



US007332591B2

(12) **United States Patent**
Sheffield et al.

(10) **Patent No.:** **US 7,332,591 B2**
(45) **Date of Patent:** **Feb. 19, 2008**

(54) **BARDET-BIEDL SUSCEPTIBILITY GENE AND USES THEREOF**

(75) Inventors: **Val C. Sheffield**, Iowa City, IA (US);
Edwin Stone, Iowa City, IA (US);
Thomas Casavant, Iowa City, IA (US);
Terry Braun, Iowa City, IA (US);
Darryl Nishimura, Coralville, IA (US)

(73) Assignee: **The University of Iowa Research Foundation**, Iowa City, IA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **11/019,081**

(22) Filed: **Dec. 21, 2004**

(65) **Prior Publication Data**

US 2006/0134649 A1 Jun. 22, 2006

(51) **Int. Cl.**
C07H 21/04 (2006.01)

(52) **U.S. Cl.** **536/23.1**; 536/23.5; 536/24.1

(58) **Field of Classification Search** None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,474,796 A * 12/1995 Brennan 427/2.13
2005/0246794 A1 * 11/2005 Khvorova et al. 800/286

FOREIGN PATENT DOCUMENTS

WO WO 01/12659 A2 * 2/2001

OTHER PUBLICATIONS

Chou, C-C., et al. 2004 *Nucleic Acids Research* 32(12): e99(pp. 1-8).*

Sequence results (2 pages).*

Ansley et al., "Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome," *Nature*, 425:628-633, 2003.

Badano et al., "Identification of a novel bardet-biedl syndrome protein, BBS7, that shares structural features with BBS1 and BBS2," *Am J Hum Genet*, 72:650-658, 2003.

Bhamidipati et al., "ADP ribosylation factor-like protein 2 (Arf12) regulates the interaction of tubulin-folding cofactor d with native tubulin," *J Cell Biol*, 149:1087-1096, 2000.

Bruford et al., "Linkage mapping in 29 bardet-biedl syndrome families confirms loci in chromosomal regions 11q13, 15q22.3-q23, and 16q21," *Genomics*, 41:93-99, 1997.

Chiang, et al., "Comparative genomic analysis identifies an ADP-ribosylation factor-like gene as the cause of bardet-biedl syndrome (BBS3)" *Am J Hum Genet*, 75:475-484, 2004.

GenBank Accession No. NP_115522.

GenBank Accession No. NP_611421.

Ingley et al., "A novel-adp-ribosylation like factor (arf-6), interacts with the protein-conduction channel sec61 β subunit," *FEBS Lett*, 459:69-74, 1999.

Jacobs et al., "ADP-ribosylation factor(arf)-like 4,6, and 7 represent a subgroup of the arf family characterized by rapid nucleotide exchange and a nuclear localization signal," *FEBS Lett*, 456:384-388, 1999.

Katsanis et al., "Mutations in MKKS cause obesity, retinal dystrophy and renal malformations associated with bardet-biedl syndrome," *Nat Genet*, 26:67-70, 2000.

Li et al., "Comparative genomics identifies a flagellar and basal body proteome that includes the bbs5 human disease gene," *Cell*, 117:541-552, 2004.

Mykytyn et al., "Identification of the gene that, when mutated, causes the human obesity syndrome bbs4," *Nat Genet*, 28:188-191, 2001.

Mykytyn et al., "Identification of the gene (bbs1) most commonly involved in bardet-biedl syndrome, a complex human obesity syndrome," *Nat Genet*, 31:435-438, 2002.

Mykytyn et al., "Bardet-biedl syndrome type 4 (bbs4)-null mice implicate bbs4 in flagella formation but not global cilia assembly," *Proc Natl Acad Sci USA*, 101:8664-8669, 2004.

Nishimura et al., "Positional cloning of novel gene on chromosome 16q causing bardet-biedl syndrome (bbs2)," *Hum Mol Genet*, 10:865-874, 2001.

Pasqualato et al., "Arf, arl, arp and sar proteins: a family of gtp-binding proteins with a structural device for 'front-back' communication," *EMBO Rep*, 3:1035-1041, 2002.

Slavotinek et al., "Mutations in mkks cause bardet-biedl syndrome," *Nature Genet*, 26:15-16, 2000.

* cited by examiner

Primary Examiner—Maryam Monshipouri

Assistant Examiner—Marsha Tsay

(74) *Attorney, Agent, or Firm*—Fulbright & Jaworski L.L.P.

(57) **ABSTRACT**

The present invention relates to the designation of ADP-ribosylation factor-like 6 as the BBS3 gene, that is involved in the genetic disease Bardet Biedl Syndrome (BBS), which is characterized by such diverse symptoms as obesity, diabetes, hypertension, mental retardation, renal cancer and other abnormalities, retinopathy and hypogonadism. Methods of use for the gene, for example in diagnosis and therapy of BBS and in drug screening, also are described.

10 Claims, 4 Drawing Sheets

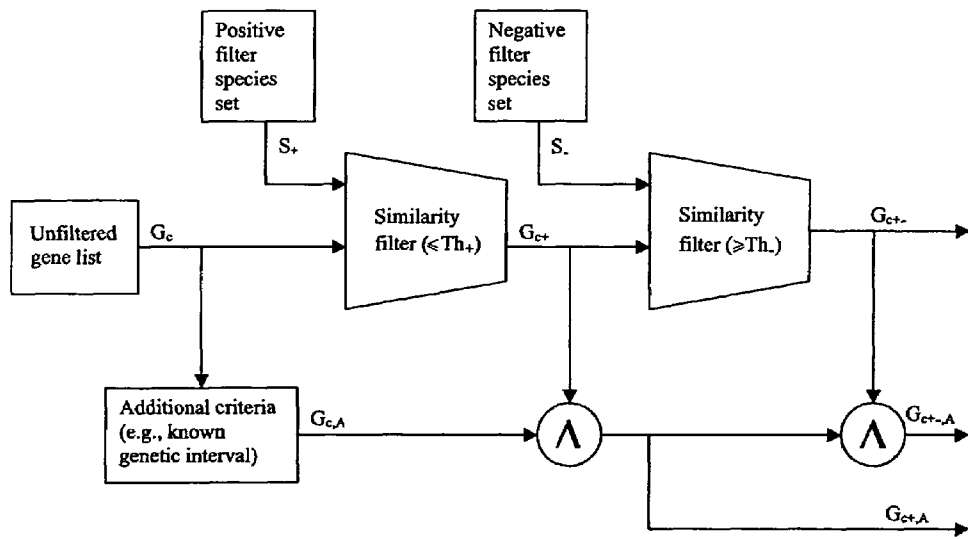


FIG. 1







<u>Marker</u>	<u>Locus</u>	<u>Genetic (cM)</u>	<u>Physical (bp)</u>	<u>IV-7</u>	<u>V-27</u>
GAT ABBE12	D3S3049	109.22	78,830,250		
Mfd233A	D3S1254	110.82	82,860,300		
AFM161xg11	D3S1276	111.89	85,177,250		
UT674	D3S1663	111.89	85,538,500		
AFM294z19	D3S1595	112.42	86,092,000		
AFMb350ze1	D3S3671	112.96	86,888,500		
GAT A13H08	D3S2386	114.02	87,839,500		
Mfd210A	D3S1251	114.02	95,598,000		
ATC3D09	D3S1752	114.02	99,066,250		
AFM126zc5	D3S1271	117.76	102,055,750		
AFMb327 yb5	D3S3655	117.76	103,025,500		
GAT A11 F06	D3S1753	117.76	103,173,750		
AFM222xb12	D3S1302	124.83	109,873,500		

FIG. 2

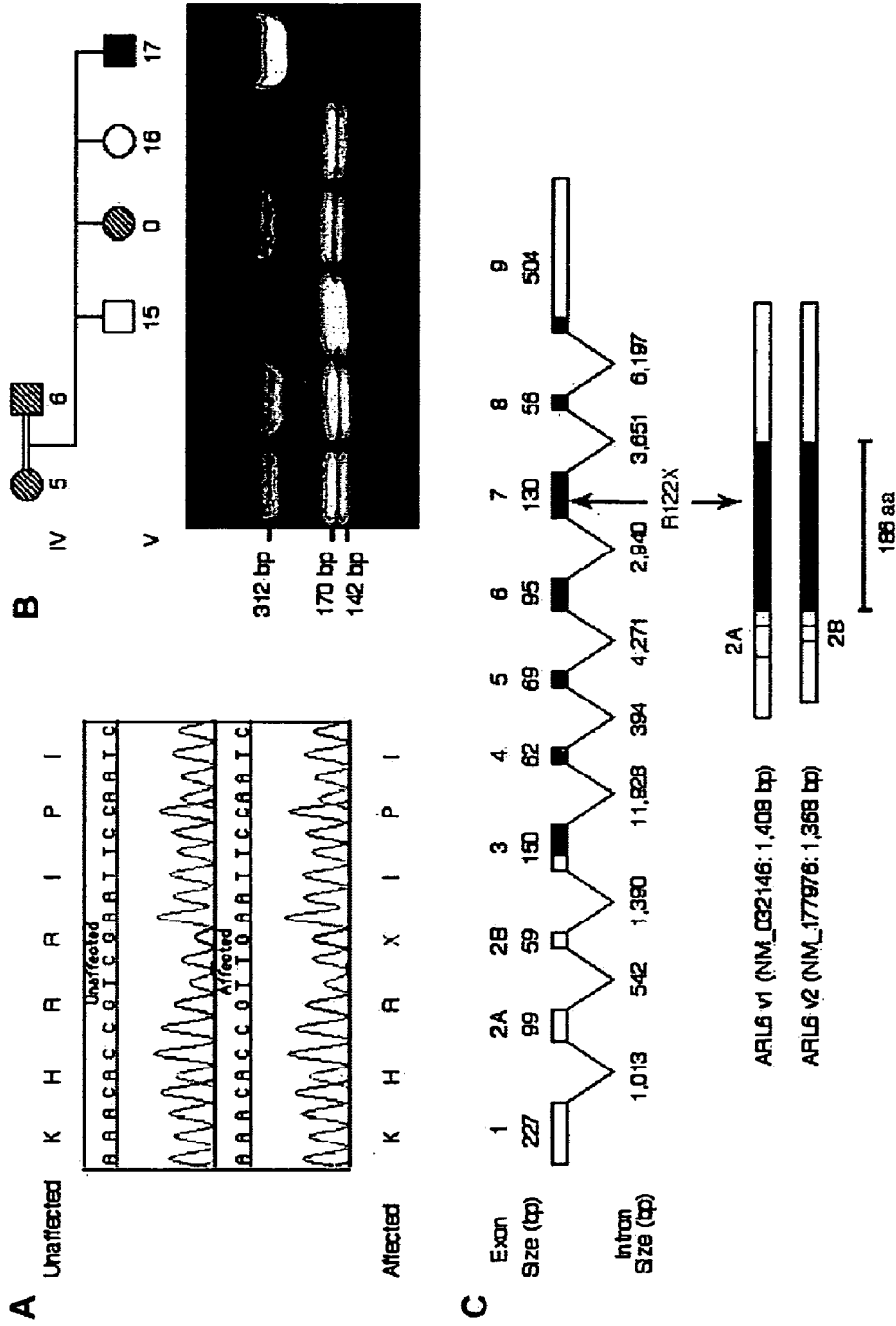


FIG. 3

HS NP 115522	1	MGLLDRLSVLLGLKKKEVHVLCGLDNGSKTTI INKLPNSNAQSQNLTPTIGFSIEKFKS	60
MM NP 062639	1	MVLLDRLSGLLGLKKKEVHVLCGLDNGSKTTI INKLPNSNAQSQDIVPTIGFSIEKFKS	60
RN ARL6	1	MGLLDRLSGLLGLKKKEVHVLCGLDNGSKTTI INKLPNSNAQSQDIVPTIGFSIEKFKS	60
CI ci0100130514	1	MGIMNKLFGWLNKGGKKEANVLCVGLDNGSKTTI INYLPNDAAQVDIVPTVGFVVEKVTM	60
DM NP 611421	1	MGLMLHNLADLIRIKKDKRMTIILVGLDNGSKSSI INHFKKSSSQTSIVVPTVGFVVEQFYI	60
CE NP 497993	1	MGFFSSLSLFLGLGKKDQVNTIVVGLDNGSKTTI ILNQLKTPETRSQQIVPTVGHVVTNFTS	60
TB 228.m00796	1	MC-----QSKTKLQVVMCGLDNSGKTTI INQVKPAQSSSKHITATVGVNVEVFER	50
TC 7892.m00006	1	MC-----QSKTKLHIITCGLDNSGKSSI INRLKPTALQSEHISATVGVNVEVFER	50
CR 171782	1	MGFFDKLLSLFCMSGKRVNVLVGLDNGSKTTI IERLKP RPRAAEVAPT VGFVDEVER	60
AN EAA66244	1	MGLAISKLFDRLWGGKEMRIIMVGLDAAAGKTTI ILYKLLGGEIVTT--IPTIGFVVEVVEY	58
AT CAB87634	1	MGLNFTLFLFSRLEFAKKEIRILMVGDLAAGKTTI ILYKLLGGEIVTT--IPTIGFVVEVVEY	58
SC CAA98769	1	MGLFASKLFSNLFENKEMRIIMVGLDGAAGKTTI ILYKLLGGEIVTT--IPTIGFVVEVVEY	58

HS NP 115522	61	--SSLSFTVFDMSGQGRYRNLEWHYYKGGQAI IFVIDSSDRLRMVVAKKEELDTLLNHFDI	118
MM NP 062639	61	--SSLSFTVFDMSGQGRYRNLEWHYYKGGQAI IFVIDSSDKLRMVVAKKEELDTLLNHFDI	118
RN ARL6	61	--SSLSFTVFDMSGQGRYRNLEWHYYKGGQAI IFVVDSSDKLRMVVAKKEELDTLLNHFDI	118
CI ci0100130514	61	--TNLSFTVFDMSGQGRYRVLWSHYYKGTQCVI FVVDSSADKLRMAVAKDELDTLLKHQTI	118
DM NP 611421	61	GMSGVSKAIDMSGATRYRNLEWHQFQKCHGI IYVIDSSDRMRVFWVVDLDTLLVQHPDL	120
CE NP 497993	61	--QNLSPHAFDMAGQMKYRSTWESYFHSQGVIFVLDSSDRLRMELLKDELMVMVHRDV	118
TB 228.m00796	51	--GRVAFIVFDMSGAKKFRGLWETYYDNIDAVIFVVDSSDHLRLCVVKEIQAMLKHEDI	108
TC 7892.m00006	51	--GSAKFTVFDMSGAKKFRGLWETYYENINGIIFVVDSSDELRLCVVKEIETALQHPDI	108
CR 171782	61	--GPLTFIVFDMSGAGRYRTLWEQYYREADAVFVVDSSADKLRMVVAREMEHMLKSNM	118
AN EAA66244	59	--KNIQFTVWDVGGQDKIRPLMRHYFQNTQGI IFVVDSSNDRDRIVEARELQRMINEDEL	116
AT CAB87634	59	--KNISFTVWDVGGQDKIRPLMRHYFQNTQGLIFVVDSSNDRDRIVEARELHRMLNEDEL	116
SC CAA98769	59	--KNISFTVWDVGGQDRIRSLMRHYRYNTEGVI FVVDSSNDRSRIGEAREVMQRMNEDEL	116



HS NP 115522	119	KHR-----RIPILFFANKMDLRDAVTSVKVSQLLCLENIK-DKPWHICASDAIKGEGE	170
MM NP 062639	119	KHR-----RIPILFFANKMDLRDSVTSVKVSQLLCLESIK-DKPWHICASDAIKGEGE	170
RN ARL6	119	KHR-----RIPILFFANKMDLRDAVTSVKVSQLLCLENIK-DKPWHICASDAIKGEGE	170
CI ci0100130514	119	MNK-----RIPILFFANKMDVQNSLSAVKCSQLLGLLENIK-EKPWHICASNAIKGEGE	170
DM NP 611421	121	CNR-----IVPILFYCNKMDMEDSLSSVKIAAALRLLENIK-DKPWHICSSSAISGEGE	172
CE NP 497993	119	VSR-----GIPIVILANKMDIPGAMTASDITVALGILNLYR-SGTWSIHSTCALTGDL	170
TB 228.m00796	109	RRELPL--GGGRVPPFLFFANKMDAAGAKTAAELVEILDLTTLMGDHPFVIFASNALKGTGV	166
TC 7892.m00006	109	ARELPKRENGAKIPFLFYANKMDLRFNAKTAELVDLLDLTTLMDRPFNIFASNALKGTGV	168
CR 171782	119	R-----KVPILLYFANKKDLFVAMPFVEIQAALGLDDIK--DRPWFQIVPSNGLTGEV	168
AN EAA66244	117	RDA-----LLLVFANKQDLFNAMSFAEITQQLGLQSLT-RRFWYIQSTCATGDL	166
AT CAB87634.1	117	RDA-----VLLLVFANKQDLFNAMNAEITDKLGLHSIR-QRHWYIQSTCATSGEGL	166
SC CAA98769	117	RNA-----AMLVFNANKQDLPEAMSAEITEKLGHSIR-NRPWFYIQTATCATSGEGL	166

HS NP 115522	171	QEGVDWLQDQIQIVKT-----	186
MM NP 062639	171	QEGVDWLQDQIQIVKT-----	186
RN ARL6	171	QEGVDWLQDQIQIVKT-----	186
CI ci0100130514	171	SDGMHWLSDQLQIVK-----	185
DM NP 611421	173	GEGVQWLIQQMRFAMLNKNAKRSRKHSK	202
CE NP 497993	171	DKAMQQLSAEITKME SRRT-----	190
TB 228.m00796	167	HEGFSWLQETASRQSC--FAGTKRG-----	189
TC 7892.m00006	169	NEGFFWLQNALRQSNTIISSSSRGARQRN	198
CR 171782	169	DKGIDWLAERLS-----	180
AN EAA66244	167	YEGLEWLAETLRKTGRD-----	183
AT CAB87634	167	YEGLDWLSNNIATKVNNLNIPF-----	188
SC CAA98769	167	YEGLEWLSNSLKNST-----	181

FIG. 4

1

BARDET-BIEDL SUSCEPTIBILITY GENE AND USES THEREOF

The government owns rights in the present invention pursuant to grant numbers P50-HL-5506 and R01-EY-11298 from the National Institutes of Health.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to the fields of genetics and molecular biology. More particular the invention relates to the identification of ADP-ribosylation factor-like 6 as being involved in Bardet-Biedl Syndrome (BBS), and designated here as BBS3. Defects in this gene are associated with a variety of clinical symptoms including diabetes, hypogonadism, high blood pressure, renal cancer and other defects, retinal degeneration, congenital heart defects, limb deformity or polydactyly, mental retardation and obesity. Also provided are methods of therapy and methods of screening for therapeutic compositions.

2. Description of Related Art

Bardet-Biedl syndrome (BBS [MIM 209900]) is a pleiotropic autosomal recessive disorder characterized by obesity, pigmentary retinopathy, polydactyly, renal abnormalities, learning disabilities, and hypogonadism (Bardet 1920; Biedl 1922; Green et al., 1989). The disorder is also associated with an increased susceptibility to diabetes mellitus, hypertension, and congenital heart disease (Harnett et al., 1988; Green et al., 1989; Elbedour et al., 1994). The disorder shows variable expressivity within and between families, a finding suggesting genetic complexity. BBS displays extensive genetic heterogeneity and a requirement for full penetrance of two mutations at one locus and a third mutation at a second locus has been suggested (Katsanis et al., 2001). To date, eight BBS loci have been mapped, and seven BBS genes have been identified. The first three BBS genes to be discovered (BBS6, BBS2, and BBS4) are not homologous to one another and were identified using positional cloning (Katsanis et al., 2000; Slavotinek et al., 2000; Mykytyn et al., 2001; Nishimura et al., 2001). Subsequently, bioinformatics comparisons of protein sequences have aided in the identification of additional BBS genes. For example, the positional cloning of the BBS1 gene was aided by the finding of limited sequence homology to BBS2 (Mykytyn et al., 2002). BBS7 and BBS8 were identified by database searches for proteins with partial homology to BBS2 and BBS4, respectively (Ansley et al., 2003; Badano et al., 2003). BBS5 was recently identified using a combination of comparative genomics and positional cloning (Li et al., 2004).

The BBS3 locus was initially mapped to chromosome 3 in a large, inbred Israeli Bedouin kindred in a study that showed the utility of using pooled DNA samples for genetic mapping of human disorders (Sheffield et al., 1994). Genotyping of the Bedouin DNA samples initially localized the disease interval to an approximately 10 cM region. Few additional BBS3-linked families have been identified, and this has limited extensive refinement of the disease locus. However, as of yet the identity of the BBS3 gene has not been elucidated.

SUMMARY OF THE INVENTION

Thus, in accordance with the present invention, there is provided an isolated and purified nucleic acid encoding a human ADP-ribosylation factor-like 6 polypeptide compris-

2

ing the sequence of SEQ ID NO:2 or 4. The nucleic acid may comprise the sequence of SEQ ID NO:1 or 3. The nucleic acid may further comprise a promoter, such as a constitutive or inducible promoter, an origin of replication, as in part of a replicable vector. The replicable vector may be a viral vector, such as a retroviral vector, an adenoviral vector, an adeno-associated viral vector, a herpesviral vector, a polyoma viral vector, a vaccinia viral vector or a lentiviral vector, including a viral vector located within a viral particle, or the replicable vector may be a non-viral vector.

In another embodiment, there is provided an oligonucleotide of about 10 to about 50 bases comprising at least 10 consecutive bases including base 364 or 499 of SEQ ID NO:1 or 3 or the complement thereof. The oligonucleotide may be 10, 15, 20, 25, 30, 35, 40, 45 or 50 bases in length. The number of consecutive bases may be 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50.

In yet another embodiment, there is provided an isolated and purified human ADP-ribosylation factor-like 6 polypeptide comprising the sequence of SEQ ID NO:2 or 4. The human ADP-ribosylation factor-like 6 polypeptide may be fused to another polypeptide.

In still yet another embodiment, there is provided a peptide of 8 to 50 residues comprising at least 5 consecutive residues including residue 122 of SEQ ID NO:2 or 4. The peptide may be 10, 15, 20, 25, 30, 35, 40, 45 or 50 residues in length. The number of consecutive residues may be 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50. The peptide may be bound to a carrier molecule, for example, by a linker.

In a further embodiment, there is provided a method of diagnosing Bardet-Biedl Syndrome (BBS) comprising identifying a mutation in a human ADP-ribosylation factor-like 6 polypeptide or nucleic acid. The method may comprise immunologic analysis using a human ADP-ribosylation factor-like 6 polypeptide-binding monoclonal antibody or polyclonal antiserum, such as by using ELISA, RIA, or Western blot. The method may comprise identifying a mutation of R₁₂₂→X or G₁₆₇→R. The method may also comprise identifying a mutation in a human ADP-ribosylation factor-like 6 mRNA or genomic sequence, such as C364T or G499A. The method may comprise amplification of the nucleic acid. The method may also comprise hybridization of the nucleic acid to a labeled nucleic acid probe. The method may also comprise sequencing of a human ADP-ribosylation factor-like 6 nucleic acid.

In still further embodiments, there are provided:

- a method of identifying an individual genetically predisposed to obesity comprising identifying a mutation in a human ADP-ribosylation factor-like 6 polypeptide or nucleic acid;
- a method of identifying an individual genetically predisposed to diabetes mellitus comprising identifying a mutation in a human ADP-ribosylation factor-like 6 polypeptide or nucleic acid;
- a method of identifying an individual genetically predisposed to renal defects comprising identifying a mutation in a human ADP-ribosylation factor-like 6 polypeptide or nucleic acid;
- a method of identifying an individual genetically predisposed to retinopathy comprising identifying a mutation in a human ADP-ribosylation factor-like 6 polypeptide or nucleic acid;

a method of identifying an individual genetically predisposed to hypogonadism comprising identifying a mutation in a human ADP-ribosylation factor-like 6 polypeptide or nucleic acid;

a method of identifying an individual genetically predisposed to polydactyly comprising identifying a mutation in a human ADP-ribosylation factor-like 6 polypeptide or nucleic acid; and

a method of identifying an individual genetically predisposed to mental retardation comprising identifying a mutation in a human ADP-ribosylation factor-like 6 polypeptide or nucleic acid;

In yet a further embodiment, there is provided a method of screening for a modulator of human ADP-ribosylation factor-like 6 polypeptide comprising (a) providing a cell expressing a human ADP-ribosylation factor-like 6 polypeptide; (b) contacting the cell with a candidate modulator; (c) measuring human ADP-ribosylation factor-like 6 expression; and (d) comparing the human ADP-ribosylation factor-like 6 expression in the presence of the candidate modulator with the expression of human ADP-ribosylation factor-like 6 in the absence of the candidate modulator; wherein a difference in the expression of human ADP-ribosylation factor-like 6 in the presence of the candidate modulator, as compared with the expression of human ADP-ribosylation factor-like 6 in the absence of the candidate modulator, identifies the candidate modulator as a modulator of human ADP-ribosylation factor-like 6 expression.

In still yet a further embodiment, there is provided a method of screening for a modulator of human ADP-ribosylation factor-like 6 expression comprising (a) providing a cell that comprises an expression construct encoding an indicator polypeptide under the control of a human ADP-ribosylation factor-like 6 promoter; (b) contacting the cell with a candidate modulator; (c) measuring expression of the indicator polypeptide; and (d) comparing the expression of the indicator polypeptide in the presence of the candidate modulator with the expression of the indicator polypeptide in the absence of the candidate modulator; wherein a difference in the expression of the indicator polypeptide in the presence of the candidate modulator, as compared with the expression of the indicator polypeptide in the absence of the candidate modulator, identifies the candidate modulator as a modulator of human ADP-ribosylation factor-like 6 expression.

An additional embodiment comprises a method of producing a modulator of human ADP-ribosylation factor-like 6 expression comprising (a) providing a cell expressing a human ADP-ribosylation factor-like 6 polypeptide; (b) contacting the cell with a candidate modulator; (c) measuring human ADP-ribosylation factor-like 6 expression; (d) comparing the human ADP-ribosylation factor-like 6 expression in the presence of the candidate modulator with the expression of human ADP-ribosylation factor-like 6 in the absence of the candidate modulator; wherein a difference in the expression of human ADP-ribosylation factor-like 6 in the presence of the candidate modulator, as compared with the expression of human ADP-ribosylation factor-like 6 in the absence of the candidate modulator, identifies the candidate modulator as a modulator of human ADP-ribosylation factor-like 6 expression; and (e) producing the modulator.

In yet an additional embodiment, there is provided a modulator of human ADP-ribosylation factor-like 6 polypeptide produced according to the method comprising (a) providing a cell expressing a human ADP-ribosylation factor-like 6 polypeptide; (b) contacting the cell with a candidate modulator; (c) measuring human ADP-ribosyla-

tion factor-like 6 expression; (d) comparing the human ADP-ribosylation factor-like 6 expression in the presence of the candidate modulator with the expression of human ADP-ribosylation factor-like 6 in the absence of the candidate modulator; wherein a difference in the expression of human ADP-ribosylation factor-like 6 in the presence of the candidate modulator, as compared with the expression of human ADP-ribosylation factor-like 6 in the absence of the candidate modulator, identifies the candidate modulator as a modulator of human ADP-ribosylation factor-like 6 polypeptide; and (e) producing the modulator.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

FIG. 1—General outline of the candidate ranking process. A list of unfiltered genes (G_c) can be prioritized through a set of positive filter species (S_+) on the basis of a similarity filter threshold (Th_+), which yields a subset of genes (G_{c+}) found in all positive filter species (S_+). These can be further screened with similarity filter threshold (Th_-) in a set of negative filter species (S_-) to yield a more restrictive subset of genes (G_{c+}). Further filtering continues by intersection (Λ) with additional criteria (A) to generate ($G_{c+,A}$). An even more refined set ($G_{c+,A}$) can be obtained by intersecting with $G_{c+,A}$ with G_{c+} .

FIG. 2—Refinement of genetic localization of the BBS3 candidate interval. The genetic map was obtained from the Marshfield Medical Clinic Web site, and the physical map distances were obtained from the UCSC Genome Browser on the basis of the July 2003 data release. Critical recombination events are also illustrated. The patient identification terminology is the same as was published previously (Sheffield et al., 1994).

FIGS. 3A-C—BBS3 mutation (R122X) detected in a large Bedouin kindred. FIG. 3A, Sequence from an affected individual from the Bedouin family and a control sample, showing the homozygous CrT change that results in premature termination at codon 122. FIG. 3B, An example of the TaqI restriction enzyme digest that was used to confirm the R122X mutation. The mutation results in the abolition of a TaqI site within exon 7. Following TaqI digestion of a PCR fragment containing exon 7, the wild-type allele is observed as two bands (142 bp and 170 bp), whereas the uncut mutant allele produces a 312-bp fragment. For the pedigree, the hatched symbols represent BBS carriers, as determined by genetic analysis; the filled symbol denotes an individual with BBS; and the open symbols are unaffected individuals. The patient-identification terminology is the same as was published previously (Sheffield et al., 1994), with the exception of the 0, which denotes a sample that was not previously available. FIG. 3C, The genomic structure of the ARL6 gene is shown, with the blue shading representing the translated

region. The two ARL6 isoforms that are produced by alternative splicing are shown below. The location of the R122X mutation within the ARL6 gene is indicated in red.

FIG. 4—Multiple alignment of ARL6 (HS_NP_115522) and the corresponding best BLAST hit in 11 other model organisms. The mutation (R122X) is denoted by an arrow. Each sequence is denoted by the first letter of the genus-species name, followed by a GenBank accession number (whenever possible) or a unique identifier (e.g., “DM_NP_611421” refers to the protein represented by NP_611421 in the genome of *Drosophila melanogaster*). Consensus residues are shown in red; conserved residues are shown in blue. Numbers flanking sequences correspond to the position of the residue within each sequence (excluding gaps). Abbreviations are as follows: HS, *Homo sapiens*; MM, *Mus musculus*; RN, *Rattus norvegicus*; CI, *Ciona intestinalis*; DM, *Drosophila melanogaster*; CE, *Caenorhabditis elegans*; TB, *Trypanosoma brucei*; TC, *Trypanosoma cruzi*; CR, *Chlamydomonas reinhardtii*; AN, *Aspergillus nidulans*; AT, *Arabidopsis thaliana*; and SC, *Saccharomyces cerevisiae*.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

Bardet Biedl Syndrome (BBS) is a debilitating genetic disorder that is prevalent in Bedouin populations, probably due to the high consanguinity observed therein. The present inventors reasoned that organisms containing orthologues to the known BBS genes were likely to also contain orthologues to as-yet-unidentified BBS genes. The sequencing of the human genome (Lander et al., 2001; Venter et al., 2001), as well as the sequencing of other model organisms, has made it possible to perform cross-species comparisons of genomes (comparative genomics). Using comparisons of the human genome with the genomes of other model organisms, particularly lower eukaryotic organisms with relatively fewer genes, the inventors prioritized BBS candidate genes for mutation screening and subsequently identified of the BBS3 gene.

1. Proteins

The protein sequence for human ADP-ribosylation factor-like 6 is provided in SEQ ID NO:2 or 4. In addition to the entire human ADP-ribosylation factor-like 6 polypeptide, the present invention also relates to fragments of the polypeptides that may or may not retain various of the functions described below. Fragments, including the N-terminus of the molecule, may be generated by genetic engineering of translation stop sites within the coding region (discussed below). Alternatively, treatment of the human ADP-ribosylation factor-like 6 polypeptide with proteolytic enzymes, known as proteases, can produce a variety of N-terminal, C-terminal and internal fragments. Peptides range from 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, and 50 residues, such as those made synthetically, up to 100, 150, 200, 250, 300, 350, 400, 450, 500 and more residues, which are conveniently produced by recombinant means or by proteolytic digestion of full length human ADP-ribosylation factor-like 6 polypeptide. Examples of fragments may include contiguous residues of SEQ ID NO:2 or 4 of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 75, 80, 85, 90, 95, 100, 200, 300, 400 or more amino acids in length. These fragments may be purified according to known methods, such as precipitation (e.g., ammonium sulfate), HPLC, ion exchange chromatography, affinity chromatography (including immu-

noaffinity chromatography) or various size separations (sedimentation, gel electrophoresis, gel filtration).

A. Variants of BBS3

Amino acid sequence variants of the human ADP-ribosylation factor-like 6 polypeptide can be substitutional, insertional or deletion variants. Deletion variants lack one or more residues of the native protein which are not essential for function or immunogenic activity, and are exemplified by the variants lacking a transmembrane sequence described above. Another common type of deletion variant is one lacking secretory signal sequences or signal sequences directing a protein to bind to a particular part of a cell. Insertional mutants typically involve the addition of material at a non-terminal point in the polypeptide. This may include the insertion of an immunoreactive epitope or simply a single residue. Terminal additions, called fusion proteins, are discussed below.

Substitutional variants typically contain the exchange of one amino acid for another at one or more sites within the protein, and may be designed to modulate one or more properties of the polypeptide, such as stability against proteolytic cleavage, without the loss of other functions or properties. Substitutions of this kind preferably are conservative, that is, one amino acid is replaced with one of similar shape and charge. Conservative substitutions are well known in the art and include, for example, the changes of: alanine to serine; arginine to lysine; asparagine to glutamine or histidine; aspartate to glutamate; cysteine to serine; glutamine to asparagine; glutamate to aspartate; glycine to proline; histidine to asparagine or glutamine; isoleucine to leucine or valine; leucine to valine or isoleucine; lysine to arginine; methionine to leucine or isoleucine; phenylalanine to tyrosine, leucine or methionine; serine to threonine; threonine to serine; tryptophan to tyrosine; tyrosine to tryptophan or phenylalanine; and valine to isoleucine or leucine.

The following is a discussion based upon changing of the amino acids of a protein to create an equivalent, or even an improved, second-generation molecule. For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid substitutions can be made in a protein sequence, and its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the DNA sequences of genes without appreciable loss of their biological utility or activity, as discussed below. Table 1 shows the codons that encode particular amino acids.

In making such changes, the hydrophobic index of amino acids may be considered. The importance of the hydrophobic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982). It is accepted that the relative hydrophobic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like.

Each amino acid has been assigned a hydrophobic index on the basis of their hydrophobicity and charge characteristics (Kyte and Doolittle, 1982), these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8);

cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydrophobic index or score and still result in a protein with similar biological activity, i.e., still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydrophobic indices are within ± 2 is preferred, those which are within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Pat. No. 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein. As detailed in U.S. Pat. No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 \pm 1); glutamate (+3.0 \pm 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 \pm 1); alanine (-0.5); histidine *-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent and immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ± 2 is preferred, those that are within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

As outlined above, amino acid substitutions are generally based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions that take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

Another embodiment for the preparation of polypeptides according to the invention is the use of peptide mimetics. Mimetics are peptide-containing molecules that mimic elements of protein secondary structure (Johnson et al, 1993). The underlying rationale behind the use of peptide mimetics is that the peptide backbone of proteins exists chiefly to orient amino acid side chains in such a way as to facilitate molecular interactions, such as those of antibody and antigen. A peptide mimetic is expected to permit molecular interactions similar to the natural molecule. These principles may be used, in conjunction with the principles outlined above, to engineer second generation molecules having many of the natural properties of human ADP-ribosylation factor-like 6 polypeptide, but with altered and even improved characteristics.

B. Domain Switching

An interesting series of mutants can be created by substituting homologous regions of various proteins. This is known, in certain contexts, as "domain switching." Domain switching involves the generation of chimeric molecules using different but, in this case, related polypeptides. By comparing various ADP-ribosylation factor-like 6 proteins,

one can make predictions as to the functionally significant regions of these molecules. It is possible, then, to switch related domains of these molecules in an effort to determine the criticality of these regions to human ADP-ribosylation factor-like 6 function. These molecules may have additional value in that these "chimeras" can be distinguished from natural molecules, while possibly providing the same function.

C. Fusion Proteins

A specialized kind of insertional variant is the fusion protein. This molecule generally has all or a substantial portion of the native molecule, linked at the N- or C-terminus, to all or a portion of a second polypeptide. For example, fusions typically employ leader sequences from other species to permit the recombinant expression of a protein in a heterologous host. Another useful fusion includes the addition of an immunologically active domain, such as an antibody epitope, to facilitate purification of the fusion protein. Inclusion of a cleavage site at or near the fusion junction will facilitate removal of the extraneous polypeptide after purification. Other useful fusions include linking of functional domains, such as active sites from enzymes, glycosylation domains, cellular targeting signals or transmembrane regions.

D. Purification of Proteins

It will be desirable to purify human ADP-ribosylation factor-like 6 polypeptides or variants thereof. Protein purification techniques are well known to those of skill in the art. These techniques involve, at one level, the crude fractionation of the cellular milieu to polypeptide and non-polypeptide fractions. Having separated the polypeptide from other proteins, the polypeptide of interest may be further purified using chromatographic and electrophoretic techniques to achieve partial or complete purification (or purification to homogeneity). Analytical methods particularly suited to the preparation of a pure peptide are ion-exchange chromatography, exclusion chromatography; polyacrylamide gel electrophoresis; isoelectric focusing. A particularly efficient method of purifying peptides is fast protein liquid chromatography or even HPLC.

Certain aspects of the present invention concern the purification, and in particular embodiments, the substantial purification, of an encoded protein or peptide. The term "purified protein or peptide" as used herein, is intended to refer to a composition, isolatable from other components, wherein the protein or peptide is purified to any degree relative to its naturally-obtainable state. A purified protein or peptide therefore also refers to a protein or peptide, free from the environment in which it may naturally occur.

Generally, "purified" will refer to a protein or peptide composition that has been subjected to fractionation to remove various other components, and which composition substantially retains its expressed biological activity. Where the term "substantially purified" is used, this designation will refer to a composition in which the protein or peptide forms the major component of the composition, such as constituting about 50%, about 60%, about 70%, about 80%, about 90%, about 95% or more of the proteins in the composition.

Various methods for quantifying the degree of purification of the protein or peptide will be known to those of skill in the art in light of the present disclosure. These include, for example, determining the specific activity of an active fraction, or assessing the amount of polypeptides within a fraction by SDS/PAGE analysis. A preferred method for assessing the purity of a fraction is to calculate the specific activity of the fraction, to compare it to the specific activity

of the initial extract, and to thus calculate the degree of purity, herein assessed by a “-fold purification number.” The actual units used to represent the amount of activity will, of course, be dependent upon the particular assay technique chosen to follow the purification and whether or not the expressed protein or peptide exhibits a detectable activity.

Various techniques suitable for use in protein purification will be well known to those of skill in the art. These include, for example, precipitation with ammonium sulphate, PEG, antibodies and the like or by heat denaturation, followed by centrifugation; chromatography steps such as ion exchange, gel filtration, reverse phase, hydroxylapatite and affinity chromatography; isoelectric focusing; gel electrophoresis; and combinations of such and other techniques. As is generally known in the art, it is believed that the order of conducting the various purification steps may be changed, or that certain steps may be omitted, and still result in a suitable method for the preparation of a substantially purified protein or peptide.

There is no general requirement that the protein or peptide always be provided in their most purified state. Indeed, it is contemplated that less substantially purified products will have utility in certain embodiments. Partial purification may be accomplished by using fewer purification steps in combination, or by utilizing different forms of the same general purification scheme. For example, it is appreciated that a cation-exchange column chromatography performed utilizing an HPLC apparatus will generally result in a greater “-fold” purification than the same technique utilizing a low pressure chromatography system. Methods exhibiting a lower degree of relative purification may have advantages in total recovery of protein product, or in maintaining the activity of an expressed protein.

It is known that the migration of a polypeptide can vary, sometimes significantly, with different conditions of SDS/PAGE (Capaldi et al., 1977). It will therefore be appreciated that under differing electrophoresis conditions, the apparent molecular weights of purified or partially purified expression products may vary.

High Performance Liquid Chromatography (HPLC) is characterized by a very rapid separation with extraordinary resolution of peaks. This is achieved by the use of very fine particles and high pressure to maintain an adequate flow rate. Separation can be accomplished in a matter of minutes, or at most an hour. Moreover, only a very small volume of the sample is needed because the particles are so small and close-packed that the void volume is a very small fraction of the bed volume. Also, the concentration of the sample need not be very great because the bands are so narrow that there is very little dilution of the sample.

Gel chromatography, or molecular sieve chromatography, is a special type of partition chromatography that is based on molecular size. The theory behind gel chromatography is that the column, which is prepared with tiny particles of an inert substance that contain small pores, separates larger molecules from smaller molecules as they pass through or around the pores, depending on their size. As long as the material of which the particles are made does not adsorb the molecules, the sole factor determining rate of flow is the size. Hence, molecules are eluted from the column in decreasing size, so long as the shape is relatively constant. Gel chromatography is unsurpassed for separating molecules of different size because separation is independent of all other factors such as pH, ionic strength, temperature, etc. There also is virtually no adsorption, less zone spreading and the elution volume is related in a simple matter to molecular weight.

Affinity Chromatography is a chromatographic procedure that relies on the specific affinity between a substance to be isolated and a molecule that it can specifically bind to. This is a receptor-ligand type interaction. The column material is synthesized by covalently coupling one of the binding partners to an insoluble matrix. The column material is then able to specifically adsorb the substance from the solution. Elution occurs by changing the conditions to those in which binding will not occur (alter pH, ionic strength, temperature, etc.).

A particular type of affinity chromatography useful in the purification of carbohydrate containing compounds is lectin affinity chromatography. Lectins are a class of substances that bind to a variety of polysaccharides and glycoproteins. Lectins are usually coupled to agarose by cyanogen bromide. Concanavalin A coupled to Sepharose was the first material of this sort to be used and has been widely used in the isolation of polysaccharides and glycoproteins other lectins that have been include lentil lectin, wheat germ agglutinin which has been useful in the purification of N-acetyl glucosaminyl residues and *Helix pomatia* lectin. Lectins themselves are purified using affinity chromatography with carbohydrate ligands. Lactose has been used to purify lectins from castor bean and peanuts; maltose has been useful in extracting lectins from lentils and jack bean; N-acetyl-D galactosamine is used for purifying lectins from soybean; N-acetyl glucosaminyl binds to lectins from wheat germ; D-galactosamine has been used in obtaining lectins from clams and L-fucose will bind to lectins from lotus.

The matrix should be a substance that itself does not adsorb molecules to any significant extent and that has a broad range of chemical, physical and thermal stability. The ligand should be coupled in such a way as to not affect its binding properties. The ligand should also provide relatively tight binding. And it should be possible to elute the substance without destroying the sample or the ligand. One of the most common forms of affinity chromatography is immunoaffinity chromatography. The generation of antibodies that would be suitable for use in accord with the present invention is discussed below.

E. Synthetic Peptides

The present invention also describes smaller human ADP-ribosylation factor-like 6-related peptides for use in various embodiments of the present invention. Because of their relatively small size, the peptides of the invention can also be synthesized in solution or on a solid support in accordance with conventional techniques. Various automatic synthesizers are commercially available and can be used in accordance with known protocols. See, for example, Stewart and Young (1984); Tam et al. (1983); Merrifield (1986); and Barany and Merrifield (1979), each incorporated herein by reference. Short peptide sequences, or libraries of overlapping peptides, usually from about 6 up to about 35 to 50 amino acids, which correspond to the selected regions described herein, can be readily synthesized and then screened in screening assays designed to identify reactive peptides. Alternatively, recombinant DNA technology may be employed wherein a nucleotide sequence which encodes a peptide of the invention is inserted into an expression vector, transformed or transfected into an appropriate host cell and cultivated under conditions suitable for expression.

F. Antigen Compositions

The present invention also provides for the use of human ADP-ribosylation factor-like 6 polypeptides or peptides as antigens for the immunization of animals relating to the production of antibodies. It is envisioned that human ADP-ribosylation factor-like 6 polypeptide or portions thereof,

will be coupled, bonded, bound, conjugated or chemically-linked to one or more agents via linkers, polylinkers or derivatized amino acids. This may be performed such that a bispecific or multivalent composition or vaccine is produced. It is further envisioned that the methods used in the preparation of these compositions will be familiar to those of skill in the art and should be suitable for administration to animals, i.e., pharmaceutically acceptable. Preferred agents are the carriers are keyhole limpet hemocyanin (KLH) or bovine serum albumin (BSA).

G. Antibody Production

In certain embodiments, the present invention provides antibodies that bind with high specificity to the human ADP-ribosylation factor-like 6 polypeptides provided herein. Thus, antibodies that bind to the polypeptide of SEQ ID NO:2 or 4 are provided. In addition to antibodies generated against the full length proteins, antibodies may also be generated in response to smaller constructs comprising epitopic core regions, including wild-type and mutant epitopes.

As used herein, the term "antibody" is intended to refer broadly to any immunologic binding agent such as IgG, IgM, IgA, IgD and IgE. Generally, IgG and/or IgM are preferred because they are the most common antibodies in the physiological situation and because they are most easily made in a laboratory setting.

Monoclonal antibodies (MAbs) are recognized to have certain advantages, e.g., reproducibility and large-scale production, and their use is generally preferred. The invention thus provides monoclonal antibodies of the human, murine, monkey, rat, hamster, rabbit and even chicken origin. Due to the ease of preparation and ready availability of reagents, murine monoclonal antibodies will often be preferred.

However, "humanized" antibodies are also contemplated, as are chimeric antibodies from mouse, rat, or other species, bearing human constant and/or variable region domains, bispecific antibodies, recombinant and engineered antibodies and fragments thereof. Methods for the development of antibodies that are "custom-tailored" to the patient's dental disease are likewise known and such custom-tailored antibodies are also contemplated.

The term "antibody" is used to refer to any antibody-like molecule that has an antigen binding region, and includes antibody fragments such as Fab', Fab, F(ab')₂, single domain antibodies (DABs), Fv, scFv (single chain Fv), and the like. The techniques for preparing and using various antibody-based constructs and fragments are well known in the art. Means for preparing and characterizing antibodies are also well known in the art (See, e.g., *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988; incorporated herein by reference).

The methods for generating monoclonal antibodies (MAbs) generally begin along the same lines as those for preparing polyclonal antibodies. Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogenic BBS3 composition in accordance with the present invention and collecting antisera from that immunized animal.

A wide range of animal species can be used for the production of antisera. Typically the animal used for production of antisera is a rabbit, a mouse, a rat, a hamster, a guinea pig or a goat. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

As is well known in the art, a given composition may vary in its immunogenicity. It is often necessary therefore to boost the host immune system, as may be achieved by

coupling a peptide or polypeptide immunogen to a carrier. Exemplary and preferred carriers are keyhole limpet hemocyanin (KLH) and bovine serum albumin (BSA). Other albumins such as ovalbumin, mouse serum albumin or rabbit serum albumin can also be used as carriers. Means for conjugating a polypeptide to a carrier protein are well known in the art and include glutaraldehyde, m-maleimido-benzoyl-N-hydroxysuccinimide ester, carbodiimide and bis-biazotized benzidine.

As is also well known in the art, the immunogenicity of a particular immunogen composition can be enhanced by the use of non-specific stimulators of the immune response, known as adjuvants. Suitable adjuvants include all acceptable immunostimulatory compounds, such as cytokines, toxins or synthetic compositions.

Adjuvants that may be used include IL-1, IL-2, IL-4, IL-7, IL-12, γ -interferon, GMCSF, BCG, aluminum hydroxide, MDP compounds, such as thur-MDP and nor-MDP, CGP (MTP-PE), lipid A, and monophosphoryl lipid A (MPL). RIBI, which contains three components extracted from bacteria, MPL, trehalose dimycolate (TDM) and cell wall skeleton (CWS) in a 2% squalene/Tween 80 emulsion is also contemplated. MHC antigens may even be used. Exemplary, often preferred adjuvants include complete Freund's adjuvant (a non-specific stimulator of the immune response containing killed *Mycobacterium tuberculosis*), incomplete Freund's adjuvants and aluminum hydroxide adjuvant.

In addition to adjuvants, it may be desirable to coadminister biologic response modifiers (BRM), which have been shown to upregulate T cell immunity or downregulate suppressor cell activity. Such BRMs include, but are not limited to, Cimetidine (CIM; 1200 mg/d) (Smith/Kline, PA); low-dose Cyclophosphamide (CYP; 300 mg/m²) (Johnson/Mead, NJ), cytokines such as γ -interferon, IL-2, or IL-12 or genes encoding proteins involved in immune helper functions, such as B-7.

The amount of immunogen composition used in the production of polyclonal antibodies varies upon the nature of the immunogen as well as the animal used for immunization. A variety of routes can be used to administer the immunogen (subcutaneous, intramuscular, intradermal, intravenous and intraperitoneal). The production of polyclonal antibodies may be monitored by sampling blood of the immunized animal at various points following immunization.

A second, booster injection, may also be given. The process of boosting and titering is repeated until a suitable titer is achieved. When a desired level of immunogenicity is obtained, the immunized animal can be bled and the serum isolated and stored, and/or the animal can be used to generate MAbs.

For production of rabbit polyclonal antibodies, the animal can be bled through an ear vein or alternatively by cardiac puncture. The removed blood is allowed to coagulate and then centrifuged to separate serum components from whole cells and blood clots. The serum may be used as is for various applications or else the desired antibody fraction may be purified by well-known methods, such as affinity chromatography using another antibody, a peptide bound to a solid matrix, or by using, e.g., protein A or protein G chromatography.

MAbs may be readily prepared through use of well-known techniques, such as those exemplified in U.S. Pat. No. 4,196,265, incorporated herein by reference. Typically, this technique involves immunizing a suitable animal with a selected immunogen composition, e.g., a purified or partially purified BBS3 protein, polypeptide, peptide or domain, be it

a wild-type or mutant composition. The immunizing composition is administered in a manner effective to stimulate antibody producing cells.

The methods for generating monoclonal antibodies (MAbs) generally begin along the same lines as those for preparing polyclonal antibodies. Rodents such as mice and rats are preferred animals, however, the use of rabbit, sheep or frog cells is also possible. The use of rats may provide certain advantages (Goding, 1986, pp. 60-61), but mice are preferred, with the BALB/c mouse being most preferred, as this is most routinely used and generally gives a higher percentage of stable fusions.

The animals are injected with antigen, generally as described above. The antigen may be coupled to carrier molecules such as keyhole limpet hemocyanin if necessary. The antigen would typically be mixed with adjuvant, such as Freund's complete or incomplete adjuvant. Booster injections with the same antigen would occur at approximately two-week intervals.

Following immunization, somatic cells with the potential for producing antibodies, specifically B lymphocytes (B cells), are selected for use in the MAb generating protocol. These cells may be obtained from biopsied spleens, tonsils or lymph nodes, or from a peripheral blood sample. Spleen cells and peripheral blood cells are preferred, the former because they are a rich source of antibody-producing cells that are in the dividing plasmablast stage, and the latter because peripheral blood is easily accessible.

Often, a panel of animals will have been immunized and the spleen of an animal with the highest antibody titer will be removed and the spleen lymphocytes obtained by homogenizing the spleen with a syringe. Typically, a spleen from an immunized mouse contains approximately 5×10^7 to 2×10^8 lymphocytes.

The antibody-producing B lymphocytes from the immunized animal are then fused with cells of an immortal myeloma cell, generally one of the same species as the animal that was immunized. Myeloma cell lines suited for use in hybridoma-producing fusion procedures preferably are non-antibody-producing, have high fusion efficiency, and enzyme deficiencies that render them incapable of growing in certain selective media which support the growth of only the desired fused cells (hybridomas).

Any one of a number of myeloma cells may be used, as are known to those of skill in the art (Goding, pp. 65-66, 1986; Campbell, 1984). For example, where the immunized animal is a mouse, one may use P3-X63/Ag8, X63-Ag8.653, NS1/1.Ag 4 1, Sp210-Ag14, FO, NSO/U, MPC-11, MPC11-X45-GTG 1.7 and S194/5XX0 Bul; for rats, one may use R210.RCY3, Y3-Ag 1.2.3, IR983F and 4B210; and U-266, GM1500-GRG2, LICR-LON-HMy2 and UC729-6 are all useful in connection with human cell fusions.

One preferred murine myeloma cell is the NS-1 myeloma cell line (also termed P3-NS-1-Ag4-1), which is readily available from the NIGMS Human Genetic Mutant Cell Repository by requesting cell line repository number GM3573. Another mouse myeloma cell line that may be used is the 8-azaguanine-resistant mouse myeloma SP2/0 non-producer cell line.

Methods for generating hybrids of antibody-producing spleen or lymph node cells and myeloma cells usually comprise mixing somatic cells with myeloma cells in a 2:1 proportion, though the proportion may vary from about 20:1 to about 1:1, respectively, in the presence of an agent or agents (chemical or electrical) that promote the fusion of cell membranes. Fusion methods using Sendai virus have been described by Kohler and Milstein (1975; 1976), and those

using polyethylene glycol (PEG), such as 37% (v/v) PEG, by Geffter et al. (1977). The use of electrically induced fusion methods is also appropriate (Goding pp. 71-74, 1986).

Fusion procedures usually produce viable hybrids at low frequencies, about 1×10^{-6} to 1×10^{-8} . However, this does not pose a problem, as the viable, fused hybrids are differentiated from the parental, unfused cells (particularly the unfused myeloma cells that would normally continue to divide indefinitely) by culturing in a selective medium. The selective medium is generally one that contains an agent that blocks the de novo synthesis of nucleotides in the tissue culture media. Exemplary and preferred agents are aminopterin, methotrexate, and azaserine. Aminopterin and methotrexate block de novo synthesis of both purines and pyrimidines, whereas azaserine blocks only purine synthesis. Where aminopterin or methotrexate is used, the media is supplemented with hypoxanthine and thymidine as a source of nucleotides (HAT medium). Where azaserine is used, the media is supplemented with hypoxanthine.

The preferred selection medium is HAT. Only cells capable of operating nucleotide salvage pathways are able to survive in HAT medium. The myeloma cells are defective in key enzymes of the salvage pathway, e.g., hypoxanthine phosphoribosyl transferase (HPRT), and they cannot survive. The B cells can operate this pathway, but they have a limited life span in culture and generally die within about two weeks. Therefore, the only cells that can survive in the selective media are those hybrids formed from myeloma and B cells.

This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants (after about two to three weeks) for the desired reactivity. The assay should be sensitive, simple and rapid, such as radioimmunoassays, enzyme immunoassays, cytotoxicity assays, plaque assays, dot immunobinding assays, and the like.

The selected hybridomas would then be serially diluted and cloned into individual antibody-producing cell lines, which clones can then be propagated indefinitely to provide MAbs. The cell lines may be exploited for MAb production in two basic ways. First, a sample of the hybridoma can be injected (often into the peritoneal cavity) into a histocompatible animal of the type that was used to provide the somatic and myeloma cells for the original fusion (e.g., a syngeneic mouse). Optionally, the animals are primed with a hydrocarbon, especially oils such as pristane (tetramethylpentadecane) prior to injection. The injected animal develops tumors secreting the specific monoclonal antibody produced by the fused cell hybrid. The body fluids of the animal, such as serum or ascites fluid, can then be tapped to provide MAbs in high concentration. Second, the individual cell lines could be cultured in vitro, where the MAbs are naturally secreted into the culture medium from which they can be readily obtained in high concentrations.

MAbs produced by either means may be further purified, if desired, using filtration, centrifugation and various chromatographic methods such as HPLC or affinity chromatography. Fragments of the monoclonal antibodies of the invention can be obtained from the monoclonal antibodies so produced by methods, which include digestion with enzymes, such as pepsin or papain, and/or by cleavage of disulfide bonds by chemical reduction. Alternatively, monoclonal antibody fragments encompassed by the present invention can be synthesized using an automated peptide synthesizer.

It is also contemplated that a molecular cloning approach may be used to generate monoclonals. For this, combinatorial immunoglobulin phagemid libraries are prepared from RNA isolated from the spleen of the immunized animal, and phagemids expressing appropriate antibodies are selected by panning using cells expressing the antigen and control cells. The advantages of this approach over conventional hybridoma techniques are that approximately 10^4 times as many antibodies can be produced and screened in a single round, and that new specificities are generated by H and L chain combination which further increases the chance of finding appropriate antibodies.

Alternatively, monoclonal antibody fragments encompassed by the present invention can be synthesized using an automated peptide synthesizer, or by expression of full-length gene or of gene fragments in *E. coli*.

H. Antibody Conjugates

The present invention further provides antibodies against human ADP-ribosylation factor-like 6 polypeptide, generally of the monoclonal type, that are linked to one or more other agents to form an antibody conjugate. Any antibody of sufficient selectivity, specificity and affinity may be employed as the basis for an antibody conjugate. Such properties may be evaluated using conventional immunological screening methodology known to those of skill in the art.

Certain examples of antibody conjugates are those conjugates in which the antibody is linked to a detectable label. "Detectable labels" are compounds or elements that can be detected due to their specific functional properties, or chemical characteristics, the use of which allows the antibody to which they are attached to be detected, and further quantified if desired. Another such example is the formation of a conjugate comprising an antibody linked to a cytotoxic or anti-cellular agent, as may be termed "immunotoxins" (described in U.S. Pat. Nos. 5,686,072, 5,578,706, 4,792,447, 5,045,451, 4,664,911 and 5,767,072, each incorporated herein by reference).

Antibody conjugates are thus preferred for use as diagnostic agents. Antibody diagnostics generally fall within two classes, those for use in *in vitro* diagnostics, such as in a variety of immunoassays, and those for use *in vivo* diagnostic protocols, generally known as "antibody-directed imaging." Again, antibody-directed imaging is less preferred for use with this invention.

Many appropriate imaging agents are known in the art, as are methods for their attachment to antibodies (see, e.g., U.S. Pat. Nos. 5,021,236 and 4,472,509, both incorporated herein by reference). Certain attachment methods involve the use of a metal chelate complex employing, for example, an organic chelating agent such as DTPA attached to the antibody (U.S. Pat. No. 4,472,509). Monoclonal antibodies may also be reacted with an enzyme in the presence of a coupling agent such as glutaraldehyde or periodate. Conjugates with fluorescein markers are prepared in the presence of these coupling agents or by reaction with an isothiocyanate.

In the case of paramagnetic ions, one might mention by way of example ions such as chromium (III), manganese (II), iron (III), iron (II), cobalt (II), nickel (II), copper (II), neodymium (III), samarium (III), ytterbium (III), gadolinium (III), vanadium (II), terbium (III), dysprosium (III), holmium (III) and erbium (III), with gadolinium being particularly preferred. Ions useful in other contexts, such as X-ray imaging, include but are not limited to lanthanum (III), gold (III), lead (II), and especially bismuth (III).

In the case of radioactive isotopes for therapeutic and/or diagnostic application, one might mention astatine²¹¹, ¹⁴carbon, ⁵¹chromium, ³⁶chlorine, ⁵⁷cobalt, ⁵⁸cobalt, copper⁶⁷, ¹⁵²Eu, gallium⁶⁷, ³hydrogen, iodine¹²³, iodine¹²⁵, iodine¹³¹, indium¹¹¹, ⁵⁹iron, ³²phosphorus, rhenium¹⁸⁶, rhenium¹⁸⁸, ⁷⁵selenium, ³⁵sulphur, technetium^{99m} and yttrium⁹⁰. ¹²⁵I is often being preferred for use in certain embodiments, and technetium^{99m} and indium¹¹¹ are also often preferred due to their low energy and suitability for long range detection.

Radioactively labeled monoclonal antibodies of the present invention may be produced according to well-known methods in the art. For instance, monoclonal antibodies can be iodinated by contact with sodium or potassium iodide and a chemical oxidizing agent such as sodium hypochlorite, or an enzymatic oxidizing agent, such as lactoperoxidase. Monoclonal antibodies according to the invention may be labeled with technetium^{99m} by ligand exchange process, for example, by reducing pertechnetate with stannous solution, chelating the reduced technetium onto a Sephadex column and applying the antibody to this column or by direct labeling techniques, e.g., by incubating pertechnetate, a reducing agent such as SNCl_2 , a buffer solution such as sodium-potassium phthalate solution, and the antibody. Intermediary functional groups which are often used to bind radioisotopes which exist as metallic ions to antibody are diethylenetriaminepentaacetic acid (DTPA) and ethylene diaminetetraacetic acid (EDTA). Also contemplated for use are fluorescent labels, including rhodamine, fluorescein isothiocyanate and renographin.

The much preferred antibody conjugates of the present invention are those intended primarily for use *in vitro*, where the antibody is linked to a secondary binding ligand or to an enzyme (an enzyme tag) that will generate a colored product upon contact with a chromogenic substrate. Examples of suitable enzymes include urease, alkaline phosphatase, (horseradish) hydrogen peroxidase and glucose oxidase. Preferred secondary binding ligands are biotin and avidin or streptavidin compounds. The use of such labels is well known to those of skill in the art and is described, for example, in U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241; each incorporated herein by reference.

2. Nucleic Acids

Important aspects of the present invention concern isolated DNA segments and recombinant vectors encoding human ADP-ribosylation factor-like 6 polypeptides or peptides, and the creation and use of recombinant host cells through the application of DNA technology, that express a wild-type, polymorphic or mutant human ADP-ribosylation factor-like 6, using the sequence of SEQ ID NO:1 or 3, and biologically functional equivalents thereof.

The present invention concerns DNA segments, isolatable from mammalian cells, such as mouse, rat or human cells, that are free from total genomic DNA and that are capable of expressing a protein, polypeptide or peptide. As used herein, the term "DNA segment" refers to a DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding human ADP-ribosylation factor-like 6 refers to a DNA segment that contains wild-type, polymorphic or mutant human ADP-ribosylation factor-like 6 coding sequences yet is isolated away from, or purified free from, total mammalian genomic DNA. Included within the term "DNA segment" are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phage, viruses, and the like.

Similarly, a DNA segment comprising an isolated or purified human ADP-ribosylation factor-like 6 gene refers to a DNA segment encoding human ADP-ribosylation factor-like 6 polypeptide or peptide coding sequences and, in certain aspects, regulatory sequences, isolated substantially away from other naturally-occurring genes or protein encoding sequences. In this respect, the term “gene” is used for simplicity to refer to a functional protein, polypeptide or peptide encoding unit. As will be understood by those in the art, this functional term includes both genomic sequences, cDNA sequences and engineered segments that express, or may be adapted to express, proteins, polypeptides, domains, peptides, fusion proteins and mutants of human ADP-ribosylation factor-like 6 sequences.

“Isolated substantially away from other coding sequences” means that the gene of interest, in this case the human ADP-ribosylation factor-like 6 gene, forms the significant part of the coding region of the DNA segment, and that the DNA segment does not contain large portions of naturally-occurring coding DNA, such as large chromosomal fragments or other functional genes or cDNA coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

A. Variants

In particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating DNA sequences that encode a human ADP-ribosylation factor-like 6 protein, polypeptide or peptide that includes within its amino acid sequence a contiguous amino acid sequence in accordance with, or essentially as set forth in, SEQ ID NO:2 or 4, corresponding to the human ADP-ribosylation factor-like 6.

The term “a sequence essentially as set forth in SEQ ID NO:2 or 4” means that the sequence substantially corresponds to a portion of SEQ ID NO:2 or 4 and has relatively few amino acids that are not identical to, or a biologically functional equivalent of, the amino acids of SEQ ID NO:2 or 4.

The term “biologically functional equivalent” is well understood in the art and is further defined in detail herein. Accordingly, sequences that have about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99%, and any range derivable therein, such as, for example, about 70% to about 80%, and more preferably about 81% and about 90%; or even more preferably, between about 91% and about 99%; of amino acids that are identical or functionally equivalent to the amino acids of SEQ ID NO:2 or 4 will be sequences that are “essentially as set forth in SEQ ID NO:2 or 4,” provided the biological activity of the protein is maintained. In particular embodiments, the biological activity of a human ADP-ribosylation factor-like 6 protein, polypeptide or peptide, or a biologically functional equivalent, comprises binding to one or more proteases, particularly serine proteases. In specific embodiments, the biological activity of a human ADP-ribosylation factor-like 6 protein, polypeptide or peptide, or a biologically functional equivalent, comprises inhibition of the activity of one or more proteases, particularly serine proteases, through binding. A preferred protease activity that may be inhibited by a human ADP-ribosylation factor-like 6 protein, polypeptide or peptide, or a biologi-

cally functional equivalent, is inhibition of the ability or rate of proteolytic cleavage catalyzed by the protease.

In certain other embodiments, the invention concerns isolated DNA segments and recombinant vectors that include within their sequence a nucleic acid sequence essentially as set forth in SEQ ID NO:1 or 3. The term “essentially as set forth in SEQ ID NO:1 or 3” is used in the same sense as described above and means that the nucleic acid sequence substantially corresponds to a portion of SEQ ID NO:1 or 3 and has relatively few codons that are not identical, or functionally equivalent, to the codons of SEQ ID NO:1 or 3.

The term “functionally equivalent codon” is used herein to refer to codons that encode the same amino acid, such as the six codons for arginine and serine, and also refers to codons that encode biologically equivalent amino acids. For optimization of expression of human ADP-ribosylation factor-like 6 in human cells, the codons are shown in Table 1 in preference of use from left to right. Thus, the most preferred codon for alanine is thus “GCC”, and the least is “GCG” (see Table 1 below). Codon usage for various organisms and organelles can be found at the website www.kazusa.or.jp/codon/, incorporated herein by reference, allowing one of skill in the art to optimize codon usage for expression in various organisms using the disclosures herein. Thus, it is contemplated that codon usage may be optimized for other animals, as well as other organisms such as a prokaryote (e.g., an eubacteria, an archaea), an eukaryote (e.g., a protist, a plant, a fungi, an animal), a virus and the like, as well as organelles that contain nucleic acids, such as mitochondria or chloroplasts, based on the preferred codon usage as would be known to those of ordinary skill in the art.

TABLE 1

Preferred Human DNA Codons									
Amino Acids			Codons						
Alanine	Ala	A	GCC	GCT	GCA	GCG			
Cysteine	Cys	C	TGC	TGT					
Aspartic acid	Asp	D	GAC	GAT					
Glutamic acid	Glu	E	GAG	GAA					
Phenylalanine	Phe	F	TTC	TTT					
Glycine	Gly	G	GGC	GGG	GGA	GGT			
Histidine	His	H	CAC	CAT					
Isoleucine	Ile	I	ATC	ATT	ATA				
Lysine	Lys	K	AAG	AAA					
Leucine	Leu	L	CTG	CTC	TTG	CTT	CTA	TTA	
Methionine	Met	M	ATG						
Asparagine	Asn	N	AAC	AAT					
Proline	Pro	P	CCC	CCT	CCA	CCG			
Glutamine	Gln	Q	CAG	CAA					
Arginine	Arg	R	CGC	AGG	CGG	AGA	CGA	CGT	
Serine	Ser	S	AGC	TCC	TCT	AGT	TCA	TCG	
Threonine	Thr	T	ACC	ACA	ACT	ACG			
Valine	Val	V	GTG	GTC	GTT	GTA			

TABLE 1-continued

<u>Preferred Human DNA Codons</u>			
Amino Acids		Codons	
Tryptophan	Trp	W	TGG
Tyrosine	Tyr	Y	TAC TAT

It will also be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5' or 3' sequences, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth above, including the maintenance of biological protein, polypeptide or peptide activity where an amino acid sequence expression is concerned. The addition of terminal sequences particularly applies to nucleic acid sequences that may, for example, include various non-coding sequences flanking either of the 5' or 3' portions of the coding region or may include various internal sequences, i.e., introns, which are known to occur within genes.

Excepting intronic or flanking regions, and allowing for the degeneracy of the genetic code, sequences that have about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99%, and any range derivable therein, such as, for example, about 70% to about 80%, and more preferably about 81% and about 90%; or even more preferably, between about 91% and about 99%; of nucleotides that are identical to the nucleotides of SEQ ID NO:1 or 3 will be sequences that are "essentially as set forth in SEQ ID NO:1 or 3."

B. Nucleic Acid Hybridization

The nucleic acid sequences disclosed herein also have a variety of uses, such as for example, utility as probes or primers in nucleic acid hybridization embodiments. Naturally, the present invention also encompasses DNA segments that are complementary, or essentially complementary, to the sequence set forth in SEQ ID NO:1 or 3. Nucleic acid sequences that are "complementary" are those that are capable of base-pairing according to the standard Watson-Crick complementarity rules. As used herein, the term "complementary sequences" means nucleic acid sequences that are substantially complementary, as may be assessed by the same nucleotide comparison set forth above, or as defined as being capable of hybridizing to the nucleic acid segment of SEQ ID NO:1 or 3 under stringent conditions such as those described herein.

As used herein, "hybridization", "hybridizes" or "capable of hybridizing" is understood to mean the forming of a double or triple stranded molecule or a molecule with partial double or triple stranded nature. The term "hybridization", "hybridize(s)" or "capable of hybridizing" encompasses the terms "stringent condition(s)" or "high stringency" and the terms "low stringency" or "low stringency condition(s)."

As used herein "stringent condition(s)" or "high stringency" are those conditions that allow hybridization between or within one or more nucleic acid strand(s) containing complementary sequence(s), but precludes hybridization of random sequences. Stringent conditions tolerate little, if any, mismatch between a nucleic acid and a target

strand. Such conditions are well known to those of ordinary skill in the art, and are preferred for applications requiring high selectivity. Non-limiting applications include isolating a nucleic acid, such as a gene or a nucleic acid segment thereof, or detecting at least one specific mRNA transcript or a nucleic acid segment thereof, and the like.

Stringent conditions may comprise low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.15 M NaCl at temperatures of about 50° C. to about 70° C. It is understood that the temperature and ionic strength of a desired stringency are determined in part by the length of the particular nucleic acid(s), the length and nucleobase content of the target sequence(s), the charge composition of the nucleic acid(s), and to the presence or concentration of formamide, tetramethylammonium chloride or other solvent(s) in a hybridization mixture.

It is also understood that these ranges, compositions and conditions for hybridization are mentioned by way of non-limiting examples only, and that the desired stringency for a particular hybridization reaction is often determined empirically by comparison to one or more positive or negative controls. Depending on the application envisioned it is preferred to employ varying conditions of hybridization to achieve varying degrees of selectivity of a nucleic acid towards a target sequence. In a non-limiting example, identification or isolation of a related target nucleic acid that does not hybridize to a nucleic acid under stringent conditions may be achieved by hybridization at low temperature and/or high ionic strength. For example, a medium stringency condition could be provided by about 0.1 to 0.25 M NaCl at temperatures of about 37° C. to about 55° C. Under these conditions, hybridization may occur even though the sequences of probe and target strand are not perfectly complementary, but are mismatched at one or more positions. In another example, a low stringency condition could be provided by about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20° C. to about 55° C. Of course, it is within the skill of one in the art to further modify the low or high stringency conditions to suite a particular application. For example, in other embodiments, hybridization may be achieved under conditions of, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 1.0 mM dithiothreitol, at temperatures between approximately 20° C. to about 37° C. Other hybridization conditions utilized could include approximately 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, at temperatures ranging from approximately 40° C. to about 72° C.

Accordingly, the nucleotide sequences of the disclosure may be used for their ability to selectively form duplex molecules with complementary stretches of genes or RNAs or to provide primers for amplification of DNA or RNA from tissues. Depending on the application envisioned, it is preferred to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence.

The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, enhancers, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

For example, nucleic acid fragments may be prepared that include a contiguous stretch of nucleotides identical to or

complementary to SEQ ID NO:1 or 3, such as, for example, about 8, about 10 to about 14, or about 15 to about 20 nucleotides, and that are chromosome sized pieces, up to about 1,000,000, about 750,000, about 500,000, about 250,000, about 100,000, about 50,000, about 20,000, or about 10,000, or about 5,000 base pairs in length, with segments of about 3,000 being preferred in certain cases, as well as DNA segments with total lengths of about 1,000, about 500, about 200, about 100 and about 50 base pairs in length (including all intermediate lengths of these lengths listed above, i.e., any range derivable therein and any integer derivable therein such a range) are also contemplated to be useful.

For example, it will be readily understood that "intermediate lengths", in these contexts, means any length between the quoted ranges, such as 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 130, 140, 150, 160, 170, 180, 190, including all integers through the 200-500; 500-1,000; 1,000-2,000; 2,000-3,000; 3,000-5,000; 5,000-10,000 ranges, up to and including sequences of about 12,001, 12,002, 13,001, 13,002, 15,000, 20,000 and the like.

Various nucleic acid segments may be designed based on a particular nucleic acid sequence, and may be of any length. By assigning numeric values to a sequence, for example, the first residue is 1, the second residue is 2, etc., an algorithm defining all nucleic acid segments can be created:

$$n \text{ to } n+y$$

where n is an integer from 1 to the last number of the sequence and y is the length of the nucleic acid (SEQ ID NO:1 or 3) segment minus one, where $n+y$ does not exceed the last number of the sequence. Thus, for a 10-mer, the nucleic acid segments correspond to bases 1 to 10, 2 to 11, 3 to 12 . . . and/or so on. For a 15-mer, the nucleic acid segments correspond to bases 1 to 15, 2 to 16, 3 to 17 . . . and/or so on. For a 20-mer, the nucleic segments correspond to bases 1 to 20, 2 to 21, 3 to 22 . . . and/or so on. In certain embodiments, the nucleic acid segment may be a probe or primer. As used herein, a "probe" generally refers to a nucleic acid used in a detection method or composition. As used herein, a "primer" generally refers to a nucleic acid used in an extension or amplification method or composition.

The use of a hybridization probe of between 17 and 100 nucleotides in length, or in some aspect of the invention even up to 1-2 Kb or more in length, allows the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over stretches greater than 20 bases in length are generally preferred, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of particular hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having stretches of 20 to 30 nucleotides, or even longer where desired. Such fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means or by introducing selected sequences into recombinant vectors for recombinant production.

In general, it is envisioned that the hybridization probes described herein will be useful both as reagents in solution hybridization, as in PCRTM, for detection of expression of corresponding genes, as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the

test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to hybridization with selected probes under desired conditions. The selected conditions will depend on the particular circumstances based on the particular criteria required (depending, for example, on the "G+C" content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, etc.). Following washing of the hybridized surface to remove non-specifically bound probe molecules, hybridization is detected, or even quantified, by means of the label.

C. Nucleic Acid Amplification

Nucleic acid used as a template for amplification is isolated from cells contained in the biological sample, according to standard methodologies (Sambrook et al., 1989). The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to convert the RNA to a complementary DNA. In one embodiment, the RNA is whole cell RNA and is used directly as the template for amplification.

Pairs of primers that selectively hybridize to nucleic acids corresponding to BBS3 genes are contacted with the isolated nucleic acid under conditions that permit selective hybridization. The term "primer," as defined herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty or thirty base pairs in length, but longer sequences can be employed. Primers may be provided in double-stranded or single-stranded form, although the single-stranded form is preferred.

Once hybridized, the nucleic acid:primer complex is contacted with one or more enzymes that facilitate template-dependent nucleic acid synthesis. Multiple rounds of amplification, also referred to as "cycles," are conducted until a sufficient amount of amplification product is produced.

Next, the amplification product is detected. In certain applications, the detection may be performed by visual means. Alternatively, the detection may involve indirect identification of the product via chemiluminescence, radioactive scintigraphy of incorporated radiolabel or fluorescent label or even via a system using electrical or thermal impulse signals (Affymax technology).

A number of template dependent processes are available to amplify the marker sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (referred to as PCRTM) which is described in detail in U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159, each incorporated herein by reference in entirety.

Briefly, in PCRTM, two primer sequences are prepared that are complementary to regions on opposite complementary strands of the marker sequence. An excess of deoxynucleoside triphosphates are added to a reaction mixture along with a DNA polymerase, e.g., Taq polymerase. If the marker sequence is present in a sample, the primers will bind to the marker and the polymerase will cause the primers to be extended along the marker sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the marker to form reaction products, excess primers will bind to the marker and to the reaction products and the process is repeated.

A reverse transcriptase PCR amplification procedure may be performed in order to quantify the amount of mRNA amplified. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook et al., 1989.

Alternative methods for reverse transcription utilize thermostable, RNA-dependent DNA polymerases. These methods are described in WO 90/07641, filed Dec. 21, 1990, incorporated herein by reference. Polymerase chain reaction methodologies are well known in the art.

Another method for amplification is the ligase chain reaction ("LCR"), disclosed in EPA No. 320 308, incorporated herein by reference in its entirety. In LCR, two complementary probe pairs are prepared, and in the presence of the target sequence, each pair will bind to opposite complementary strands of the target such that they abut. In the presence of a ligase, the two probe pairs will link to form a single unit. By temperature cycling, as in PCR™, bound ligated units dissociate from the target and then serve as "target sequences" for ligation of excess probe pairs. U.S. Pat. No. 4,883,750 describes a method similar to LCR for binding probe pairs to a target sequence.

Qbeta Replicase, described in PCT Application No. PCT/US87/00880, incorporated herein by reference, may also be used as still another amplification method in the present invention. In this method, a replicative sequence of RNA that has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence that can then be detected.

An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[alpha-thio]-triphosphates in one strand of a restriction site may also be useful in the amplification of nucleic acids in the present invention.

Strand Displacement Amplification (SDA) is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, i.e., nick translation. A similar method, called Repair Chain Reaction (RCR), involves annealing several probes throughout a region targeted for amplification, followed by a repair reaction in which only two of the four bases are present. The other two bases can be added as biotinylated derivatives for easy detection. A similar approach is used in SDA. Target specific sequences can also be detected using a cyclic probe reaction (CPR). In CPR, a probe having 3' and 5' sequences of non-specific DNA and a middle sequence of specific RNA is hybridized to DNA that is present in a sample. Upon hybridization, the reaction is treated with RNase H, and the products of the probe identified as distinctive products that are released after digestion. The original template is annealed to another cycling probe and the reaction is repeated.

Still another amplification methods described in GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety, may be used in accordance with the present invention. In the former application, "modified" primers are used in a PCR-like, template- and enzyme-dependent synthesis. The primers may be modified by labeling with a capture moiety (e.g., biotin) and/or a detector moiety (e.g., enzyme). In the latter application, an excess of labeled probes are added to a sample. In the presence of the target sequence, the probe binds and is cleaved catalytically. After cleavage, the target sequence is released intact to be bound by excess probe. Cleavage of the labeled probe signals the presence of the target sequence.

Other nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR (Gingeras et al., PCT Application WO 88/10315, incor-

porated herein by reference). In NASBA, the nucleic acids can be prepared for amplification by standard phenol/chloroform extraction, heat denaturation of a clinical sample, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer which has target specific sequences. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target specific primer, followed by polymerization. The double-stranded DNA molecules are then multiply transcribed by an RNA polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNA's are reverse transcribed into single stranded DNA, which is then converted to double stranded DNA, and then transcribed once again with an RNA polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target specific sequences.

Davey et al., EP 329 822 (incorporated herein by reference in its entirety) disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention. The ssRNA is a template for a first primer oligonucleotide, which is elongated by reverse transcriptase (RNA-dependent DNA polymerase). The RNA is then removed from the resulting DNA:RNA duplex by the action of ribonuclease H (RNase H, an RNase specific for RNA in duplex with either DNA or RNA). The resultant ssDNA is a template for a second primer, which also includes the sequences of an RNA polymerase promoter (exemplified by T7 RNA polymerase) 5' to its homology to the template. This primer is then extended by DNA polymerase (exemplified by the large "Klenow" fragment of *E. coli* DNA polymerase I), resulting in a double-stranded DNA ("dsDNA") molecule, having a sequence identical to that of the original RNA between the primers and having additionally, at one end, a promoter sequence. This promoter sequence can be used by the appropriate RNA polymerase to make many RNA copies of the DNA. These copies can then re-enter the cycle leading to very swift amplification. With proper choice of enzymes, this amplification can be done isothermally without addition of enzymes at each cycle. Because of the cyclical nature of this process, the starting sequence can be chosen to be in the form of either DNA or RNA.

Miller et al., PCT Application WO 89/06700 (incorporated herein by reference in its entirety) disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic, i.e., new templates are not produced from the resultant RNA transcripts. Other amplification methods include "RACE" and "one-sided PCR" (Frohman, 1990, incorporated herein by reference).

Methods based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby amplifying the di-oligonucleotide, may also be used in the amplification step of the present invention.

D. Nucleic Acid Detection

In certain embodiments, it will be advantageous to employ nucleic acid sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate

indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of being detected. In preferred embodiments, one may desire to employ a fluorescent label or an enzyme tag such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmentally undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known that can be employed to provide a detection means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

In embodiments wherein nucleic acids are amplified, it may be desirable to separate the amplification product from the template and the excess primer for the purpose of determining whether specific amplification has occurred. In one embodiment, amplification products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using standard methods (Sambrook et al., 1989).

Alternatively, chromatographic techniques may be employed to effect separation. There are many kinds of chromatography which may be used in the present invention: adsorption, partition, ion-exchange and molecular sieve, and many specialized techniques for using them including column, paper, thin-layer and gas chromatography.

Amplification products must be visualized in order to confirm amplification of the marker sequences. One typical visualization method involves staining of a gel with ethidium bromide and visualization under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the amplification products can then be exposed to x-ray film or visualized under the appropriate stimulating spectra, following separation.

In one embodiment, visualization is achieved indirectly. Following separation of amplification products, a labeled, nucleic acid probe is brought into contact with the amplified marker sequence. The probe preferably is conjugated to a chromophore but may be radiolabeled. In another embodiment, the probe is conjugated to a binding partner, such as an antibody or biotin, and the other member of the binding pair carries a detectable moiety.

In one embodiment, detection is by Southern blotting and hybridization with a labeled probe. The techniques involved in Southern blotting are well known to those of skill in the art and can be found in many standard books on molecular protocols. See Sambrook et al., 1989. Briefly, amplification products are separated by gel electrophoresis. The gel is then contacted with a membrane, such as nitrocellulose, permitting transfer of the nucleic acid and non-covalent binding. Subsequently, the membrane is incubated with a chromophore-conjugated probe that is capable of hybridizing with a target amplification product. Detection is by exposure of the membrane to x-ray film or ion-emitting detection devices.

One example of the foregoing is described in U.S. Pat. No. 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and transfer of nucleic acids. The apparatus permits electrophoresis and blotting without external manipulation of the gel and is ideally suited to carrying out methods according to the present invention.

Other methods for genetic screening to accurately detect mutations in genomic DNA, cDNA or RNA samples may be employed, depending on the specific situation.

Historically, a number of different methods have been used to detect point mutations, including denaturing gradient

gel electrophoresis ("DGGE"), restriction enzyme polymorphism analysis, chemical and enzymatic cleavage methods, and others. The more common procedures currently in use include direct sequencing of target regions amplified by PCR™ (see above) and single-strand conformation polymorphism analysis ("SSCP").

Another method of screening for point mutations is based on RNase cleavage of base pair mismatches in RNA/DNA and RNA/RNA heteroduplexes. As used herein, the term "mismatch" is defined as a region of one or more unpaired or mispaired nucleotides in a double-stranded RNA/RNA, RNA/DNA or DNA/DNA molecule. This definition thus includes mismatches due to insertion/deletion mutations, as well as single and multiple base point mutations.

U.S. Pat. No. 4,946,773 describes an RNase A mismatch cleavage assay that involves annealing single-stranded DNA or RNA test samples to an RNA probe, and subsequent treatment of the nucleic acid duplexes with RNase A. After the RNase cleavage reaction, the RNase is inactivated by proteolytic digestion and organic extraction, and the cleavage products are denatured by heating and analyzed by electrophoresis on denaturing polyacrylamide gels. For the detection of mismatches, the single-stranded products of the RNase A treatment, electrophoretically separated according to size, are compared to similarly treated control duplexes. Samples containing smaller fragments (cleavage products) not seen in the control duplex are scored as positive.

Currently available RNase mismatch cleavage assays, including those performed according to U.S. Pat. No. 4,946,773, require the use of radiolabeled RNA probes. Myers and Maniatis in U.S. Pat. No. 4,946,773 describe the detection of base pair mismatches using RNase A. Other investigators have described the use of an *E. coli* enzyme, RNase I, in mismatch assays. Because it has broader cleavage specificity than RNase A, RNase I would be a desirable enzyme to employ in the detection of base pair mismatches if components can be found to decrease the extent of non-specific cleavage and increase the frequency of cleavage of mismatches. The use of RNase I for mismatch detection is described in literature from Promega Biotech. Promega markets a kit containing RNase I that is shown in their literature to cleave three out of four known mismatches, provided the enzyme level is sufficiently high.

The RNase protection assay was first used to detect and map the ends of specific mRNA targets in solution. The assay relies on being able to easily generate high specific activity radiolabeled RNA probes complementary to the mRNA of interest by in vitro transcription. Originally, the templates for in vitro transcription were recombinant plasmids containing bacteriophage promoters. The probes are mixed with total cellular RNA samples to permit hybridization to their complementary targets, then the mixture is treated with RNase to degrade excess unhybridized probe. Also, as originally intended, the RNase used is specific for single-stranded RNA, so that hybridized double-stranded probe is protected from degradation. After inactivation and removal of the RNase, the protected probe (which is proportional in amount to the amount of target mRNA that was present) is recovered and analyzed on a polyacrylamide gel.

The RNase Protection assay was adapted for detection of single base mutations. In this type of RNase A mismatch cleavage assay, radiolabeled RNA probes transcribed in vitro from wild-type sequences, are hybridized to complementary target regions derived from test samples. The test target generally comprises DNA (either genomic DNA or DNA amplified by cloning in plasmids or by PCR™), although RNA targets (endogenous mRNA) have occasion-

ally been used. If single nucleotide (or greater) sequence differences occur between the hybridized probe and target, the resulting disruption in Watson-Crick hydrogen bonding at that position ("mismatch") can be recognized and cleaved in some cases by single-strand specific ribonuclease. To date, RNase A has been used almost exclusively for cleavage of single-base mismatches, although RNase I has recently been shown as useful also for mismatch cleavage. There are recent descriptions of using the MutS protein and other DNA-repair enzymes for detection of single-base mismatches.

E. Cloning of Additional BBS3 Genes

The present invention contemplates cloning BBS3 genes or cDNAs from animal (e.g., mammalian) organisms. A technique often employed by those skilled in the art of protein production today is to obtain a so-called "recombinant" version of the protein, to express it in a recombinant cell and to obtain the protein, polypeptide or peptide from such cells. These techniques are based upon the "cloning" of a DNA molecule encoding the protein from a DNA library, i.e., on obtaining a specific DNA molecule distinct from other portions of DNA. This can be achieved by, for example, cloning a cDNA molecule, or cloning a genomic-like DNA molecule.

The first step in such cloning procedures is the screening of an appropriate DNA library. The screening protocol may utilize nucleotide segments or probes derived from SEQ ID NO:1 or 3. Additionally, antibodies designed to bind to the expressed BBS3 proteins, polypeptides, or peptides may be used as probes to screen an appropriate mammalian DNA expression library. Alternatively, activity assays may be employed. The operation of such screening protocols are well known to those of skill in the art and are described in detail in the scientific literature, for example, in Sambrook et al. (1989), incorporated herein by reference. Moreover, as the present invention encompasses the cloning of genomic segments as well as cDNA molecules, it is contemplated that suitable genomic cloning methods, as known to those in the art, may also be used.

As used herein "designed to hybridize" means a sequence selected for its likely ability to hybridize to a mammalian BBS3 gene, for example due to the expected high degree of homology between the human, rat, or mouse BBS3 gene and the BBS3 genes from other mammals. Also included are segments or probes altered to enhance their ability to hybridize to or bind to a mammalian BBS3 gene. Additionally, these regions of homology also include amino acid

sequences of 4 or more consecutive amino acids selected and/or altered to increase conservation of the amino acid sequences in comparison to the same or similar region of residues in the same or related genes in one or more species. Such amino acid sequences may be derived from amino acid sequences encoded by the BBS3 gene, and more particularly from the isolated sequences of SEQ ID NO:2 or 4.

Designing probe sequences may involve selection of regions of highly conserved nucleotide sequences between various species for a particular gene or related genes, relative to the general conservation of nucleotides of the gene or related genes in one or more species. Comparison of the amino acid sequences conserved between one or more species for a particular gene may also be used to determine a group of 4 or more consecutive amino acids that are conserved relative to the protein encoded by the gene or related genes. The nucleotide probe or primers may then be designed from the region of the gene that encodes the conserved sequence of amino acids.

One may also prepare fusion proteins, polypeptides and peptides, e.g., where the BBS3 proteinaceous material coding regions are aligned within the same expression unit with other proteins, polypeptides or peptides having desired functions, such as for purification or immunodetection purposes (e.g., proteinaceous compositions that may be purified by affinity chromatography and enzyme label coding regions, respectively).

Encompassed by the invention are DNA segments encoding relatively small peptides, such as, for example, peptides of from about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, about 31, about 32, about 33, about 34, about 35, about 35, about 40, about 45, to about 50 amino acids in length, and more preferably, of from about 15 to about 30 amino acids in length; as set forth in SEQ ID NO:2 or 4 and also larger polypeptides up to and including proteins corresponding to the full-length sequences set forth in SEQ ID NO:2 or 4, and any range derivable therein and any integer derivable therein such a range.

In addition to the "standard" DNA and RNA nucleotide bases, modified bases are also contemplated for use in particular applications of the present invention. A table of exemplary, but not limiting, modified bases is provided herein below.

TABLE 2

Modified Bases			
Abbr.	Modified base description	Abbr.	Modified base description
ac4c	4-acetylcytidine	Mam5s2u	5-methoxyaminomethyl-2-thiouridine
Chm5u	5-(carboxyhydroxymethyl)uridine	Man q	Beta,D-mannosylqueosine
Cm	2'-O-methylcytidine	Mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
Cmm5s2u	5-carboxymethylaminomethyl-2-thiouridine	Mcm5u	5-methoxycarbonylmethyluridine
Cmm5u	5-carboxymethylaminomethyluridine	Mo5u	5-methoxyuridine
D	Dihydrouridine	Ms216a	2-methylthio-N6-isopentenyladenosine
Fm	2'-O-methylpseudouridine	Ms216a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine
gal q	Beta,D-galactosylqueosine	Mt6a	N-((9-beta-D-ribofuranosyl)purine-6-yl) N-methyl-carbamoyl)threonine
Gm	2'-O-methylguanosine	Mv	Uridine-5-oxyacetic acid methylester
I	Inosine	o5u	Uridine-5-oxyacetic acid (v)
I6a	N6-isopentenyladenosine	Osyw	Wybutoxosine
m1a	1-methyladenosine	P	Pseudouridine
m1f	1-methylpseudouridine	Q	Queosine

TABLE 2-continued

<u>Modified Bases</u>			
Abbr.	Modified base description	Abbr.	Modified base description
m1g	1-methylguanosine	s2c	2-thiocytidine
m1I	1-methylinosine	s2t	5-methyl-2-thiouridine
m22g	2,2-dimethylguanosine	s2u	2-thiouridine
m2a	2-methyladenosine	s4u	4-thiouridine
m2g	2-methylguanosine	T	5-methyluridine
m3c	3-methylcytidine	t6a	N-((9-beta-D-ribofuranosyl)purine-6-yl)carbamoyl)threonine
m5c	5-methylcytidine	Tm	2'-O-methyl-5-methyluridine
m6a	N6-methyladenosine	Um	2'-O-methyluridine
m7g	7-methylguanosine	Yw	Wybutosine
Mam5u	5-methylaminomethyluridine	X	3-(3-amino-3-carboxypropyl)uridine, (acp3)u

F. Mutagenesis, Peptidomimetics and Rational Drug Design

It will also be understood that this invention is not limited to the particular nucleic acid and amino acid sequences of SEQ ID NO:2 or 4. Recombinant vectors and isolated DNA segments may therefore variously include these coding regions themselves, coding regions bearing selected alterations or modifications in the basic coding region, or they may encode larger polypeptides that nevertheless include such coding regions or may encode biologically functional equivalent proteins, polypeptides or peptides that have variant amino acids sequences.

The DNA segments of the present invention encompass biologically functional equivalent human ADP-ribosylation factor-like 6/BBS3 proteins, polypeptides, and peptides. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteinaceous compositions thus encoded. Alternatively, functionally equivalent proteins, polypeptides or peptides may be created via the application of recombinant DNA technology, in which changes in the protein, polypeptide or peptide structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes may be introduced, for example, through the application of site-directed mutagenesis techniques as discussed herein below, e.g., to introduce improvements to the antigenicity of the proteinaceous composition or to test mutants in order to examine human ADP-ribosylation factor-like 6/BBS3 activity at the molecular level.

Site-specific mutagenesis is a technique useful in the preparation of individual peptides, or biologically functional equivalent proteins, polypeptides or peptides, through specific mutagenesis of the underlying DNA. The technique further provides a ready ability to prepare and test sequence variants, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art. As will be appreciated, the technique

typically employs a bacteriophage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage vectors are commercially available and their use is generally well known to those skilled in the art. Double-stranded plasmids are also routinely employed in site directed mutagenesis, which eliminates the step of transferring the gene of interest from a phage to a plasmid.

In general, site-directed mutagenesis is performed by first obtaining a single-stranded vector, or melting of two strands of a double stranded vector, which includes within its sequence a DNA sequence encoding the desired proteinaceous molecule. An oligonucleotide primer bearing the desired mutated sequence is synthetically prepared. This primer is then annealed with the single-stranded DNA preparation, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected that include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected gene using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting, as there are other ways in which sequence variants of genes may be obtained. For example, recombinant vectors encoding the desired gene may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

As modifications and changes may be made in the structure of the BBS3 genes, nucleic acids (e.g., nucleic acid segments) and proteinaceous molecules of the present invention, and still obtain molecules having like or otherwise desirable characteristics, such biologically functional equivalents are also encompassed within the present invention.

For example, certain amino acids may be substituted for other amino acids in a proteinaceous structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies, binding sites on substrate molecules or receptors, or such like. Since it is the interactive capacity and nature of a proteinaceous molecule that defines that proteinaceous molecule's biological functional activity, certain amino acid sequence substitutions can be made in a proteinaceous molecule sequence (or, of course, its underlying DNA cod-

ing sequence) and nevertheless obtain a proteinaceous molecule with like (agonistic) properties. It is thus contemplated that various changes may be made in the sequence of human ADP-ribosylation factor-like 6/BBS3 proteins, polypeptides or peptides, or the underlying nucleic acids, without appreciable loss of their biological utility or activity.

Equally, the same considerations may be employed to create a protein, polypeptide or peptide with countervailing, e.g., antagonistic properties. This is relevant to the present invention in which human ADP-ribosylation factor-like 6/BBS3 mutants or analogues may be generated. For example, a human ADP-ribosylation factor-like 6/BBS3 mutant may be generated and tested for BBS3 activity to identify those residues important for human ADP-ribosylation factor-like 6/BBS3 activity. human ADP-ribosylation factor-like 6/BBS3 mutants may also be synthesized to reflect a human ADP-ribosylation factor-like 6/BBS3 mutant that occurs in the human population and that is linked to the development of cancer. Such mutant proteinaceous molecules are particularly contemplated for use in generating mutant-specific antibodies and such mutant DNA segments may be used as mutant-specific probes and primers.

While discussion has focused on functionally equivalent polypeptides arising from amino acid changes, it will be appreciated that these changes may be effected by alteration of the encoding DNA; taking into consideration also that the genetic code is degenerate and that two or more codons may code for the same amino acid. A table of amino acids and their codons is presented herein above for use in such embodiments, as well as for other uses, such as in the design of probes and primers and the like.

In terms of functional equivalents, it is well understood by the skilled artisan that, inherent in the definition of a "biologically functional equivalent" protein, polypeptide, peptide, gene or nucleic acid, is the concept that there is a limit to the number of changes that may be made within a defined portion of the molecule and still result in a molecule with an acceptable level of equivalent biological activity. Biologically functional equivalent peptides are thus defined herein as those peptides in which certain, not most or all, of the amino acids may be substituted.

In particular, where shorter length peptides are concerned, it is contemplated that fewer amino acids changes should be made within the given peptide. Longer domains may have an intermediate number of changes. The full length protein will have the most tolerance for a larger number of changes. Of course, a plurality of distinct proteins/polypeptide/peptides with different substitutions may easily be made and used in accordance with the invention.

It is also well understood that where certain residues are shown to be particularly important to the biological or structural properties of a protein, polypeptide or peptide, e.g., residues in binding regions or active sites, such residues may not generally be exchanged. In this manner, functional equivalents are defined herein as those peptides which maintain a substantial amount of their native biological activity.

Amino acid substitutions are generally based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. An analysis of the size, shape and type of the amino acid side-chain substituents reveals that arginine, lysine and histidine are all positively charged residues; that alanine, glycine and serine are all a similar size; and that phenylalanine, tryptophan and tyrosine all have a generally similar shape. Therefore, based upon these considerations, arginine, lysine and histidine; alanine, glycine and serine;

and phenylalanine, tryptophan and tyrosine; are defined herein as biologically functional equivalents.

To effect more quantitative changes, the hydrophatic index of amino acids may be considered. Each amino acid has been assigned a hydrophatic index on the basis of their hydrophobicity and charge characteristics, these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

The importance of the hydrophatic amino acid index in conferring interactive biological function on a proteinaceous molecule is generally understood in the art (Kyte & Doolittle, 1982, incorporated herein by reference). It is known that certain amino acids may be substituted for other amino acids having a similar hydrophatic index or score and still retain a similar biological activity. In making changes based upon the hydrophatic index, the substitution of amino acids whose hydrophatic indices are within ± 2 is preferred, those which are within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity, particularly where the biological functional equivalent protein, polypeptide or peptide thereby created is intended for use in immunological embodiments, as in certain embodiments of the present invention. U.S. Pat. No. 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a proteinaceous molecule, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigenicity, i.e., with a biological property of the proteinaceous molecule.

As detailed in U.S. Pat. No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 \pm 1); glutamate (+3.0 \pm 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 \pm 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

In making changes based upon similar hydrophilicity values, the substitution of amino acids whose hydrophilicity values are within ± 2 is preferred, those which are within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

In addition to the human ADP-ribosylation factor-like 6/BBS3 peptidyl compounds described herein, it is contemplated that other sterically similar compounds may be formulated to mimic the key portions of the peptide structure. Such compounds, which may be termed peptidomimetics, may be used in the same manner as the peptides of the invention and hence are also functional equivalents.

Certain mimetics that mimic elements of proteinaceous molecule's secondary structure are described in Johnson et al. (1993). The underlying rationale behind the use of peptide mimetics is that the peptide backbone of proteinaceous molecules exists chiefly to orientate amino acid side chains in such a way as to facilitate molecular interactions, such as those of antibody and antigen. A peptide mimetic is thus designed to permit molecular interactions similar to the natural molecule.

Some successful applications of the peptide mimetic concept have focused on mimetics of β -turns within pro-

teinaceous molecules, which are known to be highly antigenic. Likely β -turn structure within a polypeptide can be predicted by computer-based algorithms, as discussed herein. Once the component amino acids of the turn are determined, mimetics can be constructed to achieve a similar spatial orientation of the essential elements of the amino acid side chains.

The generation of further structural equivalents or mimetics may be achieved by the techniques of modeling and chemical design known to those of skill in the art. The art of receptor modeling is now well known, and by such methods a chemical that binds BBS3 can be designed and then synthesized. It will be understood that all such sterically designed constructs fall within the scope of the present invention.

In addition to the 20 "standard" amino acids provided through the genetic code, modified or unusual amino acids are also contemplated for use in the present invention. A table of exemplary, but not limiting, modified or unusual amino acids is provided herein below.

TABLE 3

Modified and Unusual Amino Acids			
Abbr.	Amino Acid	Abbr.	Amino Acid
Aad	2-Aminoadipic acid	EtAsn	N-Ethylasparagine
Baad	3-Aminoadipic acid	Hyl	Hydroxylysine
Bala	Beta-alanine, beta-Amino-propionic acid	aHyl	Allo-Hydroxylysine
Abu	2-Aminobutyric acid	3Hyp	3-Hydroxyproline
4Abu	4-Aminobutyric acid, piperidinic acid	4Hyp	4-Hydroxyproline
Acp	6-Aminocaproic acid	Ide	Isodesmosine
Ahe	2-Aminoheptanoic acid	alle	Allo-Isoleucine
Aib	2-Aminoisobutyric acid	MeGly	N-Methylglycine, sarcosine
Baib	3-Aminoisobutyric acid	Melle	N-Methylisoleucine
Apm	2-Aminopimelic acid	MeLys	6-N-Methyllysine
Dbu	2,4-Diaminobutyric acid	MeVal	N-Methylvaline
Des	Desmosine	Nva	Norvaline
Dpm	2,2'-Diaminopimelic acid	Nle	Norleucine
Dpr	2,3-Diaminopropionic acid	Orn	Ornithine
EtGly	N-Ethylglycine		

In one aspect, a compound may be designed by rational drug design to function as a BBS3 inhibitor or as a serine protease inhibitor. The goal of rational drug design is to produce structural analogs of biologically active compounds. By creating such analogs, it is possible to fashion drugs, which are more active or stable than the natural molecules, which have different susceptibility to alteration or which may affect the function of various other molecules. In one approach, one would generate a three-dimensional structure for the BBS3 protein of the invention or a fragment thereof. This could be accomplished by X-ray crystallography, computer modeling or by a combination of both approaches. An alternative approach, involves the random replacement of functional groups throughout the BBS3 protein, polypeptides or peptides, and the resulting effect on function determined.

It also is possible to isolate a BBS3 protein, polypeptide or peptide specific antibody, selected by a functional assay, and then solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of anti-idiotypic antibody would be expected to be an analog of the original antigen. The anti-idiotypic antibody could then be used to identify and

isolate peptides from banks of chemically- or biologically-produced peptides. Selected peptides would then serve as the pharmacore. Anti-idiotypes may be generated using the methods described herein for producing antibodies, using an antibody as the antigen.

Thus, one may design drugs which have enhanced and improved biological activity, for example, serine protease or tumor growth or metastasis inhibition, relative to a starting BBS3 proteinaceous sequences. By virtue of the ability to recombinantly produce sufficient amounts of the BBS3 proteins, polypeptides or peptides, crystallographic studies may be performed to determine the most likely sites for mutagenesis and chemical mimicry. In addition, knowledge of the chemical characteristics of these compounds permits computer employed predictions of structure-function relationships. Computer models of various polypeptide and peptide structures are also available in the literature or computer databases. In a non-limiting example, the Entrez database (www.ncbi.nlm.nih.gov/Entrez/) may be used by one of ordinary skill in the art to identify target sequences and regions for mutagenesis.

3. Diagnosing BBS and Related Conditions

As discussed above, the present inventors have determined that alterations in the BBS3 gene are associated with BBS. Therefore, BBS3 and the corresponding gene may be employed as a diagnostic or prognostic indicator of BBS in general, and of related disorders such as diabetes, hypertension, retinal degeneration, renal carcinoma, renal malformation, congenital heart defects, limb deformity, polydactyly, hypogonadism, mental retardation and obesity. More specifically, point mutations, deletions, insertions or regulatory perturbations relating to BBS3 will be identified. The present invention contemplates further the diagnosis of disease states by detecting changes in the levels of BBS3 expression.

A. Genetic Diagnosis

One embodiment of the instant invention comprises a method for detecting variation in the expression of BBS3. This may comprise determining the level of BBS3 expressed, or determining specific alterations in the expressed product. Obviously, this sort of assay has importance in the diagnosis of related BBS, but it also is relevant to other disease states such as diabetes, retinal degeneration, renal carcinoma (cancers), renal malformation, congenital heart defects, limb deformity, hypertension, polydactyly, hypogonadism, mental retardation and obesity.

The biological sample can be any tissue or fluid. Various embodiments include cells of the skin, muscle, fascia, brain, prostate, breast, endometrium, lung, head & neck, pancreas, small intestine, blood cells, liver, testes, ovaries, colon, rectum, skin, stomach, esophagus, spleen, lymph nodes, bone marrow or kidney. Other embodiments include fluid samples such as peripheral blood, lymph fluid, ascites, serous fluid, pleural effusion, sputum, cerebrospinal fluid, lacrimal fluid, stool urine or amniotic fluid.

Nucleic acids used are isolated from cells contained in the biological sample, according to standard methodologies (Sambrook et al., 1989). The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to convert the RNA to a complementary DNA (cDNA). In one embodiment, the RNA is whole cell RNA; in another, it is poly-A RNA. Normally, the nucleic acid is amplified.

Depending on the format, the specific nucleic acid of interest is identified in the sample directly using amplification or with a second, known nucleic acid following ampli-

fication. Next, the identified product is detected. In certain applications, the detection may be performed by visual means (e.g., ethidium bromide staining of a gel). Alternatively, the detection may involve indirect identification of the product via chemiluminescence, radioactive scintigraphy of radiolabel or fluorescent label or even via a system using electrical or thermal impulse signals (Affymax Technology; Bellus, 1994).

Following detection, one may compare the results seen in a given patient with a statistically significant reference group of normal patients and patients that have BBS or BBS-related pathologies. In this way, it is possible to correlate the amount or kind of BBS detected with various clinical states.

Various types of defects have been identified by the present inventors. Thus, "alterations" should be read as including deletions, insertions, point mutations and duplications. Point mutations result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those occurring in non-germline tissues. Germ-line tissue can occur in any tissue and are inherited. Mutations in and outside the coding region also may affect the amount of BBS3 produced, both by altering the transcription of the gene or in destabilizing or otherwise altering the processing of either the transcript (mRNA) or protein.

The following table provides a summary of the changes identified in the BBS3 protein:

TABLE 4

Protein change	Characterization
Arg122Stp	homozygous, large consanguineous family
Gly167Arg	homozygous, single proband

It is contemplated that other mutations in the BBS3 gene may be identified in accordance with the present invention by detecting a nucleotide change in particular nucleic acids (U.S. Pat. No. 4,988,617, incorporated herein by reference). A variety of different assays are contemplated in this regard, including but not limited to, fluorescent in situ hybridization (FISH; U.S. Pat. No. 5,633,365 and U.S. Pat. No. 5,665,549, each incorporated herein by reference), direct DNA sequencing, PFGE analysis, Southern or Northern blotting, single-stranded conformation analysis (SSCA), RNase protection assay, allele-specific oligonucleotide (ASO, e.g., U.S. Pat. No. 5,639,611), dot blot analysis, denaturing gradient gel electrophoresis (e.g., U.S. Pat. No. 5,190,856 incorporated herein by reference), RFLP (e.g., U.S. Pat. No. 5,324,631 incorporated herein by reference) and PCRTM-SSCP. Methods for detecting and quantitating gene sequences, such as mutated genes and oncogenes, in for example biological fluids are described in U.S. Pat. No. 5,496,699, incorporated herein by reference.

a. Primers and Probes

The term primer, as defined herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty base pairs in length, but longer sequences can be employed. Primers may be provided in double-stranded or single-stranded form, although the single-stranded form is preferred. Probes are defined differently, although they may act as primers. Probes, while perhaps capable of priming, are designed to binding to the target DNA or RNA and need not be used in an amplification process.

In preferred embodiments, the probes or primers are labeled with radioactive species (³²P, ¹⁴C, ³⁵S, ³H, or other label), with a fluorophore (rhodamine, fluorescein) or a chemiluminescent (luciferase).

b. Template Dependent Amplification Methods

A number of template dependent processes are available to amplify the marker sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (referred to as PCRTM) which is described in detail in U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159, and in Innis et al., 1990, each of which is incorporated herein by reference in its entirety.

Briefly, in PCRTM, two primer sequences are prepared that are complementary to regions on opposite complementary strands of the marker sequence. An excess of deoxynucleoside triphosphates are added to a reaction mixture along with a DNA polymerase, e.g., Taq polymerase. If the marker sequence is present in a sample, the primers will bind to the marker and the polymerase will cause the primers to be extended along the marker sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the marker to form reaction products, excess primers will bind to the marker and to the reaction products and the process is repeated.

A reverse transcriptase PCRTM amplification procedure may be performed in order to quantify the amount of mRNA amplified. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook et al., 1989. Alternative methods for reverse transcription utilize thermostable, RNA-dependent DNA polymerases. These methods are described in WO 90/07641 filed Dec. 21, 1990. Polymerase chain reaction methodologies are well known in the art.

Another method for amplification is the ligase chain reaction ("LCR" U.S. Pat. Nos. 5,494,810, 5,484,699, EP 320 308, each incorporated herein by reference). In LCR, two complementary probe pairs are prepared, and in the presence of the target sequence, each pair will bind to opposite complementary strands of the target such that they abut. In the presence of a ligase, the two probe pairs will link to form a single unit. By temperature cycling, as in PCRTM, bound ligated units dissociate from the target and then serve as "target sequences" for ligation of excess probe pairs. U.S. Pat. No. 4,883,750 describes a method similar to LCR for binding probe pairs to a target sequence.

Qbeta Replicase an RNA-directed RNA polymerase, also may be used as still another amplification method in the present invention. In this method, a replicative sequence of RNA that has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence that can then be detected. Similar methods also are described in U.S. Pat. No. 4,786,600, incorporated herein by reference, which concerns recombinant RNA molecules capable of serving as a template for the synthesis of complementary single-stranded molecules by RNA-directed RNA polymerase. The product molecules so formed also are capable of serving as a template for the synthesis of additional copies of the original recombinant RNA molecule.

An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[alpha-thio]-triphosphates in one strand of a restriction site also may be useful in the amplification of nucleic acids in the present invention (Walker et al., 1992; U.S. Pat. No. 5,270,184, incorporated herein by reference). U.S. Pat. No. 5,747,

255 (incorporated herein by reference) describes an isothermal amplification using cleavable oligonucleotides for polynucleotide detection. In the method described therein, separated populations of oligonucleotides are provided that contain complementary sequences to one another and that contain at least one scissile linkage which is cleaved whenever a perfectly matched duplex is formed containing the linkage. When a target polynucleotide contacts a first oligonucleotide cleavage occurs and a first fragment is produced which can hybridize with a second oligonucleotide. Upon such hybridization, the second oligonucleotide is cleaved releasing a second fragment that can, in turn, hybridize with a first oligonucleotide in a manner similar to that of the target polynucleotide.

Strand Displacement Amplification (SDA) is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, i.e., nick translation (e.g., U.S. Pat. Nos. 5,744,311; 5,733,752; 5,733,733; 5,712,124). A similar method, called Repair Chain Reaction (RCR), involves annealing several probes throughout a region targeted for amplification, followed by a repair reaction in which only two of the four bases are present. The other two bases can be added as biotinylated derivatives for easy detection. A similar approach is used in SDA. Target specific sequences can also be detected using a cyclic probe reaction (CPR). In CPR, a probe having 3' and 5' sequences of non-specific DNA and a middle sequence of specific RNA is hybridized to DNA that is present in a sample. Upon hybridization, the reaction is treated with RNase H, and the products of the probe identified as distinctive products that are released after digestion. The original template is annealed to another cycling probe and the reaction is repeated.

Still another amplification methods described in GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety, may be used in accordance with the present invention. In the former application, "modified" primers are used in a PCRTM-like, template- and enzyme-dependent synthesis. The primers may be modified by labeling with a capture moiety (e.g., biotin) and/or a detector moiety (e.g., enzyme). In the latter application, an excess of labeled probes are added to a sample. In the presence of the target sequence, the probe binds and is cleaved catalytically. After cleavage, the target sequence is released intact to be bound by excess probe. Cleavage of the labeled probe signals the presence of the target sequence.

Other nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR (Kwoh et al., 1989; Gingeras et al., PCT Application WO 88/10315, incorporated herein by reference in their entirety). In NASBA, the nucleic acids can be prepared for amplification by standard phenol/chloroform extraction, heat denaturation of a clinical sample, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer which has target specific sequences. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target specific primer, followed by polymerization. The double-stranded DNA molecules are then multiply transcribed by an RNA polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNA's are reverse transcribed into single stranded DNA, which is then

converted to double stranded DNA, and then transcribed once again with an RNA polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target specific sequences.

Davey et al., EP 329 822 (incorporated herein by reference in its entirety) disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention. The ssRNA is a template for a first primer oligonucleotide, which is elongated by reverse transcriptase (RNA-dependent DNA polymerase). The RNA is then removed from the resulting DNA:RNA duplex by the action of ribonuclease H (RNase H, an RNase specific for RNA in duplex with either DNA or RNA). The resultant ssDNA is a template for a second primer, which also includes the sequences of an RNA polymerase promoter (exemplified by T7 RNA polymerase) 5' to its homology to the template. This primer is then extended by DNA polymerase (exemplified by the large "Klenow" fragment of *E. coli* DNA polymerase I), resulting in a double-stranded DNA ("dsDNA") molecule, having a sequence identical to that of the original RNA between the primers and having additionally, at one end, a promoter sequence. This promoter sequence can be used by the appropriate RNA polymerase to make many RNA copies of the DNA. These copies can then re-enter the cycle leading to very swift amplification. With proper choice of enzymes, this amplification can be done isothermally without addition of enzymes at each cycle. Because of the cyclical nature of this process, the starting sequence can be chosen to be in the form of either DNA or RNA.

Miller et al., PCT Application WO 89/06700 (incorporated herein by reference in its entirety) disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic, i.e., new templates are not produced from the resultant RNA transcripts. Other amplification methods include "RACE" and "one-sided PCRTM" (Frohman, 1990; Ohara et al., 1989; each herein incorporated by reference in their entirety).

Methods based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby amplifying the di-oligonucleotide, also may be used in the amplification step of the present invention. Wu et al., (1989), incorporated herein by reference in its entirety.

C. Southern/Northern Blotting

Blotting techniques are well known to those of skill in the art. Southern blotting involves the use of DNA as a target, whereas Northern blotting involves the use of RNA as a target. Each provide different types of information, although cDNA blotting is analogous, in many aspects, to blotting or RNA species.

Briefly, a probe is used to target a DNA or RNA species that has been immobilized on a suitable matrix, often a filter of nitrocellulose. The different species should be spatially separated to facilitate analysis. This often is accomplished by gel electrophoresis of nucleic acid species followed by "blotting" on to the filter.

Subsequently, the blotted target is incubated with a probe (usually labeled) under conditions that promote denaturation and rehybridization. Because the probe is designed to base pair with the target, the probe will binding a portion of the

target sequence under renaturing conditions. Unbound probe is then removed, and detection is accomplished as described above.

d. Separation Methods

It normally is desirable, at one stage or another, to separate the amplification product from the template and the excess primer for the purpose of determining whether specific amplification has occurred. In one embodiment, amplification products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using standard methods. See Sambrook et al., 1989.

Alternatively, chromatographic techniques may be employed to effect separation. There are many kinds of chromatography which may be used in the present invention: adsorption, partition, ion-exchange and molecular sieve, and many specialized techniques for using them including column, paper, thin-layer and gas chromatography (Freifelder, 1982).

e. Detection Methods

Products may be visualized in order to confirm amplification of the marker sequences. One typical visualization method involves staining of a gel with ethidium bromide and visualization under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the amplification products can then be exposed to x-ray film or visualized under the appropriate stimulating spectra, following separation.

In one embodiment, visualization is achieved indirectly. Following separation of amplification products, a labeled nucleic acid probe is brought into contact with the amplified marker sequence. The probe preferably is conjugated to a chromophore but may be radiolabeled. In another embodiment, the probe is conjugated to a binding partner, such as an antibody or biotin, and the other member of the binding pair carries a detectable moiety.

In one embodiment, detection is by a labeled probe. The techniques involved are well known to those of skill in the art and can be found in many standard books on molecular protocols. See Sambrook et al., 1989. For example, chromophore or radiolabel probes or primers identify the target during or following amplification.

One example of the foregoing is described in U.S. Pat. No. 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and transfer of nucleic acids. The apparatus permits electrophoresis and blotting without external manipulation of the gel and is ideally suited to carrying out methods according to the present invention.

In addition, the amplification products described above may be subjected to sequence analysis to identify specific kinds of variations using standard sequence analysis techniques. Within certain methods, exhaustive analysis of genes is carried out by sequence analysis using primer sets designed for optimal sequencing (Pignon et al, 1994). The present invention provides methods by which any or all of these types of analyses may be used. Using the sequences disclosed herein, oligonucleotide primers may be designed to permit the amplification of sequences throughout the BBS3 gene that may then be analyzed by direct sequencing.

f. Kit Components

All the essential materials and reagents required for detecting and sequencing BBS3 and variants thereof may be assembled together in a kit. This generally will comprise preselected primers and probes. Also included may be enzymes suitable for amplifying nucleic acids including various polymerases (RT, Taq, Sequenase™, etc.), deoxy-nucleotides and buffers to provide the necessary reaction

mixture for amplification. Such kits also generally will comprise, in suitable means, distinct containers for each individual reagent and enzyme as well as for each primer or probe.

g. Design and Theoretical Considerations for Relative Quantitative RT-PCR™

Reverse transcription (RT) of RNA to cDNA followed by relative quantitative PCR™ (RT-PCR™) can be used to determine the relative concentrations of specific mRNA species isolated from patients. By determining that the concentration of a specific mRNA species varies, it is shown that the gene encoding the specific mRNA species is differentially expressed.

In PCR™, the number of molecules of the amplified target DNA increase by a factor approaching two with every cycle of the reaction until some reagent becomes limiting. Thereafter, the rate of amplification becomes increasingly diminished until there is no increase in the amplified target between cycles. If a graph is plotted in which the cycle number is on the X axis and the log of the concentration of the amplified target DNA is on the Y axis, a curved line of characteristic shape is formed by connecting the plotted points. Beginning with the first cycle, the slope of the line is positive and constant. This is said to be the linear portion of the curve. After a reagent becomes limiting, the slope of the line begins to decrease and eventually becomes zero. At this point the concentration of the amplified target DNA becomes asymptotic to some fixed value. This is said to be the plateau portion of the curve.

The concentration of the target DNA in the linear portion of the PCR™ amplification is directly proportional to the starting concentration of the target before the reaction began. By determining the concentration of the amplified products of the target DNA in PCR™ reactions that have completed the same number of cycles and are in their linear ranges, it is possible to determine the relative concentrations of the specific target sequence in the original DNA mixture. If the DNA mixtures are cDNAs synthesized from RNAs isolated from different tissues or cells, the relative abundances of the specific mRNA from which the target sequence was derived can be determined for the respective tissues or cells. This direct proportionality between the concentration of the PCR™ products and the relative mRNA abundances is only true in the linear range of the PCR™ reaction.

The final concentration of the target DNA in the plateau portion of the curve is determined by the availability of reagents in the reaction mix and is independent of the original concentration of target DNA. Therefore, the first condition that must be met before the relative abundances of a mRNA species can be determined by RT-PCR™ for a collection of RNA populations is that the concentrations of the amplified PCR™ products must be sampled when the PCR™ reactions are in the linear portion of their curves.

The second condition that must be met for an RT-PCR™ experiment to successfully determine the relative abundances of a particular mRNA species is that relative concentrations of the amplifiable cDNAs must be normalized to some independent standard. The goal of an RT-PCR™ experiment is to determine the abundance of a particular mRNA species relative to the average abundance of all mRNA species in the sample. In the experiments described below, mRNAs for β -actin, asparagine synthetase and lipocortin II were used as external and internal standards to which the relative abundance of other mRNAs are compared.

Most protocols for competitive PCR™ utilize internal PCR™ standards that are approximately as abundant as the

target. These strategies are effective if the products of the PCRTM amplifications are sampled during their linear phases. If the products are sampled when the reactions are approaching the plateau phase, then the less abundant product becomes relatively over represented. Comparisons of relative abundances made for many different RNA samples, such as is the case when examining RNA samples for differential expression, become distorted in such a way as to make differences in relative abundances of RNAs appear less than they actually are. This is not a significant problem if the internal standard is much more abundant than the target. If the internal standard is more abundant than the target, then direct linear comparisons can be made between RNA samples.

The above discussion describes theoretical considerations for an RT-PCRTM assay for clinically derived materials. The problems inherent in clinical samples are that they are of variable quantity (making normalization problematic), and that they are of variable quality (necessitating the co-amplification of a reliable internal control, preferably of larger size than the target). Both of these problems are overcome if the RT-PCRTM is performed as a relative quantitative RT-PCRTM with an internal standard in which the internal standard is an amplifiable cDNA fragment that is larger than the target cDNA fragment and in which the abundance of the mRNA encoding the internal standard is roughly 5-100-fold higher than the mRNA encoding the target. This assay measures relative abundance, not absolute abundance of the respective mRNA species.

Other studies may be performed using a more conventional relative quantitative RT-PCRTM assay with an external standard protocol. These assays sample the PCRTM products in the linear portion of their amplification curves. The number of PCRTM cycles that are optimal for sampling must be empirically determined for each target cDNA fragment. In addition, the reverse transcriptase products of each RNA population isolated from the various tissue samples must be carefully normalized for equal concentrations of amplifiable cDNAs. This consideration is very important since the assay measures absolute mRNA abundance. Absolute mRNA abundance can be used as a measure of differential gene expression only in normalized samples. While empirical determination of the linear range of the amplification curve and normalization of cDNA preparations are tedious and time consuming processes, the resulting RT-PCRTM assays can be superior to those derived from the relative quantitative RT-PCRTM assay with an internal standard.

One reason for this advantage is that without the internal standard/competitor, all of the reagents can be converted into a single PCRTM product in the linear range of the amplification curve, thus increasing the sensitivity of the assay. Another reason is that with only one PCRTM product, display of the product on an electrophoretic gel or another display method becomes less complex, has less background and is easier to interpret.

h. Chip Technologies

Specifically contemplated by the present inventors are chip-based DNA technologies such as those described by Hacia et al. (1996) and Shoemaker et al. (1996). Briefly, these techniques involve quantitative methods for analyzing large numbers of genes rapidly and accurately. By tagging genes with oligonucleotides or using fixed probe arrays, one can employ chip technology to segregate target molecules as high density arrays and screen these molecules on the basis of hybridization. See also Pease et al., (1994); Fodor et al., (1991).

B. Immunodiagnosis

Antibodies can be used in characterizing the BBS3 content of healthy and diseased tissues, through techniques such as ELISAs and Western blotting. This may provide a prenatal screen or in counseling for those individuals seeking to have children.

The use of antibodies of the present invention, in an ELISA assay is contemplated. For example, anti-BBS3 antibodies are immobilized onto a selected surface, preferably a surface exhibiting a protein affinity such as the wells of a polystyrene microtiter plate. After washing to remove incompletely adsorbed material, it is desirable to bind or coat the assay plate wells with a non-specific protein that is known to be antigenically neutral with regard to the test antisera such as bovine serum albumin (BSA), casein or solutions of powdered milk. This allows for blocking of non-specific adsorption sites on the immobilizing surface and thus reduces the background caused by non-specific binding of antigen onto the surface.

After binding of antibody to the well, coating with a non-reactive material to reduce background, and washing to remove unbound material, the immobilizing surface is contacted with the sample to be tested in a manner conducive to immune complex (antigen/antibody) formation.

Following formation of specific immunocomplexes between the test sample and the bound antibody, and subsequent washing, the occurrence and even amount of immunocomplex formation may be determined by subjecting same to a second antibody having specificity for BBS3 that differs the first antibody. Appropriate conditions preferably include diluting the sample with diluents such as BSA, bovine gamma globulin (BGG) and phosphate buffered saline (PBS)/Tween[®]. These added agents also tend to assist in the reduction of nonspecific background. The layered antisera is then allowed to incubate for from about 2 to about 4 hr, at temperatures preferably on the order of about 25° to about 27° C. Following incubation, the antisera-contacted surface is washed so as to remove non-immunocomplexed material. A preferred washing procedure includes washing with a solution such as PBS/Tween[®], or borate buffer.

To provide a detecting means, the second antibody will preferably have an associated enzyme that will generate a color development upon incubating with an appropriate chromogenic substrate. Thus, for example, one will desire to contact and incubate the second antibody-bound surface with a urease or peroxidase-conjugated anti-human IgG for a period of time and under conditions which favor the development of immunocomplex formation (e.g., incubation for 2 hr at room temperature in a PBS-containing solution such as PBS/Tween[®]).

After incubation with the second enzyme-tagged antibody, and subsequent to washing to remove unbound material, the amount of label is quantified by incubation with a chromogenic substrate such as urea and bromocresol purple or 2,2'-azino-di-(3-ethyl-benzthiazoline)-6-sulfonic acid (ABTS) and H₂O₂, in the case of peroxidase as the enzyme label. Quantitation is then achieved by measuring the degree of color generation, e.g., using a visible spectrum spectrophotometer.

The preceding format may be altered by first binding the sample to the assay plate. Then, primary antibody is incubated with the assay plate, followed by detecting of bound primary antibody using a labeled second antibody with specificity for the primary antibody.

The steps of various other useful immunodetection methods have been described in the scientific literature, such as, e.g., Nakamura et al., (1987). Immunoassays, in their most

simple and direct sense, are binding assays. Certain preferred immunoassays are the various types of radioimmunoassays (RIA) and immunobead capture assay. Immunohistochemical detection using tissue sections also is particularly useful. However, it will be readily appreciated that detection is not limited to such techniques, and Western blotting, dot blotting, FACS analyses, and the like also may be used in connection with the present invention.

The antibody compositions of the present invention will find great use in immunoblot or Western blot analysis. The antibodies may be used as high-affinity primary reagents for the identification of proteins immobilized onto a solid support matrix, such as nitrocellulose, nylon or combinations thereof. In conjunction with immunoprecipitation, followed by gel electrophoresis, these may be used as a single step reagent for use in detecting antigens against which secondary reagents used in the detection of the antigen cause an adverse background. Immunologically-based detection methods for use in conjunction with Western blotting include enzymatically-, radiolabel-, or fluorescently-tagged secondary antibodies against the toxin moiety are considered to be of particular use in this regard. U.S. Patents concerning the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241, each incorporated herein by reference. Of course, one may find additional advantages through the use of a secondary binding ligand such as a second antibody or a biotin/avidin ligand binding arrangement, as is known in the art.

4. Methods for Screening Active Compounds

The present invention also contemplates the use of BBS3 and active fragments, and nucleic acids coding therefor, in the screening of compounds for activity in either stimulating BBS3 activity, overcoming the lack of BBS3 or blocking the effect of a mutant BBS3 molecule. These assays may make use of a variety of different formats and may depend on the kind of "activity" for which the screen is being conducted.

A. In Vitro Assays

In one embodiment, the invention is to be applied for the screening of compounds that bind to the BBS3 polypeptide or fragment thereof. The polypeptide or fragment may be either free in solution, fixed to a support, expressed in or on the surface of a cell. Either the polypeptide or the compound may be labeled, thereby permitting determining of binding.

In another embodiment, the assay may measure the inhibition of binding of BBS3 to a natural or artificial substrate or binding partner. Competitive binding assays can be performed in which one of the agents (BBS3, binding partner or compound) is labeled. Usually, the polypeptide will be the labeled species. One may measure the amount of free label versus bound label to determine binding or inhibition of binding.

Another technique for high throughput screening of compounds is described in WO 84/03564. Large numbers of small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with BBS3 and washed. Bound polypeptide is detected by various methods.

Purified BBS3 can be coated directly onto plates for use in the aforementioned drug screening techniques. However, non-neutralizing antibodies to the polypeptide can be used to immobilize the polypeptide to a solid phase. Also, fusion proteins containing a reactive region (preferably a terminal region) may be used to link the BBS3 active region to a solid phase.

Various cell lines containing wild-type or natural or engineered mutations in BBS3 gene can be used to study various functional attributes of BBS3 and how a candidate compound affects these attributes. Methods for engineering mutations are described elsewhere in this document, as are naturally-occurring mutations in BBS3 that lead to, contribute to and/or otherwise cause BBS. In such assays, the compound would be formulated appropriately, given its biochemical nature, and contacted with a target cell. Depending on the assay, culture may be required. The cell may then be examined by virtue of a number of different physiologic assays. Alternatively, molecular analysis may be performed in which the function of BBS3, or related pathways, may be explored.

B. In Vivo Assays

The present invention also encompasses the use of various animal models. Thus, any identity seen between human and other animal BBS3 provides an excellent opportunity to examine the function of BBS3 in a whole animal system where it is normally expressed. By developing or isolating mutant cells lines that fail to express normal BBS3, one can generate models in mice that will be highly predictive of BBS and related syndromes in humans and other mammals.

Treatment of animals with test compounds will involve the administration of the compound, in an appropriate form, to the animal. Administration will be by any route the could be utilized for clinical or non-clinical purposes, including but not limited to oral, nasal, buccal, rectal, vaginal or topical. Alternatively, administration may be by intratracheal instillation, bronchial instillation, intradermal, subcutaneous, intramuscular, intraperitoneal or intravenous injection. Specifically contemplated are systemic intravenous injection, regional administration via blood or lymph supply and intratumoral injection.

Determining the effectiveness of a compound in vivo may involve a variety of different criteria. Such criteria include, but are not limited to, survival, reduction of tumor burden or mass, arrest or slowing of tumor progression, elimination of tumors, inhibition or prevention of metastasis, increased activity level, improvement in immune effector function and improved food intake.

C. Rational Drug Design

The goal of rational drug design is to produce structural analogs of biologically active polypeptides or compounds with which they interact (agonists, antagonists, inhibitors, binding partners, etc.). By creating such analogs, it is possible to fashion drugs which are more active or stable than the natural molecules, which have different susceptibility to alteration or which may affect the function of various other molecules. In one approach, one would generate a three-dimensional structure for BBS3 or a fragment thereof. This could be accomplished by x-ray crystallography, computer modeling or by a combination of both approaches. An alternative approach, "alanine scan," involves the random replacement of residues throughout molecule with alanine, and the resulting affect on function determined.

It also is possible to isolate a BBS3-specific antibody, selected by a functional assay, and then solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of anti-idiotypic would be expected to be an analog of the original antigen. The anti-idiotypic could then be used to identify and isolate peptides from banks of

chemically- or biologically-produced peptides. Selected peptides would then serve as the pharmacore. Anti-idiotypes may be generated using the methods described herein for producing antibodies, using an antibody as the antigen.

Thus, one may design drugs which have improved BBS3 activity or which act as stimulators, inhibitors, agonists, antagonists of BBS3 or molecules affected by BBS3 function. By virtue of the availability of cloned BBS3 gene sequences, sufficient amounts of BBS3 can be produced to perform crystallographic studies. In addition, knowledge of the polypeptide sequences permits computer employed predictions of structure-function relationships.

D. Transgenic Animals/Knockout Animals

In one embodiment of the invention, transgenic animals are produced which contain a functional transgene encoding a functional BBS3 polypeptide or variants thereof. Transgenic animals expressing BBS3 transgenes, recombinant cell lines derived from such animals and transgenic embryos may be useful in methods for screening for and identifying agents that induce or repress function of BBS3. Transgenic animals of the present invention also can be used as models for studying disease states.

In one embodiment of the invention, a BBS3 transgene is introduced into a non-human host to produce a transgenic animal expressing a human or murine BBS3 gene. The transgenic animal is produced by the integration of the transgene into the genome in a manner that permits the expression of the transgene. Methods for producing transgenic animals are generally described by Wagner and Hoppe (U.S. Pat. No. 4,873,191; which is incorporated herein by reference), Brinster et al., 1985; which is incorporated herein by reference in its entirety) and in "Manipulating the Mouse Embryo; A Laboratory Manual" 2nd edition (1994); which is incorporated herein by reference in its entirety).

It may be desirable to replace the endogenous BBS3 by homologous recombination between the transgene and the endogenous gene; or the endogenous gene may be eliminated by deletion as in the preparation of "knock-out" animals. Typically, a BBS3 gene flanked by genomic sequences is transferred by microinjection into a fertilized egg. The microinjected eggs are implanted into a host female, and the progeny are screened for the expression of the transgene. Transgenic animals may be produced from the fertilized eggs from a number of animals including, but not limited to reptiles, amphibians, birds, mammals, and fish. Within a particularly preferred embodiment, transgenic mice are generated which overexpress BBS3 or express a mutant form of the polypeptide. Alternatively, the absence of a BBS3 in "knock-out" mice permits the study of the effects that loss of BBS3 protein has on a cell in vivo. Knock-out mice also provide a model for the development of BBS3-related disease.

As noted above, transgenic animals and cell lines derived from such animals may find use in certain testing experiments. In this regard, transgenic animals and cell lines capable of expressing wild-type or mutant BBS3 may be exposed to test substances. These test substances can be screened for the ability to enhance wild-type BBS3 expression and/or function or impair the expression or function of mutant BBS3.

5. Methods for Treating BBS

The present invention also contemplates the treatment of BBS and related symptoms such as obesity, diabetes, renal cancer or other abnormalities, retinal degeneration and hypertension by providing a BBS3 protein to cells of an affected individual.

A. Genetic Based Therapies

Specifically, the present inventors intend to provide, to a cell, an expression construct capable of providing BBS3 to that cell. Because the sequence homology between the human, and other BBS3, any of these nucleic acids could be used in human therapy, as could any of the gene sequence variants discussed above which would encode the same, or a biologically equivalent polypeptide. The lengthy discussion of expression vectors and the genetic elements employed therein is incorporated into this section by reference. Particularly preferred expression vectors are viral vectors such as adenovirus, adeno-associated virus, herpesvirus, vaccinia virus and retrovirus. Also preferred is liposomally-encapsulated expression vector.

Those of skill in the art are well aware of how to apply gene delivery to in vivo and ex vivo situations. For viral vectors, one generally will prepare a viral vector stock. Depending on the kind of virus and the titer attainable, one will deliver 1×10^4 , 1×10^5 , 1×10^6 , 1×10^7 , 1×10^8 , 1×10^9 , 1×10^{10} , 1×10^{11} or 1×10^{12} infectious particles to the patient. Similar figures may be extrapolated for liposomal or other non-viral formulations by comparing relative uptake efficiencies. Formulation as a pharmaceutically acceptable composition is discussed below.

B. Protein Therapy

Another therapy approach is the provision, to a subject, of BBS3 polypeptide, active fragments, synthetic peptides, mimetics or other analogs thereof. The protein may be produced by recombinant expression means. Formulations would be selected based on the route of administration and purpose including, but not limited to, liposomal formulations and classic pharmaceutical preparations.

6. Engineering Expression Constructs

In certain embodiments, the present invention involves the manipulation of genetic material to produce expression constructs that encode BBS3 gene. Such methods involve the generation of expression constructs containing, for example, a heterologous DNA encoding a gene of interest and a means for its expression, replicating the vector in an appropriate helper cell, obtaining viral particles produced therefrom, and infecting cells with the recombinant virus particles.

The gene will be a normal BBS3 gene discussed herein above. In the context of gene therapy, the gene will be a heterologous DNA, meant to include DNA derived from a source other than the viral genome which provides the backbone of the vector. The gene may be derived from a prokaryotic or eukaryotic source such as a bacterium, a virus, a yeast, a parasite, a plant, or even an animal. The heterologous DNA also may be derived from more than one source, i.e., a multigene construct or a fusion protein. The heterologous DNA also may include a regulatory sequence which may be derived from one source and the gene from a different source.

A. Selectable Markers

In certain embodiments of the invention, the therapeutic expression constructs of the present invention contain nucleic acid constructs whose expression may be identified in vitro or in vivo by including a marker in the expression construct. Such markers would confer an identifiable change to the cell permitting easy identification of cells containing the expression construct. Usually the inclusion of a drug selection marker aids in cloning and in the selection of transformants. For example, genes that confer resistance to neomycin, puromycin, hygromycin, DHFR, GPT, zeocin and histidinol are useful selectable markers. Alternatively,

enzymes such as herpes simplex virus thymidine kinase (tk) may be employed. Immunologic markers also can be employed. The selectable marker employed is not believed to be important, so long as it is capable of being expressed simultaneously with the nucleic acid encoding a gene product. Further examples of selectable markers are well known to one of skill in the art and include reporters such as EGFP, β -gal or chloramphenicol acetyltransferase (CAT).

B. Control Regions

a. Promoters

Throughout this application, the term "expression construct" is meant to include any type of genetic construct containing a nucleic acid coding for gene products in which part or all of the nucleic acid encoding sequence is capable of being transcribed. The transcript may be translated into a protein, but it need not be. In certain embodiments, expression includes both transcription of a gene and translation of mRNA into a gene product. In other embodiments, expression only includes transcription of the nucleic acid encoding genes of interest.

The nucleic acid encoding a gene product is under transcriptional control of a promoter. A "promoter" refers to a DNA sequence recognized by the synthetic machinery of the cell, or introduced synthetic machinery, required to initiate the specific transcription of a gene. The phrase "under transcriptional control" means that the promoter is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the gene.

The term promoter will be used here to refer to a group of transcriptional control modules that are clustered around the initiation site for RNA polymerase II. Much of the thinking about how promoters are organized derives from analyses of several viral promoters, including those for the HSV thymidine kinase (tk) and SV40 early transcription units. These studies, augmented by more recent work, have shown that promoters are composed of discrete functional modules, each consisting of approximately 7-20 bp of DNA, and containing one or more recognition sites for transcriptional activator or repressor proteins.

At least one module in each promoter functions to position the start site for RNA synthesis. The best known example of this is the TATA box, but in some promoters lacking a TATA box, such as the promoter for the mammalian terminal deoxynucleotidyl transferase gene and the promoter for the SV40 late genes, a discrete element overlying the start site itself helps to fix the place of initiation.

Additional promoter elements regulate the frequency of transcriptional initiation. Typically, these are located in the region 30-110 bp upstream of the start site, although a number of promoters have recently been shown to contain functional elements downstream of the start site as well. The spacing between promoter elements frequently is flexible, so that promoter function is preserved when elements are inverted or moved relative to one another. In the tk promoter, the spacing between promoter elements can be increased to 50 bp apart before activity begins to decline. Depending on the promoter, it appears that individual elements can function either cooperatively or independently to activate transcription.

The particular promoter employed to control the expression of a nucleic acid sequence of interest is not believed to be important, so long as it is capable of directing the expression of the nucleic acid in the targeted cell. Thus, where a human cell is targeted, it is preferable to position the nucleic acid coding region adjacent to and under the control of a promoter that is capable of being expressed in a human

cell. Generally speaking, such a promoter might include either a human or viral promoter.

In various embodiments, the human cytomegalovirus (CMV) immediate early gene promoter, the SV40 early promoter, the Rous sarcoma virus long terminal repeat, β -actin, rat insulin promoter and glyceraldehyde-3-phosphate dehydrogenase can be used to obtain high-level expression of the coding sequence of interest. The use of other viral or mammalian cellular or bacterial phage promoters which are well-known in the art to achieve expression of a coding sequence of interest is contemplated as well, provided that the levels of expression are sufficient for a given purpose. By employing a promoter with well-known properties, the level and pattern of expression of the protein of interest following transfection or transformation can be optimized.

Selection of a promoter that is regulated in response to specific physiologic or synthetic signals can permit inducible expression of the gene product. For example in the case where expression of a transgene, or transgenes when a multicistronic vector is utilized, is toxic to the cells in which the vector is produced in, it may be desirable to prohibit or reduce expression of one or more of the transgenes. Examples of transgenes that may be toxic to the producer cell line are pro-apoptotic and cytokine genes. Several inducible promoter systems are available for production of viral vectors where the transgene product may be toxic.

The ecdysone system (Invitrogen, Carlsbad, Calif.) is one such system. This system is designed to allow regulated expression of a gene of interest in mammalian cells. It consists of a tightly regulated expression mechanism that allows virtually no basal level expression of the transgene, but over 200-fold inducibility. The system is based on the heterodimeric ecdysone receptor of *Drosophila*, and when ecdysone or an analog such as muristerone A binds to the receptor, the receptor activates a promoter to turn on expression of the downstream transgene high levels of mRNA transcripts are attained. In this system, both monomers of the heterodimeric receptor are constitutively expressed from one vector, whereas the ecdysone-responsive promoter which drives expression of the gene of interest is on another plasmid. Engineering of this type of system into the gene transfer vector of interest would therefore be useful. Cotransfection of plasmids containing the gene of interest and the receptor monomers in the producer cell line would then allow for the production of the gene transfer vector without expression of a potentially toxic transgene. At the appropriate time, expression of the transgene could be activated with ecdysone or muristerone A.

Another inducible system that would be useful is the Tet-Off™ or Tet-On™ system (Clontech, Palo Alto, Calif.) originally developed by Gossen and Bujard (Gossen and Bujard, 1992; Gossen et al., 1995). This system also allows high levels of gene expression to be regulated in response to tetracycline or tetracycline derivatives such as doxycycline. In the Tet-On™ system, gene expression is turned on in the presence of doxycycline, whereas in the Tet-Off™ system, gene expression is turned on in the absence of doxycycline. These systems are based on two regulatory elements derived from the tetracycline resistance operon of *E. coli*. The tetracycline operator sequence to which the tetracycline repressor binds, and the tetracycline repressor protein. The gene of interest is cloned into a plasmid behind a promoter that has tetracycline-responsive elements present in it. A second plasmid contains a regulatory element called the tetracycline-controlled transactivator, which is composed, in the Tet-Off™ system, of the VP16 domain from the herpes

simplex virus and the wild-type tetracycline repressor. Thus in the absence of doxycycline, transcription is constitutively on. In the Tet-On™ system, the tetracycline repressor is not wild type and in the presence of doxycycline activates transcription. For gene therapy vector production, the Tet-Off™ system would be preferable so that the producer cells could be grown in the presence of tetracycline or doxycycline and prevent expression of a potentially toxic transgene, but when the vector is introduced to the patient, the gene expression would be constitutively on.

In some circumstances, it may be desirable to regulate expression of a transgene in a gene therapy vector. For example, different viral promoters with varying strengths of activity may be utilized depending on the level of expression desired. In mammalian cells, the CMV immediate early promoter if often used to provide strong transcriptional activation. Modified versions of the CMV promoter that are less potent have also been used when reduced levels of expression of the transgene are desired. When expression of a transgene in hematopoietic cells is desired, retroviral promoters such as the LTRs from MLV or MMTV are often used. Other viral promoters that may be used depending on the desired effect include SV40, RSV LTR, HIV-1 and HIV-2 LTR, adenovirus promoters such as from the E1A, E2A, or MLP region, AAV LTR, cauliflower mosaic virus, HSV-TK, and avian sarcoma virus.

Similarly tissue specific promoters may be used to effect transcription in specific tissues or cells so as to reduce potential toxicity or undesirable effects to non-targeted tissues. For example, promoters such as the PSA, probasin, prostatic acid phosphatase or prostate-specific glandular kallikrein (hK2) may be used to target gene expression in the prostate. Similarly, the following promoters may be used to target gene expression in other tissues (Table 5).

TABLE 5

<u>Tissue Specific Promoters</u>	
Tissue	Promoter
Pancreas	insulin
	elastin
	amylase
Liver	pdr-1 pdx-1
	glucokinase
	albumin PEPCK
	HBV enhancer
	alpha fetoprotein
	apolipoprotein C
	alpha-1 antitrypsin
vitellogenin, NF-AB	
Skeletal muscle	Transthyretin
	myosin H chain
	muscle creatine kinase
	dystrophin
	calpain p94
	skeletal alpha-actin
	fast troponin 1
Skin	keratin K6
	keratin K1
	CFTR
Lung	human cytokeratin 18 (K18)
	pulmonary surfactant proteins A, B and C
	CC-10
Smooth muscle	P1
	sm22 alpha
Endothelium	SM-alpha-actin
	endothelin-1
	E-selectin
	von Willebrand factor
	TIE (Korhonen et al., 1995)
	KDR/flk-1

TABLE 5-continued

<u>Tissue Specific Promoters</u>	
Tissue	Promoter
Melanocytes	tyrosinase
Adipose tissue	lipoprotein lipase (Zechner et al., 1988)
	adipsin (Spiegelman et al., 1989)
	acetyl-CoA carboxylase (Pape and Kim, 1989)
	glycerophosphate dehydrogenase (Dani et al., 1989)
	adipocyte P2 (Hunt et al., 1986)
Blood	β-globin

In certain indications, it may be desirable to activate transcription at specific times after administration of the gene therapy vector. This may be done with such promoters as those that are hormone or cytokine regulatable. For example in gene therapy applications where the indication is a gonadal tissue where specific steroids are produced or routed to, use of androgen or estrogen regulated promoters may be advantageous. Such promoters that are hormone regulatable include MMTV, MT-1, ecdysone and RuBisco. Other hormone regulated promoters such as those responsive to thyroid, pituitary and adrenal hormones are expected to be useful in the present invention. Cytokine and inflammatory protein responsive promoters that could be used include K and T Kininogen (Kageyama et al., 1987), c-fos, TNF-alpha, C-reactive protein (Arcone et al., 1988), haptoglobin (Oliviero et al., 1987), serum amyloid A2, C/EBP alpha, IL-1, IL-6 (Poli and Cortese, 1989), Complement C3 (Wilson et al., 1990), IL-8, alpha-1 acid glycoprotein (Prowse and Baumann, 1988), alpha-1 antitrypsin, lipoprotein lipase (Zechner et al., 1988), angiotensinogen (Ron et al., 1991), fibrinogen, c-jun (inducible by phorbol esters, TNF-alpha, UV radiation, retinoic acid, and hydrogen peroxide), collagenase (induced by phorbol esters and retinoic acid), metallothionein (heavy metal and glucocorticoid inducible), Stromelysin (inducible by phorbol ester, interleukin-1 and EGF), alpha-2 macroglobulin and alpha-1 antichymotrypsin.

It is envisioned that cell cycle regulatable promoters may be useful in the present invention. For example, in a bicistronic gene therapy vector, use of a strong CMV promoter to drive expression of a first gene such as p16 that arrests cells in the G1 phase could be followed by expression of a second gene such as p53 under the control of a promoter that is active in the G1 phase of the cell cycle, thus providing a "second hit" that would push the cell into apoptosis. Other promoters such as those of various cyclins, PCNA, galectin-3, E2F1, p53 and BRCA1 could be used.

Promoters that could be used according to the present invention include Lac-regulatable, chemotherapy inducible (e.g. MDR), and heat (hyperthermia) inducible promoters, Radiation-inducible (e.g., EGR (Joki et al., 1995)), Alpha-inhibin, RNA pol III tRNA met and other amino acid promoters, U1 snRNA (Bartlett et al., 1996), MC-1, PGK, alpha-actin and alpha-globin. Many other promoters that may be useful are listed in Walther and Stein (1996).

It is envisioned that any of the above promoters alone or in combination with another may be useful according to the present invention depending on the action desired. In addition, this list of promoters should not be construed to be exhaustive or limiting, those of skill in the art will know of other promoters that may be used in conjunction with the promoters and methods disclosed herein.

b. Enhancers

Enhancers are genetic elements that increase transcription from a promoter located at a distant position on the same molecule of DNA. Enhancers are organized much like promoters. That is, they are composed of many individual elements, each of which binds to one or more transcriptional proteins. The basic distinction between enhancers and promoters is operational. An enhancer region as a whole must be able to stimulate transcription at a distance; this need not be true of a promoter region or its component elements. On the other hand, a promoter must have one or more elements that direct initiation of RNA synthesis at a particular site and in a particular orientation, whereas enhancers lack these specificities. Promoters and enhancers are often overlapping and contiguous, often seeming to have a very similar modular organization.

Below is a list of promoters additional to the tissue specific promoters listed above, cellular promoters/enhancers and inducible promoters/enhancers that could be used in combination with the nucleic acid encoding a gene of interest in an expression construct (Tables 6 and 7). Additionally, any promoter/enhancer combination (as per the Eukaryotic Promoter Data Base EPDB) could also be used to drive expression of the gene. Eukaryotic cells can support cytoplasmic transcription from certain bacterial promoters if the appropriate bacterial polymerase is provided, either as part of the delivery complex or as an additional genetic expression construct.

In preferred embodiments of the invention, the expression construct comprises a virus or engineered construct derived from a viral genome. The ability of certain viruses to enter cells via receptor-mediated endocytosis and to integrate into host cell genome and express viral genes stably and efficiently have made them attractive candidates for the transfer of foreign genes into mammalian cells (Ridgeway, 1988; Nicolas and Rubenstein, 1988; Baichwal and Sugden, 1986; Temin, 1986). The first viruses used as gene vectors were DNA viruses including the papovaviruses (simian virus 40, bovine papilloma virus, and polyoma) (Ridgeway, 1988; Baichwal and Sugden, 1986) and adenoviruses (Ridgeway, 1988; Baichwal and Sugden, 1986). These have a relatively low capacity for foreign DNA sequences and have a restricted host spectrum. Furthermore, their oncogenic potential and cytopathic effects in permissive cells raise safety concerns. They can accommodate only up to 8 kB of foreign genetic material but can be readily introduced in a variety of cell lines and laboratory animals (Nicolas and Rubenstein, 1988; Temin, 1986).

c. Polyadenylation Signals

Where a cDNA insert is employed, one will typically desire to include a polyadenylation signal to effect proper polyadenylation of the gene transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed such as human or bovine growth hormone and SV40 polyadenylation signals. Also contemplated as an element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

TABLE 6

ENHANCER	
Immunoglobulin Heavy Chain	
Immunoglobulin Light Chain	
T-Cell Receptor	

TABLE 6-continued

ENHANCER	
5	HLA DQ α and DQ β
	β -Interferon
	Interleukin-2
	Interleukin-2 Receptor
	MHC Class II 5
	MHC Class II HLA-DR α
10	β -Actin
	Muscle Creatine Kinase
	Prealbumin (Transthyretin)
	Elastase I
	Metallothionein
	Collagenase
15	Albumin Gene
	α -Fetoprotein
	τ -Globin
	β -Globin
	c-fos
	c-HA-ras
	Insulin
20	Neural Cell Adhesion Molecule (NCAM)
	α 1-Antitrypsin
	H2B (TH2B) Histone
	Mouse or Type I Collagen
	Glucose-Regulated Proteins (GRP94 and GRP78)
	Rat Growth Hormone
25	Human Serum Amyloid A (SAA)
	Troponin I (TN I)
	Platelet-Derived Growth Factor
	Duchenne Muscular Dystrophy
	SV40
	Polyoma
30	Retroviruses
	Papilloma Virus
	Hepatitis B Virus
	Human Immunodeficiency Virus
	Cytomegalovirus
	Gibbon Ape Leukemia Virus

TABLE 7

Element	Inducer
40	MT II
	Phorbol Ester (TPA)
	Heavy metals
	MMTV (mouse mammary tumor virus)
	Glucocorticoids
	β -Interferon
	poly(rI)X
	poly(rc)
45	Adenovirus 5 E2
	Ela
	c-jun
	Phorbol Ester (TPA), H ₂ O ₂
	Collagenase
	Phorbol Ester (TPA)
	Stromelysin
	Phorbol Ester (TPA), IL-1
	SV40
	Phorbol Ester (TPA)
	Murine MX Gene
	Interferon, Newcastle Disease Virus
50	GRP78 Gene
	A23187
	α -2-Macroglobulin
	IL-6
	Vimentin
	Serum
	MHC Class I Gene H-2 kB
	Interferon
	HSP70
	Ela, SV40 Large T Antigen
	Proliferin
55	Tumor Necrosis Factor
	Phorbol Ester-TPA
	FMA
	Thyroid Stimulating Hormone α Gene
	Thyroid Hormone
	Insulin E Box
	Glucose

7. Methods of Gene Transfer

In order to mediate the effect transgene expression in a cell, it will be necessary to transfer the therapeutic expression constructs of the present invention into a cell. Such transfer may employ viral or non-viral methods of gene transfer. This section provides a discussion of methods and compositions of gene transfer.

A. Viral Vector-Mediated Transfer

In certain embodiments, the BBS3 gene is incorporated into a viral particle to mediate gene transfer to a cell. Typically, the virus simply will be exposed to the appropriate host cell under physiologic conditions, permitting uptake of the virus. The present methods may be advantageously employed using a variety of viral vectors, as discussed below.

a. Adenovirus

Adenovirus is particularly suitable for use as a gene transfer vector because of its mid-sized DNA genome, ease of manipulation, high titer, wide target-cell range, and high infectivity. The roughly 36 kB viral genome is bounded by 100-200 base pair (bp) inverted terminal repeats (ITR), in which are contained cis-acting elements necessary for viral DNA replication and packaging. The early (E) and late (L) regions of the genome that contain different transcription units are divided by the onset of viral DNA replication.

The E1 region (E1A and E1B) encodes proteins responsible for the regulation of transcription of the viral genome and a few cellular genes. The expression of the E2 region (E2A and E2B) results in the synthesis of the proteins for viral DNA replication. These proteins are involved in DNA replication, late gene expression, and host cell shut off (Renan, 1990). The products of the late genes (L1, L2, L3, L4 and L5), including the majority of the viral capsid proteins, are expressed only after significant processing of a single primary transcript issued by the major late promoter (MLP). The MLP (located at 16.8 map units) is particularly efficient during the late phase of infection, and all the mRNAs issued from this promoter possess a 5' tripartite leader (TL) sequence which makes them preferred mRNAs for translation.

In order for adenovirus to be optimized for gene therapy, it is necessary to maximize the carrying capacity so that large segments of DNA can be included. It also is very desirable to reduce the toxicity and immunologic reaction associated with certain adenoviral products. The two goals are, to an extent, coterminous in that elimination of adenoviral genes serves both ends. By practice of the present invention, it is possible to achieve both these goals while retaining the ability to manipulate the therapeutic constructs with relative ease.

The large displacement of DNA is possible because the cis elements required for viral DNA replication all are localized in the inverted terminal repeats (ITR) (100-200 bp) at either end of the linear viral genome. Plasmids containing ITR's can replicate in the presence of a non-defective adenovirus (Hay et al., 1984). Therefore, inclusion of these elements in an adenoviral vector should permit replication.

In addition, the packaging signal for viral encapsidation is localized between 194-385 bp (0.5-1.1 map units) at the left end of the viral genome (Hearing et al., 1987). This signal mimics the protein recognition site in bacteriophage λ DNA where a specific sequence close to the left end, but outside the cohesive end sequence, mediates the binding to proteins that are required for insertion of the DNA into the head structure. E1 substitution vectors of Ad have demonstrated that a 450 bp (0-1.25 map units) fragment at the left end of the viral genome could direct packaging in 293 cells (Levvero et al., 1991).

Previously, it has been shown that certain regions of the adenoviral genome can be incorporated into the genome of mammalian cells and the genes encoded thereby expressed. These cell lines are capable of supporting the replication of an adenoviral vector that is deficient in the adenoviral function encoded by the cell line. There also have been

reports of complementation of replication deficient adenoviral vectors by "helping" vectors, e.g., wild-type virus or conditionally defective mutants.

Replication-deficient adenoviral vectors can be complemented, in trans, by helper virus. This observation alone does not permit isolation of the replication-deficient vectors, however, since the presence of helper virus, needed to provide replicative functions, would contaminate any preparation. Thus, an additional element was needed that would add specificity to the replication and/or packaging of the replication-deficient vector. That element, as provided for in the present invention, derives from the packaging function of adenovirus.

It has been shown that a packaging signal for adenovirus exists in the left end of the conventional adenovirus map (Tibbetts, 1977). Later studies showed that a mutant with a deletion in the E1A (194-358 bp) region of the genome grew poorly even in a cell line that complemented the early (E1A) function (Hearing and Shenk, 1983). When a compensating adenoviral DNA (0-353 bp) was recombined into the right end of the mutant, the virus was packaged normally. Further mutational analysis identified a short, repeated, position-dependent element in the left end of the Ad5 genome. One copy of the repeat was found to be sufficient for efficient packaging if present at either end of the genome, but not when moved towards the interior of the Ad5 DNA molecule (Hearing et al., 1987).

By using mutated versions of the packaging signal, it is possible to create helper viruses that are packaged with varying efficiencies. Typically, the mutations are point mutations or deletions. When helper viruses with low efficiency packaging are grown in helper cells, the virus is packaged, albeit at reduced rates compared to wild-type virus, thereby permitting propagation of the helper. When these helper viruses are grown in cells along with virus that contains wild-type packaging signals, however, the wild-type packaging signals are recognized preferentially over the mutated versions. Given a limiting amount of packaging factor, the virus containing the wild-type signals are packaged selectively when compared to the helpers. If the preference is great enough, stocks approaching homogeneity should be achieved.

b. Retrovirus

The retroviruses are a group of single-stranded RNA viruses characterized by an ability to convert their RNA to double-stranded DNA in infected cells by a process of reverse-transcription (Coffin, 1990). The resulting DNA then stably integrates into cellular chromosomes as a provirus and directs synthesis of viral proteins. The integration results in the retention of the viral gene sequences in the recipient cell and its descendants. The retroviral genome contains three genes—gag, pol and env—that code for capsid proteins, polymerase enzyme, and envelope components, respectively. A sequence found upstream from the gag gene, termed Ψ , functions as a signal for packaging of the genome into virions. Two long terminal repeat (LTR) sequences are present at the 5' and 3' ends of the viral genome. These contain strong promoter and enhancer sequences and also are required for integration in the host cell genome (Coffin, 1990).

In order to construct a retroviral vector, a nucleic acid encoding a promoter is inserted into the viral genome in the place of certain viral sequences to produce a virus that is replication-defective. In order to produce virions, a packaging cell line containing the gag, pol and env genes but without the LTR and Ψ components is constructed (Mann et al., 1983). When a recombinant plasmid containing a human

cDNA, together with the retroviral LTR and Ψ sequences is introduced into this cell line (by calcium phosphate precipitation for example), the Ψ sequence allows the RNA transcript of the recombinant plasmid to be packaged into viral particles, which are then secreted into the culture media (Nicolas and Rubenstein, 1988; Temin, 1986; Mann et al., 1983). The media containing the recombinant retroviruses is collected, optionally concentrated, and used for gene transfer. Retroviral vectors are able to infect a broad variety of cell types. However, integration and stable expression of many types of retroviruses require the division of host cells (Paskind et al., 1975).

An approach designed to allow specific targeting of retrovirus vectors recently was developed based on the chemical modification of a retrovirus by the chemical addition of galactose residues to the viral envelope. This modification could permit the specific infection of cells such as hepatocytes via asialoglycoprotein receptors, should this be desired.

A different approach to targeting of recombinant retroviruses was designed in which biotinylated antibodies against a retroviral envelope protein and against a specific cell receptor were used. The antibodies were coupled via the biotin components by using streptavidin (Roux et al., 1989). Using antibodies against major histocompatibility complex class I and class II antigens, the infection of a variety of human cells that bore those surface antigens was demonstrated with an ecotropic virus *in vitro* (Roux et al., 1989).

c. Adeno-associated Virus

AAV utilizes a linear, single-stranded DNA of about 4700 base pairs. Inverted terminal repeats flank the genome. Two genes are present within the genome, giving rise to a number of distinct gene products. The first, the cap gene, produces three different virion proteins (VP), designated VP-1, VP-2 and VP-3. The second, the rep gene, encodes four non-structural proteins (NS). One or more of these rep gene products is responsible for transactivating AAV transcription.

The three promoters in AAV are designated by their location, in map units, in the genome. These are, from left to right, p5, p19 and p40. Transcription gives rise to six transcripts, two initiated at each of three promoters, with one of each pair being spliced. The splice site, derived from map units 42-46, is the same for each transcript. The four non-structural proteins apparently are derived from the longer of the transcripts, and three virion proteins all arise from the smallest transcript.

AAV is not associated with any pathologic state in humans. Interestingly, for efficient replication, AAV requires "helping" functions from viruses such as herpes simplex virus I and II, cytomegalovirus, pseudorabies virus and, of course, adenovirus. The best characterized of the helpers is adenovirus, and many "early" functions for this virus have been shown to assist with AAV replication. Low level expression of AAV rep proteins is believed to hold AAV structural expression in check, and helper virus infection is thought to remove this block.

The terminal repeats of the AAV vector can be obtained by restriction endonuclease digestion of AAV or a plasmid such as p201, which contains a modified AAV genome (Samulski et al., 1987), or by other methods known to the skilled artisan, including but not limited to chemical or enzymatic synthesis of the terminal repeats based upon the published sequence of AAV. The ordinarily skilled artisan can determine, by well-known methods such as deletion analysis, the minimum sequence or part of the AAV ITRs which is required to allow function, i.e., stable and site-

specific integration. The ordinarily skilled artisan also can determine which minor modifications of the sequence can be tolerated while maintaining the ability of the terminal repeats to direct stable, site-specific integration.

AAV-based vectors have proven to be safe and effective vehicles for gene delivery *in vitro*, and these vectors are being developed and tested in pre-clinical and clinical stages for a wide range of applications in potential gene therapy, both *ex vivo* and *in vivo* (Carter and Flotte, 1996; Chatterjee et al., 1995; Ferrari et al., 1996; Fisher et al., 1996; Flotte et al., 1993; Goodman et al., 1994; Kaplitt et al., 1994; 1996; Kessler et al., 1996; Koeberl et al., 1997; Mizukami et al., 1996).

AAV-mediated efficient gene transfer and expression in the lung has led to clinical trials for the treatment of cystic fibrosis (Carter and Flotte, 1996; Flotte et al., 1993). Similarly, the prospects for treatment of muscular dystrophy by AAV-mediated gene delivery of the dystrophin gene to skeletal muscle, of Parkinson's disease by tyrosine hydroxylase gene delivery to the brain, of hemophilia B by Factor IX gene delivery to the liver, and potentially of myocardial infarction by vascular endothelial growth factor gene to the heart, appear promising since AAV-mediated transgene expression in these organs has recently been shown to be highly efficient (Fisher et al., 1996; Flotte et al., 1993; Kaplitt et al., 1994; 1996; Koeberl et al., 1997; McCown et al., 1996; Ping et al., 1996; Xiao et al., 1996).

d. Other Viral Vectors

Other viral vectors may be employed as expression constructs in the present invention. Vectors derived from viruses such as vaccinia virus (Ridgeway, 1988; Baichwal and Sugden, 1986; Coupar et al., 1988) canary pox virus, and herpes viruses may be employed. These viruses offer several features for use in gene transfer into various mammalian cells.

B. Non-viral Transfer

Several non-viral methods for the transfer of expression constructs into cultured mammalian cells are contemplated by the present invention. These include calcium phosphate precipitation (Graham and Van Der Eb, 1973; Chen and Okayama, 1987; Rippe et al., 1990) DEAE-dextran (Gopal, 1985), electroporation (Tur-Kaspa et al., 1986; Potter et al., 1984), direct microinjection (Harland and Weintraub, 1985), DNA-loaded liposomes (Nicolau and Sene, 1982; Fraley et al., 1979), cell sonication (Fechheimer et al., 1987), gene bombardment using high velocity microprojectiles (Yang et al., 1990), and receptor-mediated transfection (Wu and Wu, 1987; Wu and Wu, 1988).

Once the construct has been delivered into the cell the nucleic acid encoding the therapeutic gene may be positioned and expressed at different sites. In certain embodiments, the nucleic acid encoding the therapeutic gene may be stably integrated into the genome of the cell. This integration may be in the cognate location and orientation via homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle. How the expression construct is delivered to a cell and where in the cell the nucleic acid remains is dependent on the type of expression construct employed.

In a particular embodiment of the invention, the expression construct may be entrapped in a liposome. Liposomes are vesicular structures characterized by a phospholipid

bilayer membrane and an inner aqueous medium. Multilamellar liposomes have multiple lipid layers separated by aqueous medium. They form spontaneously when phospholipids are suspended in an excess of aqueous solution. The lipid components undergo self-rearrangement before the formation of closed structures and entrap water and dissolved solutes between the lipid bilayers (Ghosh and Bachawat, 1991). The addition of DNA to cationic liposomes causes a topological transition from liposomes to optically birefringent liquid-crystalline condensed globules (Radler et al., 1997). These DNA-lipid complexes are potential non-viral vectors for use in gene therapy.

Liposome-mediated nucleic acid delivery and expression of foreign DNA in vitro has been very successful. Using the β -lactamase gene, Wong et al., (1980) demonstrated the feasibility of liposome-mediated delivery and expression of foreign DNA in cultured chick embryo, HeLa, and hepatoma cells. Nicolau et al., (1987) accomplished successful liposome-mediated gene transfer in rats after intravenous injection. Also included are various commercial approaches involving "lipofection" technology.

In certain embodiments of the invention, the liposome may be complexed with a hemagglutinating virus (HVJ). This has been shown to facilitate fusion with the cell membrane and promote cell entry of liposome-encapsulated DNA (Kaneda et al., 1989). In other embodiments, the liposome may be complexed or employed in conjunction with nuclear nonhistone chromosomal proteins (HMG-1) (Kato et al., 1991). In yet further embodiments, the liposome may be complexed or employed in conjunction with both HVJ and HMG-1. In that such expression constructs have been successfully employed in transfer and expression of nucleic acid in vitro and in vivo, then they are applicable for the present invention.

Other vector delivery systems which can be employed to deliver a nucleic acid encoding a therapeutic gene into cells are receptor-mediated delivery vehicles. These take advantage of the selective uptake of macromolecules by receptor-mediated endocytosis in almost all eukaryotic cells. Because of the cell type-specific distribution of various receptors, the delivery can be highly specific (Wu and Wu, 1993).

Receptor-mediated gene targeting vehicles generally consist of two components: a cell receptor-specific ligand and a DNA-binding agent. Several ligands have been used for receptor-mediated gene transfer. The most extensively characterized ligands are asialoorosomucoid (ASOR) (Wu and Wu, 1987) and transferrin (Wagner et al., 1990). Recently, a synthetic neoglycoprotein, which recognizes the same receptor as ASOR, has been used as a gene delivery vehicle (Ferkol et al., 1993; Perales et al., 1994) and epidermal growth factor (EGF) has also been used to deliver genes to squamous carcinoma cells (Myers, EPO 0273085).

In other embodiments, the delivery vehicle may comprise a ligand and a liposome. For example, Nicolau et al., (1987) employed lactosyl-ceramide, a galactose-terminal asialganglioside, incorporated into liposomes and observed an increase in the uptake of the insulin gene by hepatocytes. Thus, it is feasible that a nucleic acid encoding a therapeutic gene also may be specifically delivered into a cell type such as prostate, epithelial or tumor cells, by any number of receptor-ligand systems with or without liposomes. For example, the human prostate-specific antigen (Watt et al., 1986) may be used as the receptor for mediated delivery of a nucleic acid in prostate tissue.

In another embodiment of the invention, the expression construct may simply consist of naked recombinant DNA or plasmids. Transfer of the construct may be performed by any

of the methods mentioned above which physically or chemically permeabilize the cell membrane. This is applicable particularly for transfer in vitro, however, it may be applied for in vivo use as well. Dubensky et al., (1984) successfully injected polyomavirus DNA in the form of CaPO_4 precipitates into liver and spleen of adult and newborn mice demonstrating active viral replication and acute infection. Benvenisty and Neshif (1986) also demonstrated that direct intraperitoneal injection of CaPO_4 precipitated plasmids results in expression of the transfected genes. It is envisioned that DNA encoding a CAM also may be transferred in a similar manner in vivo and express CAM.

Another embodiment of the invention for transferring a naked DNA expression construct into cells may involve particle bombardment. This method depends on the ability to accelerate DNA coated microprojectiles to a high velocity allowing them to pierce cell membranes and enter cells without killing them (Klein et al., 1987). Several devices for accelerating small particles have been developed. One such device relies on a high voltage discharge to generate an electrical current, which in turn provides the motive force (Yang et al., 1990). The microprojectiles used have consisted of biologically inert substances such as tungsten or gold beads.

8. Formulations and Routes for Administration to Patients

Where clinical applications are contemplated, it will be necessary to prepare pharmaceutical compositions—expression vectors, virus stocks, proteins, antibodies and drugs—in a form appropriate for the intended application. Generally, this will entail preparing compositions that are essentially free of pyrogens, as well as other impurities that could be harmful to humans or animals.

One will generally desire to employ appropriate salts and buffers to render delivery vectors stable and allow for uptake by target cells. Buffers also will be employed when recombinant cells are introduced into a patient. Aqueous compositions of the present invention comprise an effective amount of the vector to cells, dissolved or dispersed in a pharmaceutically acceptable carrier or aqueous medium. Such compositions also are referred to as inocula. The phrase "pharmaceutically or pharmacologically acceptable" refer to molecular entities and compositions that do not produce adverse, allergic, or other untoward reactions when administered to an animal or a human. As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the vectors or cells of the present invention, its use in therapeutic compositions is contemplated. Supplementary active ingredients also can be incorporated into the compositions.

The active compositions of the present invention may include classic pharmaceutical preparations. Administration of these compositions according to the present invention will be via any common route so long as the target tissue is available via that route. This includes oral, nasal, buccal, rectal, vaginal or topical. Alternatively, administration may be by orthotopic, intradermal, subcutaneous, intramuscular, intraperitoneal or intravenous injection. Such compositions would normally be administered as pharmaceutically acceptable compositions, described supra.

The active compounds also may be administered parenterally or intraperitoneally. Solutions of the active compounds as free base or pharmacologically acceptable salts can be

prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. 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. The prevention of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the 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 techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

For oral administration the polypeptides of the present invention may be incorporated with excipients and used in the form of non-ingestible mouthwashes and dentifrices. A mouthwash may be prepared incorporating the active ingredient in the required amount in an appropriate solvent, such as a sodium borate solution (Dobell's Solution). Alternatively, the active ingredient may be incorporated into an antiseptic wash containing sodium borate, glycerin and potassium bicarbonate. The active ingredient also may be dispersed in dentifrices, including: gels, pastes, powders and slurries. The active ingredient may be added in a therapeutically effective amount to a paste dentifrice that may include water, binders, abrasives, flavoring agents, foaming agents, and humectants.

The compositions of the present invention may be formulated in a neutral or salt form. Pharmaceutically-acceptable salts include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like.

Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as injectable solutions, drug release capsules and the like. For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, sterile aqueous media which can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage could be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject. Moreover, for human administration, preparations should meet sterility, pyrogenicity, general safety and purity standards as required by FDA Office of Biologics standards.

EXAMPLES

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

Example 1

Materials and Methods

Subjects. Signed informed-consent forms, approved by the Institutional Review Board at the University of Iowa and collaborating institutions, were obtained from all study participants on entry into the study. The diagnosis of BBS in the large Bedouin Arab family was based on the presence of at least two of the cardinal features of BBS (obesity, polydactyly, renal anomalies, retinopathy, hypogonadism, and learning disabilities). The diagnosis of BBS in individuals without a family history of the disorder was based on the presence of at least three of the cardinal features of BBS, most commonly obesity, polydactyly, and retinal degeneration. The clinical features of BBS in the Bedouin family

have been reported elsewhere (Kwitek-Black et al., 1993; Sheffield et al., 1994; Carmi et al., 1995).

Bioinformatic Analysis. Identification and prioritization of BBS candidate genes in this study used a computational comparative genomics technique with similarities to previous methods (Avidor-Reiss et al., 2004; Li et al., 2004). This approach is briefly described, in general terms, in this section, followed by a description of its implementation. Specific details of the application of this method to the study of BBS can be found in the "Results" section. A basic principle of this approach is that specific biological features are manifest in an evolutionarily conserved structural and/or functional characteristic shared among a subset of species. In addition, a subset of species is also identifiable that explicitly lacks this same characteristic.

Together, these two subsets of species serve as positive and negative filters, respectively. Employing computational sequence-similarity tools and available protein and genomic sequences, these filters are applied to a candidate set of genes to prioritize them for experimental analysis, such as mutation screening. FIG. 1 depicts this process, which is outlined as follows:

1. Initial determination of candidate gene set G_c .
2. Selection of positive and negative sets of species: S_+ and S_- .
3. Determination of thresholds for similarity filters: Th_+ and Th_- .
4. Computational application of filters S_+ and S_- to candidate gene set G_c .
 - a. Similarity analysis of G_c relative to S_+ , and retention of candidates exceeding Th_+ . This subset is referred to as " G_{c+} ".
 - b. Similarity analysis of relative to S_- , and rejection of candidates falling below Th_- , referred to as " $G_{c\pm}$ ".
5. Application of additional criteria, A , for candidate gene ranking.
 - a. Utilization of known or suspected linkage interval.
 - i. Intersection of the subset of genes in the interval, G_{c+A} , with G_{c+} and $G_{c\pm}$, referred to as " G_{c+A} " and " $G_{c\pm, A}$ ".
 - b. Utilization of available expression information.
 - c. Utilization of other annotations (e.g., Gene Ontology Consortium [GO] biological process, molecular function, cellular component, etc. [Ashburner et al., 2000]).

The implementation of this process for identification of BBS3 candidate genes employed BLAST (Altschul et al., 1990) as the similarity-analysis tool. The thresholds (Th_+ and Th_-) were expressed as BLAST e-values. These values were determined based on similarity scores of the known BBS genes. The test for was implemented as S_- a "less than" comparison with Th_+ ; the test for S_+ was a "greater than" comparison with Th_- . The linkage interval for BBS3 was applied as additional criteria A . The selection of sets S_+ and S_- and was based upon the presence or absence of ciliated structures, as further described in the Results section.

Data Sources. Genomic information was obtained from Ensembl (*Caenorhabditis elegans*, *Danio rerio*, *Drosophila melanogaster*, and *Mus musculus*), from the U.S. Department of Energy Joint Genome Institute (JGI) (*Ciona intestinalis* and *Chlamydomonas reinhardtii*), from The Institute for Genomic Research (TIGR) (*Trypanosoma brucei* and *T. cruzi*), from the *Saccharomyces* Genome Database (SGD) (*Saccharomyces cerevisiae*), from The *Arabidopsis* Information Resource (TAIR) (*Arabidopsis thaliana*), and from the Broad Institute (*Aspergillus nidulans*). Predicted

Ensembl genes for humans (Release 22.34a) were retrieved for human genomewide analysis.

Genotyping. PCR amplification for the analysis of STRPs was performed using 40 ng genomic DNA in 8.4 ml reactions containing 1.25 ml 10# PCR buffer (100 mM Tris-HCl [pH 8.8], 500 mM KCl, 15 mM MgCl₂, 0.01% gelatin [w/v]), 200 uM each of dATP, dCTP, dGTP and dTTP, 2.5 pmol of each primer, and 0.2 U of Taq polymerase. Samples were subjected to 35 cycles of 94° C. for 30 s; 50, 52, 55, or 57° C. (as required) for 30 sec; and 72° C. for 30 sec. Amplification products were electrophoresed on 6% polyacrylamide gels containing 7.7 M urea at 60 W for about 2 h. The bands were visualized by silver staining. Oligonucleotide primers for the STRPs were obtained as MapPairs (Research Genetics) or were custom designed and synthesized commercially.

DNA Sequencing and Mutation Screening. PCR products for sequencing were amplified in a 25-ml reaction volume and were visualized on a 1.2% agarose gel. The corresponding bands were excised and purified using the QIAquick gel extraction kit (Qiagen). 4.5 ml of purified PCR product was used as template for sequencing reactions using dye-terminator chemistry (Applied Biosystems). PCR product sequencing reactions were precipitated in the presence of glycogen and isopropanol. The reactions were analyzed on an ABI 3730 DNA Sequencer. All sequence variants were verified bidirectionally by direct DNA sequencing and/or by restriction enzyme digestion. Primer sequences used to screen the entire ARL6 gene are available in table A1 (online only).

In some cases, the coding sequence of candidate genes was screened by SSCP analysis. Amplicons for SSCP analysis were designed to be about 200 bp in size. For SSCP, PCR products were electrophoresed on SSCP gels (7 ml 50% glycerol, 3.5 ml 5# TBE, 8.8 ml 37.5:1 acrylamide/bisacrylamide, and 50.7 ml ddH₂O) for 3-4 h in 0.5# TBE at room temperature, with the temperature controlled by a cooling fan. The gels were silver stained to visualize DNA bands. Abnormal variants were sequenced and compared with a control sample (CEPH sample 1331-01) to detect any changes from that of the normal sequence.

Example 2

Results

Genetic Fine Mapping. The BBS3 pedigree from a large Israeli Bedouin kindred has been previously published (Sheffield et al., 1994). DNA samples were obtained from the 13 affected individuals in this pedigree and from numerous other individuals in this Bedouin population, including 21 unaffected siblings and 12 parents (some of whom were also siblings of affected individuals). To narrow the disease interval, genotyping with STRPs was performed across the linked interval, focusing on two affected individuals who were not homozygous for all markers across the genetic interval initially reported (Sheffield et al., 1994). Mapping of the homozygous interval in these two individuals allowed us to refine the disease interval to approximately 5.3 cM (FIG. 2). This interval proved to be a region of lower-than-average recombination, in part because the 16.9-Mb region between the flanking markers (D3S1595 and D3S3655) crosses the centromere. Analysis of the human genome (UCSC Genome Browser) across the BBS3 interval revealed a minimum of 67 UniGene clusters. In an attempt to further narrow the BBS3 interval, the inventors genotyped 54 small kindreds, including many that had a single affected individual, looking

for families consistent with linkage to the BBS3 locus and/or for isolated affected individuals that were homozygous across this interval. Only one small pedigree was identified that was consistent with linkage on the basis of a single homozygous affected individual, indicating that BBS3 is a rare cause of BBS.

Comparative Genomic Analysis to Identify BBS Candidate Genes. The existence of several known BBS genes, previously identified by positional cloning, provided the opportunity to use bioinformatics methods to identify additional BBS genes. Since prior studies of BBS4 (Kim et al., 2004), BBS8 (Ansley et al., 2003), and the BBS4 mouse model (Mykityn et al., 2004) implicated the involvement of BBS proteins in ciliary function, the inventors included within the set of genomes to be analyzed, the genomes of well-studied eukaryotic ciliated organisms (S_+), as well as some non-ciliated organisms (S_-) for which complete genome sequence was available.

The value of α was determined on the basis of Th_+ BLAST analysis of five known BBS proteins (BBS1, BBS2, BBS4, BBS7, and BBS8) against available genomic sequence from ciliated organisms, including nematode, zebrafish, fruit fly, laboratory mouse, sea squirt, biflagellated green algae, and flagellated parasites. Putative BBS orthologues were identified for many but not all organisms at a significance level of $\leq e^{-35}$. The genome of *M. musculus* and *D. rerio* contain BBS orthologues with highly significant expected values, sequence percent identity (>63%), and similarity (>75%) to five known BBS proteins. The genome of *D. melanogaster* contains only orthologues to BBS1, BBS4, and BBS8, whereas the remaining ciliated organisms (*T. brucei*, *T. cruzi*, *C. reinhardtii*, and *C. intestinalis*) showed significant expected values ($\leq e^{-40}$), percent identity (>10%), and similarity (>40%) to BBS1, BBS2, BBS4, BBS7, and BBS8. The inventors reasoned that the smaller genomes (compared with those of humans) of organisms such as *T. brucei*, *T. cruzi*, *C. reinhardtii*, *C. intestinalis*, and *C. elegans*, which contain orthologues to BBS genes, should be enriched for additional BBS genes. To further enhance for cilia-related genes, multiple ciliated organisms (*T. brucei*, *T. cruzi*, *C. reinhardtii*, and *C. intestinalis*) were chosen for comparison with the human genome, to define a set of BBS candidate genes. Similarly, the genomes of *S. cerevisiae* and *A. thaliana*, which do not contain putative orthologues of the known BBS genes (S_- organisms), S_- were selected to aid in exclusion of genes as potential BBS candidates. A Th_- significance level of $\leq e^{-35}$ was selected on the basis of these comparisons.

The central step in the bioinformatics methods was the application of the positive and negative filter sets (S_+ and S_- ; Th_+ and Th_-). A BLAST analysis was performed of the set of 21,184 candidate genes (G_c), as identified in Ensembl (Curwen et al., 2004) against the genomes of *C. intestinalis*, *C. reinhardtii*, *T. brucei*, and *T. cruzi*. These comparisons resulted in the identification of 1,588 genes (G_{c+}) in common between the known human genome (G_c) and these model organisms (S_+), as shown in Table 8. To remove genes that were less likely to be BBS genes, the inventors compared this set of 1,588 genes against the S_- genomes of *S. cerevisiae* and *A. thaliana* (two genomes that do not contain orthologues to the known BBS genes). This analysis resulted in a refined set of 114 BBS candidates ($G_{c\pm}$) that were present in the S_+ genomes of *C. intestinalis*, *C. reinhardtii*, *T. brucei*, and *T. cruzi* but were absent in the genomes of *S. cerevisiae* and *A. thaliana*. None of these 114 genes mapped to the BBS3 interval, as shown in Table 8. However, two genes (Hs.388438 and Hs.446399) mapped to the BBS5

locus on chromosome 2q31. During the course of this study, the BBS5 gene was identified by others using a similar comparative genomic strategy (Li et al., 2004). It is of note that the BBS5 gene was one of the two BBS5 candidate genes (Hs.446399) identified in our study. The complete list of 1,588 genes in the G_{c+} set and the 114 genes in the $G_{c\pm}$ set are available at the authors' Web site.

TABLE 8

	Ensembl Genes from the Comparative Genomics Approach		
	All Ensembl Genes	S_+ Filter Only	S_+ and S_- Filters
All chromosomes	21,184 (G_c)	1,588 (G_{c+})	114 ($G_{c\pm}$)
BBS3 interval	62 (G_{cA})	4 (G_{c+A})	0 ($G_{c\pm A}$)

In addition to the whole human genome analysis described above, the inventors performed an intersection of the 1,588 G_{c+} candidates with the 62 Ensembl genes contained within the human BBS3-linked interval. This analysis identified four BBS3 candidate genes (G_{c+A}), as shown in Table 8. The inventors then prioritized the four G_{c+A} genes for mutation analysis using two criteria: (1) a broad tissue pattern of expression similar to that of known BBS genes, on the basis of representation in dbEST, and (2) putative function, on the basis of GO data. On the basis of these criteria, candidate genes from the BBS3-linked interval, including a gene known as "ARL6," were selected for mutation analysis. ARL6 (ADP-ribosylation factor [ARF]-like 6) is a member of a subgroup of the ARF family, proteins that regulate diverse cellular functions including regulation of intracellular traffic (Kahn and Gilman 1984; Price et al., 1988; Sewell and Kahn 1988; Pasqualato et al., 2002).

Mutation Analysis in the Bedouin Family. The coding sequence and splice sites from BBS3 candidate genes identified in the phylogenetic comparisons were initially sequenced in DNA amplified from two affected members of the BBS3-linked kindred. No nonsynonymous sequence variants were identified in the coding sequence of Hs.145172, Hs.16986 (FLJ11046), Hs.370688 (TOMM70A), Hs.23294 (MINA), Hs.369885, and Hs.223341. However, sequencing of ARL6 revealed a homozygous stop mutation in exon 7 (R122X) resulting in a predicted truncation of the protein from 186 amino acids to 121 amino acids. This mutation segregates completely with BBS in the Bedouin kindred in that all 13 affected individuals were homozygous for the mutation and each obligate carrier (12 parents of affected individuals) was heterozygous for the mutation. Twenty-one unaffected siblings of BBS patients either were heterozygous for the mutation or were homozygous for the normal allele (FIG. 3). To determine whether R122X is a rare polymorphism, the inventors genotyped 100 Arab control individuals (200 chromosomes) from the Middle East and 90 additional control individuals (180 chromosomes) from diverse ethnicities. No R122X alleles were detected in these 380 chromosomes. To exclude the possibility that a short, functional ARL6 isoform exists that excludes exon 7 (the exon containing the R122X mutation) as a functional component, the inventors analyzed the 47 available ARL6 ESTs in dbEST. All ARL6 isoforms identified in this analysis include the complete exon 7 sequence.

In addition to the mutation analysis performed in the BBS3 Bedouin kindred, the inventors performed mutation screening of 90 unrelated BBS patients, including a single

patient that was shown to be homozygous across the BBS3 genetic interval. No nonsynonymous coding sequences were identified, consistent with previous genotyping data that indicate that BBS3 is a rare cause of the disorder.

Example 3

Discussion

The identification of the BBS3 gene presented challenges, because of the extensive genetic heterogeneity of BBS, the paucity of families mapping to the BBS3 locus, and the diversity of BBS genes. Analysis of the BBS3 genetic interval in an attempt to find candidate genes, on the basis of homology to the known BBS genes, failed to produce attractive candidate genes. Genetic mapping narrowed the disease interval to a region containing at least 62 genes. To identify and prioritize the best BBS candidate genes, the inventors used a comparative genomics approach to select genes shared by organisms harboring orthologues to known BBS genes. Sequencing of DNA from BBS patients in the large Bedouin kindred that was initially used to map the BBS3 locus, revealed that each affected individual harbored a homozygous stop mutation (R122X) in one of the genes identified in this manner, ARL6. R122X results in the predicted truncation of the protein from 186 to 121 residues, including many highly conserved amino acids (FIG. 4) (SEQ ID NOS:5-16). These data convincingly demonstrate that ARL6 is the BBS3 gene. The identification of mutations only in the large previously linked BBS3 Bedouin family is consistent with previous reports of the rarity of BBS3 as a cause of this syndrome (Bruford et al., 1997). The reported tissue expression pattern of ARL6 (Jacobs et al., 1999) and analysis of ESTs from dbEST indicate that ARL6 is widely expressed in human tissues including brain, eye, heart, and kidney, further supporting this gene as a cause of BBS.

The computational comparative genomics approach used in our study is similar to the approach used recently to identify the BBS5 gene (Li et al., 2004). In that study, genes that were conserved in the genome from the model organism *C. reinhardtii* and not in *A. thaliana* and that mapped to the BBS5 locus were selected as BBS candidate genes. Of note, the BBS5 gene was found within the comparative genomics set of 114 G_{c+} genes in our study. When this set of 114 genes was intersected with the set of 62 Ensembl genes in the linked BBS3 interval, no genes were identified. It is possible that different choices for S_{-} and Th_{-} for would have resulted in a nonempty set of candidates in the set $G_{c+,A}$. However, one of the strengths of this method is the ability to employ additional criteria, such as a known linked interval. As can be seen in Table 8, both dimensions of filtering allow a distillation of candidates to a number, which then may be individually examined for functional annotation and/or direct mutation screening. Since other BBS genes likely exist, the comparative genomics approach should prove useful in identifying additional genes causing rare cases of BBS.

Little is known about the pathophysiology of BBS and the specific functions of BBS proteins. BBS6 is caused by mutations in the MKKS gene (Katsanis et al., 2000; Slavotinek et al., 2000), mutations that also cause McKusick-Kaufman syndrome (hydrometrocolpos, post-axial polydactyly, and congenital heart defects) (Robinow and Shaw, 1979; Stone et al., 2000). MKKS has sequence homology to the a subunit of the *Thermoplasma acidophilum* thermosome (Stone et al., 2000), a prokaryotic chaperonin complex with similarity to a eukaryotic chaperonin

called “tailless complex polypeptide ring complex” (TRiC) (Frydman et al., 1992). However, the other known BBS proteins have no significant similarity to chaperonins or other proteins of known function. BBS4 and BBS8 contain tetratricopeptide repeat (TPR) domains, indicating that they may interact with other proteins. Recently, it was hypothesized that BBS8 and other BBS proteins are involved in cilia function, on the basis of localization of BBS8 to the basal body of ciliated cells (Ansley et al., 2003). In addition, BBS4 was recently shown to localize to the centriolar satellites of centrosomes and basal bodies of primary cilia where it interacts with components of the dynein transport machinery (Kim et al., 2004). The inventors recently demonstrated that BBS4 knockout mice have features of the human disorder and that absence of this protein leads to failure of spermatozoa flagellar formation but not to the failure of the formation of cilia in general (Mykytyn et al., 2004). Furthermore, the inventors demonstrated that absence of this protein did not disrupt initial formation of photoreceptor outer segments, including the connecting cilia; rather, photoreceptors underwent cell death due to apoptosis. Collectively, these results support the hypothesis that BBS proteins are involved in ciliary function and/or intracellular transport.

The comparative genomic analysis of the known BBS genes performed to identify the BBS3 gene provides some insight into the function of BBS genes. A striking finding is that, in general, ciliated organisms have orthologous sequences to BBS proteins, whereas nonciliated organisms do not (Mykytyn et al., 2004). For example, even lower, single-cell ciliated organisms—such as *T. brucei*, *T. cruzi*, and *C. reinhardtii*—have homologous sequences to the known BBS proteins. The ciliated organism *C. intestinalis* also has homologous sequences to BBS proteins. In contrast, organisms without cilia (e.g., *S. cerevisiae*, *S. pombe*, *A. nidulans*, and *A. thaliana*) do not have sequences that are homologous to BBS proteins. These data support a role for BBS genes in cilia function or some other specialized function common to ciliated organisms. This hypothesis is supported by two previous studies, in which comparative genomics were used to identify genes found in ciliated organisms (Avidor-Reiss et al., 2004; Li et al., 2004). The finding that some organisms do not have orthologues to BBS genes indicates that the BBS genes do not play a role in a process required of all eukaryotic cells—or, at a minimum, that BBS genes add specificity to a general cellular process found in all eukaryotic organisms. This finding is consistent with the BBS genes playing a role in ciliary function and/or a role in intracellular transport.

Although the precise function of ARL6 is not known (Ingleby et al., 1999; Jacobs et al., 1999; Pasqualato et al., 2002), the identification of this gene as the BBS3 gene supports a general role for BBS genes in intracellular transport. The ADP-ribosylation factor group of proteins is part of the Ras superfamily, which is subdivided into ARF and ARF-like (ARL) subgroups (Pasqualato et al., 2002). ARF proteins were initially identified because of their ability to stimulate ADP-ribosyltransferase activity of cholera toxin (Kahn and Gilman 1984; Price et al., 1988; Sewell and Kahn 1988; Pasqualato et al., 2002). The ARL proteins were identified on the basis of their similarity to ARF proteins. These proteins are thought to have diverse roles, including regulation of intracellular vesicle and membrane trafficking and microtubule assembly (Bhamidipati et al., 2000; Pasqualato et al., 2002). Further elucidation of the precise function of ARL6 and the other BBS proteins, as well as their interactions, should provide important clues to the

pathophysiology of diverse phenotypes, including obesity, diabetes, and retinal degeneration.

All of the composition and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference:

U.S. Pat. No. 3,817,837
 U.S. Pat. No. 5,021,236
 U.S. Pat. No. 3,850,752
 U.S. Pat. No. 3,939,350
 U.S. Pat. No. 3,996,345
 U.S. Pat. No. 4,196,265
 U.S. Pat. No. 4,275,149
 U.S. Pat. No. 4,277,437
 U.S. Pat. No. 4,366,241
 U.S. Pat. No. 4,472,509
 U.S. Pat. No. 4,554,101
 U.S. Pat. No. 4,664,911
 U.S. Pat. No. 4,683,195
 U.S. Pat. No. 4,683,202
 U.S. Pat. No. 4,786,600
 U.S. Pat. No. 4,792,447
 U.S. Pat. No. 4,800,159
 U.S. Pat. No. 4,873,191
 U.S. Pat. No. 4,883,750
 U.S. Pat. No. 4,946,773
 U.S. Pat. No. 4,988,617
 U.S. Pat. No. 5,045,451
 U.S. Pat. No. 5,190,856
 U.S. Pat. No. 5,270,184
 U.S. Pat. No. 5,279,721
 U.S. Pat. No. 5,324,631
 U.S. Pat. No. 5,484,699
 U.S. Pat. No. 5,494,810
 U.S. Pat. No. 5,496,699
 U.S. Pat. No. 5,578,706
 U.S. Pat. No. 5,633,365
 U.S. Pat. No. 5,639,611
 U.S. Pat. No. 5,665,549
 U.S. Pat. No. 5,686,072
 U.S. Pat. No. 5,712,124
 U.S. Pat. No. 5,733,733
 U.S. Pat. No. 5,733,752
 U.S. Pat. No. 5,744,311
 U.S. Pat. No. 5,747,255
 U.S. Pat. No. 5,767,072
 GB Application 2 202 328
 EPO No. 320 308

EPO No. 329 822
 PCT/US87/00880
 PCT/US89/01025
 PCT Appln. WO 89/06700
 5 PCT Appln. WO 84/03564
 PCT Appln. WO 88/10315
 PCT Appln. WO 89/06700
 PCT Appln. WO 90/07641
 PCT Appln. WO 01/00825
 10 PCT Appln. WO 88/10315
 Altschul et al., *J. Mol. Biol.*, 215:403-410, 1990.
 Ansley et al., *Nature*, 425:628-633, 2003.
 Arcone et al., *Nucl. Acids Res.*, 16(8): 3195-3207, 1988.
 Ashburner et al., *Nat. Genet.*, 25:25-29, 2000.
 15 Avidor-Reiss et al., *Cell*, 117:527-539, 2004.
 Badano et al., *Am. J. Hum. Genet.*, 72:650-658, 2003.
 Baichwal and Sugden, In: *Gene Transfer*, 117-148, 1986.
 Barany and Merrifield, *The Peptides*, 1-284, 1979.
 Bardet, Sur un syndrome d'obesite infantile avec polydac-
 20 tylie et retinite pigmentaire (contribution a l'etude des
 formes cliniques de l'obesite hypophysaire), Thesis, Uni-
 versity of Paris, Paris, 1920.
 Bartlett et al., *Proc. Natl. Acad. Sci. USA*, 93:8852-8857,
 1996.
 25 Benvenisty and Neshif, *Proc. Natl. Acad. Sci. USA*,
 83:9551-9555, 1986.
 Bhamidipati et al., *J. Cell Biol.*, 149:1087-1096, 2000.
 Biedl, *Dtsch. Med. Wschr.*, 48:1630, 1922.
 Brinster et al., *Proc. Natl. Acad. Sci. USA*, 82: 4438-4442,
 30 1985.
 Bruford et al., *Genomics*, 41:93-99, 1997.
 Campbell, In: *Monoclonal Antibody Technology, Laboratory
 Techniques in Biochemistry and Molecular Biology*,
 13:75-83, Amsterdam, Elsevier, 1984.
 35 Capaldi et al., *Biochem. Biophys. Res. Comm.*, 76:425, 1977.
 Carmi et al., *Am. J. Med. Genet.*, 59:199-203, 1995.
 Carter and Flotte, *Curr. Top Microbiol. Immunol.*, 218:119-
 144, 1996.
 Chatterjee et al., *Ann. N.Y. Acad. Sci.*, 770:79-90, 1995.
 40 Chen and Okayama, *Mol. Cell Biol.*, 7:2745-2752, 1987.
 Coffin, In: *Virology*, Fields et al. (Eds.), Raven Press, NY,
 1437-1500, 1990.
 Coupar et al., *Gene*, 68:1-10, 1988.
 Curwen et al., *Genome Res.*, 14:942-950, 2004.
 45 Dani et al., *Mol. Cell Endocrinol.*, 63(1-2):199-208, 1989.
 Dubensky et al., *Proc. Natl. Acad. Sci. USA*, 81:7529-7533,
 1984.
 Elbedour et al., *Am. J. Med. Genet.*, 52:164-169, 1994.
 Fechheimer et al., *Proc. Natl. Acad. Sci. USA*, 84:8463-
 50 8467, 1987.
 Ferkol et al., *FASEB J.*, 7:1081-1091, 1993.
 Ferrari et al., *J. Virol.*, 70:3227-3234, 1996.
 Fisher et al., *J. Virol.*, 70:520-532, 1996.
 Flotte et al., *Proc. Natl. Acad. Sci. USA*, 90:10613-10617,
 55 1993.
 Fodor et al., *Science*, 251:767-773, 1991.
 Fraley et al., *Proc. Natl. Acad. Sci. USA*, 76:3348-3352,
 1979.
 Freifelder, In: *Physical Biochemistry Applications to Bio-
 60 chemistry and Molecular Biology*, 2nd Ed. Wm. Freeman
 and Co., N.Y., 1982.
 Frohman, In: *PCR Protocols: A Guide To Methods And
 Applications*, Academic Press, N.Y., 1990.
 Frydman et al., *Embo J.*, 11:4767-4778, 1992.
 65 Gefter et al., *Somatic Cell Genet*, 3(2):231-6, 1977.
 Ghosh and Bachhawat, In: *Liver Diseases, Targeted Diag-
 nosis and Therapy Using Specific Receptors and Ligands*.

Goding, 1986, In: *Monoclonal Antibodies: Principles and Practice*, pp. 60-61, and 71-74, 1986.

Goodman et al., *Blood*, 84:1492-1500, 1994.

Gopal, *Mol. Cell Biol.*, 5:1188-1190, 1985.

Gossen and Bujard, *Proc. Natl. Acad. Sci. USA*, 89:5547-5551, 1992.

Gossen et al., *Science*, 268:1766-1769, 1995.

Graham and van der Eb, *Virology*, 52:456-467, 1973.

Green et al., *N. Engl. J. Med.*, 321:1002-1009, 1989.

Hacia et al., *Nature Genetics*, 14:441-447, 1996.

Harland and Weintraub, *J. Cell Biol.*, 101:1094-1099, 1985.

Harnett et al., *N. Engl. J. Med.*, 319:615-618, 1988.

Hay et al., *J. Mol. Biol.*, 175:493-510, 1984.

Hearing and Shenk, *J. Mol. Biol.* 167:809-822, 1983.

Hearing et al., *J. Virol.*, 67:2555-2558, 1987.

Hunt et al., *Proc. Natl. Acad. Sci. USA*, 83(11):3786-3790, 1986.

Ingleby et al., *FEBS Lett.*, 459:69-74, 1999.

Innis et al., *Proc. Natl. Acad. Sci. USA*, 85(24):9436-9440, 1988.

Jacobs et al., *FEBS Lett.*, 456:384-388, 1999.

Johnson et al., In: *Biotechnology And Pharmacy*, Pezzuto et al. (Eds.), Chapman and Hall, N.Y., 1993.

Joki et al., *Human Gene Ther.*, 6:1507-1513, 1995.

Kageyama et al., *J. Biol. Chem.*, 262(5):2345-2351, 1987.

Kahn and Gilman, *J. Biol. Chem.*, 259:6228-6234, 1984.

Kaneda et al., *Science*, 243:375-378, 1989.

Kaplitt et al., *Natl. Genet.*, 8:148-153, 1994.

Kato et al., *J. Biol. Chem.*, 266:3361-3364, 1991.

Katsanis et al., *Nat. Genet.*, 26:67-70, 2000.

Katsanis et al., *Science*, 293:2256-2259, 2001.

Kessler et al., *Proc. Natl. Acad. Sci. USA*, 93:14082-14087, 1996.

Kim et al., *Nat. Genet.*, 36:462-470, 2004.

Klein et al., *Nature*, 327:70-73, 1987.

Koeberl et al., *Proc. Natl. Acad. Sci. USA*, 94:1426-1431, 1997.

Kohler and Milstein, *Eur. J. Immunol.*, 6:511-519, 1976.

Kohler and Milstein, *Nature*, 256:495-497, 1975.

Korhonen et al., *Blood*, 86(5):1828-1835, 1995.

Kwitek-Black et al., *Nat. Genet.*, 5:392-396, 1993.

Kwoh et al., *Proc. Natl. Acad. Sci. USA*, 86: 1173, 1989.

Kyte and Doolittle, *J. Mol. Biol.*, 157(1):105-32, 1982.

Lander et al., *Nature*, 409:860-921, 2001.

Levrero et al., *Gene*, 101: 195-202, 1991.

Li et al., *Cell*, 117:541-552, 2004.

Manipulating the Mouse Embryo, A Laboratory Manual, 2nd Edition Hogan, Beedington, Costantini and Long, (Eds), Cold Spring Harbor Laboratory Press, 1994.

Mann et al., *Cell*, 33:153-159, 1983.

McCown et al., *Brain Res.*, 713:99-107, 1996.

Merrifield, *Science*, 232: 341-347, 1986.

Mizukami et al., *Virology*, 217:124-130, 1996.

Myers, EPO 0273085

Mykytyn et al., *Nat. Genet.*, 28:188-191, 2001.

Mykytyn et al., *Nat. Genet.*, 31:435-438, 2002.

Mykytyn et al., *Proc. Natl. Acad. Sci. USA*, 101:8664-8669, 2004.

Nakamura et al., In: *Handbook of Experimental Immunology* (4th Ed.), Weir et al. (Eds), 1:27, Blackwell Scientific Publ., Oxford, 1987.

Nicolau and Rubenstein, In: *Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez and Denhardt (Eds.), Stoneham: Butterworth, 494-513, 1988.

Nicolau and Sene, *Biochim. Biophys. Acta*, 721:185-190, 1982.

Nicolau et al., *Methods Enzymol.*, 149:157-176, 1987.

Nishimura et al., *Hum. Mol. Genet.*, 10:865-874, 2001.

Ohara et al., *Proc. Natl. Acad. Sci. USA*, 86: 5673-5677, 1989.

Oliviero et al., *EMBO J*, 6(7):1905-1912, 1987.

Pape and Kim, *Mol. Cell Biol.*, 9(3):974-982, 1989.

Paskind et al., *Virology*, 67:242-248, 1975.

Pasqualato et al., *EMBO Rep.*, 3:1035-1041, 2002.

Pease et al., *Proc. Natl. Acad. Sci. USA*, 91:5022-5026, 1994.

Perales et al., *Proc. Natl. Acad. Sci. USA*, 91:4086-4090, 1994.

Pignon et al., *Hum. Mutat.*, 3: 126-132, 1994.

Ping et al., *Microcirculation*, 3:225-228, 1996.

Poli and Cortese, *Proc. Natl. Acad. Sci. USA*, 86:8202-8206, 1989.

Potter et al., *Proc. Natl. Acad. Sci. USA*, 81:7161-7165, 1984.

Price et al., *Proc. Natl. Acad. Sci. USA*, 85:5488-5491, 1988.

Prowse and Baumann, *Mol. Cell Biol.*, 8(1):42-51, 1988.

Radler et al., *Science*, 275:810-814, 1997.

Remington's Pharmaceutical Sciences, 15th ed., pages 1035-1038 and 1570-1580, Mack Publishing Company, Easton, Pa., 1980.

Ridgeway, In: *Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez and Denhardt (Eds.), Stoneham: Butterworth, 467-492, 1988.

Rippe et al., *Mol. Cell Biol.*, 10:689-695, 1990.

Robinow and Shaw, *J. Pediatr.*, 94:776-778, 1979.

Ron et al., *Mol. Cell Biol.*, 2887-2895, 1991.

Roux et al., *Proc. Natl. Acad. Sci. USA*, 86:9079-9083, 1989.

Sambrook et al., In: *Molecular cloning*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

Samulski et al., *J. Virol.*, 61(10):3096-3101, 1987.

Sewell and Kahn, *Proc. Natl. Acad. Sci. USA*, 85:4620-4624, 1988.

Sheffield et al., *Hum. Mol. Genet.*, 3:1331-1335, 1994.

Shoemaker et al., *Nature Genet.*, 14:450-456, 1996.

Slavotinek et al., *Nature Genet.*, 26:15-16, 2000.

Spiegelman et al., *J. Biol. Chem.*, 264(3):1811-1815, 1989.

Stewart and Young, *Solid Phase Peptide Synthesis*, 2d Ed., Pierce Chemical Co., 1984.

Stone et al., *Nat. Genet.*, 25:79-82, 2000.

Tam et al., *J. Am. Chem. Soc.*, 105:6442, 1983.

Temin, In: *Gene Transfer*, Kucherlapati (Ed.), N.Y., Plenum Press, 149-188, 1986.

Tibbetts, *Cell*, 12:243-249, 1977.

Tur-Kaspa et al., *Mol. Cell Biol.*, 6:716-718, 1986.

Venter et al., *Science*, 291:1304-1351, 2001.

Wagner et al., *Proc. Natl. Acad. Sci. USA*, 87:9:3410-3414, 1990.

Walker et al., *Proc. Natl. Acad. Sci. USA*, 89:392-396, 1992.

Walther and Stein, *J. Mol. Med.*, 74:379-392, 1996.

Watt et al., *Proc. Natl. Acad. Sci. USA*, 83(2): 3166-3170, 1986.

Wilson et al., *Mol. Cell Biol.*, 6:6181-6191, 1990.

Wong et al., *Gene*, 10:87-94, 1980.

Wu and Wu, *Adv. Drug Delivery Rev.*, 12:159-167, 1993.

Wu and Wu, *Biochem.*, 27:887-892, 1988.

Wu and Wu, *J. Biol. Chem.*, 262:4429-4432, 1987.

Wu et al., *Genomics*, 4:560, 1989.

Xiao et al., *J. Virol.*, 70:8098-8108, 1996.

Yang et al., *Proc. Natl. Acad. Sci. USA*, 87:9568-9572, 1990.

Zechner et al., *Mol. Cell Biol.*, 2394-2401, 1988.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 16

<210> SEQ ID NO 1
 <211> LENGTH: 561
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 Primer

<400> SEQUENCE: 1

```
atgggattgc tagacagact ttcagtcttg cttggcctga agaagaagga ggttcattgtt    60
ttgtgccttg ggctagataa tagtggcaaa acgacgatca ttaacaaact taaaccttca    120
aatgtcaat ctcaaaatatt ccttccaaca ataggattca gcatagagaa attcaaatca    180
tccagtttgt catttacagt gtttgacatg tcaggtcaag gaagatacag aaatctctgg    240
gaacactatt ataaagaagg ccaagctatt atttttgtca ttgatagtag tgatagatta    300
agaatggttg tggccaaaga agaactcgat actcttctga atcatccaga tattaaacac    360
cgtcgaatc caatcttatt ctttgcaaat aaaatggatc ttagagatgc agtgacatct    420
gtaaaagtgt ctcagttgct gtgtttagag aacatcaaag ataaacctg gcatatttgt    480
gctagtgatg ccataaaagg agaaggcttg caagaaggty tagactggct tcaagatcag    540
atccagactg tgaagacatg a                                         561
```

<210> SEQ ID NO 2
 <211> LENGTH: 186
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 peptide

<400> SEQUENCE: 2

```
Met Gly Leu Leu Asp Arg Leu Ser Val Leu Leu Gly Leu Lys Lys Lys
  1           5           10           15
Glu Val His Val Leu Cys Leu Gly Leu Asp Asn Ser Gly Lys Thr Thr
  20           25           30
Ile Ile Asn Lys Leu Lys Pro Ser Asn Ala Gln Ser Gln Asn Ile Leu
  35           40           45
Pro Thr Ile Gly Phe Ser Ile Glu Lys Phe Lys Ser Ser Ser Leu Ser
  50           55           60
Phe Thr Val Phe Asp Met Ser Gly Gln Gly Arg Tyr Arg Asn Leu Trp
  65           70           75           80
Glu His Tyr Tyr Lys Glu Gly Gln Ala Ile Ile Phe Val Ile Asp Ser
  85           90           95
Ser Asp Arg Leu Arg Met Val Val Ala Lys Glu Glu Leu Asp Thr Leu
  100          105          110
Leu Asn His Pro Asp Ile Lys His Arg Arg Ile Pro Ile Leu Phe Phe
  115          120          125
Ala Asn Lys Met Asp Leu Arg Asp Ala Val Thr Ser Val Lys Val Ser
  130          135          140
Gln Leu Leu Cys Leu Glu Asn Ile Lys Asp Lys Pro Trp His Ile Cys
  145          150          155          160
Ala Ser Asp Ala Ile Lys Gly Glu Gly Leu Gln Glu Gly Val Asp Trp
  165          170          175
```

-continued

Leu Gln Asp Gln Ile Gln Thr Val Lys Thr
180 185

<210> SEQ ID NO 3
<211> LENGTH: 582
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
Primer

<400> SEQUENCE: 3

```
atgggattgc tagacagact ttcagtcttg cttggcctga agaagaagga ggttcattgtt    60
ttgtgccttg ggctagataa tagtggcaaa acgacgatca ttaacaaact taaaccttca    120
aatgtccaat ctcaaaatat ccttccaaca ataggattca gcatagagaa attcaaatca    180
tccagtttgt catttacagt gtttgacatg tcaggccaag gaagatacag aaatctctgg    240
gaacactatt ataaagaagg ccaagctatt atttttgtca ttgatagtag tgatagatta    300
agaatggttg tggccaaaga agaactcgtat actcttctga atcatccaga tattaaacac    360
cgtcgaattc caatcttatt ctttgcaaat aaaatggatc ttagagatgc agtgacatct    420
gtaaaagtgt ctcaagttgt gtgtttagag aacatcaaag ataaaccttg gcatatttgt    480
gctagtgatg ccataaaagg agaagccttg caagaaggtg tagactggct tcaagaaaaa    540
acaatacaat cagatccaga ctgtgaagac atgaaaagat aa                        582
```

<210> SEQ ID NO 4
<211> LENGTH: 193
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 4

```
Met Gly Leu Leu Asp Arg Leu Ser Val Leu Leu Gly Leu Lys Lys Lys
  1           5           10           15
Glu Val His Val Leu Cys Leu Gly Leu Asp Asn Ser Gly Lys Thr Thr
          20           25           30
Ile Ile Asn Lys Leu Lys Pro Ser Asn Ala Gln Ser Gln Asn Ile Leu
          35           40           45
Pro Thr Ile Gly Phe Ser Ile Glu Lys Phe Lys Ser Ser Ser Leu Ser
          50           55           60
Phe Thr Val Phe Asp Met Ser Gly Gln Gly Arg Tyr Arg Asn Leu Trp
          65           70           75           80
Glu His Tyr Tyr Lys Glu Gly Gln Ala Ile Ile Phe Val Ile Asp Ser
          85           90           95
Ser Asp Arg Leu Arg Met Val Val Ala Lys Glu Glu Leu Asp Thr Leu
          100          105          110
Leu Asn His Pro Asp Ile Lys His Arg Arg Ile Pro Ile Leu Phe Phe
          115          120          125
Ala Asn Lys Met Asp Leu Arg Asp Ala Val Thr Ser Val Lys Val Ser
          130          135          140
Gln Leu Leu Cys Leu Glu Asn Ile Lys Asp Lys Pro Trp His Ile Cys
          145          150          155          160
Ala Ser Asp Ala Ile Lys Gly Glu Gly Leu Gln Glu Gly Val Asp Trp
          165          170          175
Leu Gln Glu Lys Thr Ile Gln Ser Asp Pro Asp Cys Glu Asp Met Lys
```

-continued

180 185 190

Arg

<210> SEQ ID NO 5
 <211> LENGTH: 186
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide

<400> SEQUENCE: 5

Met Gly Leu Leu Asp Arg Leu Ser Val Leu Leu Gly Leu Lys Lys Lys
 1 5 10 15

Glu Val His Val Leu Cys Leu Gly Leu Asp Asn Ser Gly Lys Thr Thr
 20 25 30

Ile Ile Asn Lys Leu Lys Pro Ser Asn Ala Gln Ser Gln Asn Ile Leu
 35 40 45

Pro Thr Ile Gly Phe Ser Ile Glu Lys Phe Lys Ser Ser Ser Leu Ser
 50 55 60

Phe Thr Val Phe Asp Met Ser Gly Gln Gly Arg Tyr Arg Asn Leu Trp
 65 70 75 80

Glu His Tyr Tyr Lys Glu Gly Gln Ala Ile Ile Phe Val Ile Asp Ser
 85 90 95

Ser Asp Arg Leu Arg Met Val Val Ala Lys Glu Glu Leu Asp Thr Leu
 100 105 110

Leu Asn His Pro Asp Ile Lys His Arg Arg Ile Pro Ile Leu Phe Phe
 115 120 125

Ala Asn Lys Met Asp Leu Arg Asp Ala Val Thr Ser Val Lys Val Ser
 130 135 140

Gln Leu Leu Cys Leu Glu Asn Ile Lys Asp Lys Pro Trp His Ile Cys
 145 150 155 160

Ala Ser Asp Ala Ile Lys Gly Glu Gly Leu Gln Glu Gly Val Asp Trp
 165 170 175

Leu Gln Asp Gln Ile Gln Thr Val Lys Thr
 180 185

<210> SEQ ID NO 6
 <211> LENGTH: 186
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide

<400> SEQUENCE: 6

Met Val Leu Leu Asp Arg Leu Ser Gly Leu Leu Gly Leu Lys Lys Lys
 1 5 10 15

Glu Val His Val Leu Cys Leu Gly Leu Asp Asn Ser Gly Lys Thr Thr
 20 25 30

Ile Ile Asn Lys Leu Lys Pro Ser Asn Ala Gln Ser Gln Asp Ile Val
 35 40 45

Pro Thr Ile Gly Phe Ser Ile Glu Lys Phe Lys Ser Ser Ser Leu Ser
 50 55 60

Phe Thr Val Phe Asp Met Ser Gly Gln Gly Arg Tyr Arg Asn Leu Trp
 65 70 75 80

Glu His Tyr Tyr Lys Asp Gly Gln Ala Ile Ile Phe Val Ile Asp Ser
 85 90 95

-continued

```

Ser Asp Lys Leu Arg Met Val Val Ala Lys Glu Glu Leu Asp Thr Leu
    100                               105                               110

Leu Asn His Pro Asp Ile Lys His Arg Arg Ile Pro Ile Leu Phe Phe
    115                               120                               125

Ala Asn Lys Met Asp Leu Arg Asp Ser Val Thr Ser Val Lys Val Ser
    130                               135                               140

Gln Leu Leu Cys Leu Glu Ser Ile Lys Asp Lys Pro Trp His Ile Cys
    145                               150                               155                               160

Ala Ser Asp Ala Ile Lys Gly Glu Gly Leu Gln Glu Gly Val Asp Trp
    165                               170                               175

Leu Gln Asp Gln Ile Gln Ala Val Lys Thr
    180                               185

```

```

<210> SEQ ID NO 7
<211> LENGTH: 186
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
    Peptide

```

```

<400> SEQUENCE: 7

```

```

Met Gly Leu Leu Asp Arg Leu Ser Gly Leu Leu Gly Leu Lys Lys Lys
    1                               5                               10                               15

Glu Val His Val Leu Cys Leu Gly Leu Asp Asn Ser Gly Lys Thr Thr
    20                               25                               30

Ile Ile Asn Lys Leu Lys Pro Ser Asn Ala Gln Val Gln Asp Ile Val
    35                               40                               45

Pro Thr Ile Gly Phe Ser Ile Glu Lys Phe Lys Ser Ser Ser Leu Ser
    50                               55                               60

Phe Thr Val Phe Asp Met Ser Gly Gln Gly Arg Tyr Arg Asn Leu Trp
    65                               70                               75                               80

Glu His Tyr Tyr Lys Asp Gly Gln Ala Ile Ile Phe Val Val Asp Ser
    85                               90                               95

Ser Asp Lys Leu Arg Asn Val Val Ala Lys Glu Glu Leu Asp Thr Leu
    100                               105                               110

Leu Asn His Pro Asp Ile Lys His Arg Arg Ile Pro Ile Leu Phe Phe
    115                               120                               125

Ala Asn Lys Met Asp Leu Arg Asp Ala Val Thr Ser Val Lys Val Ser
    130                               135                               140

Gln Leu Leu Cys Leu Glu Asn Ile Lys Asp Lys Pro Trp His Ile Cys
    145                               150                               155                               160

Ala Ser Asp Ala Leu Lys Gly Glu Gly Leu Gln Glu Gly Val Asp Trp
    165                               170                               175

Leu Gln Asp Gln Ile Gln Ala Val Lys Thr
    180                               185

```

```

<210> SEQ ID NO 8
<211> LENGTH: 185
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
    Peptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (131)
<223> OTHER INFORMATION: xaa = can be any naturally occurring amino acid

```

-continued

<400> SEQUENCE: 8

```

Met Gly Ile Met Asn Lys Leu Phe Gly Trp Leu Lys Gly Lys Lys Lys
 1           5           10           15
Glu Ala Asn Val Leu Cys Val Gly Leu Asp Asn Ser Gly Lys Thr Thr
          20           25           30
Ile Ile Asn Tyr Leu Lys Pro Asn Asp Ala Gln Gln Val Asp Ile Val
          35           40           45
Phe Thr Val Gly Phe Asn Val Glu Lys Val Thr Met Thr Asn Leu Ser
          50           55           60
Phe Thr Val Phe Asp Met Ser Gly Gln Gly Arg Tyr Arg Val Leu Trp
          65           70           75           80
Ser His Tyr Tyr Lys Glu Thr Gln Gly Val Ile Phe Val Val Asp Ser
          85           90           95
Ala Asp Lys Leu Arg Met Ala Val Ala Lys Asp Glu Leu Asp Gln Leu
          100          105          110
Leu Lys His Gln Thr Ile Met Asn Lys Arg Ile Pro Ile Leu Phe Phe
          115          120          125
Ala Asn Xaa Met Asp Val Gln Asn Ser Leu Ser Ala Val Lys Cys Ser
          130          135          140
Gln Leu Leu Gly Leu Glu Asn Thr Lys Glu Lys Pro Trp His Ile Cys
          145          150          155          160
Ala Ser Asn Ala Lys Thr Gly Glu Gly Leu Ser Asp Gly Met His Trp
          165          170          175
Leu Ser Asp Gln Leu Gln Thr Val Lys
          180          185

```

<210> SEQ ID NO 9

<211> LENGTH: 202

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide

<400> SEQUENCE: 9

```

Met Gly Met Leu His Asn Leu Ala Asp Leu Ile Lys Ile Lys Lys Asp
 1           5           10           15
Lys Met Thr Ile Leu Val Leu Gly Leu Asn Asn Ser Gly Lys Ser Ser
          20           25           30
Ile Ile Asn His Phe Lys Lys Ser Ser Glu Gln Thr Ser Ile Val Val
          35           40           45
Phe Thr Val Gly Phe Met Val Glu Gln Phe Tyr Ile Gly Met Ser Gly
          50           55           60
Val Ser Ile Lys Ala Ile Asp Met Ser Gly Ala Thr Arg Tyr Arg Asn
          65           70           75           80
Leu Trp Glu His Gln Phe Lys Asn Cys His Gly Ile Ile Tyr Val Ile
          85           90           95
Asp Ser Ser Asp Arg Met Arg Phe Val Val Lys Asp Glu Leu Asp
          100          105          110
Leu Val Leu Gln Asn Pro Gly Leu Cys Asn Arg Ile Val Pro Ile Leu
          115          120          125
Phe Tyr Gly Asn Lys Met Asp Met Glu Asp Ser Leu Ser Ser Val Lys
          130          135          140
Ile Ala Ala Ala Leu Arg Leu Glu Asn Ile Lys Asp Lys Pro Trp His
          145          150          155          160

```

-continued

```

Ile Cys Ser Ser Ser Ala Thr Ser Gly Glu Gly Leu Gly Glu Gly Val
      165                      170                      175

Gln Trp Leu Ile Gln Gln Met Arg Phe Ala Met Leu Asn Asn Lys Asn
      180                      185                      190

Ala Ala Lys Ser Arg Ser Lys His Ser Lys
      195                      200

```

```

<210> SEQ ID NO 10
<211> LENGTH: 190
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      Peptide

```

```

<400> SEQUENCE: 10

```

```

Met Gly Phe Phe Ser Ser Leu Ser Ser Leu Phe Gly Leu Gly Lys Lys
  1                      5                      10                      15

Asp Val Asn Ile Val Val Val Gly Leu Asp Asn Ser Gly Lys Thr Thr
      20                      25                      30

Ile Leu Asn Gln Leu Lys Thr Pro Glu Thr Arg Ser Gln Gln Ile Val
      35                      40                      45

Pro Thr Val Gly His Val Val Thr Asn Ser Phe Thr Gln Asn Leu Ser
      50                      55                      60

Phe His Ala Phe Asp Met Ala Gly Gln Met Lys Tyr Arg Ser Thr Trp
      65                      70                      75                      80

Glu Ser Tyr Phe His Ser Ser Gln Gly Val Ile Phe Val Leu Asp Ser
      85                      90                      95

Ser Asp Arg Leu Arg Met Glu Leu Leu Lys Asp Glu Leu Met Met Val
      100                     105                     110

Met Glu His Lys Asp Val Val Ser Arg Gly Ile Pro Ile Val Ile Leu
      115                     120

Ala Asn Lys Met Asp Ile Pro Gly Ala Met Thr Ala Ser Asp Ile Thr
      130                     135                     140

Val Ala Leu Gly Leu Asn Leu Tyr Arg Ser Gly Thr Trp Ser Ile His
      145                     150                     155                     160

Ser Thr Cys Ala Leu Thr Gly Asp Gly Leu Asp Lys Ala Met Gln Gln
      165                     170                     175

Leu Ser Ala Glu Ile Thr Lys Tyr Met Glu Ser Arg Arg Thr
      180                     185                     190

```

```

<210> SEQ ID NO 11
<211> LENGTH: 189
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      Peptide

```

```

<400> SEQUENCE: 11

```

```

Met Gly Gln Ser Lys Thr Lys Leu Gln Val Val Met Cys Gly Leu Asp
  1                      5                      10                      15

Asn Ser Gly Lys Thr Thr Ile Ile Asn Gln Val Lys Pro Ala Gln Ser
      20                      25                      30

Ser Ser Lys His Ile Thr Ala Thr Val Gly Tyr Asn Val Glu Thr Phe
      35                      40                      45

Glu Lys Gly Arg Val Ala Phe Thr Val Pro Asp Met Gly Gly Ala Lys
      50                      55                      60

```

-continued

Lys Phe Arg Gly Leu Trp Glu Thr Tyr Tyr Asp Asn Ile Asp Ala Val
65 70 75 80

Ile Phe Val Val Asp Ser Ser Asp His Leu Arg Leu Cys Val Val Lys
85 90 95

Ser Glu Ile Gln Ala His Leu Lys His Glu Asp Ile Arg Arg Glu Leu
100 105 110

Pro Gly Gly Gly Arg Val Pro Phe Leu Phe Phe Ala Asn Lys Met Asp
115 120 125

Ala Ala Gly Ala Lys Thr Ala Ala Glu Leu Val Glu Ile Leu Asp Leu
130 135 140

Thr Thr Leu Met Gly Asp His Pro Glu Val Ile Phe Ala Ser Asn Ala
145 150 155 160

Leu Lys Gly Thr Gly Val His Glu Gly Phe Ser Trp Leu Gln Glu Thr
165 170 175

Ala Ser Arg Gln Ser Gly Lys Ala Gly Thr Lys Arg Gly
180 185

<210> SEQ ID NO 12
<211> LENGTH: 198
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide

<400> SEQUENCE: 12

Met Gly Gln Ser Lys Thr Lys Leu His Ile Ile Thr Cys Gly Leu Asp
1 5 10 15

Asn Ser Gly Lys Ser Ser Ile Ile Asn Arg Leu Lys Pro Thr Ala Leu
20 25 30

Gln Ser Glu His Ile Ser Ala Thr Val Gly Tyr Asn Val Glu Val Phe
35 40 45

Glu Lys Gly Ser Ala Lys Phe Thr Val Phe Asp Met Gly Gly Ala Lys
50 55 60

Lys Phe Arg Gly Leu Trp Glu Thr Tyr Tyr Glu Asn Ile Asn Gly Ile
65 70 75 80

Ile Phe Val Val Asp Ser Ser Asp Glu Leu Arg Leu Cys Val Val Lys
85 90 95

Glu Glu Ile Glu Leu Met Leu Gln His Pro Asp Ile Ala Arg Glu Leu
100 105 110

Pro Lys Thr Asn Gly Ala Lys Ile Pro Phe Leu Phe Tyr Ala Asn Lys
115 120 125

Met Asp Leu Pro Asn Ala Lys Thr Ala Ala Glu Leu Val Asp Leu Leu
130 135 140

Asp Leu Thr Thr Leu Met Ala Asp Arg Pro Phe Asn Ile Phe Ala Ser
145 150 155 160

Asn Ala Leu Lys Gly Thr Gly Val Asn Glu Gly Phe Glu Trp Leu Gln
165 170 175

Asn Ala Leu Leu Arg Gln Ser Asn Thr Ile Ser Ser Ser Ser Arg
180 185 190

Gly Ala Arg Gln Arg Asn
195

<210> SEQ ID NO 13
<211> LENGTH: 179
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide

<400> SEQUENCE: 13

Met Gly Phe Phe Asp Lys Leu Leu Ser Leu Phe Gly Met Ser Gly Lys
 1 5 10 15

Lys Val Asn Val Leu Val Val Gly Leu Asp Asn Ser Gly Lys Thr Thr
 20 25 30

Ile Ile Glu Arg Leu Lys Pro Arg Pro Arg Gln Ala Ala Glu Val Ala
 35 40 45

Pro Thr Val Gly Phe Thr Asp Asp Glu Val Glu Lys Gly Pro Leu Thr
 50 55 60

Phe Thr Val Phe Asp Met Ser Gly Ala Gly Arg Tyr Arg Thr Leu Trp
 65 70 75 80

Glu Gln Tyr Tyr Arg Glu Ala Asp Ala Val Val Phe Val Val Asp Ser
 85 90 95

Ala Asp Lys Leu Arg Met Val Val Ala Arg Asp Glu Met Glu His Met
 100 105 110

Leu Lys His Ser Asn Met Arg Lys Val Pro Ile Leu Tyr Phe Ala Asn
 115 120 125

Lys Lys Asp Leu Pro Val Ala Met Pro Pro Val Glu Ile Ala Gln Ala
 130 135 140

Leu Gly Leu Asp Asp Ile Lys Asp Arg Pro Gln Ile Val Pro Ser Asn
 145 150 155 160

Gly Leu Thr Gly Glu Gly Val Asp Lys Gly Ile Asp Trp Leu Ala Glu
 165 170 175

Arg Leu Ser

<210> SEQ ID NO 14
 <211> LENGTH: 183
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide

<400> SEQUENCE: 14

Met Gly Asp Ala Ile Ser Lys Leu Phe Asp Arg Leu Trp Gly Lys Lys
 1 5 10 15

Glu Met Arg Ile Leu Met Val Gly Leu Asp Ala Ala Gly Lys Thr Thr
 20 25 30

Ile Leu Tyr Lys Leu Lys Leu Gly Glu Ile Val Thr Thr Ile Pro Thr
 35 40 45

Ile Gly Phe Asn Val Glu Thr Val Glu Tyr Lys Asn Ile Gln Phe Thr
 50 55 60

Val Trp Asp Val Gly Gly Gln Asp Lys Ile Arg Pro Leu Trp Arg His
 65 70 75 80

Tyr Phe Gln Asn Thr Gln Gly Ile Ile Phe Val Val Asp Ser Asn Asp
 85 90 95

Arg Asp Arg Ile Val Glu Ala Arg Glu Leu Gln Arg Met Leu Asn
 100 105 110

Glu Asp Glu Leu Arg Asp Ala Leu Leu Val Phe Ala Asn Lys Gln
 115 120 125

Asp Leu Pro Asn Ala Met Ser Pro Ala Glu Ile Thr Gln Gln Leu Gly
 130 135 140

-continued

Leu Gln Ser Leu Thr Arg Arg Pro Trp Tyr Ile Gln Ser Thr Cys Ala
 145 150 155 160

Thr Thr Gly Asp Gly Leu Tyr Glu Gly Leu Glu Trp Leu Ala Glu Thr
 165 170 175

Leu Arg Lys Thr Gly Arg Asp
 180

<210> SEQ ID NO 15
 <211> LENGTH: 188
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 Peptide

<400> SEQUENCE: 15

Met Gly Leu Asn Phe Thr Lys Leu Phe Ser Arg Leu Phe Ala Lys Lys
 1 5 10 15

Glu Met Arg Ile Leu Met Val Gly Leu Asp Ala Ala Gly Lys Thr Thr
 20 25 30

Ile Leu Tyr Lys Leu Lys Ile Gly Glu Ile Val Thr Thr Ile Pro Thr
 35 40 45

Ile Gly Phe Asn Val Glu Thr Val Glu Tyr Lys Asn Ile Ser Phe Thr
 50 55 60

Val Trp Asp Val Gly Gly Gln Asp Lys Ile Arg Pro Leu Trp Arg His
 65 70 75 80

Tyr Phe Gln Asn Thr Gln Gly Leu Ile Phe Val Val Asp Ser Asn Asp
 85 90 95

Arg Asp Arg Val Val Glu Ala Arg Asp Glu Leu His Arg Met Leu Asn
 100 105 110

Glu Asp Glu Leu Arg Asp Ala Val Leu Leu Val Phe Ala Asn Lys Gln
 115 120 125

Asp Leu Pro Asn Ala Met Asn Ala Ala Glu Ile Thr Asp Lys Leu Gly
 130 135 140

Leu His Ser Leu Arg Gln Arg His Trp Tyr Ile Gln Ser Thr Cys Ala
 145 150 155 160

Thr Ser Gly Glu Gly Leu Tyr Glu Gly Leu Asp Trp Leu Ser Asn Asn
 165 170 175

Ile Ala Thr Lys Val Asn Asn Leu Asn Ile Pro Phe
 180 185

<210> SEQ ID NO 16
 <211> LENGTH: 181
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 Peptide

<400> SEQUENCE: 16

Met Gly Leu Phe Ala Ser Lys Leu Phe Ser Asn Leu Phe Gly Asn Lys
 1 5 10 15

Glu Met Arg Ile Leu Met Val Gly Leu Asp Gly Ala Gly Lys Thr Thr
 20 25 30

Val Leu Tyr Lys Leu Lys Leu Gly Glu Val Ile Thr Thr Ile Pro Thr
 35 40 45

Ile Gly Phe Asn Val Glu Thr Val Gln Tyr Lys Asn Ile Ser Phe Thr
 50 55 60

-continued

Val	Trp	Asp	Val	Gly	Gly	Gln	Asp	Arg	Ile	Arg	Ser	Leu	Trp	Arg	His
65					70					75					80
Tyr	Tyr	Arg	Asn	Thr	Glu	Gly	Val	Ile	Phe	Val	Val	Asp	Ser	Asn	Asp
			85						90					95	
Arg	Ser	Arg	Ile	Gly	Glu	Ala	Arg	Glu	Val	Met	Gln	Arg	Met	Leu	Asn
			100					105					110		
Glu	Asp	Glu	Leu	Arg	Asn	Ala	Ala	Trp	Leu	Val	Phe	Ala	Asn	Lys	Gln
		115					120					125			
Asp	Leu	Pro	Glu	Ala	Met	Ser	Ala	Ala	Glu	Ile	Thr	Glu	Lys	Leu	Gly
		130				135					140				
Leu	His	Ser	Ile	Arg	Asn	Arg	Pro	Trp	Phe	Ile	Gln	Ala	Thr	Cys	Ala
145					150					155					160
Thr	Ser	Gly	Glu	Gly	Leu	Tyr	Glu	Gly	Leu	Glu	Trp	Leu	Ser	Asn	Ser
				165					170					175	
Leu	Lys	Asn	Ser	Thr											
			180												

What is claimed is:

1. An isolated and purified nucleic acid encoding a human ADP-ribosylation factor-like 6 polypeptide comprising the sequence of SEQ ID NO:4.
2. The nucleic acid of claim 1, wherein the nucleic acid comprises the sequence of SEQ ID NO:3.
3. The nucleic acid of claim 1, further comprising a promoter.
4. The nucleic acid of claim 3, wherein the promoter is constitutive or inducible.
5. The nucleic acid of claim 1, further comprising an origin of replication.
6. The nucleic acid of claim 5, wherein said nucleic acid is part of a replicable vector.

7. The nucleic acid of claim 6, wherein said vector is a viral vector.
8. The nucleic acid of claim 7, wherein said viral vector is selected from the group consisting of a retroviral vector, an adenoviral vector, an adeno-associated viral vector, a herpesviral vector, a polyoma viral vector, a vaccinia viral vector and a lentiviral vector.
9. The nucleic acid of claim 7, wherein said viral vector is located within a viral particle.
10. The nucleic acid of claim 6, wherein said vector is a non-viral vector.

* * * * *