" (also used interchangeably with the terms, "hyperproliferative" and "neoplastic") refers to cells having the capacity for autonomous growth, *i.e.*, an abnormal state or condition characterized by rapidly proliferating cell growth. Cancerous disease states may be categorized as pathologic, *i.e.*, characterizing or constituting a disease state, *e.g.*, malignant tumor growth, or may be categorized as non-pathologic, *i.e.*, a deviation from normal but not associated with a disease state, *e.g.*, cell proliferation associated with wound repair. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. The term "cancer" includes malignancies of the various organ systems, such as those affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus. The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term "carcinoma" also includes carcinosarcomas, *e.g.*, which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures. The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

[0071] The MID 9002 molecules of the invention can be used to monitor, treat and/or diagnose a variety of proliferative disorders. Such disorders include hematopoietic neoplastic disorders. As used herein, the term "hematopoietic neoplastic disorders" includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, *e.g.*, arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Preferably, the diseases arise from poorly differentiated acute leukemias, *e.g.*, erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus (1991) *Crit Rev. in Oncol./Hemotol.* 11:267-97); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

[0072] Thus, the MID 9002 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more proliferative and/or differentiative disorders or other sulfatase disorders. As used herein, "sulfatase disorders" are diseases or disorders whose pathogenesis is caused by, is related to, or is associated with aberrant or deficient sulfatase protein function or expression. Examples of such disorders, *e.g.*, sulfatase-associated or other MID 9002-associated disorders, include but are not limited to, cellular proliferative and/or differentiative disorders (*e.g.*, adenoma, adenocarcinoma), disorders associated with bone metabolism, immune *e.g.*, inflammatory, disorders, cardiovascular disorders, endothelial cell disorders, liver disorders, viral diseases, pain or metabolic disorders.

[0073] The MID 9002 molecules can be used to treat gastrointestinal neoplastic disorders (*e.g.*, adenoma, adenocarcinoma) in part because sulfatase family members are found in the colon, liver, and pancreas.

[0074] Examples of cellular proliferative and/or differentiative disorders include cancer, *e.g.*, carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic disorders, *e.g.*, leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of prostate, colon, lung, breast, and liver origin.

[0075] As used herein, the term "cancer" (also used interchangeably with the terms, "hyperproliferative" and "neoplastic") refers to cells having the capacity for autonomous growth, *i.e.*, an abnormal state or condition characterized by rapidly proliferating cell growth. Cancerous disease states may be categorized as pathologic, i.e., characterizing or constituting a disease state, *e.g.*, malignant tumor growth, or may be categorized as non-pathologic, i.e., a deviation from normal but not associated with a disease state, *e.g.*, cell proliferation associated with wound repair. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. The term "cancer" includes malignancies of the various organ systems, such as those affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus. The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term "carcinoma" also includes carcinosarcomas, *e.g.*, which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures. The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

[0076] The MID 9002 molecules of the invention can be used to monitor, treat and/or diagnose a variety of proliferative disorders. Such disorders include hematopoietic neoplastic disorders. As used herein, the term "hematopoietic neoplastic disorders" includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, *e.g.*, arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Preferably, the diseases arise from poorly differentiated acute leukemias, *e.g.*, erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus (1991) *Crit Rev. in Oncol./Hemotol.* 11:267-97); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

[0077] Aberrant expression and/or activity of MID 9002 molecules can mediate disorders associated with bone metabolism. "Bone metabolism" refers to direct or indirect effects in the formation or degeneration of bone structures, *e.g.*, bone formation, bone resorption, etc., which can ultimately affect the concentrations in serum of calcium and phosphate. This term also includes activities mediated by MID 9002 molecules in bone cells, *e.g.* osteoclasts and osteoblasts, that can in turn result in bone formation and degeneration. For example, MID 9002 molecules can support different activities of bone resorbing osteoclasts such as the stimulation of differentiation of monocytes and mononuclear phagocytes into osteoclasts. Accordingly, MID 9002 molecules that modulate the production of bone cells can influence bone formation and degeneration, and thus can be used to treat bone disorders. Examples of such disorders include, but are not limited to, osteoporosis, osteodystrophy, osteomalacia, rickets, osteitis fibrosa cystica, renal osteodystrophy, osteosclerosis, anti-convulsant treatment, osteopenia, fibrogenesis-imperfecta ossium, secondary hyperparathyrodism, hypoparathyroidism, hyperparathyroidism, cirrhosis, obstructive jaundice, drug induced metabolism, medullary carcinoma, chronic renal disease, rickets, sarcoidosis, glucocorticoid antagonism, malabsorption syndrome, steatorrhea, tropical sprue, idiopathic hypercalcemia and milk fever.

[0078] The MID 9002 nucleic acid and protein of the invention can be used to treat and/or diagnose a variety of immune, *e.g.*, inflammatory, (*e.g.* respiratory inflammatory) disorders. Examples of immune disorders or diseases include, but are not limited to, autoimmune diseases (including, for example, diabetes mellitus, arthritis (including rheumatoid arthritis, juvenile rheumatoid arthritis, osteoarthritis, psoriatic arthritis), multiple sclerosis, encephalomyelitis, myasthenia gravis, systemic lupus erythematosis, autoimmune thyroiditis, dermatitis (including atopic dermatitis and eczematous dermatitis), psoriasis, Sjögren's Syndrome, inflammatory bowel disease, *e.g.* Crohn's disease and ulcerative colitis, aphthous ulcer, iritis, conjunctivitis, keratoconjunctivitis, asthma, allergic asthma, chronic obstructive pulmonary disease, cutaneous lupus erythematosus, scleroderma, vaginitis, proctitis, drug eruptions, leprosy reversal reactions, erythema nodosum leprosum, autoimmune uveitis, allergic encephalomyelitis, acute necrotizing hemorrhagic encephalopathy, idiopathic bilateral progressive sensorineural hearing loss, aplastic anemia, pure red cell anemia, idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis, chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue, lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis, uveitis posterior, and interstitial lung fibrosis), graft-versus-host disease, cases of transplantation, and allergy such as, atopic allergy.

[0079] As used herein, disorders involving the heart, or "cardiovascular disease" or a "cardiovascular disorder" includes a disease or disorder which affects the cardiovascular system, e.g., the heart, the blood vessels, and/or the blood. A cardiovascular disorder can be caused by an imbalance in arterial pressure, a malfunction of the heart, or an occlusion of a blood vessel, *e.g.*, by a thrombus. A cardiovascular disorder includes, but is not limited to disorders such as arteriosclerosis, atherosclerosis, cardiac hypertrophy, ischemia reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism, tachycardia, bradycardia, pressure overload, aortic bending, coronary artery ligation, vascular heart disease, valvular disease, including but not limited to, valvular degeneration caused by calcification, rheumatic heart disease, endocarditis, or complications of artificial valves; atrial fibrillation, long-QT syndrome, congestive heart failure, sinus node dysfunction, angina, heart failure, hypertension, atrial fibrillation, atrial flutter, pericardial disease, including but not limited to, pericardial effusion and pericarditis; cardiomyopathies, *e.g.*, dilated cardiomyopathy or idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, ischemic disease, arrhythmia, sudden cardiac death, and cardiovascular developmental disorders (*e.g.*, arteriovenous malformations, arteriovenous fistulae, raynaud's syndrome, neurogenic thoracic outlet syndrome, causalgia/reflex sympathetic dystrophy, hemangioma, aneurysm, cavernous angioma, aortic valve stenosis, atrial septal defects, atrioventricular canal, coarctation of the aorta, ebsteins anomaly, hypoplastic left heart syndrome, interruption of the aortic arch, mitral valve prolapse, ductus arteriosus, patent foramen ovale, partial anomalous pulmonary venous return, pulmonary atresia with ventricular septal defect, pulmonary atresia without ventricular septal defect, persistance of the fetal circulation, pulmonary valve stenosis, single ventricle, total anomalous pulmonary venous return, transposition of the great vessels, tricuspid atresia, truncus arteriosus, ventricular septal defects). A cardiovascular disease or disorder also can include an endothelial cell disorder.

[0080] As used herein, an "endothelial cell disorder" includes a disorder characterized by aberrant, unregulated, or unwanted endothelial cell activity, *e.g.*, proliferation, migration, angiogenesis, or vascularization; or aberrant expression of cell surface adhesion molecules or genes associated with angiogenesis, *e.g.*, TIE-2, FLT and FLK. Endothelial cell disorders include tumorigenesis, tumor metastasis, psoriasis, diabetic retinopathy, endometriosis, Grave's disease, ischemic disease (*e.g.*, atherosclerosis), and chronic inflammatory diseases (*e.g.*, rheumatoid arthritis).

[0081] Disorders which can be treated or diagnosed by methods described herein include, but are not limited to, disorders associated with an accumulation in the liver of fibrous tissue, such as that resulting from an imbalance between production and degradation of the extracellular matrix accompanied by the collapse and condensation of preexisting fibers. The methods described herein can be used to diagnose or treat hepatocellular necrosis or injury induced by a wide variety of agents including processes which disturb homeostasis, such as an inflammatory process, tissue damage resulting from toxic injury or altered hepatic blood flow, and infections (*e.g.*, bacterial, viral and parasitic). For example, the methods can be used for the early detection of hepatic injury, such as portal hypertension or hepatic fibrosis. In addition, the methods can be employed to detect liver fibrosis attributed to inborn errors of metabolism, for example, fibrosis resulting from a storage disorder such as Gaucher's disease (lipid abnormalities) or a glycogen storage disease, A1-antitrypsin deficiency; a disorder mediating the accumulation (*e.g.*, storage) of an exogenous substance, for example, hemochromatosis (iron-overload syndrome) and copper storage diseases (Wilson's disease), disorders resulting in the accumulation of a toxic metabolite (*e.g.*, tyrosinemia, fructosemia and galactosemia) and peroxisomal disorders (*e.g.*, Zellweger syndrome). Additionally, the methods described herein can be used for the early detection and treatment of liver injury associated with the administration of various chemicals or drugs, such as for example, methotrexate, isonizaid, oxyphenisatin, methyldopa, chlorpromazine, tolbutamide or alcohol, or which represents a hepatic manifestation of a vascular disorder such as obstruction of either the intrahepatic or extrahepatic bile flow or an alteration in hepatic circulation resulting, for example, from chronic heart failure, veno-occlusive disease, portal vein thrombosis or Budd-Chiari syndrome.

[0082] Additionally, MID 9002 molecules can play an important role in the etiology of certain viral diseases, including but not limited to Hepatitis B, Hepatitis C and Herpes Simplex Virus (HSV). Modulators of MID 9002 activity could be used to control viral diseases. The modulators can be used in the treatment and/or diagnosis of viral infected tissue or virus-associated tissue fibrosis, especially liver and liver fibrosis. Also, MID 9002 modulators can be used in the treatment and/or diagnosis of virus-associated carcinoma, especially hepatocellular cancer.

[0083] Additionally, MID 9002 can play an important role in the regulation of metabolism or pain disorders. Diseases of metabolic imbalance include, but are not limited to, obesity, anorexia nervosa, cachexia, lipid disorders, and diabetes. Examples of pain disorders include, but are not limited to, pain response elicited during various forms of tissue injury, *e.g.*, inflammation, infection, and ischemia, usually referred to as hyperalgesia (described in, for example, Fields (1987) *Pain*, New York:McGraw-Hill); pain associated with musculoskeletal disorders, *e.g.*, joint pain; tooth pain; headaches; pain associated with surgery; pain related to irritable bowel syndrome; or chest pain.

[0084] The MID 9002 protein, fragments thereof, and derivatives and other variants of the sequence in SEQ ID NO:2 thereof are collectively referred to as "polypeptides or proteins of the invention" or "MID 9002 polypeptides or proteins". Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as "nucleic acids of the invention" or "MID 9002 nucleic acids."

[0085] As used herein, the term "nucleic acid molecule" includes DNA molecules (*e.g.*, a cDNA or genomic DNA) and RNA molecules (*e.g.*, an mRNA) and analogs of the DNA or RNA generated, *e.g.*, by the use of nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

[0086] The term "isolated or purified nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[0087] As used herein, the term "hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions" describes conditions for hybridization and washing. Guidance for performing hybridization reactions can be found in *Current Protocols in Molecular Biology* (1989) John Wiley & Sons, N.Y., 6.3.1-6.3.6, which is incorporated by reference. Aqueous and nonaqueous methods are described in that reference and either can be used. Specific hybridization conditions referred to herein are as follows: 1) low stringency hybridization conditions in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by two washes in 0.2X SSC, 0.1% SDS at least at 50°C (the temperature of the washes can be increased to 55°C for low stringency conditions); 2) medium stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C; 3) high stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C; and preferably 4) very high stringency hybridization conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Very high stringency conditions (4) are the preferred conditions and the ones that should be used unless otherwise specified.

[0088] As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

[0089] As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a MID 9002 protein, preferably a mammalian MID 9002 protein, and can further include non-coding regulatory sequences, and introns.

[0090] An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language "substantially free" means preparation of MID 9002 protein having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-MID 9002 protein (also referred to herein as a "contaminating protein"), or of chemical precursors or non-MID 9002 chemicals. When the MID 9002 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The invention includes isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

[0091] A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of MID 9002 (*e.g.*, the sequence of SEQ ID NO:1 or 3) without abolishing or more preferably, without substantially altering a biological activity, whereas an "essential" amino acid residue results in such a change. For example, amino acid residues that are conserved among the polypeptides of the present invention, *e.g.*, those present in the sulfatase domain, are predicted to be particularly unamenable to alteration.

[0092] A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a MID 9002 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a MID 9002 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for MID 9002 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1 or SEQ ID NO:3, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

[0093] As used herein, a "biologically active portion" of a MID 9002 protein includes a fragment of a MID 9002 protein which participates in an interaction between a MID 9002 molecule and a non-MID 9002 molecule. Biologically active portions of a MID 9002 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the MID 9002 protein, *e.g.*, the amino acid sequence shown in SEQ ID NO:2, which include fewer amino acids than the full length MID 9002 protein, and exhibit at least one activity of a MID 9002 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the MID 9002 protein, *e.g.*, the ability to catalyze the hydrolysis of a sulfate ester or the ability to degrade the extracellular matrix. A biologically active portion of a MID 9002 protein can be a polypeptide which is, for example, 10, 25, 50, 100, 200, or more amino acids in length. Biologically active portions of a MID 9002 protein can be used as targets for developing agents which modulate a MID 9002 mediated activity, *e.g.*, the ability to catalyze the hydrolysis of a sulfate ester or the ability to degrade the extracellular matrix.

[0094] Calculations of homology or sequence identity (the terms “homology” and “identity” are used interchangeably herein) between sequences are performed as follows:

[0095] To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence (*e.g.*, when aligning a second sequence to the MID 9002 amino acid sequence of SEQ ID NO:2 having 589 amino acid residues, at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid “homology”). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[0096] The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (1970) *J. Mol. Biol.* 48:444-453 algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity or homology limitation of the invention) are a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[0097] The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers and Miller ((1989) CABIOS, 4:11-17) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

[0098] The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to MID 9002 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to MID 9002 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25:3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

[0099] Particular MID 9002 polypeptides of the present invention have an amino acid sequence substantially identical to the amino acid sequence of SEQ ID NO:2. In the context of an amino acid sequence, the term "substantially identical" is used herein to refer to a first amino acid that contains a sufficient or minimum number of amino acid residues that are i) identical to, or ii) conservative substitutions of aligned amino acid residues in a second amino acid sequence such that the first and second amino acid sequences can have a common structural domain and/or common functional activity. For example, amino acid sequences that contain a common structural domain having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%. 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to SEQ ID NO:2 are termed substantially identical.

[00100] In the context of nucleotide sequence, the term "substantially identical" is used herein to refer to a first nucleic acid sequence that contains a sufficient or minimum number of nucleotides that are identical to aligned nucleotides in a second nucleic acid sequence such that the first and second nucleotide sequences encode a polypeptide having common functional activity, or encode a common structural polypeptide domain or a common functional polypeptide activity. For example, nucleotide sequences having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%. 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to SEQ ID NO:1 or 3 are termed substantially identical.

[00101] "Misexpression or aberrant expression", as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, *e.g.*, increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-transitional modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, *e.g.*, a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

[00102] "Subject", as used herein, can refer to a mammal, *e.g.*, a human, or to an experimental or animal or disease model. The subject can also be a non-human animal, *e.g.*, a horse, cow, goat, or other domestic animal.

[00103] A "purified preparation of cells", as used herein, refers to, in the case of plant or animal cells, an *in vitro* preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

[00104] Various aspects of the invention are described in further detail below.

Isolated Nucleic Acid Molecules

[00105] In one aspect, the invention provides, an isolated or purified, nucleic acid molecule that encodes a MID 9002 polypeptide described herein, *e.g.*, a full length MID 9002 protein or a fragment thereof, *e.g.*, a biologically active portion of MID 9002 protein. Also included is a nucleic acid fragment suitable for use as a hybridization probe, which can be used, *e.g.*, to identify a nucleic acid molecule encoding a polypeptide of the invention, MID 9002 mRNA, and fragments suitable for use as primers, *e.g.*, PCR primers for the amplification or mutation of nucleic acid molecules.

[00106] In one embodiment, an isolated nucleic acid molecule of the invention includes the nucleotide sequence shown in SEQ ID NO:1, or a portion of any of this nucleotide sequence. In one embodiment, the nucleic acid molecule includes sequences encoding the human MID 9002 protein (i.e., "the coding region" of SEQ ID NO:1, as shown in SEQ ID NO:3), as well as 5' untranslated sequences (nucleotides 1 to 67 of SEQ ID NO:1) and 3' untranslated sequences (nucleotides 1835 to 1858 of SEQ ID NO:1). Alternatively, the nucleic acid molecule can include only the coding region of SEQ ID NO:1 (*e.g.*, SEQ ID NO:3) and, *e.g.*, no flanking sequences which normally accompany the subject sequence. In another embodiment, the nucleic acid molecule encodes a sequence corresponding to a fragment of the protein from about amino acid 38 to 520 of SEQ ID NO:2 or a mature protein from about amino acid 32 to 589 of SEQ ID NO:2.

[00107] In another embodiment, an isolated nucleic acid molecule of the invention includes a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, or a portion of any of these nucleotide sequences. In other embodiments, the nucleic acid molecule of the invention is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3 such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1 or 3, thereby forming a stable duplex.

[00108] In one embodiment, an isolated nucleic acid molecule of the present invention includes a nucleotide sequence which is at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the entire length of the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, or a portion, preferably of the same length, of any of these nucleotide sequences.

MID 9002 Nucleic Acid Fragments

[00109] A nucleic acid molecule of the invention can include only a portion of the nucleic acid sequence of SEQ ID NO:1 or 3. For example, such a nucleic acid molecule can include a fragment which can be used as a probe or primer or a fragment encoding a portion of a MID 9002 protein, *e.g.*, an immunogenic or biologically active portion of a MID 9002 protein. A fragment can comprise those nucleotides of

SEQ ID NO:1, which encode a sulfatase domain of human MID 9002. The nucleotide sequence determined from the cloning of the MID 9002 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other MID 9002 family members, or fragments thereof, as well as MID 9002 homologs, or fragments thereof, from other species.

[00110] In another embodiment, a nucleic acid includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5' or 3' noncoding region. Other embodiments include a fragment which includes a nucleotide sequence encoding an amino acid fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof, particularly fragments thereof which are at least 100 amino acids in length. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments should not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

[00111] A nucleic acid fragment can include a sequence corresponding to a domain, region, or functional site described herein. A nucleic acid fragment can also include one or more domain, region, or functional site described herein. Thus, for example, a MID 9002 nucleic acid fragment can include a sequence corresponding to a sulfatase domain, as described herein.

[00112] MID 9002 probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense or antisense sequence of SEQ ID NO:1 or SEQ ID NO:3, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1 or SEQ ID NO:3.

[00113] In a preferred embodiment the nucleic acid is a probe which is at least 5 or 10, and less than 200, more preferably less than 100, or less than 50, base pairs in length. It should be identical, or differ by 1, or less than in 5 or 10 bases, from a sequence disclosed herein. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[00114] A probe or primer can be derived from the sense or anti-sense strand of a nucleic acid which encodes a sulfatase domain from about amino acids 38 to 520 of

SEQ ID NO:2; a sulfatase domain from about amino acids 37 to 589 of SEQ ID NO:2; a mature protein from about amino acids 32 to 589 of SEQ ID NO:2; a N-terminal luminal domain from about amino acids 1 to 227 of SEQ ID NO:2; a transmembrane domain from about amino acids 227 to 249 of SEQ ID NO:2; a loop domain from about amino acids 249 to 291 of SEQ ID NO:2; a transmembrane domain from about amino acids 291 to 307 of SEQ ID NO:2; and a C-terminal luminal domain from about amino acids 307 to 589 of SEQ ID NO:2.

[00115] In another embodiment a set of primers is provided, *e.g.*, primers suitable for use in a PCR, which can be used to amplify a selected region of a MID 9002 sequence, *e.g.*, a domain, region, site or other sequence described herein. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differ by one base from a sequence disclosed herein or from a naturally occurring variant. For example, primers suitable for amplifying all or a portion of any of the following regions are provided: a sulfatase domain from about amino acids 38 to 520 of SEQ ID NO:2; a sulfatase domain from about amino acids 37 to 589 of SEQ ID NO:2; a mature protein from about amino acids 32 to 589 of SEQ ID NO:2; a N-terminal luminal domain from about amino acids 1 to 227 of SEQ ID NO:2; a transmembrane domain from about amino acids 227 to 249 of SEQ ID NO:2; a loop domain from about amino acids 249 to 291 of SEQ ID NO:2; a transmembrane domain from about amino acids 291 to 307 of SEQ ID NO:2; and a C-terminal luminal domain from about amino acids 307 to 589 of SEQ ID NO:2.

[00116] A nucleic acid fragment can encode an epitope bearing region of a polypeptide described herein.

[00117] A nucleic acid fragment encoding a "biologically active portion of a MID 9002 polypeptide" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1 or 3, which encodes a polypeptide having a MID 9002 biological activity (*e.g.*, the biological activities of the MID 9002 proteins are described herein), expressing the encoded portion of the MID 9002 protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the MID 9002 protein. For example, a nucleic acid fragment encoding a biologically active portion of MID 9002 includes a sulfatase domain, *e.g.*, amino acid residues about 38 to 520 of SEQ ID NO:2. A nucleic acid fragment encoding a biologically active portion of a

MID 9002 polypeptide, can comprise a nucleotide sequence which is greater than 300 or more nucleotides in length.

[00118] In preferred embodiments, a nucleic acid includes a nucleotide sequence which is about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, or more nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:1 or SEQ ID NO:3.

MID 9002 Nucleic Acid Variants

[00119] The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3. Such differences can be due to degeneracy of the genetic code (and result in a nucleic acid which encodes the same MID 9002 proteins as those encoded by the nucleotide sequence disclosed herein. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence which differs, by at least 1, but less than 5, 10, 20, 50, or 100 amino acid residues that shown in SEQ ID NO:2. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[00120] Nucleic acids of the inventor can be chosen for having codons, which are preferred, or non-preferred, for a particular expression system. *E.g.*, the nucleic acid can be one in which at least one codon, at preferably at least 10%, or 20% of the codons has been altered such that the sequence is optimized for expression in *E. coli*, yeast, human, insect, or CHO cells.

[00121] Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism) or can be non naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product).

[00122] In a preferred embodiment, the nucleic acid differs from that of SEQ ID NO: 1 or 3, *e.g.*, as follows: by at least one but less than 10, 20, 30, or 40 nucleotides; at least one but less than 1%, 5%, 10% or 20% of the nucleotides in the subject nucleic acid. If necessary for this analysis the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[00123] Orthologs, homologs, and allelic variants can be identified using methods known in the art. These variants comprise a nucleotide sequence encoding a polypeptide that is 50%, at least about 55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more identical to the nucleotide sequence shown in SEQ ID NO:2 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO 2 or a fragment of the sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of the MID 9002 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the MID 9002 gene.

[00124] Preferred variants include those that are correlated with one or more of the following sulfatase activities: (1) the ability to catalyze the hydrolysis of a sulfate ester (*e.g.,* a aryl sulfate ester); (2) the ability to degrade extracellular matrix; (3) the ability to bind a chemical compound comprising a sulfate ester; (4) the ability to affect bone and/or cartilage metabolism/formation/destruction; and (5) the ability to affect cancer progression (*e.g.*, invasion, metastasis).

[00125] Allelic variants of MID 9002, *e.g.*, human MID 9002, include both functional and non-functional proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the MID 9002 protein within a population that maintain (1) the ability to catalyze the hydrolysis of a sulfate ester (*e.g.,* a aryl sulfate ester); (2) the ability to degrade extracellular matrix; (3) the ability to bind a chemical compound comprising a sulfate ester; (4) the ability to affect bone and/or cartilage metabolism/formation/destruction; and/or (5) the ability to affect cancer progression (*e.g.*, invasion, metastasis). Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein. Non-functional allelic variants are naturally-occurring amino acid sequence variants of the MID 9002, *e.g.*, human MID 9002, protein within a population that do not have (1) the ability to catalyze the hydrolysis of a sulfate ester (*e.g.,* a aryl sulfate ester); (2) the ability to degrade extracellular matrix; (3) the ability to bind a chemical compound comprising a sulfate ester; (4) the ability to affect bone and/or cartilage metabolism/formation/destruction; and/or (5) the ability to affect cancer progression (*e.g.*, invasion, metastasis). Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion, or premature truncation of the amino acid sequence of SEQ ID NO:2, or a substitution, insertion, or deletion in critical residues or critical regions of the protein.

[00126] Moreover, nucleic acid molecules encoding other MID 9002 family members and, thus, which have a nucleotide sequence which differs from the MID 9002 sequences of SEQ ID NO:1 or SEQ ID NO:3 are intended to be within the scope of the invention.

Antisense Nucleic Acid Molecules, Ribozymes and Modified MID 9002 Nucleic Acid Molecules

[00127] In another aspect, the invention features, an isolated nucleic acid molecule which is antisense to MID 9002. An "antisense" nucleic acid can include a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire MID 9002 coding strand, or to only a portion thereof (*e.g.*, the coding region of human MID 9002 corresponding to SEQ ID NO:3). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding MID 9002 (*e.g.*, the 5' and 3' untranslated regions).

[00128] An antisense nucleic acid can be designed such that it is complementary to the entire coding region of MID 9002 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of MID 9002 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of MID 9002 mRNA, *e.g.*, between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length.

[00129] An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. The antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[00130] The antisense nucleic acid molecules of the invention are typically administered to a subject (*e.g.*, by direct injection at a tissue site), or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a MID 9002 protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For systemic administration, antisense molecules can be modified such that they specifically or selectively bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

[00131] In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

[00132] In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. A ribozyme having specificity for a MID 9002-encoding nucleic acid can include one or more sequences complementary to the nucleotide sequence of a MID

9002 cDNA disclosed herein (i.e., SEQ ID NO:1 or SEQ ID NO:3), and a sequence having known catalytic sequence responsible for mRNA cleavage (see U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach (1988) *Nature* 334:585-591). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a MID 9002-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, MID 9002 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel and Szostak (1993) *Science* 261:1411-1418.

[00133] MID 9002 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the MID 9002 (*e.g.*, the MID 9002 promoter and/or enhancers) to form triple helical structures that prevent transcription of the MID 9002 gene in target cells. See generally, Helene (1991) *Anticancer Drug Des.* 6:569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14:807-15. The potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[00134] The invention also provides detectably labeled oligonucleotide primer and probe molecules. Typically, such labels are chemiluminescent, fluorescent, radioactive, or colorimetric.

[00135] A MID 9002 nucleic acid molecule can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4: 5-23). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic, *e.g.*, a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) *supra*; Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci.* 93: 14670-675.

[00136] PNAs of MID 9002 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of MID 9002 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (*e.g.*, by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (*e.g.*, S1 nucleases (Hyrup *et al.* (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

[00137] In other embodiments, the oligonucleotide can include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, *e.g.*, Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (see, *e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide can be conjugated to another molecule, (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[00138] The invention also includes molecular beacon oligonucleotide primer and probe molecules having at least one region which is complementary to a MID 9002 nucleic acid of the invention, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantitating the presence of the MID 9002 nucleic acid of the invention in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi *et al.*, U.S. Patent No. 5,854,033; Nazarenko *et al.*, U.S. Patent No. 5,866,336, and Livak *et al.*, U.S. Patent 5,876,930.

Isolated MID 9002 Polypeptides

[00139] In another aspect, the invention features, an isolated MID 9002 protein, or fragment, *e.g.*, a biologically active portion, for use as immunogens or antigens to raise or test (or more generally to bind) anti-MID 9002 antibodies. MID 9002 protein can be isolated from cells or tissue sources using standard protein purification techniques. MID 9002 protein or fragments thereof can be produced by recombinant DNA techniques or synthesized chemically.

[00140] Polypeptides of the invention include those which arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing events, and alternative translational and post-translational events. The polypeptide can be expressed in systems, *e.g.*, cultured cells, which result in substantially the same post-translational modifications present when the polypeptide is expressed in a native cell, or in systems which result in the alteration or omission of post-translational modifications, *e.g.*, glycosylation or cleavage, present in a native cell.

[00141] In a preferred embodiment, a MID 9002 polypeptide has one or more of the following characteristics:

[00142] it has the ability to catalyze the hydrolysis of a sulfate ester (*e.g.,* a aryl sulfate ester);

[00143] it has the ability to degrade extracellular matrix or components thereof;

[00144] it has the ability to bind a chemical compound comprising a sulfate ester;

[00145] it has the ability to affect bone and/or cartilage metabolism/formation/destruction; and

[00146] it has the ability to affect cancer progression (*e.g.*, invasion, metastasis);

[00147] it has a molecular weight, *e.g.*, a deduced molecular weight, preferably ignoring any contribution of post translational modifications, amino acid composition or other physical characteristic of a MID 9002 polypeptide, *e.g.*, a polypeptide of SEQ ID NO:2;

[00148] it has an overall sequence similarity of at least 60%, preferably at least 70%, more preferably at least 80, 90, or 95%, with a polypeptide of SEQ ID NO:2;

[00149] it is expressed in at least the following human tissues and cell lines: at high levels in kidney, pancreas, colon, liver, and at medium levels in the dorsal root ganglion;

[00150] it has a sulfatase domain which is preferably about 70%, 80%, 90% or 95% identical to amino acid residues about 38 to 520 and 37 to 589 of SEQ ID NO:2;

[00151] it has a transmembrane domain which is preferably about 70%, 80%, 90% or 95% identical to amino acid residues about 227 to 249 or 291 to 307 of SEQ ID NO:2; and

[00152] it has at least one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, preferably sixteen, and most preferably seventeen of the cysteines found in the amino acid sequence of the native protein.

[00153] In a preferred embodiment the MID 9002 protein, or fragment thereof, differs from the corresponding sequence in SEQ ID NO:2. In one embodiment it differs by at least one but by less than 15, 10 or 5 amino acid residues. In another it differs from the corresponding sequence in SEQ ID NO:2 by at least one residue but less than 20%, 15%, 10% or 5% of the residues in it differ from the corresponding sequence in SEQ ID NO:2. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) The differences are, preferably, differences or changes at a non-essential residue or a conservative substitution. In a preferred embodiment the differences are not in the sulfatase domain at about residues 38 to 520 of SEQ ID NO:2. In another embodiment one or more differences are in the sulfatase domain at about residues 38 to 520 of SEQ ID NO:2.

[00154] Other embodiments include a protein that contains one or more changes in amino acid sequence, *e.g.*, a change in an amino acid residue which is not essential for activity. Such MID 9002 proteins differ in amino acid sequence from SEQ ID NO:2, yet retain biological activity.

[00155] In one embodiment, the protein includes an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to SEQ ID NO:2.

[00156] A MID 9002 protein or fragment is provided which varies from the sequence of SEQ ID NO:2 in regions defined by amino acids about 1 to 38 and/or 520 to 589 by at least one but by less than 15, 10 or 5 amino acid residues in the protein or fragment but which does not differ from SEQ ID NO:2 in regions defined by amino acids about 38 to 520. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) In some embodiments the difference is at a non-essential residue or is a conservative substitution, while in others the difference is at an essential residue or is a non-conservative substitution.

[00157] In one embodiment, a biologically active portion of a MID 9002 protein includes a sulfatase domain. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques

and evaluated for one or more of the functional activities of a native MID 9002 protein.

[00158] In a preferred embodiment, the MID 9002 protein has an amino acid sequence shown in SEQ ID NO:2. In other embodiments, the MID 9002 protein is sufficiently or substantially identical to SEQ ID NO:2. In yet another embodiment, the MID 9002 protein is sufficiently or substantially identical to SEQ ID NO:2 and retains the functional activity of the protein of SEQ ID NO:2, as described in detail in the subsections above.

MID 9002 Chimeric or Fusion Proteins

[00159] In another aspect, the invention provides MID 9002 chimeric or fusion proteins. As used herein, a MID 9002 "chimeric protein" or "fusion protein" includes a MID 9002 polypeptide linked to a non-MID 9002 polypeptide. A "non-MID 9002 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the MID 9002 protein, *e.g.*, a protein which is different from the MID 9002 protein and which is derived from the same or a different organism. The MID 9002 polypeptide of the fusion protein can correspond to all or a portion *e.g.*, a fragment described herein of a MID 9002 amino acid sequence. In a preferred embodiment, a MID 9002 fusion protein includes at least one (or two) biologically active portion of a MID 9002 protein. The non-MID 9002 polypeptide can be fused to the N-terminus or C-terminus of the MID 9002 polypeptide.

[00160] The fusion protein can include a moiety which has a high affinity for a ligand. For example, the fusion protein can be a GST-MID 9002 fusion protein in which the MID 9002 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant MID 9002. Alternatively, the fusion protein can be a MID 9002 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of MID 9002 can be increased through use of a heterologous signal sequence.

[00161] Fusion proteins can include all or a part of a serum protein, *e.g.*, a portion of an immunoglobulin (*e.g.*, IgG, IgA, or IgE), *e.g.*, an Fc region and/or the hinge C1 and C2 sequences of an immunoglobulin or human serum albumin.

[00162] The MID 9002 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The MID 9002 fusion proteins can be used to affect the bioavailability of a MID 9002 substrate. MID 9002 fusion proteins can be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a MID 9002 protein; (ii) mis-regulation of the MID 9002 gene; and (iii) aberrant post-translational modification of a MID 9002 protein.

[00163] Moreover, the MID 9002-fusion proteins of the invention can be used as immunogens to produce anti-MID 9002 antibodies in a subject, to purify MID 9002 ligands and in screening assays to identify molecules which inhibit the interaction of MID 9002 with a MID 9002 substrate.

[00164] Expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A MID 9002-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the MID 9002 protein.

Variants of MID 9002 Proteins

[00165] In another aspect, the invention also features a variant of a MID 9002 polypeptide, *e.g.*, which functions as an agonist (mimetics) or as an antagonist. Variants of the MID 9002 proteins can be generated by mutagenesis, *e.g.*, discrete point mutation, the insertion or deletion of sequences or the truncation of a MID 9002 protein. An agonist of the MID 9002 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a MID 9002 protein. An antagonist of a MID 9002 protein can inhibit one or more of the activities of the naturally occurring form of the MID 9002 protein by, for example, competitively modulating a MID 9002-mediated activity of a MID 9002 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Preferably, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the MID 9002 protein.

[00166] Variants of a MID 9002 protein can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of a MID 9002 protein for agonist or antagonist activity.

[00167] Libraries of fragments *e.g.*, N terminal, C terminal, or internal fragments, of a MID 9002 protein coding sequence can be used to generate a variegated population of fragments for screening and subsequent selection of variants of a MID 9002 protein.

[00168] Variants in which a cysteine residues is added or deleted or in which a residue which is glycosylated is added or deleted are particularly preferred.

[00169] Methods for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property are known in the art. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify MID 9002 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6:327-331).

[00170] Cell based assays can be exploited to analyze a variegated MID 9002 library. For example, a library of expression vectors can be transfected into a cell line, *e.g.*, a cell line, which ordinarily responds to MID 9002 in a substrate-dependent manner. The transfected cells are then contacted with MID 9002 and the effect of the expression of the mutant on signaling by the MID 9002 substrate can be detected, *e.g.*, by measuring the hydrolysis of sulfate esters which may be provided endogenously or exogenously. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the MID 9002 substrate, and the individual clones further characterized.

[00171] In another aspect, the invention features a method of making a MID 9002 polypeptide, *e.g.*, a peptide having a non-wild type activity, *e.g.*, an antagonist, agonist, or super agonist of a naturally occurring MID 9002 polypeptide, *e.g.*, a naturally occurring MID 9002 polypeptide. The method includes altering the sequence of a MID 9002 polypeptide, *e.g.*, altering the sequence, *e.g.*, by substitution or deletion of one or more residues of a non-conserved region, a domain or residue disclosed herein, and testing the altered polypeptide for the desired activity.

[00172] In another aspect, the invention features a method of making a fragment or analog of a MID 9002 polypeptide a biological activity of a naturally occurring MID 9002 polypeptide. The method includes altering the sequence, *e.g.*, by substitution or deletion of one or more residues, of a MID 9002 polypeptide, *e.g.*, altering the sequence of a non-conserved region, or a domain or residue described herein, and testing the altered polypeptide for the desired activity.

Anti-MID 9002 Antibodies

[00173] In another aspect, the invention provides an anti-MID 9002 antibody. The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include scFV and dcFV fragments, Fab and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as papain or pepsin, respectively.

[00174] The antibody can be a polyclonal, monoclonal, recombinant, *e.g.*, a chimeric or humanized, fully human, non-human, *e.g.*, murine, or single chain antibody. In a preferred embodiment it has effector function and can fix complement. The antibody can be coupled to a toxin or imaging agent.

[00175] A full-length MID 9002 protein or, antigenic peptide fragment of MID 9002 can be used as an immunogen or can be used to identify anti-MID 9002 antibodies made with other immunogens, *e.g.*, cells, membrane preparations, and the like. The antigenic peptide of MID 9002 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of MID 9002. Preferably, the antigenic peptide includes at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

[00176] Fragments of MID 9002 which include residues about 50 to 70, about 145 to 193, about 250 to 265, about 275 to 285, about 315 to 325, about 345 to 385, about 435 to 450, about 455 to 470, about 505 to 515, or about 525 to 545 of SEQ ID NO:2 can be used to make, *e.g.*, used as immunogens or used to characterize the specificity of an antibody, antibodies against hydrophilic regions of the MID 9002 protein (see Figure 2). Similarly, fragments of MID 9002 which include residues about 1 to 22, about 32 to 45, about 185 to 248, about 265 to 275, about 282 to 300, about 428 to 438, or about 566 to 575 of SEQ ID NO:2 can be used to make an antibody against a hydrophobic region of the MID 9002 protein; fragments of MID 9002 which include residues about 1 to 227, about 249 to 291, about 307 to 589, or a subset thereof, *e.g.* about residues 1 to 31, about residues 32 to 227, about residues 32 to 100, about residues 100 to 227, about residues 307 to 400, about residues 400 to 589 of SEQ ID

NO:2, can be used to make an antibody against an luminal or extraluminal region of the MID 9002 protein; a fragment of MID 9002 which includes residues about 38 to 520, or about 37 to 589 of SEQ ID NO:2 can be used to make an antibody against the sulfatase region of the MID 9002 protein.

[00177] Antibodies reactive with, or specific or selective for, any of these regions, or other regions or domains described herein are provided.

[00178] Preferred epitopes encompassed by the antigenic peptide are regions of MID 9002 located on the surface of the protein, *e.g.*, hydrophilic regions, as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human MID 9002 protein sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the MID 9002 protein and are thus likely to constitute surface residues useful for targeting antibody production.

[00179] In a preferred embodiment the antibody can bind to the luminal portion of the MID 9002 protein. In another embodiment, the antibody binds an extraluminal portion of the MID 9002 protein.

[00180] In a preferred embodiment the antibody binds an epitope on any domain or region on MID 9002 proteins described herein.

[00181] Additionally, chimeric, humanized, and completely human antibodies are also within the scope of the invention. Chimeric, humanized, but most preferably, completely human antibodies are desirable for applications which include repeated administration, *e.g.*, therapeutic treatment of human patients, and some diagnostic applications.

[00182] Chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, can be made using standard recombinant DNA techniques. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA*

84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559).

[00183] A humanized or complementarity determining region (CDR)-grafted antibody will have at least one or two, but generally all three recipient CDR's (of heavy and or light immuoglobulin chains) replaced with a donor CDR. The antibody may be replaced with at least a portion of a non-human CDR or only some of the CDR's may be replaced with non-human CDR's. It is only necessary to replace the number of CDR's required for binding of the humanized antibody to a MID 9002 or a fragment thereof. Preferably, the donor will be a rodent antibody, *e.g.*, a rat or mouse antibody, and the recipient will be a human framework or a human consensus framework. Typically, the immunoglobulin providing the CDR's is called the "donor" and the immunoglobulin providing the framework is called the "acceptor." In one embodiment, the donor immunoglobulin is a non-human (*e.g.*, rodent). The acceptor framework is a naturally-occurring (*e.g.*, a human) framework or a consensus framework, or a sequence about 85% or higher, preferably 90%, 95%, 99% or higher identical thereto.

[00184] As used herein, the term “consensus sequence” refers to the sequence formed from the most frequently occurring amino acids (or nucleotides) in a family of related sequences (See *e.g.*, Winnaker, (1987) *From Genes to Clones* (Verlagsgesellschaft, Weinheim, Germany). In a family of proteins, each position in the consensus sequence is occupied by the amino acid occurring most frequently at that position in the family. If two amino acids occur equally frequently, either can be included in the consensus sequence. A “consensus framework” refers to the framework region in the consensus immunoglobulin sequence.

[00185] An antibody can be humanized by methods known in the art. Humanized antibodies can be generated by replacing sequences of the Fv variable region which are not directly involved in antigen binding with equivalent sequences from human Fv variable regions. General methods for generating humanized antibodies are provided by Morrison (1985) *Science* 229:1202-1207, by Oi *et al.* (1986) *BioTechniques* 4:214, and by Queen *et al.* US patent Nos. 5,585,089, 5,693,761 and 5,693,762, the contents of all of which are hereby incorporated by reference. Those methods include isolating, manipulating, and expressing the nucleic acid sequences that encode all or part of immunoglobulin Fv variable regions from at least one of a heavy or light chain. Sources of such nucleic acid are well known to those skilled in the art and, for example, may be obtained from a hybridoma producing an antibody against a MID 9002 polypeptide or fragment thereof. The recombinant DNA encoding the humanized antibody, or fragment thereof, can then be cloned into an appropriate expression vector.

[00186] Humanized or CDR-grafted antibodies can be produced by CDR-grafting or CDR substitution, wherein one, two, or all CDR's of an immunoglobulin chain can be replaced. See *e.g.*, U.S. Patent No. 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060; Winter US patent No. 5,225,539, the contents of all of which are hereby expressly incorporated by reference. Winter describes a CDR-grafting method which may be used to prepare the humanized antibodies of the present invention (UK Patent Application GB 2188638A, filed on March 26, 1987; Winter US patent No. 5,225,539), the contents of which is expressly incorporated by reference.

[00187] Also within the scope of the invention are humanized antibodies in which specific amino acids have been substituted, deleted or added. Preferred humanized antibodies have amino acid substitutions in the framework region, such as to improve binding to the antigen. For example, a humanized antibody will have framework residues identical to the donor framework residue or to another amino acid other than the recipient framework residue. To generate such antibodies, a selected, small number of acceptor framework residues of the humanized immunoglobulin chain can be replaced by the corresponding donor amino acids. Preferred locations of the substitutions include amino acid residues adjacent to the CDR, or which are capable of interacting with a CDR (see *e.g.*, US patent No. 5,585,089). Criteria for selecting amino acids from the donor are described in US 5,585,089, *e.g.*, columns 12-16 of US 5,585,089, the *e.g.*, columns 12-16 of US 5,585,089, the contents of which are hereby incorporated by reference. Other techniques for humanizing antibodies are described in Padlan *et al.* EP 519596 A1, published on December 23, 1992.

[00188] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice that are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. *See*, for example, Lonberg and Huszar (1995) *Int. Rev. Immunol. 13*:65-93); and U.S. Patent Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, CA) and Medarex, Inc. (Princeton, NJ), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

[00189] Completely human antibodies that recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, *e.g.*, a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. This technology is described by Jespers *et al.* (1994) *Bio/Technology 12*:899-903).

[00190] The anti-MID 9002 antibody can be a single chain antibody. A single-chain antibody (scFV) can be engineered as described in, for example, Colcher *et al.* (1999) *Ann. N Y Acad. Sci.* 880:263-80; and Reiter (1996) *Clin. Cancer Res.* 2:245-52. The single chain antibody can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target MID 9002 protein.

[00191] In a preferred embodiment, the antibody has reduced or no ability to bind an Fc receptor. For example, it is an isotype or subtype, fragment or other mutant, which does not support binding to an Fc receptor, *e.g.*, it has a mutagenized or deleted Fc receptor binding region.

[00192] An antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, puromycin, maytansinoids, *e.g.*, maytansinol (see US Patent No. 5,208,020), CC-1065 (see US Patent Nos. 5,475,092, 5,585,499, 5,846,545) and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, CC-1065, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine, vinblastine, taxol and maytansinoids). Radioactive ions include, but are not limited to iodine, yttrium and praseodymium.

[00193] The conjugates of the invention can be used for modifying a given biological response, the therapeutic moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the therapeutic moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, α-interferon, β-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophase colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

[00194] Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[00195] An anti-MID 9002 antibody (*e.g.*, monoclonal antibody) can be used to isolate MID 9002 by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an anti-MID 9002 antibody can be used to detect MID 9002 protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-MID 9002 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labelling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

[00196] In preferred embodiments, an antibody can be made by immunizing with a purified MID 9002 antigen, or a fragment thereof, *e.g.*, a fragment described herein, a membrane associated antigen, tissues, *e.g.*, crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions, *e.g.*, membrane fractions.

[00197] Antibodies which bind only a native MID 9002 protein, only denatured or otherwise non-native MID 9002 protein, or which bind both, are within the invention. Antibodies with linear or conformational epitopes are within the invention. Conformational epitopes sometimes can be identified by identifying antibodies which bind to native but not denatured MID 9002 protein.

Recombinant Expression Vectors, Host Cells and Genetically Engineered Cells

[00198] In another aspect, the invention includes, vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide described herein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors include, *e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses.

[00199] A vector can include a MID 9002 nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably the recombinant expression vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or polypeptides, including fusion proteins or polypeptides, encoded by nucleic acids as described herein (*e.g.*, MID 9002 proteins, mutant forms of MID 9002 proteins, fusion proteins, and the like).

[00200] The recombinant expression vectors of the invention can be designed for expression of MID 9002 proteins in prokaryotic or eukaryotic cells. For example, polypeptides of the invention can be expressed in *E. coli*, insect cells (*e.g.*, using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA . Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[00201] Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

[00202] Purified fusion proteins can be used in MID 9002 activity assays, (*e.g.*, direct assays or competitive assays described in detail below), or to generate antibodies specific or selective for MID 9002 proteins. In a preferred embodiment, a fusion protein expressed in a retroviral expression vector of the present invention can be used to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (*e.g.*, six weeks).

[00203] To maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman (1990) *Gene Expression Technology: Methods in*

Enzymology 185, Academic Press, San Diego, California 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[00204] The MID 9002 expression vector can be a yeast expression vector, a vector for expression in insect cells, *e.g.*, a baculovirus expression vector or a vector suitable for expression in mammalian cells.

[00205] When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

[00206] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example, the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

[00207] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. Regulatory sequences (*e.g.*, viral promoters and/or enhancers) operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes see Weintraub *et al.*, (1986) *Reviews - Trends in Genetics* 1:1.

[00208] Another aspect the invention provides a host cell which includes a nucleic acid molecule described herein, *e.g.*, a MID 9002 nucleic acid molecule within a recombinant expression vector or a MID 9002 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications can occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[00209] A host cell can be any prokaryotic or eukaryotic cell. For example, a MID 9002 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary (CHO) cells or CV-1 origin, SV-40 (COS) cells). Other suitable host cells are known to those skilled in the art.

[00210] Vector DNA can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

[00211] A host cell of the invention can be used to produce (i.e., express) a MID 9002 protein. Accordingly, the invention further provides methods for producing a MID 9002 protein using the host cells of the invention. In one embodiment, the method includes culturing the host cell of the invention (into which a recombinant expression vector encoding a MID 9002 protein has been introduced) in a suitable medium such that a MID 9002 protein is produced. In another embodiment, the method further includes isolating a MID 9002 protein from the medium or the host cell.

[00212] In another aspect, the invention features, a cell or purified preparation of cells which include a MID 9002 transgene, or which otherwise misexpress MID 9002. The cell preparation can consist of human or non-human cells, *e.g.*, rodent cells, *e.g.*, mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a MID 9002 transgene, *e.g.*, a heterologous form of a MID 9002, *e.g.*, a gene derived from humans (in the case of a non-human cell). The MID 9002 transgene can be misexpressed, *e.g.*, overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpresses an endogenous MID 9002, *e.g.*, a gene the expression of which is disrupted, *e.g.*, a knockout. Such cells can serve as a model for studying disorders which are related to mutated or misexpressed MID 9002 alleles or for use in drug screening.

[00213] In another aspect, the invention features, a human cell, *e.g.*, a hematopoietic stem cell, transformed with nucleic acid which encodes a subject MID 9002 polypeptide.

[00214] Also provided are cells, preferably human cells, *e.g.*, human hematopoietic or fibroblast cells, in which an endogenous MID 9002 is under the control of a regulatory sequence that does not normally control the expression of the endogenous MID 9002 gene. The expression characteristics of an endogenous gene within a cell, *e.g.*, a cell line or microorganism, can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous MID 9002 gene. For example, an endogenous MID 9002 gene which is "transcriptionally silent," *e.g.*, not normally expressed, or expressed only at very low levels, can be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, *e.g.*, Chappel, US 5,272,071; WO 91/06667, published in May 16, 1991.

Transgenic Animals

[00215] The invention provides non-human transgenic animals. Such animals are useful for studying the function and/or activity of a MID 9002 protein and for identifying and/or evaluating modulators of MID 9002 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA or a rearrangement, *e.g.*, a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic animal. A transgene can direct the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, other transgenes, *e.g.*, a knockout, reduce expression. Thus, a transgenic animal can be one in which an endogenous MID 9002 gene has been altered by, *e.g.*, by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

[00216] Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a transgene of the invention to direct expression of a MID 9002 protein to particular cells. A transgenic founder animal can be identified based upon the presence of a MID 9002 transgene in its genome and/or expression of MID 9002 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a MID 9002 protein can further be bred to other transgenic animals carrying other transgenes.

[00217] MID 9002 proteins or polypeptides can be expressed in transgenic animals or plants, *e.g.*, a nucleic acid encoding the protein or polypeptide can be introduced into the genome of an animal. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, *e.g.*, a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Suitable animals are mice, pigs, cows, goats, and sheep.

[00218] The invention also includes a population of cells from a transgenic animal, as discussed, *e.g.*, below.

Uses

[00219] The nucleic acid molecules, proteins, protein homologs, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (*e.g.*, therapeutic and prophylactic).

[00220] The isolated nucleic acid molecules of the invention can be used, for example, to express a MID 9002 protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect a MID 9002 mRNA (*e.g.*, in a biological sample) or a genetic alteration in a MID 9002 gene, and to modulate MID 9002 activity, as described further below. The MID 9002 proteins can be used to treat disorders characterized by insufficient or excessive production of a MID 9002 substrate or production of MID 9002 inhibitors. In addition, the MID 9002 proteins can be used to screen for naturally occurring MID 9002 substrates, to screen for drugs or compounds which modulate MID 9002 activity, as well as to treat disorders characterized by insufficient or excessive production of MID 9002 protein or production of MID 9002 protein forms which have decreased, aberrant or unwanted activity compared to MID 9002 wild type protein (*e.g.*, aberrant or deficient sulfatase function or expression). Moreover, the anti-MID 9002 antibodies of the invention can be used to detect and isolate MID 9002 proteins, regulate the bioavailability of MID 9002 proteins, and modulate MID 9002 activity.

[00221] A method of evaluating a compound for the ability to interact with, *e.g.*, bind, a subject MID 9002 polypeptide is provided. The method includes: contacting the compound with the subject MID 9002 polypeptide; and evaluating ability of the compound to interact with, *e.g.*, to bind or form a complex with the subject MID 9002 polypeptide. This method can be performed *in vitro*, *e.g.*, in a cell free system, or *in vivo*, *e.g.*, in a two-hybrid interaction trap assay. This method can be used to identify naturally occurring molecules which interact with subject MID 9002 polypeptide. It can also be used to find natural or synthetic inhibitors of subject MID 9002 polypeptide. Screening methods are discussed in more detail below.

Screening Assays:

[00222] The invention provides methods (also referred to herein as "screening assays") for identifying modulators, i.e., candidate or test compounds or agents (*e.g.*, proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to MID 9002 proteins, have a stimulatory or inhibitory effect on, for example, MID 9002 expression or MID 9002 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a MID 9002 substrate. Compounds thus identified can be used to modulate the activity of target gene products (*e.g.*, MID 9002 genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions.

[00223] In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a MID 9002 protein or polypeptide or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a MID 9002 protein or polypeptide or a biologically active portion thereof.

[00224] The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, *e.g.*, Zuckermann *et al.* (1994) *J. Med. Chem.* 37:2678-85); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des.*12:145).

[00225] Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909-13; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422-426; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678-85; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233-51.

[00226] Libraries of compounds can be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner, USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382; Felici (1991) *J. Mol. Biol.* 222:301-310; Ladner *supra.*).

[00227] In one embodiment, an assay is a cell-based assay in which a cell which expresses a MID 9002 protein or biologically active portion thereof is contacted with a test compound, and the ability of the test compound to modulate MID 9002 activity is determined. Determining the ability of the test compound to modulate MID 9002 activity can be accomplished by monitoring, for example, the sulfatase function. The cell, for example, can be of mammalian origin, *e.g.*, human.

[00228] The ability of the test compound to modulate MID 9002 binding to a compound, *e.g.*, a MID 9002 substrate, or to bind to MID 9002 can also be evaluated. This can be accomplished, for example, by coupling the compound, *e.g.*, the substrate, with a radioisotope or enzymatic label such that binding of the compound, *e.g.*, the substrate, to MID 9002 can be determined by detecting the labeled compound, *e.g.*, substrate, in a complex. Alternatively, MID 9002 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate MID 9002 binding to a MID 9002 substrate in a complex. For example, compounds (*e.g.*, MID 9002 substrates) can be labeled with ^{125}I, ^{14}C, ^{35}S or ^{3}H., either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

[00229] The ability of a compound (*e.g.*, a MID 9002 substrate) to interact with MID 9002 with or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with MID 9002 without the labeling of either the compound or the MID 9002. McConnell *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (*e.g.*, Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and MID 9002.

[00230] In yet another embodiment, a cell-free assay is provided in which a MID 9002 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the MID 9002 protein or biologically active portion thereof is evaluated. Preferred biologically active portions of the MID 9002 proteins to be used in assays of the present invention include fragments which participate in interactions with non-MID 9002 molecules, *e.g.*, fragments with high surface probability scores.

[00231] Soluble and/or membrane-bound forms of isolated proteins (*e.g.*, MID 9002 proteins or biologically active portions thereof) can be used in the cell-free assays of the invention. When membrane-bound forms of the protein are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

[00232] Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

[00233] The interaction between two molecules can also be detected, *e.g.*, using fluorescence energy transfer (FET) (see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos, *et al.*, U.S. Patent No. 4,868,103). A fluorophore label on the first, ‘donor’ molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, ‘acceptor’ molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the ‘donor’ protein molecule can simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the ‘acceptor’ molecule label can be differentiated from that of the ‘donor’. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the ‘acceptor’ molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (*e.g.*, using a fluorimeter).

[00234] In another embodiment, determining the ability of the MID 9002 protein to bind to a target molecule can be accomplished using real-time Biomolecular

Interaction Analysis (BIA) (see, *e.g.*, Sjolander and Urbaniczky (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[00235] In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

[00236] It may be desirable to immobilize either MID 9002, an anti-MID 9002 antibody or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a MID 9002 protein, or interaction of a MID 9002 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/MID 9002 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or MID 9002 protein, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of MID 9002 binding or activity determined using standard techniques.

[00237] Other techniques for immobilizing either a MID 9002 protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated MID 9002 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

[00238] In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; *e.g.*, using a labeled antibody specific or selective for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, *e.g.*, a labeled anti-Ig antibody).

[00239] In one embodiment, this assay is performed utilizing antibodies reactive with MID 9002 protein or target molecules but which do not interfere with binding of the MID 9002 protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or MID 9002 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the MID 9002 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the MID 9002 protein or target molecule.

[00240] Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (see, for example, Rivas and Minton (1993) *Trends Biochem Sci* 18:284-7); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, *e.g.*, Ausubel *et al.*, eds. (1999) *Current Protocols in Molecular Biology*, J. Wiley, New York.); and immunoprecipitation (see, for example, Ausubel *et al.*, eds. (1999) *Current Protocols in Molecular Biology*, J. Wiley, New York). Such resins and chromatographic techniques are known to one skilled in the art (see, *e.g.*, Heegaard (1998) *J Mol Recognit* 11:141-8; Hage and Tweed (1997) *J Chromatogr B Biomed Sci Appl.* 699:499-525). Further, fluorescence energy transfer can also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[00241] In a preferred embodiment, the assay includes contacting the MID 9002 protein or biologically active portion thereof with a known compound which binds MID 9002 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a MID 9002 protein, wherein determining the ability of the test compound to interact with a MID 9002 protein includes determining the ability of the test compound to preferentially bind to MID 9002 or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

[00242] The target gene products of the invention can, *in vivo*, interact with one or more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as "binding partners." Compounds that disrupt such interactions can be useful in regulating the activity of the target gene product. Such compounds can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the MID 9002 genes herein identified. In an alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of a MID 9002 protein through modulation of the activity of a downstream effector of a MID 9002 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

[00243] To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

[00244] These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, *e.g.*, by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, *e.g.*, compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

[00245] In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (*e.g.*, a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific or selective for the species to be anchored can be used to anchor the species to the solid surface.

[00246] In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; *e.g.*, using a labeled antibody specific or selective for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, *e.g.*, a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

[00247] Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; *e.g.*, using an immobilized antibody specific or selective for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific or selective for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

[00248] In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, *e.g.*, U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

[00249] In yet another aspect, the MID 9002 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with MID 9002 ("MID 9002-binding proteins" or "MID 9002-bp") and are involved in MID 9002 activity. Such MID 9002-bps can be activators or inhibitors of signals by the MID 9002 proteins or MID 9002 targets as, for example, downstream elements of a MID 9002-mediated signaling pathway.

[00250] The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a MID 9002 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively the: MID 9002 protein can be the fused to the activator domain.) If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming a MID 9002-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, lacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the MID 9002 protein.

[00251] In another embodiment, modulators of MID 9002 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of MID 9002 mRNA or protein evaluated relative to the level of expression of MID 9002 mRNA or protein in the absence of the candidate compound. When expression of MID 9002 mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of MID 9002 mRNA or protein expression. Alternatively, when expression of MID 9002 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of MID 9002 mRNA or protein expression. The level of MID 9002 mRNA or protein expression can be determined by methods described herein for detecting MID 9002 mRNA or protein.

[00252] In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a MID 9002 protein can be confirmed *in vivo*, *e.g.*, in an animal such as an animal model for aberrant or deficient sulfatase function or expression or an animal model for colon cancer.

[00253] This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein (*e.g.*, a MID 9002 modulating agent, an antisense MID 9002 nucleic acid molecule, a MID 9002-specific antibody, or a MID 9002-binding partner) in an appropriate animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be used for treatments as described herein.

Detection Assays

[00254] Portions or fragments of the nucleic acid sequences identified herein can be used as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome *e.g.*, to locate gene regions associated with genetic disease or to associate MID 9002 with a disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

Chromosome Mapping

[00255] The MID 9002 nucleotide sequences or portions thereof can be used to map the location of the MID 9002 genes on a chromosome. This process is called chromosome mapping. Chromosome mapping is useful in correlating the MID 9002 sequences with genes associated with disease.

[00256] Briefly, MID 9002 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the MID 9002 nucleotide sequences. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the MID 9002 sequences will yield an amplified fragment.

[00257] A panel of somatic cell hybrids in which each cell line contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, can allow easy mapping of individual genes to specific human chromosomes. (D'Eustachio *et al.* (1983) *Science* 220:919-924).

[00258] Other mapping strategies *e.g.*, in situ hybridization (described in Fan *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries can be used to map MID 9002 to a chromosomal location.

[00259] Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.* (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York).

[00260] Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

[00261] Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in McKusick, *Mendelian Inheritance in Man*, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland *et al.* (1987) *Nature,* 325:783-787.

[00262] Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the MID 9002 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

[00263] MID 9002 sequences can be used to identify individuals from biological samples using, *e.g.*, restriction fragment length polymorphism (RFLP). In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, the fragments separated, *e.g.*, in a Southern blot, and probed to yield bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

[00264] Furthermore, the sequences of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the MID 9002 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it. Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences.

[00265] Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 can provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

[00266] If a panel of reagents from MID 9002 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

Use of Partial MID 9002 Sequences in Forensic Biology

[00267] DNA-based identification techniques can also be used in forensic biology. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

[00268] The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 (*e.g.*, fragments derived from the noncoding regions of SEQ ID NO:1 having a length of at least 20 bases, preferably at least 30 bases) are particularly appropriate for this use.

[00269] The MID 9002 nucleotide sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such MID 9002 probes can be used to identify tissue by species and/or by organ type.

[00270] In a similar fashion, these reagents, *e.g.*, MID 9002 primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

[00271] The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual.

[00272] Generally, the invention provides, a method of determining if a subject is at risk for a disorder related to a lesion in or the misexpression of a gene which encodes MID 9002.

[00273] Such disorders include, *e.g.*, a disorder associated with the misexpression of MID 9002 gene; a disorder of the gastrointestinal system.

[00274] The method includes one or more of the following:

[00275] detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the MID 9002 gene, or detecting the presence or absence of a mutation in a region which controls the expression of the gene, *e.g.*, a mutation in the 5' control region;

[00276] detecting, in a tissue of the subject, the presence or absence of a mutation which alters the structure of the MID 9002 gene;

[00277] detecting, in a tissue of the subject, the misexpression of the MID 9002 gene, at the mRNA level, *e.g.*, detecting a non-wild type level of an mRNA;

[00278] detecting, in a tissue of the subject, the misexpression of the gene, at the protein level, *e.g.*, detecting a non-wild type level of a MID 9002 polypeptide.

[00279] In preferred embodiments the method includes: ascertaining the existence of at least one of: a deletion of one or more nucleotides from the MID 9002 gene; an insertion of one or more nucleotides into the gene, a point mutation, *e.g.*, a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, *e.g.*, a translocation, inversion, or deletion.

[00280] For example, detecting the genetic lesion can include: (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence from SEQ ID NO:1, or naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the MID 9002 gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting, by hybridization, *e.g.*, *in situ* hybridization, of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion.

[00281] In preferred embodiments detecting the misexpression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the MID 9002 gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; or a non-wild type level of MID 9002.

[00282] Methods of the invention can be used prenatally or to determine if a subject's offspring will be at risk for a disorder.

[00283] In preferred embodiments the method includes determining the structure of a MID 9002 gene, an abnormal structure being indicative of risk for the disorder.

[00284] In preferred embodiments the method includes contacting a sample from the subject with an antibody to the MID 9002 protein or a nucleic acid, which hybridizes specifically with the gene. These and other embodiments are discussed below.

Diagnostic and Prognostic Assays

[00285] The presence, level, or absence of MID 9002 protein or nucleic acid in a biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting MID 9002 protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes MID 9002 protein such that the presence of MID 9002 protein or nucleic acid is detected in the biological sample. The term "biological sample" includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. A preferred biological sample is serum. The level of expression of the MID 9002 gene can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the MID 9002 genes; measuring the amount of protein encoded by the MID 9002 genes; or measuring the activity of the protein encoded by the MID 9002 genes.

[00286] The level of mRNA corresponding to the MID 9002 gene in a cell can be determined both by *in situ* and by *in vitro* formats.

[00287] The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length MID 9002 nucleic acid, such as the nucleic acid of SEQ ID NO:1, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to MID 9002 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays are described herein.

[00288] In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with the probes, for example, in a two-dimensional gene chip array. A skilled artisan can adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the MID 9002 genes.

[00289] The level of mRNA in a sample that is encoded by one of MID 9002 can be evaluated with nucleic acid amplification, *e.g.*, by rtPCR (Mullis (1987) U.S. Patent No. 4,683,202), ligase chain reaction (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193), self sustained sequence replication (Guatelli *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.*, (1989), *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.*, (1988) *Bio/Technology* 6:1197), rolling circle replication (Lizardi *et al.*, U.S. Patent No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

[00290] For *in situ* methods, a cell or tissue sample can be prepared/processed and immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the MID 9002 gene being analyzed.

[00291] In another embodiment, the methods further contacting a control sample with a compound or agent capable of detecting MID 9002 mRNA, or genomic DNA, and comparing the presence of MID 9002 mRNA or genomic DNA in the control sample with the presence of MID 9002 mRNA or genomic DNA in the test sample.

[00292] A variety of methods can be used to determine the level of protein encoded by MID 9002. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample. In a preferred embodiment, the antibody bears a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with a detectable substance. Examples of detectable substances are provided herein.

[00293] The detection methods can be used to detect MID 9002 protein in a biological sample *in vitro* as well as *in vivo. In vitro* techniques for detection of MID 9002 protein include enzyme linked immunosorbent assays (ELISAs), immunoprecipitations, immunofluorescence, enzyme immunoassay (EIA), radioimmunoassay (RIA), and Western blot analysis. *In vivo* techniques for detection of MID 9002 protein include introducing into a subject a labeled anti-MID 9002 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

[00294] In another embodiment, the methods further include contacting the control sample with a compound or agent capable of detecting MID 9002 protein, and comparing the presence of MID 9002 protein in the control sample with the presence of MID 9002 protein in the test sample.

[00295] The invention also includes kits for detecting the presence of MID 9002 in a biological sample. For example, the kit can include a compound or agent capable of detecting MID 9002 protein or mRNA in a biological sample; and a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect MID 9002 protein or nucleic acid.

[00296] For antibody-based kits, the kit can include: (1) a first antibody (*e.g.*, attached to a solid support) which binds to a polypeptide corresponding to a marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

[00297] For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, *e.g.*, a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also includes a buffering agent, a preservative, or a protein stabilizing agent. The kit can also includes components necessary for detecting the detectable agent (*e.g.*, an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

[00298] The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with misexpressed or aberrant or unwanted MID 9002 expression or activity. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as pain or deregulated cell proliferation.

[00299] In one embodiment, a disease or disorder associated with aberrant or unwanted MID 9002 expression or activity is identified. A test sample is obtained from a subject and MID 9002 protein or nucleic acid (*e.g.*, mRNA or genomic DNA) is evaluated, wherein the level, *e.g.*, the presence or absence, of MID 9002 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted MID 9002 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest, including a biological fluid (*e.g.*, serum), cell sample, or tissue.

[00300] The prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted MID 9002 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a proliferative/differentiative disorder such as colon cancer.

[00301] The methods of the invention can also be used to detect genetic alterations in a MID 9002 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in MID 9002 protein activity or nucleic acid expression, such as a proliferative/differentiative disorder such as colon cancer. In preferred embodiments, the methods include detecting, in a sample from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a MID 9002-protein, or the mis-expression of the MID 9002 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a MID 9002 gene; 2) an addition of one or more nucleotides to a MID 9002 gene; 3) a substitution of one or more nucleotides of a MID 9002 gene, 4) a chromosomal rearrangement of a MID 9002 gene; 5) an alteration in the level of a messenger RNA transcript of a MID 9002 gene, 6) aberrant modification of a MID 9002 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a MID 9002 gene, 8) a non-wild type level of a MID 9002-protein, 9) allelic loss of a MID 9002 gene, and 10) inappropriate post-translational modification of a MID 9002-protein.

[00302] An alteration can be detected without a probe/primer in a polymerase chain reaction, such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR), the latter of which can be particularly useful for detecting point mutations in the MID 9002-gene. This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a MID 9002 gene under conditions such that hybridization and amplification of the MID 9002 gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein. Alternatively, other amplification methods described herein or known in the art can be used.

[00303] In another embodiment, mutations in a MID 9002 gene from a sample cell can be identified by detecting alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined, *e.g.*, by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

[00304] In other embodiments, genetic mutations in MID 9002 can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, two dimensional arrays, *e.g.*, chip based arrays. Such arrays include a plurality of addresses, each of which is positionally distinguishable from the other. A different probe is located at each address of the plurality. The arrays can have a high density of addresses, *e.g.*, can contain hundreds or thousands of oligonucleotides probes (Cronin *et al.* (1996)

Human Mutation 7: 244-255; Kozal *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in MID 9002 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

[00305] In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the MID 9002 gene and detect mutations by comparing the sequence of the sample MID 9002 with the corresponding wild-type (control) sequence. Automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve *et al.* (1995) *Biotechniques* 19:448-53), including sequencing by mass spectrometry.

[00306] Other methods for detecting mutations in the MID 9002 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242; Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295).

[00307] In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in MID 9002 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662; U.S. Patent No. 5,459,039).

[00308] In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in MID 9002 genes. For example, single strand conformation polymorphism (SSCP) can be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992)

Genet. Anal. Tech. Appl. 9:73-79). Single-stranded DNA fragments of sample and control MID 9002 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments can be labeled or detected with labeled probes. The sensitivity of the assay can be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

[00309] In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

[00310] Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230).

[00311] Alternatively, allele specific amplification technology which depends on selective PCR amplification can be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification can carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification can also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189-

93). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

[00312] The methods described herein can be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which can be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a MID 9002 gene.

Use of MID 9002 Molecules as Surrogate Markers

[00313] The MID 9002 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the MID 9002 molecules of the invention can be detected, and can be correlated with one or more biological states *in vivo*. For example, the MID 9002 molecules of the invention can serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (*e.g.*, with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers can serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (*e.g.*, early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (*e.g.*, an assessment of cardiovascular disease can be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection can be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

[00314] The MID 9002 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker can be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug can be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker can be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug can be sufficient to activate multiple rounds of marker (*e.g.*, a MID 9002 marker) transcription or expression, the amplified marker can be in a quantity which is more readily detectable than the drug itself. Also, the marker can be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-MID 9002 antibodies can be employed in an immune-based detection system for a MID 9002 protein marker, or MID 9002-specific radiolabeled probes can be used to detect a MID 9002 mRNA marker. Furthermore, the use of a pharmacodynamic marker can offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

[00315] The MID 9002 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, *e.g.*, McLeod *et al.* (1999) *Eur. J. Cancer* 35:1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, can be selected. For example, based on the presence or quantity of RNA, or protein (*e.g.*, MID 9002 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment can be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in MID 9002 DNA can correlate with a MID 9002 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

Pharmaceutical Compositions

[00316] The nucleic acid and polypeptides, fragments thereof, as well as anti-MID 9002 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions. Such compositions typically include the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

[00317] A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[00318] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[00319] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[00320] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, *e.g.*, gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[00321] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

[00322] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

[00323] The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[00324] In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[00325] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[00326] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ ED_{50}. Compounds which exhibit high therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[00327] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[00328] As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors can influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody, unconjugated or conjugated as described herein, can include a single treatment or, preferably, can include a series of treatments.

[00329] For antibodies, the preferred dosage is 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (*e.g.*, into the brain). A method for lipidation of antibodies is described by Cruikshank *et al.* ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

[00330] The present invention encompasses agents which modulate expression or activity. An agent can, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics (*e.g.*, peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e.,. including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

[00331] Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (*e.g.*, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher can, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[00332] The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see *e.g.*, Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

[00333] The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Methods of Treatment:

[00334] The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted MID 9002 expression or activity. As used herein, the term "treatment" is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

[00335] With regards to both prophylactic and therapeutic methods of treatment, such treatments can be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (*e.g.*, a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the MID 9002 molecules of the present invention or MID 9002 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

[00336] In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted MID 9002 expression or activity, by administering to the subject a MID 9002 or an agent which modulates MID 9002 expression or at least one MID 9002 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted MID 9002 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the MID 9002 aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of MID 9002 aberrance, for example, a MID 9002, MID 9002 agonist or MID 9002 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

[00337] It is possible that some MID 9002 disorders can be caused, at least in part, by an abnormal level of gene product, or by the presence of a gene product exhibiting abnormal activity. As such, the reduction in the level and/or activity of such gene products would bring about the amelioration of disorder symptoms.

[00338] The MID 9002 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more of cellular proliferative and/or differentiative disorders, which are described above. The molecules of the invention also can act as novel diagnostic targets and therapeutic agents for controlling one or more of disorders associated with bone metabolism, immune, *e.g.*, inflammatory, disorders, cardiovascular disorders, endothelial cell disorders, liver disorders, viral diseases, pain disorders and metabolic disorders.

[00339] Examples of cellular proliferative and/or differentiative disorders include cancer, *e.g.*, carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic disorders, *e.g.*, leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of prostate, colon, lung, breast and liver origin.

[00340] As used herein, the term “cancer” (also used interchangeably with the terms, "hyperproliferative" and "neoplastic") refers to cells having the capacity for autonomous growth, i.e., an abnormal state or condition characterized by rapidly proliferating cell growth. Cancerous disease states may be categorized as pathologic, i.e., characterizing or constituting a disease state, *e.g.*, malignant tumor growth, or may be categorized as non-pathologic, i.e., a deviation from normal but not associated with a disease state, *e.g.*, cell proliferation associated with wound repair. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. The term “cancer” includes malignancies of the various organ systems, such as those affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus. The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term "carcinoma" also includes carcinosarcomas, *e.g.*, which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures. The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

[00341] The MID 9002 molecules of the invention can be used to monitor, treat and/or diagnose a variety of proliferative disorders. Such disorders include hematopoietic neoplastic disorders. As used herein, the term "hematopoietic neoplastic disorders" includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, *e.g.*, arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Preferably, the diseases arise from poorly differentiated acute leukemias, *e.g.*, erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus (1991) *Crit Rev. in Oncol./Hemotol.* 11:267-97); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

[00342] Aberrant expression and/or activity of MID 9002 molecules can mediate disorders associated with bone metabolism. "Bone metabolism" refers to direct or indirect effects in the formation or degeneration of bone structures, *e.g.*, bone formation, bone resorption, etc., which can ultimately affect the concentrations in serum of calcium and phosphate. This term also includes activities mediated by MID 9002 molecules in bone cells, *e.g.* osteoclasts and osteoblasts, that can in turn result in bone formation and degeneration. For example, MID 9002 molecules can support different activities of bone resorbing osteoclasts such as the stimulation of differentiation of monocytes and mononuclear phagocytes into osteoclasts. Accordingly, MID 9002 molecules that modulate the production of bone cells can influence bone formation and degeneration, and thus can be used to treat bone disorders. Examples of such disorders include, but are not limited to, osteoporosis, osteodystrophy, osteomalacia, rickets, osteitis fibrosa cystica, renal osteodystrophy, osteosclerosis, anti-convulsant treatment, osteopenia, fibrogenesis-imperfecta ossium, secondary hyperparathyrodism, hypoparathyroidism, hyperparathyroidism, cirrhosis, obstructive jaundice, drug induced metabolism, medullary carcinoma, chronic renal disease, rickets, sarcoidosis, glucocorticoid antagonism, malabsorption syndrome, steatorrhea, tropical sprue, idiopathic hypercalcemia and milk fever.

[00343] The MID 9002 nucleic acid and protein of the invention can be used to treat and/or diagnose a variety of immune, *e.g.*, inflammatory (*e.g.* respiratory inflammatory) disorders. Examples immune and inflammatory disorders or diseases include, but are not limited to, autoimmune diseases (including, for example, diabetes mellitus, arthritis (including rheumatoid arthritis, juvenile rheumatoid arthritis, osteoarthritis, psoriatic arthritis), multiple sclerosis, encephalomyelitis, myasthenia gravis, systemic lupus erythematosis, autoimmune thyroiditis, dermatitis (including atopic dermatitis and eczematous dermatitis), psoriasis, Sjögren's Syndrome, inflammatory bowel disease, *e.g.* Crohn's disease and ulcerative colitis, aphthous ulcer, iritis, conjunctivitis, keratoconjunctivitis, asthma, allergic asthma, chronic obstructive pulmonary disease, cutaneous lupus erythematosus, scleroderma, vaginitis, proctitis, drug eruptions, leprosy reversal reactions, erythema nodosum leprosum, autoimmune uveitis, allergic encephalomyelitis, acute necrotizing hemorrhagic encephalopathy, idiopathic bilateral progressive sensorineural hearing loss, aplastic anemia, pure red cell anemia, idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis, chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue, lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis, uveitis posterior, and interstitial lung fibrosis), graft-versus-host disease, cases of transplantation, and allergy such as, atopic allergy.

[00344] As used herein, disorders involving the heart, or "cardiovascular disease" or a "cardiovascular disorder" includes a disease or disorder which affects the cardiovascular system, e.g., the heart, the blood vessels, and/or the blood. A cardiovascular disorder can be caused by an imbalance in arterial pressure, a malfunction of the heart, or an occlusion of a blood vessel, *e.g.*, by a thrombus. A cardiovascular disorder includes, but is not limited to disorders such as arteriosclerosis, atherosclerosis, cardiac hypertrophy, ischemia reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism, tachycardia, bradycardia, pressure overload, aortic bending, coronary artery ligation, vascular heart disease, valvular disease, including but not limited to, valvular degeneration caused by calcification, rheumatic heart disease, endocarditis, or complications of artificial valves; atrial fibrillation, long-QT syndrome, congestive heart failure, sinus node dysfunction, angina, heart failure, hypertension, atrial fibrillation, atrial flutter, pericardial disease, including but not limited to, pericardial effusion and pericarditis; cardiomyopathies, *e.g.,* dilated cardiomyopathy or idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, ischemic disease, arrhythmia, sudden cardiac death, and cardiovascular developmental disorders (*e.g.*, arteriovenous malformations, arteriovenous fistulae, raynaud's syndrome, neurogenic thoracic outlet syndrome, causalgia/reflex sympathetic dystrophy, hemangioma, aneurysm, cavernous angioma, aortic valve stenosis, atrial septal defects, atrioventricular canal, coarctation of the aorta, ebsteins anomaly, hypoplastic left heart syndrome, interruption of the aortic arch, mitral valve prolapse, ductus arteriosus, patent foramen ovale, partial anomalous pulmonary venous return, pulmonary atresia with ventricular septal defect, pulmonary atresia without ventricular septal defect, persistance of the fetal circulation, pulmonary valve stenosis, single ventricle, total anomalous pulmonary venous return, transposition of the great vessels, tricuspid atresia, truncus arteriosus, ventricular septal defects). A cardiovascular disease or disorder also can include an endothelial cell disorder.

[00345] As used herein, an "endothelial cell disorder" includes a disorder characterized by aberrant, unregulated, or unwanted endothelial cell activity, *e.g.*, proliferation, migration, angiogenesis, or vascularization; or aberrant expression of cell surface adhesion molecules or genes associated with angiogenesis, *e.g.*, TIE-2, FLT and FLK. Endothelial cell disorders include tumorigenesis, tumor metastasis, psoriasis, diabetic retinopathy, endometriosis, Grave's disease, ischemic disease (*e.g.*, atherosclerosis), and chronic inflammatory diseases (*e.g.*, rheumatoid arthritis).

[00346] Disorders which can be treated or diagnosed by methods described herein include, but are not limited to, disorders associated with an accumulation in the liver of fibrous tissue, such as that resulting from an imbalance between production and degradation of the extracellular matrix accompanied by the collapse and condensation of preexisting fibers. The methods described herein can be used to diagnose or treat hepatocellular necrosis or injury induced by a wide variety of agents including processes which disturb homeostasis, such as an inflammatory process, tissue damage resulting from toxic injury or altered hepatic blood flow, and infections (*e.g.*, bacterial, viral and parasitic). For example, the methods can be used for the early detection of hepatic injury, such as portal hypertension or hepatic fibrosis. In addition, the methods can be employed to detect liver fibrosis attributed to inborn errors of metabolism, for example, fibrosis resulting from a storage disorder such as Gaucher's disease (lipid abnormalities) or a glycogen storage disease, A1-antitrypsin deficiency; a disorder mediating the accumulation (*e.g.*, storage) of an exogenous substance, for example, hemochromatosis (iron-overload syndrome) and copper storage diseases (Wilson's disease), disorders resulting in the accumulation of a toxic metabolite (*e.g.*, tyrosinemia, fructosemia and galactosemia) and peroxisomal disorders (*e.g.*, Zellweger syndrome). Additionally, the methods described herein can be useful for the early detection and treatment of liver injury associated with the administration of various chemicals or drugs, such as for example, methotrexate, isonizaid, oxyphenisatin, methyldopa, chlorpromazine, tolbutamide or alcohol, or which represents a hepatic manifestation of a vascular disorder such as obstruction of either the intrahepatic or extrahepatic bile flow or an alteration in hepatic circulation resulting, for example, from chronic heart failure, veno-occlusive disease, portal vein thrombosis or Budd-Chiari syndrome.

[00347] Additionally, MID 9002 molecules can play an important role in the etiology of certain viral diseases, including but not limited to Hepatitis B, Hepatitis C and Herpes Simplex Virus (HSV). Modulators of MID 9002 activity could be used to control viral diseases. The modulators can be used in the treatment and/or diagnosis of viral infected tissue or virus-associated tissue fibrosis, especially liver and liver fibrosis. Also, MID 9002 modulators can be used in the treatment and/or diagnosis of virus-associated carcinoma, especially hepatocellular cancer.

[00348] Additionally, MID 9002 can play an important role in the regulation of metabolism or pain disorders. Diseases of metabolic imbalance include, but are not limited to, obesity, anorexia nervosa, cachexia, lipid disorders, and diabetes. Examples of pain disorders include, but are not limited to, pain response elicited during various forms of tissue injury, *e.g.*, inflammation, infection, and ischemia, usually referred to as hyperalgesia (described in, for example, Fields, H.L. (1987) *Pain*, New York:McGraw-Hill); pain associated with musculoskeletal disorders, *e.g.*, joint pain; tooth pain; headaches; pain associated with surgery; pain related to irritable bowel syndrome; or chest pain.

[00349] As discussed, successful treatment of MID 9002 disorders can be brought about by techniques that serve to inhibit the expression or activity of target gene products. For example, compounds, *e.g.*, an agent identified using an assays described above, that proves to exhibit negative modulatory activity, can be used in accordance with the invention to prevent and/or ameliorate symptoms of MID 9002 disorders. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, or antibodies (including, for example, polyclonal, monoclonal, humanized, human, anti-idiotypic, chimeric or single chain antibodies, and Fab, $F(ab')_2$ and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof).

[00350] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

[00351] It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular protein, it can be preferable to co-administer normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

[00352] Another method by which nucleic acid molecules can be utilized in treating or preventing a disease characterized by MID 9002 expression is through the use of aptamer molecules specific for MID 9002 protein. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically or selectively bind to protein ligands (see, *e.g.*, Osborne *et al.* (1997) *Curr. Opin. Chem Biol.* 1: 5-9; and Patel (1997) *Curr Opin Chem Biol* 1:32-46). Since nucleic acid molecules can in many cases be more conveniently introduced into target cells than therapeutic protein molecules can be, aptamers offer a method by which MID 9002 protein activity can be specifically decreased without the introduction of drugs or other molecules which can have pluripotent effects.

[00353] Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies can, therefore, by administered in instances whereby negative modulatory techniques are appropriate for the treatment of MID 9002 disorders. For a description of antibodies, see the Antibody section above.

[00354] In circumstances wherein injection of an animal or a human subject with a MID 9002 protein or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against MID 9002 through the use of anti-idiotypic antibodies (see, for example, Herlyn (1999) *Ann Med* 31:66-78; and Bhattacharya-Chatterjee and Foon (1998) *Cancer Treat Res.* 94:51-68). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the MID 9002 protein. Vaccines directed to a disease characterized by MID 9002 expression can also be generated in this fashion.

[00355] In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies can be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (see *e.g.*, Marasco *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:7889-7893).

[00356] The identified compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat or ameliorate MID 9002 disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures as described above.

[00357] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[00358] Another example of determination of effective dose for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays can utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. The compound which is able to modulate MID 9002 activity is used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell *et al* (1996) *Current Opinion in Biotechnology* 7:89-94 and in Shea (1994) *Trends in Polymer Science* 2:166-173. Such "imprinted" affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis *et al* (1993) *Nature* 361:645-647. Through the use of isotope-labeling, the "free" concentration of compound which modulates the expression or activity of MID 9002 can be readily monitored and used in calculations of IC_{50}.

[00359] Such "imprinted" affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC_{50}. An rudimentary example of such a “biosensor” is discussed in Kriz *et al* (1995) *Analytical Chemistry* 67:2142-2144.

[00360] Another aspect of the invention pertains to methods of modulating MID 9002 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a MID 9002 or agent that modulates one or more of the activities of MID 9002 protein activity associated with the cell. An agent that modulates MID 9002 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a MID 9002 protein (*e.g.*, a MID 9002 substrate or receptor), a MID 9002 antibody, a MID 9002 agonist or antagonist, a peptidomimetic of a MID 9002 agonist or antagonist, or other small molecule.

[00361] In one embodiment, the agent stimulates one or MID 9002 activities. Examples of such stimulatory agents include active MID 9002 protein and a nucleic acid molecule encoding MID 9002. In another embodiment, the agent inhibits one or more MID 9002 activities. Examples of such inhibitory agents include antisense MID 9002 nucleic acid molecules, anti-MID 9002 antibodies, and MID 9002 inhibitors. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a MID 9002 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, up regulates or down regulates) MID 9002 expression or activity. In another embodiment, the method involves administering a MID 9002 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted MID 9002 expression or activity.

[00362] Stimulation of MID 9002 activity is desirable in situations in which MID 9002 is abnormally downregulated and/or in which increased MID 9002 activity is likely to have a beneficial effect. For example, stimulation of MID 9002 activity is desirable in situations in which a MID 9002 is downregulated and/or in which increased MID 9002 activity is likely to have a beneficial effect. Likewise, inhibition of MID 9002 activity is desirable in situations in which MID 9002 is abnormally upregulated and/or in which decreased MID 9002 activity is likely to have a beneficial effect.

Pharmacogenomics

[00363] The MID 9002 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on MID 9002 activity (*e.g.*, MID 9002 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) MID 9002-associated disorders (*e.g.*, aberrant or deficient sulfatase function or expression) associated with aberrant or unwanted MID 9002 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) can be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician can consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a MID 9002 molecule or MID 9002 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a MID 9002 molecule or MID 9002 modulator.

[00364] Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23:983-985 and Linder *et al.* (1997) *Clin. Chem.* 43:254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

[00365] One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (*e.g.*, a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.)

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

[00366] Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (*e.g.*, a MID 9002 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

[00367] Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (*e.g.*, a MID 9002 molecule or MID 9002 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

[00368] Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a MID 9002 molecule or MID 9002 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

[00369] The present invention further provides methods for identifying new agents, or combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the MID 9002 genes of the present invention, wherein these products can be associated with resistance of the cells to a therapeutic agent. Specifically, the activity of the proteins encoded by the MID 9002 genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By blocking the activity of one or more of the resistance proteins, target cells, *e.g.*, human cells, will become sensitive to treatment with an agent to which the unmodified target cells were resistant.

[00370] Monitoring the influence of agents (*e.g.*, drugs) on the expression or activity of a MID 9002 protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase MID 9002 gene expression, protein levels, or upregulate MID 9002 activity, can be monitored in clinical trials of subjects exhibiting decreased MID 9002 gene expression, protein levels, or downregulated MID 9002 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease MID 9002 gene expression, protein levels, or downregulate MID 9002 activity, can be monitored in clinical trials of subjects exhibiting increased MID 9002 gene expression, protein levels, or upregulated MID 9002 activity. In such clinical trials, the expression or activity of a MID 9002 gene, and preferably, other genes that have been implicated in, for example, a sulfatase-associated or another MID 9002-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

Other Embodiments

[00371] In another aspect, the invention features a method of analyzing a plurality of capture probes. The method is useful, *e.g.*, to analyze gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, *e.g.*, a nucleic acid or peptide sequence, wherein the capture probes are from a cell or subject which expresses MID 9002 or from a cell or subject in which a MID 9002 mediated response has been elicited; contacting the array with a MID 9002 nucleic acid (preferably purified), a MID 9002 polypeptide (preferably purified), or an anti-MID 9002 antibody, and thereby evaluating the plurality of capture probes. Binding, *e.g.*, in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, *e.g.*, by a signal generated from a label attached to the MID 9002 nucleic acid, polypeptide, or antibody.

[00372] The capture probes can be a set of nucleic acids from a selected sample, *e.g.*, a sample of nucleic acids derived from a control or non-stimulated tissue or cell.

[00373] The method can include contacting the MID 9002 nucleic acid, polypeptide, or antibody with a first array having a plurality of capture probes and a second array having a different plurality of capture probes. The results of each hybridization can be compared, *e.g.*, to analyze differences in expression between a first and second sample. The first plurality of capture probes can be from a control sample, *e.g.*, a wild type, normal, or non-diseased, non-stimulated, sample, *e.g.*, a biological fluid, tissue, or cell sample. The second plurality of capture probes can be from an experimental sample, *e.g.*, a mutant type, at risk, disease-state or disorder-state, or stimulated, sample, *e.g.*, a biological fluid, tissue, or cell sample.

[00374] The plurality of capture probes can be a plurality of nucleic acid probes each of which specifically hybridizes, with an allele of MID 9002. Such methods can be used to diagnose a subject, *e.g.*, to evaluate risk for a disease or disorder, to evaluate suitability of a selected treatment for a subject, to evaluate whether a subject has a disease or disorder.

[00375] The method can be used to detect SNPs, as described above.

[00376] In another aspect, the invention features, a method of analyzing MID 9002, *e.g.*, analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a MID 9002 nucleic acid or amino acid sequence; comparing the MID 9002 sequence with one or more preferably a plurality of sequences from a collection of sequences, *e.g.*, a nucleic acid or protein sequence database; to thereby analyze MID 9002.

[00377] The method can include evaluating the sequence identity between a MID 9002 sequence and a database sequence. The method can be performed by accessing the database at a second site, *e.g.*, over the internet. Preferred databases include GenBank™ and SwissProt.

[00378] In another aspect, the invention features, a set of oligonucleotides, useful, *e.g.*, for identifying SNP's, or identifying specific alleles of MID 9002. The set includes a plurality of oligonucleotides, each of which has a different nucleotide at an interrogation position, *e.g.*, an SNP or the site of a mutation. In a preferred embodiment, the oligonucleotides of the plurality identical in sequence with one another (except for differences in length). The oligonucleotides can be provided with differential labels, such that an oligonucleotide which hybridizes to one allele provides a signal that is distinguishable from an oligonucleotides which hybridizes to a second allele.

[00379] The sequences of MID 9002 molecules are provided in a variety of mediums to facilitate use thereof. A sequence can be provided as a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a MID 9002 molecule. Such a manufacture can provide a nucleotide or amino acid sequence, *e.g.*, an open reading frame, in a form which allows examination of the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exist in nature or in purified form.

[00380] A MID 9002 nucleotide or amino acid sequence can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact disc and CD-ROM; electrical storage media such as RAM, ROM, EPROM, EEPROM, and the like; and general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having thereon MID 9002 sequence information of the present invention.

[00381] As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus of other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as personal digital assistants (PDAs), cellular phones, pagers, and the like; and local and distributed processing systems.

[00382] As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the MID 9002 sequence information.

[00383] A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a MID 9002 nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of data processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

[00384] By providing the MID 9002 nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. A search is used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

[00385] The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has a sulfatase-associated or another MID 9002-associated disease or disorder or a pre-disposition to a sulfatase-associated or another MID 9002-associated disease or disorder, wherein the method comprises the steps of determining MID 9002 sequence information associated with the subject and based on the MID 9002 sequence information, determining whether the subject has a sulfatase-associated or another MID 9002-associated disease or disorder and/or recommending a particular treatment for the disease, disorder, or pre-disease condition.

[00386] The present invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a sulfatase-associated or another MID 9002-associated disease or disorder or a pre-disposition to a disease associated with MID 9002, wherein the method comprises the steps of determining

MID 9002 sequence information associated with the subject, and based on the MID 9002 sequence information, determining whether the subject has a sulfatase-associated or another MID 9002-associated disease or disorder or a pre-disposition to a sulfatase-associated or another MID 9002-associated disease or disorder, and/or recommending a particular treatment for the disease, disorder, or pre-disease condition. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the subject.

[00387] The present invention also provides in a network, a method for determining whether a subject has a sulfatase-associated or another MID 9002-associated disease or disorder or a pre-disposition to a sulfatase-associated or another MID 9002-associated disease or disorder, said method comprising the steps of receiving MID 9002 sequence information from the subject and/or information related thereto, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to MID 9002 and/or corresponding to a sulfatase-associated or another MID 9002-associated disease or disorder, and based on one or more of the phenotypic information, the MID 9002 information (*e.g.*, sequence information and/or information related thereto), and the acquired information, determining whether the subject has a sulfatase-associated or another MID 9002-associated disease or disorder or a pre-disposition to a sulfatase-associated or another MID 9002-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder, or pre-disease condition.

[00388] The present invention also provides a business method for determining whether a subject has a sulfatase-associated or another MID 9002-associated disease or disorder or a pre-disposition to a sulfatase-associated or another MID 9002-associated disease or disorder, said method comprising the steps of receiving information related to MID 9002 (*e.g.*, sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to MID 9002 and/or related to a sulfatase-associated or another MID 9002-associated disease or disorder, and based on one or more of the phenotypic information, the MID 9002 information, and the acquired information, determining whether the subject has a sulfatase-associated or another MID 9002-associated disease or disorder or a pre-disposition to a sulfatase-associated or another MID 9002-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder, or pre-disease condition.

[00389] The invention also includes an array comprising a MID 9002 sequence of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression, one of which can be MID 9002. This allows a profile to be developed showing a battery of genes specifically expressed in one or more tissues.

[00390] In addition to such qualitative information, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue if ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression in that tissue. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

[00391] In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of a sulfatase-associated or another MID 9002-associated disease or disorder, progression of sulfatase-associated or another MID 9002-associated disease or disorder, and processes, such a cellular transformation associated with the sulfatase-associated or another MID 9002-associated disease or disorder.

[00392] The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (*e.g.*, acertaining the effect of MID 9002 expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

[00393] The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (*e.g.*, including MID 9002) that could serve as a molecular target for diagnosis or therapeutic intervention.

[00394] As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. Typical sequence lengths of a target sequence are from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

[00395] Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI).

[00396] Thus, the invention features a method of making a computer readable record of a sequence of a MID 9002 sequence which includes recording the sequence on a computer readable matrix. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

[00397] In another aspect, the invention features a method of analyzing a sequence. The method includes: providing a MID 9002 sequence, or record, in computer readable form; comparing a second sequence to the MID 9002 sequence; thereby analyzing a sequence. Comparison can include comparing to sequences for sequence identity or determining if one sequence is included within the other, *e.g.*, determining if the MID 9002 sequence includes a sequence being compared. In a preferred embodiment the MID 9002 or second sequence is stored on a first computer, *e.g.*, at a first site and the comparison is performed, read, or recorded on a second computer, *e.g.*, at a second site. *E.g.*, the MID 9002 or second sequence can be stored in a public or proprietary database in one computer, and the results of the comparison performed, read, or recorded on a second computer. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

[00398] This invention is further illustrated by the following exemplification, which should not be construed as limiting.

EXEMPLIFICATION

Gene Expression Analysis

[00399] Total RNA was prepared from various human tissues by a single step extraction method using RNA STAT-60 according to the manufacturer's instructions (TelTest, Inc). Each RNA preparation was treated with DNase I (Ambion) at 37°C for 1 hour. DNAse I treatment was determined to be complete if the sample required at least 38 PCR amplification cycles to reach a threshold level of fluorescence using β-2 microglobulin as an internal amplicon reference. The integrity of the RNA samples following DNase I treatment was confirmed by agarose gel electrophoresis and ethidium bromide staining. After phenol extraction cDNA was prepared from the sample using the SUPERSCRIPT™ Choice System following the manufacturer's instructions (GibcoBRL). A negative control of RNA without reverse transcriptase was mock reverse transcribed for each RNA sample.

[00400] Human MID 9002 expression was measured by TaqMan® quantitative PCR (Perkin Elmer Applied Biosystems) in cDNA prepared from a variety of normal and diseased (e.g., cancerous) human tissues or cell lines.

[00401] Probes were designed by PrimerExpress software (PE Biosystems) based on the sequence of the human MID 9002 gene. Each human MID 9002 gene probe was labeled using FAM (6-carboxyfluorescein), and the β2-microglobulin reference probe was labeled with a different fluorescent dye, VIC. The differential labeling of the target gene and internal reference gene thus enabled measurement in same well. Forward and reverse primers and the probes for both β2-microglobulin and target gene were added to the TaqMan® Universal PCR Master Mix (PE Applied Biosystems). Although the final concentration of primer and probe could vary, each was internally consistent within a given experiment. A typical experiment contained 200nM of forward and reverse primers plus 100nM probe for β-2 microglobulin and 600 nM forward and reverse primers plus 200 nM probe for the target gene. TaqMan matrix experiments were carried out on an ABI PRISM 7700 Sequence Detection System (PE Applied Biosystems). The thermal cycler conditions were as follows: hold for 2 min at 50°C and 10 min at 95°C, followed by two-step PCR for 40 cycles of 95°C for 15 sec followed by 60°C for 1 min.

[00402] The following method was used to quantitatively calculate human MID 9002 gene expression in the various tissues relative to β-2 microglobulin expression in the same tissue. The threshold cycle (Ct) value is defined as the cycle at which a statistically significant increase in fluorescence is detected. A lower Ct value is indicative of a higher mRNA concentration. The Ct value of the human MID 9002 gene is normalized by subtracting the Ct value of the β-2 microglobulin gene to obtain a $_{\Delta}$Ct value using the following formula: $_{\Delta}Ct=Ct_{\text{human 59914 and 59921}} - Ct_{\beta\text{-2 microglobulin}}$. Expression is then calibrated against a cDNA sample showing a comparatively low level of expression of the human MID 9002 gene. The $_{\Delta}$Ct value for the calibrator sample is then subtracted from $_{\Delta}$Ct for each tissue sample according to the following formula: $_{\Delta\Delta}Ct={}_{\Delta}Ct_{\text{-sample}} - {}_{\Delta}Ct_{\text{-calibrator}}$. Relative expression is then calculated using the arithmetic formula given by $2^{-\Delta\Delta Ct}$. Expression of the target human MID 9002 gene in each of the tissues tested is then graphically represented as discussed in more detail below.

[00403] The results indicate significant MID 9002 expression in kidney, pancreas, dorsal root ganglion, colon, and liver as well as in colon cancer (see Table below).

MID 9002 Phase 1.3.3 Expression	
Tissue	**Expression**

Artery, normal	0
Vein, normal	0
Aortic SMC Early	0
Coronary SMC	0
Static HUVEC	0.7026
Shear HUVEC	0.3489
Heart, normal	0
Heart, CHF	0
Kidney	21.8885
Adipose, normal	0
Pancreas	45.1228
Primary osteoblasts	0.056
Osteoclasts (diff.)	0
Skin, normal	0
Spinal cord, normal	0
Brain, cortex, normal	3.3806
Brain, hypothalamus	0.7877
Nerve	0.7075
Dorsal Root Ganglion	9.1628
Glial Cells (astrocytes)	0.7478
Glioblastoma	0.1295
Breast, normal	0
Breast, tumor	0.024
Ovary, normal	1.6198
Ovary, tumor	0
Prostate, normal	0
Prostate, tumor	0
Epithelial Cells (prostate)	0
Colon, normal	2.2749
Colon, tumor	26.278
Lung, normal	0
Lung, tumor	0.0628
Lung, COPD	0.1685
Colon, IBD	0.1738
Liver, normal	16.5159
Liver, fibrosis	17.4576
Dermal cells, fibroblasts	0
Spleen, normal	0
Tonsil, normal	0
Lymph Node	0
Small Intestine	0
Skin, decubitus	0
Synovium	0
BM-MNC (Bone marrow)	0
Activated PBMC	0

9002 Expression in Oncology Phase II Plate

Tissue	Expression
PIT 400 Breast N	0.08
PIT 372 Breast N	0.06
CHT 558 Breast N	0.00
CLN 168 Breast T IDC	0.14
MDA 304 Breast T MD-DC	0.11
NDR 58 Breast T IDC	0.38
NDR 05 Breast T DC	1.71
CHT 562 Breast T DC	0.75
NDR 138Breast T ILC (LG)	0.75
CHT 184 Lymph Node	0.00
PIT58 Lung (breast)	0.00
PIT 208 OvaryN	3.09
CHT 620 OvaryN	2.46
CLN03 Ovary T	0.14
CLN17 Ovary T	0.08
MDA 25 Ovary T	0.00
MDA 216 Ovary T	0.08
CLN012 Ovary T	2.03
MDA 185 Lung N	0.06
CLN 980 Lung N	0.23
MDA 183 Lung N	0.13
MP 1215 Lung T-SmC	0.09
MDA 259 Lung T PDNSCC	0.33
CHT 832 Lung T-PDNSCC	2.99
MDA262 LungT-SCC	0.87
CHT LungT-ACA	0.18
CHT331 LungT-ACA	1.04
CHT 405 Colon N	4.98
CHT 523 Colon N	0.34
CHT 371 Colon N	4.04
CHT 382 Colon T MD	18.45
CHT 528 Colon T MD	40.95
CLN 609 Colon T	21.64
CHT 372 Colon T MD-PD	12.56
CHT 340 Colon-Liver Met	24.35
NDR 100 Colon-Liver Met	27.97
PIT 260 Liver N (female)	7.95
CHT1653 Cervix Squamous	0.37
CHT569 Cervix Squamous	0.00
A24 HMVEC-Arr	2.52
C48 HMVEC-Prol	1.42

9002 Expression in the Colonic ACA Plate	
Tissue	**Expression**
PIT 337 Colon N	0.29
CHT 410 Colon N	0.65

CHT 425 Colon N	2.56
CHT 571 Colon N	3.02
Pit 281 Colon N	3.83
NDR 211 Colon N	2.91
CHT 122 Adenomas	10.53
CHT 887 Adenomas	35.65
CHT 414 Colonic ACA-B	4.81
CHT 841 Colonic ACA-B	0.00
CHT 890 Colonic ACA-B	7.42
CHT 910 Colonic ACA-B	4.23
CHT 807 Colonic ACA-B	7.92
CHT 382 Colonic ACA-B	10.64
CHT 377 Colonic ACA-B	1.83
CHT 520 Colonic ACA-C	9.13
CHT 596 Colonic ACA-C	6.94
CHT 907 Colonic ACA-C	3.53
CHT 372 Colonic ACA-C	8.46
NDR 210 Colonic ACA-C	2.55
CHT 1365 Colonic ACA-C	12.05
CLN 740 Liver N	54.79
CLN 741 Liver N	34.92
NDR 165 Liver N	7.21
NDR 150 Liver N	16.40
PIT 236 Liver N	11.01
CHT 1878 Liver N	61.43
CHT 077 Col Liver Met	21.42
CHT 119 Col Liver Met	14.99
CHT 131 Col Liver Met	25.12
CHT 218 Col Liver Met	16.40
CHT 739 Col Liver Met	12.43
CHT 755 Col Liver Met	32.35
CHT 215 Col Abdominal Met	4.73

9002 Expression in Xenograft Panel	
Tissue	**Expression**
MCF-7 Breast T	0.00
ZR 75 Breast T	0.07
T47D Breast T	0.02
MDA 231 Breast T	2.46
MDA 435 Breast T	4.65
SKBr3 Breast	0.00
DLD1 Colon T (Stage C)	21.72
SW 480 Colon T (Stage B)	1.93
SW 620 Colon T (Stage C)	3.57
HCT 116	0.00
HT 29	7.70
Cob 205	7.73

NCIH 125	0.00
NCIH 322	0.00
NCIH 460	0.00
A 549	71.55
NIHB E	0.00
SKOV-3 ovary	0.60
OV CAR-3 ovary	0.00
293 Baby Kidney	7.95
293T Baby Kidney	0.00

9002 Expression in *in vitro* Synchronized Cell Cycle Panel	
Tissue	**Expression**
HCT 116 Aphidl t=0	0
HCT 116 Aphidl t=3	0
HCT 116 Aphidl t=6	0
HCT 116 Aphidl t=9	0
HCT 116 Aphidl t=12	0
HCT 116 Aphidl t=15	0
HCT 116 Aphidl t=18	0
HCT 116 Aphidl t=21	0
HCT 116 Aphidl t=24	0
HCT 116 Noc t=0	0
HCT 116 Noc t=3	0
HCT 116 Noc t=6	0
HCT 116 Noc t=9	0
HCT 116 Noc t=15	0
HCT 116 Noc t=18	0
HCT 116 Noc t=21	0
HCT 116 Noc t=24	0
DLD noc t=3	20.69
DLD noc t=9	32.80
DLD noc t=12	42.54
DLD noc t=15	39.42
DLD noc t=18	41.81
DLD noc t=21	44.81
A549 Mimo t=0	89.93
A549 Mimo t=3	79.66
A549 Mimo t=6	76.68
A549 Mimo t=9	64.93
A549 Mimo t=15	70.81
A549 Mimo t=18	67.45
A549 Mimo t=21	66.52
A549 Mimo t=24	56.52
MCF 10A Mimo t=0	0
MCF 10A Mimo t=3	0
MCF 10A Mimo t=6	0

MCF 10A Mimo t=9	0
MCF 10A Mimo t=12	0
MCF 10A Mimo t=18	0
MCF 10A Mimo t=21	0
MCF 10A Mimo t=24	0

[00404] The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

Equivalents

[00405] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein.

Claims

What is claimed:

1. A method of identifying a nucleic acid molecule associated with a pathological disorder selected from the group consisting of:

a) contacting a sample comprising nucleic acid molecules with a hybridization probe comprising at least 25 contiguous nucleotides of SEQ ID NO:1 OR 3; and detecting the presence of a nucleic acid molecule in said sample that hybridizes to said probe, and

b) contacting a sample comprising nucleic acid molecules with a first and a second amplification primer, said first primer comprising at least 25 contiguous nucleotides of SEQ ID NO:1 OR 3, and said second primer comprising at least 25 contiguous nucleotides from the complement of SEQ ID NO:1 OR 3, ; incubating said sample under conditions that allow nucleic acid amplification; and detecting the presence of a nucleic acid molecule in said sample that is amplified,

thereby identifying a nucleic acid molecule associated with a pathological disorder.

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

a) contacting a sample comprising polypeptides with a 9002 binding substance; and

b) detecting the presence of a polypeptide in said sample that binds to said 9002 binding substance, thereby identifying a polypeptide associated with a pathological disorder.

3. A method of identifying a subject having a pathological disorder, or at risk for developing a pathological disorder selected from the group consisting of:

a) contacting a sample obtained from said subject comprising nucleic acid molecules with a hybridization probe comprising at least 25 contiguous nucleotides of SEQ ID NO:1 OR 3, ; and detecting the presence of a nucleic acid molecule in said sample that hybridizes to said probe; and

b) contacting a sample obtained from said subject comprising nucleic acid molecules with a first and a second amplification primer, said first primer comprising at least 25 contiguous nucleotides of SEQ ID NO:1 OR 3, and said second primer comprising at least 25 contiguous nucleotides from the complement of SEQ ID NO:1 OR 3, ; incubating said sample under conditions that allow nucleic acid amplification; and detecting the presence of a nucleic acid molecule in said sample that is amplified;

thereby identifying a subject having a pathological disorder, or at risk for developing a pathological disorder.

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

a) contacting a sample obtained from said subject comprising polypeptides with a 9002 binding substance; and

b) detecting the presence of a polypeptide in said sample that binds to said 9002 binding substance, thereby identifying a subject having a pathological disorder, or at risk for developing a pathological disorder.

5. A method for treating a subject having a pathological disorder or a pathological disorder characterized by aberrant 9002 polypeptide activity or aberrant 9002 nucleic acid expression comprising administering to the subject a 9002 modulator, thereby treating said subject having a pathological disorder.

6. The method of claim 5 wherein the modulator is selected from the group consisting of small organic molecules, peptides, polynucleotides and antibodies.

7. The method of any of claims 1-5 wherein the pathological disorder is selected from the group consisting of cancer, cell proliferation disordersand disorders associated with aberrant angiogenesis.

8. A method of detecting an agent that binds a 9002 polypeptide comprising:

a) combining an agent selected from the group consisting of small organic molecules, peptides, polynucleotides and antibodies; and

b) detecting or measuring the formation of a complex between the agent and the 9002 polypeptide under conditions suitable for binding of an agent to the 9002 polypeptide, to thereby identify an agent that binds to the 9002 polypeptide.

9. The method of claim 7 wherein the agent is a modulator of 9002 polypeptide activity.

cDNA Sequence of human MID 9002 (SEQ ID NO:1)

CCTTCCTCTTCTTGATCGGGGATTCAGGAAGGAGCCCAGGAGCAGAGGAAGTAGAGAGAG
AGACAACATGTTACATCTGCACCATTCTTGTTTGTGTTTCAGGAGCTGGCTGCCAGCGAT
GCTCGCTGTACTGCTAAGTTTGGCACCATCAGCTTCCAGCGACATTTCCGCCTCCCGACC
GAACATCCTTCTTCTGATGGCGGACGACCTTGGCATTGGGGACATTGGCTGCTATGGCAA
CAACACCATGAGGACTCCGAATATTGACCGCCTTGCAGAGGACGGCGTGAAGCTGACCCA
ACACATCTCTGCCGCATCTTTGTGCACCCCAAGCAGAGCCGCCTTCCTCACGGGCAGATA
CCCTGTGCGATCAGGGATGGTTTCCAGCATTGGTTACCGTGTTCTTCAGTGGACCGGAGC
ATCTGGAGGTCTTCCAACAAATGAGACAACTTTTGCAAAAATACTGAAAGAGAAAGGCTA
TGCCACTGGACTCATTGGAAAATGGCATCTGGGTCTCAACTGTGAGTCAGCCAGTGATCA
TTGCCACCACCCTCTCCATCATGGCTTTGAGCATTTCTACGGAATGCCTTTCTCCTTGAT
GGGTGATTGCGCCCGCTGGGAACTCTCAGAGAAGCGTGTCAACCTGGAACAAAAACTCAA
CTTCCTCTTCCAAGTCCTGGCCTTGGTTGCCCTCACACTGGTAGCAGGGAAGCTCACACA
CCTGATACCCGTCTCGTGGATGCCGGTCATCTGGTCAGCCCTTTCGGCCGTCCTCCTCCT
CGCAAGCTCCTATTTTGTGGGTGCTCTGATTGTCCATGCCGATTGCTTTCTGATGAGAAA
CCACACCATCACGGAGCAGCCCATGTGCTTCCAAAGAACGACACCCCTTATTCTGCAGGA
GGTTGCGTCCTTTCTCAAAAGGAATAAGCATGGGCCTTTCCTCCTCTTTGTTTCCTTTCT
ACACGTTCACATCCCTCTTATCACTATGGAGAACTTCCTCGGGAAGAGTCTCCACGGGCT
GTATGGGGACAACGTAGAGGAGATGGACTGGATGGTAGGACGGATCCTTGACACTTTGGA
CGTGGAGGGTTTGAGCAACAGCACCCTCATTTATTTTACGTCGGATCACGGCGGTTCCCT
AGAGAATCAACTTGGAAACACCCAGTATGGTGGCTGGAATGGAATTTATAAAGGTGGGAA
GGGCATGGGAGGATGGGAAGGTGGGATCCGCGTGCCCGGGATCTTCCGCTGGCCCGGGGT
GCTCCCGGCCGGCCGAGTGATTGGCGAGCCCACGAGTCTGATGGACGTGTTCCCCACCGT
GGTCCGGCTGGCGGGCGGCGAGGTGCCCCAGGACAGAGTGATTGACGGCCAAGACCTTCT
GCCCTTGCTCCTGGGGACAGCCCAACACTCAGACCACGAGTTCCTGATGCATTATTGTGA
GAGGTTTCTGCACGCAGCCAGGTGGCATCAACGGGACAGAGGAACAATGTGGAAAGTCCA
CTTTGTGACGCCTGTGTTCCAGCCAGAGGGAGCCGGTGCCTGCTATGGAAGAAAGGTCTG
CCCGTGCTTTGGGGAAAAAGTAGTCCACCACGATCCACCTTTGCTCTTTGACCTCTCAAG
AGACCCTTCTGAGACCCACATCCTCACACCAGCCTCAGAGCCCGTGTTCTATCAGGTGAT
GGAACGAGTCCAGCAGGCGGTGTGGGAACACCAGCGGACACTCAGCCCAGTTCCTCTGCA
GCTGGACAGGCTGGGCAACATCTGGAGACCGTGGCTGCAGCCCTGCTGTGGCCCGTTCCC
CCTCTGCTGGTGCCTTAGGGAAGATGACCCACAATAAATGTCTGCAGTGAAAAGCTGG

Figure 1a

Predicted Amino Acid Sequence of MID 9002 (SEQ ID NO:2)

```
MLHLHHSCLCFRSWLPAMLAVLLSLAPSASSDISASRPNILLLMADDLGIGDIGCYGNNT
MRTPNIDRLAEDGVKLTQHISAASLCTPSRAAFLTGRYPVRSGMVSSIGYRVLQWTGASG
GLPTNETTFAKILKEKGYATGLIGKWHLGLNCESASDHCHHPLHHGFEHFYGMPFSLMGD
CARWELSEKRVNLEQKLNFLFQVLALVALTLVAGKLTHLIPVSWMPVIWSALSAVLLLAS
SYFVGALIVHADCFLMRNHTITEQPMCFQRTTPLILQEVASFLKRNKHGPFLLFVSFLHV
HIPLITMENFLGKSLHGLYGDNVEEMDWMVGRILDTLDVEGLSNSTLIYFTSDHGGSLEN
QLGNTQYGGWNGIYKGGKGMGGWEGGIRVPGIFRWPGVLPAGRVIGEPTSLMDVFPTVVR
LAGGEVPQDRVIDGQDLLPLLLGTAQHSDHEFLMHYCERFLHAARWHQRDRGTMWKVHFV
TPVFQPEGAGACYGRKVCPCFGEKVVHHDPPLLFDLSRDPSETHILTPASEPVFYQVMER
VQQAVWEHQRTLSPVPLQLDRLGNIWRPWLQPCCGPFPLCWCLREDDPQ
```

Figure 1b

Coding Sequence of MID 9002 (SEQ ID NO:3)

```
          ATGTTACATCTGCACCATTCTTGTTTGTGTTTCAGGAGCTGGCTGCCAGCGAT
GCTCGCTGTACTGCTAAGTTTGGCACCATCAGCTTCCAGCGACATTTCCGCCTCCCGACC
GAACATCCTTCTTCTGATGGCGGACGACCTTGGCATTGGGGACATTGGCTGCTATGGCAA
CAACACCATGAGGACTCCGAATATTGACCGCCTTGCAGAGGACGGCGTGAAGCTGACCCA
ACACATCTCTGCCGCATCTTTGTGCACCCCAAGCAGAGCCGCCTTCCTCACGGGCAGATA
CCCTGTGCGATCAGGGATGGTTTCCAGCATTGGTTACCGTGTTCTTCAGTGGACCGGAGC
ATCTGGAGGTCTTCCAACAAATGAGACAACTTTTGCAAAAATACTGAAAGAGAAAGGCTA
TGCCACTGGACTCATTGGAAAATGGCATCTGGGTCTCAACTGTGAGTCAGCCAGTGATCA
TTGCCACCACCCTCTCCATCATGGCTTTGAGCATTTCTACGGAATGCCTTTCTCCTTGAT
GGGTGATTGCGCCCGCTGGGAACTCTCAGAGAAGCGTGTCAACCTGGAACAAAAACTCAA
CTTCCTCTTCCAAGTCCTGGCCTTGGTTGCCCTCACACTGGTAGCAGGGAAGCTCACACA
CCTGATACCCGTCTCGTGGATGCCGGTCATCTGGTCAGCCCTTTCGGCCGTCCTCCTCCT
CGCAAGCTCCTATTTTGTGGGTGCTCTGATTGTCCATGCCGATTGCTTTCTGATGAGAAA
CCACACCATCACGGAGCAGCCCATGTGCTTCCAAAGAACGACACCCCTTATTCTGCAGGA
GGTTGCGTCCTTTCTCAAAAGGAATAAGCATGGGCCTTTCCTCCTCTTTGTTTCCTTTCT
ACACGTTCACATCCCTCTTATCACTATGGAGAACTTCCTCGGGAAGAGTCTCCACGGGCT
GTATGGGGACAACGTAGAGGAGATGGACTGGATGGTAGGACGGATCCTTGACACTTTGGA
CGTGGAGGGTTTGAGCAACAGCACCCTCATTTATTTTACGTCGGATCACGGCGGTTCCCT
AGAGAATCAACTTGGAAACACCCAGTATGGTGGCTGGAATGGAATTTATAAAGGTGGGAA
GGGCATGGGAGGATGGGAAGGTGGGATCCGCGTGCCCGGGATCTTCCGCTGGCCCGGGGT
GCTCCCGGCCGGCCGAGTGATTGGCGAGCCCACGAGTCTGATGGACGTGTTCCCCACCGT
GGTCCGGCTGGCGGGCGGCGAGGTGCCCCAGGACAGAGTGATTGACGGCCAAGACCTTCT
GCCCTTGCTCCTGGGGACAGCCCAACACTCAGACCACGAGTTCCTGATGCATTATTGTGA
GAGGTTTCTGCACGCAGCCAGGTGGCATCAACGGGACAGAGGAACAATGTGGAAAGTCCA
CTTTGTGACGCCTGTGTTCCAGCCAGAGGGAGCCGGTGCCTGCTATGGAAGAAAGGTCTG
CCCGTGCTTTGGGGAAAAAGTAGTCCACCACGATCCACCTTTGCTCTTTGACCTCTCAAG
AGACCCTTCTGAGACCCACATCCTCACACCAGCCTCAGAGCCCGTGTTCTATCAGGTGAT
GGAACGAGTCCAGCAGGCGGTGTGGGAACACCAGCGGACACTCAGCCCAGTTCCTCTGCA
GCTGGACAGGCTGGGCAACATCTGGAGACCGTGGCTGCAGCCCTGCTGTGGCCCGTTCCC
CCTCTGCTGGTGCCTTAGGGAAGATGACCCACAATAA
```

Figure 1c

Amino Acid Sequence of Mature Form of MID 9002 (SEQ ID NO:4)

DISASRPNILLLMADDLGIGDIGCYGNNTMRTPNIDRLAEDGVKLTQHISAASLCTPSRA
AFLTGRYPVRSGMVSSIGYRVLQWTGASGGLPTNETTFAKILKEKGYATGLIGKWHLGLN
CESASDHCHHPLHHGFEHFYGMPFSLMGDCARWELSEKRVNLEQKLNFLFQVLALVALTL
VAGKLTHLIPVSWMPVIWSALSAVLLLASSYFVGALIVHADCFLMRNHTITEQPMCFQRT
TPLILQEVASFLKRNKHGPFLLFVSFLHVHIPLITMENFLGKSLHGLYGDNVEEMDWMVG
RILDTLDVEGLSNSTLIYFTSDHGGSLENQLGNTQYGGWNGIYKGGKGMGGWEGGIRVPG
IFRWPGVLPAGRVIGEPTSLMDVFPTVVRLAGGEVPQDRVIDGQDLLPLLLGTAQHSDHE
FLMHYCERFLHAARWHQRDRGTMWKVHFVTPVFQPEGAGACYGRKVCPCFGEKVVHHDPP
LLFDLSRDPSETHILTPASEPVFYQVMERVQQAVWEHQRTLSPVPLQLDRLGNIWRPWLQ
PCCGPFPLCWCLREDDPQ

Figure 1d

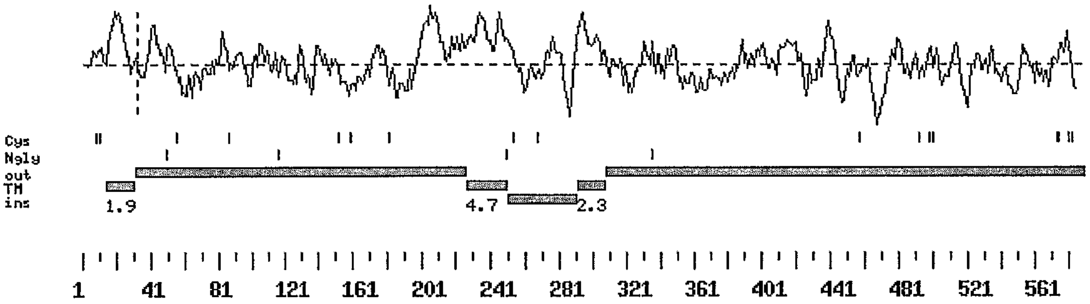

Cys
Ngly
out
TM
ins
1.9
4.7
2.3
1
41
81
121
161
201
241
281
321
361
401
441
481
521
561

Figure 2

```
Searching for complete domains in PFAM
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 /prod/ddm/seqanal/PFAM/pfam6.2/Pfam
Sequence file:            /prod/ddm/wspace/orfanal/oa-script.16928.seq
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Query:  9002

Scores for sequence family classification (score includes all domains):
Model     Description                                     Score    E-value  N
--------  -----------                                     -----    ------- ---
Sulfatase Sulfatase                                        794.2   4.9e-235   1

Parsed for domains:
Model     Domain  seq-f seq-t    hmm-f hmm-t      score  E-value
--------  ------- ----- -----    ----- -----      -----  -------
Sulfatase   1/1      38   520 ..     1   520 []   794.2 4.9e-235

Alignments of top-scoring domains:
Sulfatase: domain 1 of 1, from 38 to 520: score 794.2, E = 4.9e-235
                   *->PNvllilaDDlGigdlgcyghptirTPnldrLAeeGlrFtnhytatp
                      PN+ll++aDDlGigd+gcyg++t+rTPn+drLAe+G+++t+h +a++
        9002    38    PNILLLMADDLGIGDIGCYGNNTMRTPNIDRLAEDGVKLTQHISAAS 84

                   lCsPSRAaLlTGryphrhGmvsngrlgvlgftaksgglpldettLpelLk
                   lC+PSRAa+lTGryp+r+Gmvs+++++vl++t+ sgglp++ett++++Lk
        9002    85 LCTPSRAAFLTGRYPVRSGMVSSIGYRVLQWTGASGGLPTNETTFAKILK 134

                   eaGYaTglvGKWHlglnensdaagdgehlPlgwrGfdyfdgflygspfty
                   e GYaTgl+GKWHlgln++s  a+d++h+Pl+ +Gf++f    yg+pf++
        9002   135 EKGYATGLIGKWHLGLNCES--ASDHCHHPLH-HGFEHF----YGMPFSL 177

                   deencdngegteppeaypeqgwlp.qilgyyltdlladkalglldvasaa
                    +++c+++e++e  ++++++++++ q+l+++ ++l+a+k+++l++v+++
        9002   178 -MGDCARWELSEKRVNLEQKLNFLfQVLALVALTLVAGKLTHLIPVSWMP 226

                   gr.ll.ak.alaasrPFflyisppaphfsilfrnfkevaqpyrapqltql
                   ++ ++ ++ +l+as +F+  + ++a++f  l+rn+++++qp+++ ++t+l
        9002   227 VIwSAlSAvLLLASSYFVGALIVHADCF--LMRNHTITEQPMCFQRTTPL 274

                   fvdeaadfiernkekPfflylaflrlhvhtplfspaedleskdflgrsqr
                   +++e a+f++rnk+ Pf+l+++fl  hvh+pl +      +++flg+s +
        9002   275 ILQEVASFLKRNKHGPFLLFVSFL--HVHIPLIT------MENFLGKSLH 316

                   grYgdlveemDdlvGrvldaLedlGlldNTlvifTSDnGahlegtpe..w
                   g+Ygd+veemD++vGr+ld+L+ +Gl+++Tl++fTSD+G++le+++++++
        9002   317 GLYGDNVEEMDWMVGRILDTLDVEGLSNSTLIYFTSDHGGSLENQLGntQ 366

                   ygggngplkggKgygslyeGgiRvPllvrwPggiapagrvkekselvshv
                   ygg+ng++kggKg+g+ +eGgiRvP+++rwP g++pagrv++  e++s++
        9002   367 YGGWNGIYKGGKGMGG-WEGGIRVPGIFRWP-GVLPAGRVIG--EPTSLM 412

                   DlaPTildlAGaplPkvanGakdrplDGvsllplllggaapsrrahetlf
                   D++PT+++lAG+++P+      dr++DG +llplllg+a++s+  he+l+
        9002   413 DVFPTVVRLAGGEVPQ------DRVIDGQDLLPLLLGTAQHSD--HEFLM 454

                   hyngkgrklravrwprksgktpklkahfftpafdddtnngwecvgtvsqa
                   hy  ++r+l+a+rw++ + +++++k+hf+tp+f++ + ++  c+g
        9002   455 HY--CERFLHAARWHQRD-RGTMWKVHFVTPVFQPEGAGA--CYG----- 494

                   ddiedcrcegvetvthhdppelyDlsrDP<-*
                     + +c c+g e+v hhdpp+l+DlsrDP
        9002   495 --RKVCPCFG-EKVVHHDPPLLFDLSRDP    520
```

Figure 3

ProDom Matches

ProdomId	Start	End	Description	Score
View Prodom PD001700	37	181	p2001.1 (61) ARS(6) STS(3) IDS(2) // HYDROLASE ARYLSULFATASE PRECURSOR SIGNAL GLYCOPROTEIN LYSOSOME SULPHOHYDROLASE ARYL-SULFATE MUCOPOLYSACCHARIDOSIS	447
View Prodom PD006857	214	348	p2001.1 (17) STS(3) ARS(2) ARSA(2) // ARYLSULFATASE PRECURSOR HYDROLASE GLYCOPROTEIN SIGNAL STEROID METABOLISM PREGNANCY STERYL-SULFATASE	250
View Prodom PD208511	275	422	p2001.1 (2) // YDEN SULFATASE ARYLSULFATASE 3.1.6.- HYDROLASE PRECURSOR SIGNAL	87
View Prodom PD002587	277	358	p2001.1 (7) IDS(2) // HYDROLASE IDURONATE LYSOSOME PRECURSOR 2-SULFATASE SIGNAL GLYCOPROTEIN ZYMOGEN CG12014 CHOLINE	97
View Prodom PD105119	319	556	p2001.1 (1) // ATSG	

Figure 4a

81
View Prodom PD345058

331
366
p2001.1 (2) // ARYLSULFATASE PRECURSOR 3.1.6.- HYDROLASE SIGNAL GLYCOPROTEIN ASD LYSOSOME MUTATION

130
View Prodom PD004614

349
441
p2001.1 (15) STS(3) ARS(2) ARSA(2) // ARYLSULFATASE PRECURSOR HYDROLASE GLYCOPROTEIN STEROID SIGNAL METABOLISM STERYL-SULFATASE MICROSOME

164
View Prodom PD335606

388
441
p2001.1 (3) // ARYLSULFATASE 3.1.6.- PRECURSOR HYDROLASE SIGNAL GLYCOPROTEIN HIGHLY SIMILAR MUTATION

210
View Prodom PD105130

388
547
p2001.1 (3) // HYDROLASE SULFATASE PRECURSOR MDSA PHOSPHONATE MONOESTER SIGNAL MUCIN-DESULFATING

89
View Prodom PD013467

390
552
p2001.1 (8) ARS(2) // ARYLSULFATASE HYDROLASE SULPHOHYDROLASE ARYL-SULFATE PRECURSOR SIGNAL ATSA

95
View Prodom PD013468

442
586
p2001.1 (8) STS(3) // PRECURSOR STEROID GLYCOPROTEIN ARYLSULFATASE HYDROLASE MICROSOME TRANSMEMBRANE STERYL-

Figure 4b

```
537
 View Prodom PD016856


                              450
                                   589
                                        p2001.1 (6) ARS(2) // ARYLSULFATASE
                                        SULPHOHYDROLASE PRECURSOR ARYL-SULFATE
                                        SIGNAL GLYCOPROTEIN
                                        N-ACETYLGALACTOSAMINE-6-SULFATE SIMILAR


165
         ProdomId
                              Start
                                   End
                                                          Description

Score


View Prodom PD013468

>PD013468 p2001.1 (8) STS(3) // PRECURSOR STEROID GLYCOPROTEIN ARYLSULFATASE
           HYDROLASE MICROSOME TRANSMEMBRANE STERYL-
         Length = 150

 Score = 537 (194.1 bits), Expect = 6.8e-52, P = 6.8e-52
 Identities = 95/150 (63%), Positives = 107/150 (71%)

Query:   442 LGTAQHSDHEFLMHYCERFLHAARWHQRD-RGTMWKVHFVTPVFQPEGAGACYGRKVCPC 500
             LG AQHS+HEFL HYC  +LHAARWHQ D  G++WK H+ TP FQP+GAG CYG  VC C
Sbjct:     1 LGNAQHSEHEFLFHYCGDYLHAARWHQHDDSGSVWKWHYFTPNFQPQGAGGCYGTNVCGC 60

Query:   501 FGEK-VVHHDPPLLFDLSRDPSETHILTPASEPVFYQVMERVQQAVWEHQRTLSPVPLQL 559
             FG   V HHDPPLLFDLSRDP E H LTPASEP++ +V+ R+  AV EHQ TL P P QL
Sbjct:    61 FGHNHVTHHDPPLLFDLSRDPGERHPLTPASEPLYQEVLSRITSAVQEHQETLVPAPPQL 120

Query:   560 DRLGNIWRPWLQPCCGPFP---LCWCLRED 586
                   W+PWLQPCCG FP    CWC  ED
Sbjct:   121 SMCNWAWKPWLQPCCGHFPHPLFCWC-HED 149



View Prodom PD001700

>PD001700 p2001.1 (61) ARS(6) STS(3) IDS(2) // HYDROLASE ARYLSULFATASE
           PRECURSOR SIGNAL GLYCOPROTEIN LYSOSOME SULPHOHYDROLASE ARYL-SULFATE
           MUCOPOLYSACCHARIDOSIS
         Length = 193

 Score = 447 (162.4 bits), Expect = 2.3e-42, P = 2.3e-42
 Identities = 83/152 (54%), Positives = 112/152 (73%)

Query:    37 RPNILLLMADDLGIGDIGCYGNNTMR--TPNIDRLAEDGVKLTQHISAASLCTPSRAAFL 94
             RPNI+L++ADD+G GDIGCYGN TM+  TPN+DRLAE+G++ T + +++ LCTPSRA+ L
Sbjct:    36 RPNIILILADDMGFGDIGCYGNPTMKKLTPNLDRLAEEGIRFTNYYTSSPLCTPSRASLL 95

Query:    95 TGRYPVRSGMVSSIGYRVLQWTGAS--GGLPTNETTFAKILKEKGYATGLIG--KWHLGL 150
             TGRYP R+GM S+  Y    +  +S  GGLP NETT  ++LKE GY TG++G   WHLG+
Sbjct:    96 TGRYPHRTGMYSNTLYSEFPYGSSSYWGGLPENETTLPELLKEAGYNTGMVGGKMWHLGV 155

Query:   151 NCESASDHCHHPLHHGFEHF-YGMPFSLMGDC 181
             N  S SD  +HP +HGF++F Y +PF+   DC
Sbjct:   156 NPSSTSDFAYHPWNHGFDYFGYNLPFTNTWDC 187
```

Figure 4c

```
View Prodom PD006857

>PD006857 p2001.1 (17) STS(3) ARS(2) ARSA(2) // ARYLSULFATASE PRECURSOR
          HYDROLASE GLYCOPROTEIN SIGNAL STEROID METABOLISM PREGNANCY
          STERYL-SULFATASE
        Length = 163

 Score = 250 (93.1 bits), Expect = 9.6e-21, P = 9.6e-21
 Identities = 53/137 (38%), Positives = 72/137 (52%)

Query:    214 GKLTHLIPVSWMPVIWXXXXXXXXXXXXYFVGALIVHADCFLMRNHTITEQPMCFQRTTP 273
              GK   LI V WM +              Y        A+CFL  N TI +QPM ++  T
Sbjct:     27 GKCCGLIHVPWMVIFSMVLAAALIGAAWYSFCYYFRPANCFLYENLTIVQQPMQYKNLTQ 86

Query:    274 LILQEVASFLKR--NKHGPFLLFVSFLHVHIPLITMENFLGKSLHGLYGDNVEEMDWMVG 331
              L   E   F++R  N+  PF L+ +F HVH PL + ++F G S  G YGDNV EMDW V
Sbjct:     87 LYTDEALDFIQRQQNRDRPFFLYYAFAHVHTPLFSSQDFAGTSRRGRYGDNVMEMDWAVQ 146

Query:    332 RILDTLDVEGLSNSTLI 348
              +I+D L+   +S +T+I
Sbjct:    147 KIVDKLEENNISENTII 163


View Prodom PD335606

>PD335606 p2001.1 (3) // ARYLSULFATASE 3.1.6.- PRECURSOR HYDROLASE SIGNAL
          GLYCOPROTEIN HIGHLY SIMILAR MUTATION
        Length = 55

 Score = 210 (79.0 bits), Expect = 2.2e-16, P = 2.2e-16
 Identities = 43/54 (79%), Positives = 44/54 (81%)

Query:    388 RVPGIFRWPGVLPAGRVIGEPTSLMDVFPTVVRLAGGEVPQDRVIDGQDLLPLL 441
              R PG F WPG LPAGRVI EP SLMDVFPTVV LAGGE PQDRVIDG +LLP L
Sbjct:      2 REPGPFWWPGHLPAGRVIHEPGSLMDVFPTVVALAGGEPPQDRVIDGLNLLPTL 55


View Prodom PD016856

>PD016856 p2001.1 (6) ARS(2) // ARYLSULFATASE SULPHOHYDROLASE PRECURSOR
          ARYL-SULFATE SIGNAL GLYCOPROTEIN N-ACETYLGALACTOSAMINE-6-SULFATE
          SIMILAR
        Length = 153

 Score = 165 (63.1 bits), Expect = 1.6e-11, P = 1.6e-11
 Identities = 48/148 (32%), Positives = 69/148 (46%)

Query:    450 HEFLMHYCERFLHAARW-----HQRDRGTMWKVHFVTPV-FQPEGAGACYGRKVC-PCFG 502
              H  + +YC+  L A R      H + +   W+  +   + +  +G   C    VC  C G
Sbjct:      8 HRPIFYYCKDNLMAVRVGKYKAHFKTQTVSWQNEYGPNLDYFCQGGFPCPDYFVCNDCEG 67

Query:    503 EKVVHHDPPLLFDLSRDPSETHILTPASEPVFYQ-VMERVQQAVWEHQRTLSPVPLQLDR 561
              + V  H+PPL+FDL RDP E + L     P  YQ V+ +V+QAV EHQ +L      LD
Sbjct:     68 DCVTEHNPPLIFDLDRDPGERYPLQ--FHPCEYQDVLFQVKQAVEEHQSSLVKGKPVLDS 125

Query:    562 LGNIWRPWLQPCCGPFPLCWCLREDDPQ 589
                      + PCC P   C C    +P+
Sbjct:    126 YNQAVMHSIVPCCNPASNCVCNYVHEPE 153
```

Figure 4d

```
View Prodom PD004614

>PD004614 p2001.1 (15) STS(3) ARS(2) ARSA(2) // ARYLSULFATASE PRECURSOR
          HYDROLASE GLYCOPROTEIN STEROID SIGNAL METABOLISM STERYL-SULFATASE
          MICROSOME
        Length = 94

 Score = 164 (62.8 bits), Expect = 2.0e-11, P = 2.0e-11
 Identities = 38/94 (40%), Positives = 47/94 (50%)

Query:   349 YFTSDHGGSLENQLGNTQXXXXXXXXXXXXXXXXX-XXXXRVPGIFRWPGVLPAGRVIGE 407
             YFTSD+G   E      +                      RVP +  WPG +  G+V  E
Sbjct:     1 YFTSDNGPHAEKCPHRGEIHGGSNGVFRGGKGNTTWEGGVRVPALVYWPGHIQPGQVSHE 60

Query:   408 PTSLMDVFPTVVRLAGGEVPQDRVIDGQDLLPLL 441
               SLMD+FPT   LAG  +P DRVIDG+DL+PLL
Sbjct:    61 LVSLMDIFPTAADLAGASLPTDRVIDGKDLMPLL 94
```

```
View Prodom PD345058

>PD345058 p2001.1 (2) // ARYLSULFATASE PRECURSOR 3.1.6.- HYDROLASE SIGNAL
          GLYCOPROTEIN ASD LYSOSOME MUTATION
        Length = 45

 Score = 130 (50.8 bits), Expect = 9.4e-08, P = 9.4e-08
 Identities = 25/36 (69%), Positives = 29/36 (80%)

Query:   331 GRILDTLDVEGLSNSTLIYFTSDHGGSLENQLGNTQ 366
             GRIL+T+D  GL NST  YFTSDHGG LEN+ G+TQ
Sbjct:     1 GRILNTIDDNGLKNSTFTYFTSDHGGHLENRDGHTQ 36
```

```
View Prodom PD002587

>PD002587 p2001.1 (7) IDS(2) // HYDROLASE IDURONATE LYSOSOME PRECURSOR
          2-SULFATASE SIGNAL GLYCOPROTEIN ZYMOGEN CG12014 CHOLINE
        Length = 189

 Score = 97 (39.2 bits), Expect = 0.00076, Sum P(2) = 0.00076
 Identities = 23/64 (35%), Positives = 36/64 (56%)

Query:   297 FLHVHI--PLITMENFLGKSLHGLYGDNVEEMDWMVGRILDTLDVEGLSNSTLIYFTSDH 354
             FLH++I  P   +     + +   Y  ++  +D  +GR+L TLD  G+ + T+I FTSDH
Sbjct:   126 FLHLNIDYPYGPIPEEQVRQIRRSYFGSISYIDDQIGRLLATLDETGMLDDTIIVFTSDH 185

Query:   355 GGSL 358
             G  L
Sbjct:   186 GWML 189

 Score = 50 (22.7 bits), Expect = 0.00076, Sum P(2) = 0.00076
 Identities = 13/30 (43%), Positives = 16/30 (53%)

Query:   277 QEVASFLK-RNKHG--PFLLFVSFLHVHIP 303
             +E   +LK R KH   PF L V + H H P
Sbjct:    50 EEALRYLKGRMKHSGKPFFLTVGYHHPHDP 79
```

Figure 4e

```
View Prodom PD105119


>PD105119 p2001.1 (1) // ATSG
        Length = 336

 Score = 81 (33.6 bits), Expect = 0.033, Sum P(3) = 0.033
 Identities = 18/37 (48%), Positives = 23/37 (62%)

Query:   319 YGDNVEEMDWMVGRILDTLDVEGLSNSTLIYFTSDHG 355
             YG ++   D  VGR+LDTL   GL  ST + F +DHG
Sbjct:    71 YG-SIATADEAVGRLLDTLADTGLDASTWVVFVTDHG 106

 Score = 54 (24.1 bits), Expect = 0.033, Sum P(3) = 0.033
 Identities = 17/42 (40%), Positives = 23/42 (54%)

Query:   401 AGRVIGEPTSLMDVFPTVVRLAGGEVPQDRVIDGQDLLPLLL 442
             A RV  E  S +D+ PT++ L   EVP D  ++G    P LL
Sbjct:   136 APRVYDELFSGVDLVPTLLDLLRLEVPAD--VEGVSHAPALL 175

 Score = 44 (20.5 bits), Expect = 0.033, Sum P(3) = 0.033
 Identities = 15/47 (31%), Positives = 20/47 (42%)

Query:   513 LFDLSRDPSETHILTPASEPVFYQVMERVQQAVWEH---QRTLSPVP 556
             L+DL  DP+ET+ L    +            AV  H   QRT   +P
Sbjct:   249 LYDLRADPTETNNLLAGDDSTQGVAAIAADLAVRLHDWRQRTADVIP 295



View Prodom PD013467

>PD013467 p2001.1 (8) ARS(2) // ARYLSULFATASE HYDROLASE SULPHOHYDROLASE
          ARYL-SULFATE PRECURSOR SIGNAL ATSA
        Length = 237

 Score = 95 (38.5 bits), Expect = 0.059, P = 0.058
 Identities = 48/173 (27%), Positives = 68/173 (39%)

Query:   390 PGIFRWP-GVLPAGRVIGEPTSLMDVFPTVVRLAGGEVPQDRV-------IDGQDLLPLL 441
             P +  WP G+   G++      + DV PT+   AG   P   V       +DG   +  L
Sbjct:    68 PFMISWPNGITRHGKIDASFMHVYDVAPTLYEFAGIPEPGKSVKGKPQKPMDGVSFVRYL 127

Query:   442 LGTAQHSDHEFLMHYCERFLHAARWHQRDRGTMWKVHFVTPVFQPEGAGACYGRKVCPCF 501
              G AQ  D      Y E   H   WH    G  WK   + P   P   G    R+  P
Sbjct:   128 DGDAQAPDRHNTQ-YGEELGHRGIWHD---G-WWKCTRLPPT--PWTLGPATTRRFAP-- 178

Query:   502 GEKVVHHDPPLLFDLSRDPSETHILTPASEPVFYQVMERVQQAVWEH--QRTL 552
             G      D   L++L  DPS+ H L  A  P     ++ + + +WE   +RT+
Sbjct:   179 GTYDPDDDVWELYNLHDDPSQAHDLA-AEHP---DKLKELSE-LWERYAKRTM 226
```

Figure 4f

```
View Prodom PD105130

>PD105130 p2001.1 (3) // HYDROLASE SULFATASE PRECURSOR MDSA PHOSPHONATE
          MONOESTER SIGNAL MUCIN-DESULFATING
        Length = 165

 Score = 89 (36.4 bits), Expect = 0.11, P = 0.10
 Identities = 44/168 (26%), Positives = 73/168 (43%)

Query:    388 RVPGIFRWP-GVLPAGR-VIGEP-TSLMDVFPTVVRLAGGEVPQDRVIDGQDLLPLLL-G 443
              R P + R+P  + PA R  + E     +D  PT++ +AG E P+   +DG+  LP L  G
Sbjct:     18 RTPLVIRYPKAIKPATRGTVDEHLVQNIDFAPTILDVAGVE-PKPETMDGRSFLPFLFDG 76

Query:    444 TAQHSDHEFLMHYCER--FLHAARWHQRDRGTMWK-VHFVTPVFQPEGAGACYGRKVCPC 500
                Q        HY     F H  R H   R    K +HF             YG    P
Sbjct:     77 KPQDWRQSIYYHYDYDYPFEHHVRRHDGVRTDRCKLIHF-------------YGE---PW 120

Query:    501 -FGEKVVHHDPPLLFDLSRDPSETHILTPASEPVFYQVMERVQQAVWE 547
               +    V ++   L+D+  DP+E + L    +P + ++ +  ++A+++
Sbjct:    121 KYDHDTVDYEE--LYDMQNDPNELNNLY--GDPGYEKITKEYKKALFD 164


View Prodom PD208511

>PD208511 p2001.1 (2) // YDEN SULFATASE ARYLSULFATASE 3.1.6.- HYDROLASE
          PRECURSOR SIGNAL
        Length = 155

 Score = 87 (35.7 bits), Expect = 0.15, P = 0.14
 Identities = 42/157 (26%), Positives = 58/157 (36%)

Query:    275 ILQEVASFLKRNKHG--PFLLFVSFLHVHIPLITME------NFL-GKSLHGLYGDNVEE 325
              I  E   F+ R K    PF L+++    H+P    E      NF  G      Y   +
Sbjct:      5 ITDETFDFIDRAKRDNQPFFLWLNPNRPHLPNHNPEPYQYQRNFNNGWQTADNYYAQLYS 64

Query:    326 MDWMVGRILDTLDVEGLSNSTLIYFTSDHGGSLENQLGNTQXXXXXXXXXXXXXXXXXX 385
              +D  VGRI++ L   G  ++T+I FT+D+G        N
Sbjct:     65 VDDGVGRIMEQLKDNGFYDNTIIAFTTDNGAE------NFDWPDPGNGPFKGYKGQTYPG 118

Query:    386 XXRVPGIFRWPGVLPAGRVIGEPTSLMDVFPTVVRLA 422
                  P    WPG LP G       S MD +PT +  A
Sbjct:    119 GFHTPCFMWWPGKLPPGNYDNGIMSGMDWYPTFLDAA 155
```

Figure 4g

```
>sp|P51690|ARSE_HUMAN ARYLSULFATASE E PRECURSOR (EC 3.1.6.-) (ASE).
            >sp|P51690|ARSE_HUMAN ARYLSULFATASE E PRECURSOR (EC 3.1.6.-)
(ASE).
            >gi|791004|emb|CAA58556| (X83573) ARSE [Homo sapiens]
            >pir|I37187|I37187 arylsulfatase E (EC 3.1.6.-) - human
            Length = 589

  Plus Strand HSPs:

 Score = 3208 (1472.8 bits), Expect = 0.0, P = 0.0
 Identities = 589/589 (100%), Positives = 589/589 (100%), Frame = +2

Query:    68 MLHLHHSCLCFRSWLPAMLAVLLSLAPSASSDISASRPNILLLMADDLGIGDIGCYGNNT 247
             MLHLHHSCLCFRSWLPAMLAVLLSLAPSASSDISASRPNILLLMADDLGIGDIGCYGNNT
Sbjct:     1 MLHLHHSCLCFRSWLPAMLAVLLSLAPSASSDISASRPNILLLMADDLGIGDIGCYGNNT 60

Query:   248 MRTPNIDRLAEDGVKLTQHISAASLCTPSRAAFLTGRYPVRSGMVSSIGYRVLQWTGASG 427
             MRTPNIDRLAEDGVKLTQHISAASLCTPSRAAFLTGRYPVRSGMVSSIGYRVLQWTGASG
Sbjct:    61 MRTPNIDRLAEDGVKLTQHISAASLCTPSRAAFLTGRYPVRSGMVSSIGYRVLQWTGASG 120

Query:   428 GLPTNETTFAKILKEKGYATGLIGKWHLGLNCESASDHCHHPLHHGFEHFYGMPFSLMGD 607
             GLPTNETTFAKILKEKGYATGLIGKWHLGLNCESASDHCHHPLHHGFEHFYGMPFSLMGD
Sbjct:   121 GLPTNETTFAKILKEKGYATGLIGKWHLGLNCESASDHCHHPLHHGFEHFYGMPFSLMGD 180

Query:   608 CARWELSEKRVNLEQKLNFLFQVLALVALTLVAGKLTHLIPVSWMPVIWSALSAVLLLAS 787
             CARWELSEKRVNLEQKLNFLFQVLALVALTLVAGKLTHLIPVSWMPVIWSALSAVLLLAS
Sbjct:   181 CARWELSEKRVNLEQKLNFLFQVLALVALTLVAGKLTHLIPVSWMPVIWSALSAVLLLAS 240

Query:   788 SYFVGALIVHADCFLMRNHTITEQPMCFQRTTPLILQEVASFLKRNKHGPFLLFVSFLHV 967
             SYFVGALIVHADCFLMRNHTITEQPMCFQRTTPLILQEVASFLKRNKHGPFLLFVSFLHV
Sbjct:   241 SYFVGALIVHADCFLMRNHTITEQPMCFQRTTPLILQEVASFLKRNKHGPFLLFVSFLHV 300

Query:   968 HIPLITMENFLGKSLHGLYGDNVEEMDWMVGRILDTLDVEGLSNSTLIYFTSDHGGSLEN 1147
             HIPLITMENFLGKSLHGLYGDNVEEMDWMVGRILDTLDVEGLSNSTLIYFTSDHGGSLEN
Sbjct:   301 HIPLITMENFLGKSLHGLYGDNVEEMDWMVGRILDTLDVEGLSNSTLIYFTSDHGGSLEN 360

Query:  1148 QLGNTQYGGWNGIYKGGKGMGGWEGGIRVPGIFRWPGVLPAGRVIGEPTSLMDVFPTVVR 1327
             QLGNTQYGGWNGIYKGGKGMGGWEGGIRVPGIFRWPGVLPAGRVIGEPTSLMDVFPTVVR
Sbjct:   361 QLGNTQYGGWNGIYKGGKGMGGWEGGIRVPGIFRWPGVLPAGRVIGEPTSLMDVFPTVVR 420

Query:  1328 LAGGEVPQDRVIDGQDLLPLLLGTAQHSDHEFLMHYCERFLHAARWHQRDRGTMWKVHFV 1507
             LAGGEVPQDRVIDGQDLLPLLLGTAQHSDHEFLMHYCERFLHAARWHQRDRGTMWKVHFV
Sbjct:   421 LAGGEVPQDRVIDGQDLLPLLLGTAQHSDHEFLMHYCERFLHAARWHQRDRGTMWKVHFV 480

Query:  1508 TPVFQPEGAGACYGRKVCPCFGEKVVHHDPPLLFDLSRDPSETHILTPASEPVFYQVMER 1687
             TPVFQPEGAGACYGRKVCPCFGEKVVHHDPPLLFDLSRDPSETHILTPASEPVFYQVMER
Sbjct:   481 TPVFQPEGAGACYGRKVCPCFGEKVVHHDPPLLFDLSRDPSETHILTPASEPVFYQVMER 540

Query:  1688 VQQAVWEHQRTLSPVPLQLDRLGNIWRPWLQPCCGPFPLCWCLREDDPQ 1834
             VQQAVWEHQRTLSPVPLQLDRLGNIWRPWLQPCCGPFPLCWCLREDDPQ
Sbjct:   541 VQQAVWEHQRTLSPVPLQLDRLGNIWRPWLQPCCGPFPLCWCLREDDPQ 589
```

Figure 5