

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
12 September 2002 (12.09.2002)

PCT

(10) International Publication Number  
**WO 02/070563 A2**

- (51) International Patent Classification<sup>7</sup>: **C07K 14/705** 2NU (GB). **POTTER, Sarah, Jane** [GB/GB]; 39 Kneller Road, Brockley, London SE4 2AR (GB).
- (21) International Application Number: PCT/GB02/00986
- (22) International Filing Date: 5 March 2002 (05.03.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
0105402.2 5 March 2001 (05.03.2001) GB
- (71) Applicant (for all designated States except US): **INPHARMATICA LIMITED** [GB/GB]; 60 Charlotte Street, London W1T 2NU (GB).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **FAGAN, Richard, Joseph** [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). **PHELPS, Christopher, Benjamin** [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). **PHILLIPS, Tom** [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). **PIERON, Valerie, Nathalie** [FR/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). **ALLEN, Kathryn, Elizabeth** [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). **ALLEN, Janet, Marjorie** [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB).
- (74) Agents: **MERCER, Christopher, Paul** et al.; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



**WO 02/070563 A2**

(54) Title: NUCLEAR HORMONE RECEPTOR LIGAND BINDING DOMAIN

(57) Abstract: This invention relates to a novel protein, termed BAA22563.1, herein identified as a Nuclear Hormone Receptor Ligand Binding Domain and to the use of this protein and nucleic acid sequence from the encoding gene in the diagnosis, prevention and treatment of disease.

## NUCLEAR HORMONE RECEPTOR LIGAND BINDING DOMAIN

This invention relates to a novel protein, termed BAA22563.1 herein identified as a Nuclear Hormone Receptor Ligand Binding Domain and to the use of this protein and nucleic acid sequence from the encoding gene in the diagnosis, prevention and treatment  
5 of disease.

All publications, patents and patent applications cited herein are incorporated in full by reference.

### BACKGROUND

The process of drug discovery is presently undergoing a fundamental revolution as the  
10 era of functional genomics comes of age. The term "functional genomics" applies to an approach utilising bioinformatics tools to ascribe function to protein sequences of interest. Such tools are becoming increasingly necessary as the speed of generation of sequence data is rapidly outpacing the ability of research laboratories to assign functions to these protein sequences.

15 As bioinformatics tools increase in potency and in accuracy, these tools are rapidly replacing the conventional techniques of biochemical characterisation. Indeed, the advanced bioinformatics tools used in identifying the present invention are now capable of outputting results in which a high degree of confidence can be placed.

Various institutions and commercial organisations are examining sequence data as they  
20 become available and significant discoveries are being made on an on-going basis. However, there remains a continuing need to identify and characterise further genes and the polypeptides that they encode, as targets for research and for drug discovery.

Recently, a remarkable tool for the evaluation of sequences of unknown function has  
25 been developed by the Applicant for the present invention. This tool is a database system, termed the Biopendium search database, that is the subject of co-pending International Patent Application No. PCT/GB01/01105. This database system consists of an integrated data resource created using proprietary technology and containing information generated from an all-by-all comparison of all available protein or nucleic acid sequences.

The aim behind the integration of these sequence data from separate data resources is to

combine as much data as possible, relating both to the sequences themselves and to information relevant to each sequence, into one integrated resource. All the available data relating to each sequence, including data on the three-dimensional structure of the encoded protein, if this is available, are integrated together to make best use of the information that is known about each sequence and thus to allow the most educated  
5 predictions to be made from comparisons of these sequences. The annotation that is generated in the database and which accompanies each sequence entry imparts a biologically relevant context to the sequence information.

This data resource has made possible the accurate prediction of protein function from  
10 sequence alone. Using conventional technology, this is only possible for proteins that exhibit a high degree of sequence identity (above about 20%-30% identity) to other proteins in the same functional family. Accurate predictions are not possible for proteins that exhibit a very low degree of sequence homology to other related proteins of known function.

15 In the present case, a protein whose sequence is recorded in a publicly available database as BAA22563.1 (NCBI Genebank nucleotide accession number AB007510.1 and a Genebank protein accession number BAA22563.1), is implicated as a novel member of the Nuclear Hormone Receptor Ligand Binding Domain family.

## 20 **I. Introduction to Nuclear Hormone Receptor Ligand Binding Domains**

The Nuclear Hormone Receptor gene superfamily (see Table 1) encodes structurally related proteins that regulate the transcription of target genes. These proteins include receptors for steroid and thyroid hormones, vitamins, and other proteins for which no ligands have been found. Nuclear Receptors are composed of two key domains, a DNA-  
25 Binding Domain (DBD) and a Ligand Binding Domain (LBD). The DBD directs the receptors to bind specific DNA sequences as monomers, homodimers, or heterodimers. The DBD is a particular type of zinc-finger, found only in Nuclear Receptors. Nuclear Receptors with DBDs can be readily identified at the sequence level by searching for matches to the PROSITE consensus sequence (PS00031).

The Ligand Binding Domain (LBD) binds and responds to the cognate hormone. Ligand binding to the LBD triggers a conformational change which expels a bound “Nuclear Receptor Co-Repressor”. The site previously occupied by the Co-Repressor is then free to recruit a “Nuclear Receptor Co-Activator”. This Ligand-triggered swap of a Co-Repressor for a Co-Activator is the mechanism by which Ligand binding leads to the transcriptional activation of target genes. All ligand binding domains contain a consensus sequence, the “LBD motif” (see Table 2) which mediates Co-Repressor and Co-Activator binding. The LBD is the binding site for all Nuclear Hormone Receptor targeted drugs to date and it is thus desirable to identify novel Ligand Binding Domains since these will be attractive drug targets. Ligand Binding Domains share low sequence identity (~15%) but have very similar structures and so present ideal targets for a structure-based relationship tool such as Genome Threader.

Many protein sequences have already been annotated in the public domain as Nuclear Hormone Receptors by their possession of DBDs using basic search tools like PROSITE, and their LBDs inferred on the basis of this. Because of this it is anticipated that any novel LBDs identified by Genome Threader *which are not annotated as nuclear receptors* will lack the DBD entirely. A precedent for a protein which has an LBD but lacks a DBD is provided by DAX1. Thus we annotate these DBD-less hits not as “Nuclear Hormone Receptors” but rather as containing a “Nuclear Hormone Receptor Ligand Binding Domain”.

Family: Steroid Hormone Receptors	
Subfamilies	Glucocorticoid Receptors
	Progesterone Receptors
	Androgen Receptors
	Estrogen Receptors
<i>Family: Thyroid Hormone Receptor-like Factors</i>	
Subfamilies	Retinoic Acid Receptors (RARs)

	Retinoid X Receptors (RXRs)
	Thyroid Hormone Receptors
	Vitamin D Receptor
	NGFI-B
	FTZ-F1
	Peroxisome Proliferator Activated Receptors (PPARs)
	Ecdysone Receptors
	Retinoid Orphan Receptors (RORs)
	Tailless/COUP
	HNF-4
	CF1
	Knirps
<i>Family: DAX1</i>	
Subfamilies	DAX1

Table 1: Nuclear hormone Receptor Superfamily

1	2	3	4	5	6	7	8	9	10	11	12	13
L	Any residues	2	L	Any 3 residues (or 2 residues or 4 residues)			D	Q	Any residues (or 1 or 3 residues)	2	L	L
I			I				E	N			I	I
A			A					R			A	A
V			V					H			V	V
M			M					K			M	M
F			F					S			F	F
Y			Y					T			Y	Y
W			W								W	W

Table 2: The “LBD motif”. Numbers along the top row refer to residue position within the motif. Letters refer to amino acids by the 1-letter code. Letters within one column are all acceptable for that position within the motif. For example L, I, A, V, M, F, Y or W can occupy the first position of the “LBD motif”. Note that there is observed variation in the number of residues found between position 4 and 8, and position 9 and 12. The “LBD motif” was constructed by aligning 681 sequences of Nuclear Hormone Receptor Ligand Binding Domains, and identifying conserved patterns of residues.

## II. Nuclear Hormone Receptors and Disease

Nuclear Hormone Receptors have been shown to play a role in diverse physiological functions, many of which can play a role in disease processes (see Table 3).

Nuclear Hormone Receptor	Disease
Androgen Receptor	Androgen Insensitivity Syndrome (Lubahn <i>et al.</i> 1989 Proc. Natl. Acad. Sci. USA 86, 9534-9538).
	Reifenstein syndrome (Wooster <i>et al.</i> 1992 Nat. Genet. 2, 132-134).
	X-linked recessive spinal and bulbar muscular atrophy (MacLean <i>et al.</i> 1995 Mol. Cell. Endocrinol. 112,133-141).
	Male breast cancer ((Wooster <i>et al.</i> 1992 Nat. Genet. 2, 132-134).
Glucocorticoid Receptor	Nelson's syndrome (Karl <i>et al.</i> 1996 J. Clin. Endocrinol. Metab. 81, 124-129).
	Glucocorticoid resistant acute T-cell leukemia (Hala <i>et al.</i> 1996 Int. J. Cancer 68, 663-668).
Mineralocorticoid Receptor	Pseudohypoaldosteronism (Chung <i>et al.</i> 1995 J. Clin. Endocrinol. Metab. 80, 3341-3345).
Estrogen Receptor alpha	ER alpha expression is elevated in a subset of human breast cancers. The application of Tamoxifen is the major therapy to prevent breast tumour progression. Unfortunately 35% of ER alpha positive breast cancers are Tamoxifen resistant (Petrangeli <i>et al.</i> 1994 J. Steroid Biochem. Mol. Biol. 49, 327-331).
Vitamin D3 Receptor	Mutations in the Vitamin D3 receptor produce a hereditary disorder similar in phenotype to Vitamin D3 deficiency (Rickets) (Hughes <i>et al.</i> 1988 Science 242, 1702-1725).
Retinoic Acid Receptor alpha	Acute Myeloid Leukemia (Lavau and Dejean 1994 Leukemia 8, 9-15).
Thyroid Hormone Receptor beta	"Generalised Resistance to Thyroid Hormones" (GRTH) (Refetoff 1994 Thyroid 4, 345-349).
DAX1	X-linked Adrenal Hypoplasia Congenita (AHC) and Hypogonadism (Ito <i>et al.</i> 1997 Mol. Cell. Biol. 17, 1476-1483).

Table 3. Nuclear Hormone Receptors and disease.

Alteration of Nuclear Hormone Receptors by ligands which bind to their LBD thus provides a means to alter the disease phenotype. There is thus a great need for the identification of novel Nuclear Hormone Receptor Ligand Binding Domains, as these proteins may play a role in the diseases identified above, as well as in other disease states.

- 5 The identification of novel Nuclear Hormone Receptor Ligand Binding Domains is thus highly relevant for the treatment and diagnosis of disease, particularly those identified in Table 3.

### THE INVENTION

- 10 The invention is based on the discovery that the BAA22563.1 protein functions as a Nuclear Hormone Receptor Ligand Binding Domain.

For the BAA22563.1 protein, it has been found that a region including residues 1104-1309 of this protein sequence adopts an equivalent fold to residues 6 (Asp216) to 209 (Asp427) of the Thyroid Hormone Receptor beta (PDB code 1BSX:A). Thyroid  
15 Hormone Receptor beta is known to function as a Nuclear Hormone Receptor Ligand Binding Domain. Furthermore, the "LBD motif" residues PHE293, LEU296, ASP300, GLN301, LEU304 and LEU305 of the Thyroid Hormone Receptor beta are conserved as PHE1174, VAL1177, ASP1181, ASN1182, LEU1185 and LEU1186 in BAA22563.1, respectively. This relationship is not just to Thyroid Hormone Receptor beta, but rather to  
20 the Nuclear Hormone Receptor Ligand Binding Domain family as a whole. Thus, by reference to the Genome Threader<sup>TM</sup> alignment of BAA22563.1 with the Thyroid Hormone Receptor beta (1BSX:A) PHE1174, VAL1177, ASP1181, ASN1182, LEU1185 and LEU1186 of BAA22563.1 are predicted to form the "LBD motif" residues.

The combination of equivalent fold and conservation of "LBD motif" residues allows the  
25 functional annotation of this region of BAA22563.1, and therefore proteins that include this region, as possessing Nuclear Hormone Receptor Ligand Binding Domain activity.

In a first aspect, the invention provides a polypeptide, which polypeptide:

- (i) comprises the amino acid sequence as recited in SEQ ID NO:2;
- (ii) is a fragment thereof having Nuclear Hormone Receptor Ligand Binding Domain

activity or having an antigenic determinant in common with the polypeptides of (i); or

(iii) is a functional equivalent of (i) or (ii).

Preferably, the polypeptide:

5 (i) consists of the amino acid sequence as recited in SEQ ID NO:2;

(ii) is a fragment thereof having Nuclear Hormone Receptor Ligand Binding Domain activity or having an antigenic determinant in common with the polypeptides of (i); or

(iii) is a functional equivalent of (i) or (ii).

10 The polypeptide having the sequence recited in SEQ ID NO:2 is referred to hereafter as "the LBDG2 polypeptide".

According to this aspect of the invention, a preferred polypeptide fragment according to part ii) above includes the region of the LBDG2 polypeptide that is predicted as that responsible for Nuclear Hormone Receptor Ligand Binding Domain activity (hereafter, 15 the "LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region"), or is a variant thereof that possesses the "LBD motif" (PHE1174, VAL1177, ASP1181, ASN1182, LEU1185 and LEU1186, or equivalent residues). As defined herein, the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 1104 and residue 1309 of the LBDG2 polypeptide sequence.

20 This aspect of the invention also includes fusion proteins that incorporate polypeptide fragments and variants of these polypeptide fragments as defined above, provided that said fusion proteins possess activity as a Nuclear Hormone Receptor Ligand Binding Domain.

In a second aspect, the invention provides a purified nucleic acid molecule that encodes a 25 polypeptide of the first aspect of the invention. Preferably, the purified nucleic acid molecule has the nucleic acid sequence as recited in SEQ ID NO:1 (encoding the LBDG2 polypeptide), or is a redundant equivalent or fragment of this sequence. A preferred nucleic acid fragment is one that encodes a polypeptide fragment according to part ii) above, preferably a polypeptide fragment that includes the LBDG2 Nuclear Hormone

Receptor Ligand Binding Domain region, or that encodes a variant of these fragments as this term is defined above.

In a third aspect, the invention provides a purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule of the second  
5 aspect of the invention.

In a fourth aspect, the invention provides a vector, such as an expression vector, that contains a nucleic acid molecule of the second or third aspect of the invention.

In a fifth aspect, the invention provides a host cell transformed with a vector of the fourth aspect of the invention.

10 In a sixth aspect, the invention provides a ligand which binds specifically to, and which preferably inhibits the Nuclear Hormone Receptor Ligand Binding Domain activity of, a polypeptide of the first aspect of the invention.

In a seventh aspect, the invention provides a compound that is effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the  
15 invention or to regulate the activity of a polypeptide of the first aspect of the invention.

A compound of the seventh aspect of the invention may either increase (agonise) or decrease (antagonise) the level of expression of the gene or the activity of the polypeptide. Importantly, the identification of the function of the region defined herein as the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG2  
20 polypeptide, respectively, allows for the design of screening methods capable of identifying compounds that are effective in the treatment and/or diagnosis of diseases in which Nuclear Hormone Receptor Ligand Binding Domains are implicated.

In an eighth aspect, the invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a  
25 vector of the fourth aspect of the invention, or a ligand of the fifth aspect of the invention, or a compound of the sixth aspect of the invention, for use in therapy or diagnosis.

The inventors have discovered that the mRNA for LBDG2 is found in extracts from a variety of human tissues. The finding of high levels of the transcript in the human spleen is consistent with a role of LBDG2 in the immune system and in particular in lymphocyte

development and function and in particular in B cell development and function. It is therefore considered that the development of agonists and antagonists for LBDG2 may have a particular role in the therapeutic intervention in various human diseases of the immune system including autoimmunity, allergies and diseases associated with immunoglobulin dysfunction. These diseases include type I diabetes mellitus, rheumatoid arthritis, multiple sclerosis, psoriasis, renal failure arising from glomerulopathies, scleroderma, inflammatory bowel disease (both Crohns disease and ulcerative colitis), transplant rejection, asthma, atopic dermatitis, eczema, myelomas and in infectious diseases that require production of antibodies, for example, intracellular pathogen such as virus infected cells, tuberculosis, listeria.

Messenger RNA for LBDG2 has also been found in human B cell lines such as Daudi, IM9 and Raji cells. These findings are consistent with the discovery mentioned above, that LBDG2 mRNA is found in the spleen. Finding high levels of expression of the mRNA in U937 cells suggests a role for LBDG2 in monocyte/macrophage functions and, as such, agonists or antagonists may be valuable in treating inflammatory diseases including chronic obstructive pulmonary disease (COPD), osteoarthritis, rheumatoid arthritis, inflammatory bowel disease, fibrosis such as liver fibrosis (cirrhosis) and skin fibrosis (scarring), atherosclerosis, dementia, multiple sclerosis, inflammatory pain.

In addition, significant levels of LBDG2 have been found in adrenal, ovary, prostate and testis tissue. This indicates that the development of agonists and antagonists to LBDG2 may be of value in diseases such as benign prostatic hypertrophy, prostatic cancer, ovarian cancer and testicular cancer. In addition, agonists or antagonists for LBDG2 may be developed for treatment of diseases including but not exclusive to hypertension, responses to stress including stress of infectious diseases, regulation of salt and water homeostasis, control of fertility through regulation of ovulation (infertility and contraception), regulation of implantation (infertility and contraception) and regulation of spermatogenesis (infertility and contraception).

It has also been found that the mRNA for LBDG2 is expressed at significant levels in the human brain. This is noteworthy, as this provides a potential link to human disease states and development of agonists and antagonists for the ligand binding domain of LBDG2

offers the potential for therapeutic intervention in various human diseases including cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours; myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposi's sarcoma; autoimmune/inflammatory disorders, including autoimmunity, allergies and diseases associated with immunoglobulin dysfunction, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma and organ transplant rejection; cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia; neurological disorders including, central nervous system disease, Alzheimer's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain; developmental disorders; metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hyperparathyroidism, hypercalcemia, hypercholesterolemia, hyperlipidemia, and obesity; renal disorders, including glomerulonephritis, renovascular hypertension; dermatological disorders, including, acne, eczema, and wound healing, negative effects of aging; AIDS; infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated.

The finding of a "non-classical" nuclear hormone receptor such as LBDG2 which contains a ligand binding domain in the absence of a DNA binding domain is consistent with the known literature which has consistently reported widespread effects of steroids in the brain (known as neurosteroids) and that these effects, in general, are mediated not through the known classic steroid hormone nuclear receptors which requires transcriptional activation. For instance, neurosteroids have been shown to influence neurotransmission particularly in the field of receptors such as those for GABA and NMDA and Sigma receptors. Neurosteroids have been shown to play a neuroprotective role. Therapeutic intervention through the development of agonists (or antagonists) to LBDG2 may therefore have a role in treatment of neurodegenerative conditions such as dementia, Parkinson's disease and neurodegeneration following cerebrovascular disease such as infarction or haemorrhage (stroke) and trauma to the central nervous system and

spinal cord. In addition, neurosteroids have been shown to influence cognitive processing, spatial learning and memory, anxiety and behaviours such as craving which leads to addictive behaviour patterns. Development of agonists and antagonists to LBDG2 may therefore lead to therapeutic intervention to treat dementias, learning  
5 difficulties, anxiety, addictive behaviours such as but not exclusively alcoholism, eating disorders and drug addiction.

In a ninth aspect, the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide of the first aspect of the invention or the activity of a polypeptide of the first aspect of the  
10 invention in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease. Such a method will preferably be carried out *in vitro*. Similar methods may be used for monitoring the therapeutic treatment of disease in a patient, wherein altering the level of expression or activity of a polypeptide or nucleic acid molecule over the period  
15 of time towards a control level is indicative of regression of disease.

A preferred method for detecting polypeptides of the first aspect of the invention comprises the steps of: (a) contacting a ligand, such as an antibody, of the sixth aspect of the invention with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

20 A number of different such methods according to the ninth aspect of the invention exist, as the skilled reader will be aware, such as methods of nucleic acid hybridization with short probes, point mutation analysis, polymerase chain reaction (PCR) amplification and methods using antibodies to detect aberrant protein levels. Similar methods may be used on a short or long term basis to allow therapeutic treatment of a disease to be monitored  
25 in a patient. The invention also provides kits that are useful in these methods for diagnosing disease.

In a tenth aspect, the invention provides for the use of a polypeptide of the first aspect of the invention as a Nuclear Hormone Receptor Ligand Binding Domain. The invention also provides for the use of a nucleic acid molecule according to the second or third  
30 aspects of the invention to express a protein that possesses Nuclear Hormone Receptor

Ligand Binding Domain activity. The invention also provides a method for effecting Nuclear Hormone Receptor Ligand Binding Domain activity, said method utilising a polypeptide of the first aspect of the invention.

In an eleventh aspect, the invention provides a pharmaceutical composition comprising a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, in conjunction with a pharmaceutically-acceptable carrier.

In a twelfth aspect, the present invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, for use in the manufacture of a medicament for the diagnosis or treatment of a disease, such as cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposis' sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hyperparathyroidism, hypercalcemia, hypercholesterolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, dermatological disorders, including, acne, eczema, and wound healing, negative effects of aging, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated, as well as the other more specific diseases and conditions mentioned above in connection with the eighth aspect of the invention.

In a thirteenth aspect, the invention provides a method of treating a disease in a patient comprising administering to the patient a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a  
5 compound of the seventh aspect of the invention.

For diseases in which the expression of a natural gene encoding a polypeptide of the first aspect of the invention, or in which the activity of a polypeptide of the first aspect of the invention, is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound  
10 administered to the patient should be an agonist. Conversely, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound administered to the patient should be an antagonist. Examples of such antagonists include antisense nucleic acid molecules,  
15 ribozymes and ligands, such as antibodies.

In a fourteenth aspect, the invention provides transgenic or knockout non-human animals that have been transformed to express higher, lower or absent levels of a polypeptide of the first aspect of the invention. Such transgenic animals are very useful models for the study of disease and may also be using in screening regimes for the identification of  
20 compounds that are effective in the treatment or diagnosis of such a disease.

A summary of standard techniques and procedures which may be employed in order to utilise the invention is given below. It will be understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors and reagents described. It is also to be understood that the terminology used herein is for the purpose of describing  
25 particular embodiments only and it is not intended that this terminology should limit the scope of the present invention. The extent of the invention is limited only by the terms of the appended claims.

Standard abbreviations for nucleotides and amino acids are used in this specification.

The practice of the present invention will employ, unless otherwise indicated,  
30 conventional techniques of molecular biology, microbiology, recombinant DNA

technology and immunology, which are within the skill of those working in the art.

Such techniques are explained fully in the literature. Examples of particularly suitable texts for consultation include the following: Sambrook Molecular Cloning; A Laboratory Manual, Second Edition (1989); DNA Cloning, Volumes I and II (D.N. Glover ed. 1985);  
5 Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); Transcription and Translation (B.D. Hames & S.J. Higgins eds. 1984); Animal Cell Culture (R.I. Freshney ed. 1986); Immobilized Cells and Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984); the Methods in Enzymology series (Academic Press, Inc.), especially volumes 154 &  
10 155; Gene Transfer Vectors for Mammalian Cells (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer and Walker, eds. 1987, Academic Press, London); Scopes, (1987) Protein Purification: Principles and Practice, Second Edition (Springer Verlag, N.Y.); and Handbook of Experimental Immunology, Volumes I-IV (D.M. Weir and C. C. Blackwell  
15 eds. 1986).

As used herein, the term "polypeptide" includes any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e. peptide isosteres. This term refers both to short chains (peptides and oligopeptides) and to longer chains (proteins).

20 The polypeptide of the present invention may be in the form of a mature protein or may be a pre-, pro- or prepro- protein that can be activated by cleavage of the pre-, pro- or prepro- portion to produce an active mature polypeptide. In such polypeptides, the pre-, pro- or prepro- sequence may be a leader or secretory sequence or may be a sequence that is employed for purification of the mature polypeptide sequence.

25 The polypeptide of the first aspect of the invention may form part of a fusion protein. For example, it is often advantageous to include one or more additional amino acid sequences which may contain secretory or leader sequences, pro-sequences, sequences which aid in purification, or sequences that confer higher protein stability, for example during recombinant production. Alternatively or additionally, the mature polypeptide may be  
30 fused with another compound, such as a compound to increase the half-life of the

polypeptide (for example, polyethylene glycol).

Polypeptides may contain amino acids other than the 20 gene-encoded amino acids, modified either by natural processes, such as by post-translational processing or by chemical modification techniques which are well known in the art. Among the known  
5 modifications which may commonly be present in polypeptides of the present invention are glycosylation, lipid attachment, sulphation, gamma-carboxylation, for instance of glutamic acid residues, hydroxylation and ADP-ribosylation. Other potential modifications include acetylation, acylation, amidation, covalent attachment of flavin, covalent attachment of a haeme moiety, covalent attachment of a nucleotide or nucleotide  
10 derivative, covalent attachment of a lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulphide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, GPI anchor formation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation,  
15 racemization, selenoylation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl terminus in a polypeptide, or both, by a covalent modification is common in  
20 naturally-occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention.

The modifications that occur in a polypeptide often will be a function of how the polypeptide is made. For polypeptides that are made recombinantly, the nature and extent of the modifications in large part will be determined by the post-translational  
25 modification capacity of the particular host cell and the modification signals that are present in the amino acid sequence of the polypeptide in question. For instance, glycosylation patterns vary between different types of host cell.

The polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally-occurring polypeptides (for example purified from  
30 cell culture), recombinantly-produced polypeptides (including fusion proteins),

synthetically-produced polypeptides or polypeptides that are produced by a combination of these methods.

The functionally-equivalent polypeptides of the first aspect of the invention may be polypeptides that are homologous to the LBDG2 polypeptide. Two polypeptides are said to be "homologous", as the term is used herein, if the sequence of one of the polypeptides has a high enough degree of identity or similarity to the sequence of the other polypeptide. "Identity" indicates that at any particular position in the aligned sequences, the amino acid residue is identical between the sequences. "Similarity" indicates that, at any particular position in the aligned sequences, the amino acid residue is of a similar type between the sequences. Degrees of identity and similarity can be readily calculated (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing. Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991).

Homologous polypeptides therefore include natural biological variants (for example, allelic variants or geographical variations within the species from which the polypeptides are derived) and mutants (such as mutants containing amino acid substitutions, insertions or deletions) of the LBDG2 polypeptide. Such mutants may include polypeptides in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; among the basic residues Lys and Arg; or among the aromatic residues Phe and Tyr. Particularly preferred are variants in which several, i.e. between 5 and 10, 1 and 5, 1 and 3, 1 and 2 or just 1 amino acids are substituted, deleted or added in any combination. Especially preferred are silent substitutions, additions and deletions, which do not alter the properties and activities of the protein. Also especially preferred in this regard are conservative substitutions.

Such mutants also include polypeptides in which one or more of the amino acid residues includes a substituent group.

Typically, greater than 80% identity between two polypeptides (preferably, over a specified region) is considered to be an indication of functional equivalence. Preferably, functionally equivalent polypeptides of the first aspect of the invention have a degree of sequence identity with the LBDG2 polypeptide, or with active fragments thereof, of greater than 80%. More preferred polypeptides have degrees of identity of greater than 85%, 90%, 95%, 98% or 99%, respectively with the LBDG2 polypeptide, or with active fragments thereof.

Percentage identity, as referred to herein, is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].

In the present case, preferred active fragments of the LBDG2 polypeptide are those that include the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region and which possess the "LBD motif" of residues PHE1174, VAL1177, ASP1181, ASN1182, LEU1185 and LEU1186, or equivalent residues. By "equivalent residues" is meant residues that are equivalent to the "LBD motif" residues, provided that the Nuclear Hormone Receptor Ligand Binding Domain region retains activity as a Nuclear Hormone Receptor Ligand Binding Domain. For example PHE1174 replaced by LEU, ILE, ALA, VAL, MET, TYR or TRP. For example VAL1177 may be replaced by LEU, ILE, ALA, MET, PHE, TYR or TRP. For example ASP1181 may be replaced by GLU. For example ASN1182 may be replaced by GLN, ARG, HIS, LYS, SER, THR. For example LEU1185 may be replaced by ILE, ALA, VAL, MET, PHE, TYR or TRP. For example LEU1186 may be replaced by ILE, ALA, VAL, MET, PHE, TYR or TRP. Accordingly, this aspect of the invention includes polypeptides that have degrees of identity of greater than 80%, preferably, greater than 85%, 90%, 95%, 98% or 99%, respectively, with the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG2 polypeptide and which possess the "LBD motif" of PHE1174, VAL1177, ASP1181, ASN1182, LEU1185 and LEU1186, or equivalent residues. As discussed above, the LBDG2

Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 1104 and residue 1309 of the LBDG2 polypeptide sequence.

The functionally-equivalent polypeptides of the first aspect of the invention may also be polypeptides which have been identified using one or more techniques of structural alignment. For example, the Inpharmatica Genome Threader<sup>TM</sup> technology that forms one aspect of the search tools used to generate the Biopendium search database may be used (see co-pending International patent application PCT/GB01/01105) to identify polypeptides of presently-unknown function which, while having low sequence identity as compared to the LBDG2 polypeptide, are predicted to have Nuclear Hormone Receptor Ligand Binding Domain activity, by virtue of sharing significant structural homology with the LBDG2 polypeptide sequence.

By "significant structural homology" is meant that the Inpharmatica Genome Threader<sup>TM</sup> predicts two proteins, or protein regions, to share structural homology with a certainty of at least 10% more preferably, at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and above. The certainty value of the Inpharmatica Genome Threader<sup>TM</sup> is calculated as follows. A set of comparisons was initially performed using the Inpharmatica Genome Threader<sup>TM</sup> exclusively using sequences of known structure. Some of the comparisons were between proteins that were known to be related (on the basis of structure). A neural network was then trained on the basis that it needed to best distinguish between the known relationships and known not-relationships taken from the CATH structure classification ([www.biochem.ucl.ac.uk/bsm/cath](http://www.biochem.ucl.ac.uk/bsm/cath)). This resulted in a neural network score between 0 and 1. However, again as the number of proteins that are related and the number that are unrelated were known, it was possible to partition the neural network results into packets and calculate empirically the percentage of the results that were correct. In this manner, any genuine prediction in the Biopendium search database has an attached neural network score and the percentage confidence is a reflection of how successful the Inpharmatica Genome Threader<sup>TM</sup> was in the training/testing set.

Structural homologues of LBDG2 should share structural homology with the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region and possess the "LBD motif" residues PHE1174, VAL1177, ASP1181, ASN1182, LEU1185 and LEU1186, or

equivalent residues. Such structural homologues are predicted to have Nuclear Hormone Receptor Ligand Binding Domain activity by virtue of sharing significant structural homology with this polypeptide sequence and possessing the "LBD motif" residues.

The polypeptides of the first aspect of the invention also include fragments of the LBDG2 polypeptide, functional equivalents of the fragments of the LBDG2 polypeptide, and  
5 fragments of the functional equivalents of the LBDG2 polypeptides, provided that those functional equivalents and fragments retain Nuclear Hormone Receptor Ligand Binding Domain activity or have an antigenic determinant in common with the LBDG2 polypeptide.

10 As used herein, the term "fragment" refers to a polypeptide having an amino acid sequence that is the same as part, but not all, of the amino acid sequence of the LBDG2 polypeptides or one of its functional equivalents. The fragments should comprise at least n consecutive amino acids from the sequence and, depending on the particular sequence, n preferably is 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more). Small  
15 fragments may form an antigenic determinant.

Preferred polypeptide fragments according to this aspect of the invention are fragments that include a region defined herein as the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG2 polypeptides, respectively. These regions are the regions that have been annotated as a Nuclear Hormone Receptor Ligand Binding  
20 Domain.

For the LBDG2 polypeptide, this region is considered to extend between residue 1104 and residue 1309.

Variants of this fragment are included as embodiments of this aspect of the invention, provided that these variants possess activity as a Nuclear Hormone Receptor Ligand  
25 Binding Domain.

In one respect, the term "variant" is meant to include extended or truncated versions of this polypeptide fragment.

For extended variants, it is considered highly likely that the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG2 polypeptide will fold correctly and show

Nuclear Hormone Receptor Ligand Binding Domain activity if additional residues C terminal and/or N terminal of these boundaries in the LBDG2 polypeptide sequence are included in the polypeptide fragment. For example, an additional 5, 10, 20, 30, 40 or even 50 or more amino acid residues from the LBDG2 polypeptide sequence, or from a homologous sequence, may be included at either or both the C terminal and/or N terminal of the boundaries of the Nuclear Hormone Receptor Ligand Binding Domain regions of the LBDG2 polypeptide, without prejudicing the ability of the polypeptide fragment to fold correctly and exhibit Nuclear Hormone Receptor Ligand Binding Domain activity.

For truncated variants of the LBDG2 polypeptide, one or more amino acid residues may be deleted at either or both the C terminus or the N terminus of the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG2 polypeptide, although the "LBD motif" residues (PHE1174, VAL1177, ASP1181, ASN1182, LEU1185 and LEU1186), or equivalent residues should be maintained intact; deletions should not extend so far into the polypeptide sequence that any of these residues are deleted.

In a second respect, the term "variant" includes homologues of the polypeptide fragments described above, that possess significant sequence homology with the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG2 polypeptide and which possess the "LBD motif" residues (PHE1174, VAL1177, ASP1181, ASN1182, LEU1185 and LEU1186), or equivalent residues, provided that said variants retain activity as an Nuclear Hormone Receptor Ligand Binding Domain.

Homologues include those polypeptide molecules that possess greater than 80% identity with the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain regions, of the LBDG2 polypeptides, respectively. Percentage identity is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1]. Preferably, variant homologues of polypeptide fragments of this aspect of the invention have a degree of sequence identity with the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain regions, of the LBDG2 polypeptides, respectively, of greater than 80%. More preferred variant polypeptides have degrees of identity of greater than 85%, 90%, 95%, 98% or 99%, respectively with the

LBDG2 Nuclear Hormone Receptor Ligand Binding Domain regions of the LBDG2, polypeptides, provided that said variants retain activity as a Nuclear Hormone Receptor Ligand Binding Domain. Variant polypeptides also include homologues of the truncated forms of the polypeptide fragments discussed above, provided that said variants retain  
5 activity as a Nuclear Hormone Receptor Ligand Binding Domain.

The polypeptide fragments of the first aspect of the invention may be polypeptide fragments that exhibit significant structural homology with the structure of the polypeptide fragment defined by the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain regions, of the LBDG2 polypeptide sequence, for example, as identified by the  
10 Inpharmatica Genome Threader™. Accordingly, polypeptide fragments that are structural homologues of the polypeptide fragments defined by the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain regions of the LBDG2 polypeptide sequence should adopt the same fold as that adopted by this polypeptide fragment, as this fold is defined above.

15 Structural homologues of the polypeptide fragment defined by the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region should also retain the "LBD motif" residues PHE1174, VAL1177, ASP1181, ASN1182, LEU1185 and LEU1186, or equivalent residues.

Such fragments may be "free-standing", i.e. not part of or fused to other amino acids or  
20 polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the fragment of the invention most preferably forms a single continuous region. For instance, certain preferred embodiments relate to a fragment having a pre- and/or pro- polypeptide region fused to the amino terminus of the fragment and/or an additional region fused to the carboxyl  
25 terminus of the fragment. However, several fragments may be comprised within a single larger polypeptide.

The polypeptides of the present invention or their immunogenic fragments (comprising at least one antigenic determinant) can be used to generate ligands, such as polyclonal or monoclonal antibodies, that are immunospecific for the polypeptides. Such antibodies  
30 may be employed to isolate or to identify clones expressing the polypeptides of the

invention or to purify the polypeptides by affinity chromatography. The antibodies may also be employed as diagnostic or therapeutic aids, amongst other applications, as will be apparent to the skilled reader.

The term "immunospecific" means that the antibodies have substantially greater affinity  
5 for the polypeptides of the invention than their affinity for other related polypeptides in the prior art. As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub> and Fv, which are capable of binding to the antigenic determinant in question. Such antibodies thus bind to the polypeptides of the first aspect of the invention.

10 If polyclonal antibodies are desired, a selected mammal, such as a mouse, rabbit, goat or horse, may be immunised with a polypeptide of the first aspect of the invention. The polypeptide used to immunise the animal can be derived by recombinant DNA technology or can be synthesized chemically. If desired, the polypeptide can be conjugated to a carrier protein. Commonly used carriers to which the polypeptides may  
15 be chemically coupled include bovine serum albumin, thyroglobulin and keyhole limpet haemocyanin. The coupled polypeptide is then used to immunise the animal. Serum from the immunised animal is collected and treated according to known procedures, for example by immunoaffinity chromatography.

Monoclonal antibodies to the polypeptides of the first aspect of the invention can also be  
20 readily produced by one skilled in the art. The general methodology for making monoclonal antibodies using hybridoma technology is well known (see, for example, Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, 77-96 in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985).

25 Panels of monoclonal antibodies produced against the polypeptides of the first aspect of the invention can be screened for various properties, i.e., for isotype, epitope, affinity, etc. Monoclonal antibodies are particularly useful in purification of the individual polypeptides against which they are directed. Alternatively, genes encoding the monoclonal antibodies of interest may be isolated from hybridomas, for instance by PCR  
30 techniques known in the art, and cloned and expressed in appropriate vectors.

Chimeric antibodies, in which non-human variable regions are joined or fused to human constant regions (see, for example, Liu *et al.*, Proc. Natl. Acad. Sci. USA, 84, 3439 (1987)), may also be of use.

The antibody may be modified to make it less immunogenic in an individual, for example  
5 by humanisation (see Jones *et al.*, Nature, 321, 522 (1986); Verhoeyen *et al.*, Science, 239: 1534 (1988); Kabat *et al.*, J. Immunol., 147: 1709 (1991); Queen *et al.*, Proc. Natl Acad. Sci. USA, 86, 10029 (1989); Gorman *et al.*, Proc. Natl Acad. Sci. USA, 88: 34181 (1991); and Hodgson *et al.*, Bio/Technology 9: 421 (1991)). The term "humanised antibody", as used herein, refers to antibody molecules in which the CDR amino acids  
10 and selected other amino acids in the variable domains of the heavy and/or light chains of a non-human donor antibody have been substituted in place of the equivalent amino acids in a human antibody. The humanised antibody thus closely resembles a human antibody but has the binding ability of the donor antibody.

In a further alternative, the antibody may be a "bispecific" antibody, that is an antibody  
15 having two different antigen binding domains, each domain being directed against a different epitope.

Phage display technology may be utilised to select genes which encode antibodies with binding activities towards the polypeptides of the invention either from repertoires of PCR amplified V-genes of lymphocytes from humans screened for possessing the  
20 relevant antibodies, or from naive libraries (McCafferty, J. *et al.*, (1990), Nature 348, 552-554; Marks, J. *et al.*, (1992) Biotechnology 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, (1991) Nature 352, 624-628).

Antibodies generated by the above techniques, whether polyclonal or monoclonal, have  
25 additional utility in that they may be employed as reagents in immunoassays, radioimmunoassays (RIA) or enzyme-linked immunosorbent assays (ELISA). In these applications, the antibodies can be labelled with an analytically-detectable reagent such as a radioisotope, a fluorescent molecule or an enzyme.

Preferred nucleic acid molecules of the second and third aspects of the invention are  
30 those which encode the polypeptide sequences recited in SEQ ID NO:2, and functionally

equivalent polypeptides, including active fragments of the LBDG2 polypeptide, such as a fragment including the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG2 polypeptide sequence, or a homologue thereof.

5 Nucleic acid molecules encompassing these stretches of sequence form a preferred embodiment of this aspect of the invention.

These nucleic acid molecules may be used in the methods and applications described herein. The nucleic acid molecules of the invention preferably comprise at least n consecutive nucleotides from the sequences disclosed herein where, depending on the particular sequence, n is 10 or more (for example, 12, 14, 15, 18, 20, 25, 30, 35, 40 or  
10 more).

The nucleic acid molecules of the invention also include sequences that are complementary to nucleic acid molecules described above (for example, for antisense or probing purposes).

Nucleic acid molecules of the present invention may be in the form of RNA, such as  
15 mRNA, or in the form of DNA, including, for instance cDNA, synthetic DNA or genomic DNA. Such nucleic acid molecules may be obtained by cloning, by chemical synthetic techniques or by a combination thereof. The nucleic acid molecules can be prepared, for example, by chemical synthesis using techniques such as solid phase phosphoramidite chemical synthesis, from genomic or cDNA libraries or by separation  
20 from an organism. RNA molecules may generally be generated by the *in vitro* or *in vivo* transcription of DNA sequences.

The nucleic acid molecules may be double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

25 The term "nucleic acid molecule" also includes analogues of DNA and RNA, such as those containing modified backbones, and peptide nucleic acids (PNA). The term "PNA", as used herein, refers to an antisense molecule or an anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues, which preferably ends in lysine. The terminal lysine confers

solubility to the composition. PNAs may be pegylated to extend their lifespan in a cell, where they preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P.E. *et al.* (1993) *Anticancer Drug Des.* 8:53-63).

A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:2, or an active  
5 fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:1. These molecules also may have a different sequence which, as a result of the degeneracy of the genetic code, encodes the polypeptide SEQ ID NO:2, or an active fragment of the LBDG2 polypeptide, such as a fragment including the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The  
10 LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 1104 and residue 1309 of the LBDG2 polypeptide sequence. In SEQ ID NO:1 the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region is thus encoded by a nucleic acid molecule including nucleotide 3351 to 3968. Nucleic acid molecules encompassing this stretch of sequence, and homologues of this sequence, form  
15 a preferred embodiment of this aspect of the invention.

Such nucleic acid molecules that encode the polypeptide of SEQ ID NO:2 may include, but are not limited to, the coding sequence for the mature polypeptide by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those encoding a leader or secretory sequence, such as a pro-, pre- or prepro- polypeptide  
20 sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with further additional, non-coding sequences, including non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription (including termination signals), ribosome binding and mRNA stability. The nucleic acid molecules may also include  
25 additional sequences which encode additional amino acids, such as those which provide additional functionalities.

The nucleic acid molecules of the second and third aspects of the invention may also encode the fragments or the functional equivalents of the polypeptides and fragments of the first aspect of the invention.

30 As discussed above, a preferred fragment of the LBDG2 polypeptide is a fragment

including the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The Nuclear Hormone Receptor Ligand Binding Domain region is encoded by a nucleic acid molecule including nucleotide 3351 to 3968 of SEQ ID NO:1.

Functionally equivalent nucleic acid molecules according to the invention may be naturally-occurring variants such as a naturally-occurring allelic variant, or the molecules may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the nucleic acid molecule may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells or organisms.

Among variants in this regard are variants that differ from the aforementioned nucleic acid molecules by nucleotide substitutions, deletions or insertions. The substitutions, deletions or insertions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or insertions.

The nucleic acid molecules of the invention can also be engineered, using methods generally known in the art, for a variety of reasons, including modifying the cloning, processing, and/or expression of the gene product (the polypeptide). DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides are included as techniques which may be used to engineer the nucleotide sequences. Site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations and so forth.

Nucleic acid molecules which encode a polypeptide of the first aspect of the invention may be ligated to a heterologous sequence so that the combined nucleic acid molecule encodes a fusion protein. Such combined nucleic acid molecules are included within the second or third aspects of the invention. For example, to screen peptide libraries for inhibitors of the activity of the polypeptide, it may be useful to express, using such a combined nucleic acid molecule, a fusion protein that can be recognised by a commercially-available antibody. A fusion protein may also be engineered to contain a cleavage site located between the sequence of the polypeptide of the invention and the sequence of a heterologous protein so that the polypeptide may be cleaved and purified

away from the heterologous protein.

The nucleic acid molecules of the invention also include antisense molecules that are partially complementary to nucleic acid molecules encoding polypeptides of the present invention and that therefore hybridize to the encoding nucleic acid molecules (hybridization). Such antisense molecules, such as oligonucleotides, can be designed to  
5 recognize, specifically bind to and prevent transcription of a target nucleic acid encoding a polypeptide of the invention, as will be known by those of ordinary skill in the art (see, for example, Cohen, J.S., Trends in Pharm. Sci., 10, 435 (1989), Okano, J. Neurochem. 56, 560 (1991); O'Connor, J. Neurochem 56, 560 (1991); Lee *et al.*, Nucleic Acids Res 6,  
10 3073 (1979); Cooney *et al.*, Science 241, 456 (1988); Dervan *et al.*, Science 251, 1360 (1991).

The term "hybridization" as used here refers to the association of two nucleic acid molecules with one another by hydrogen bonding. Typically, one molecule will be fixed to a solid support and the other will be free in solution. Then, the two molecules may be  
15 placed in contact with one another under conditions that favour hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase molecule to the solid support (Denhardt's reagent or BLOTTO); the concentration of the molecules; use of compounds to increase the rate of association of  
20 molecules (dextran sulphate or polyethylene glycol); and the stringency of the washing conditions following hybridization (see Sambrook *et al.* [*supra*]).

The inhibition of hybridization of a completely complementary molecule to a target molecule may be examined using a hybridization assay, as known in the art (see, for example, Sambrook *et al.* [*supra*]). A substantially homologous molecule will then  
25 compete for and inhibit the binding of a completely homologous molecule to the target molecule under various conditions of stringency, as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods Enzymol. 152:507-511).

"Stringency" refers to conditions in a hybridization reaction that favour the association of  
30 very similar molecules over association of molecules that differ. High stringency

hybridisation conditions are defined as overnight incubation at 42°C in a solution comprising 50% formamide, 5XSSC (150mM NaCl, 15mM trisodium citrate), 50mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulphate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in  
5 0.1X SSC at approximately 65°C. Low stringency conditions involve the hybridisation reaction being carried out at 35°C (see Sambrook *et al.* [*supra*]). Preferably, the conditions used for hybridization are those of high stringency.

Preferred embodiments of this aspect of the invention are nucleic acid molecules that are at least 80% identical over their entire length to a nucleic acid molecule encoding the  
10 LBDG2 polypeptide (SEQ ID NO:2), and nucleic acid molecules that are substantially complementary to such nucleic acid molecules. A preferred active fragment is a fragment that includes an LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG2 polypeptide sequences, respectively. Accordingly, preferred nucleic acid molecules include those that are at least 80% identical over their entire length to a nucleic  
15 acid molecule encoding the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG2 polypeptide sequence.

Percentage identity, as referred to herein, is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>).

20 Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to the nucleic acid molecule having the sequence given in SEQ ID NO:1 to a region including nucleotides 3351-3968 of this sequence; or a nucleic acid molecule that is complementary to any one of these regions of nucleic acid. In this regard, nucleic acid molecules at least 90%, preferably at  
25 least 95%, more preferably at least 98% or 99% identical over their entire length to the same are particularly preferred. Preferred embodiments in this respect are nucleic acid molecules that encode polypeptides which retain substantially the same biological function or activity as the LBDG2 polypeptide.

The invention also provides a process for detecting a nucleic acid molecule of the  
30 invention, comprising the steps of: (a) contacting a nucleic probe according to the

invention with a biological sample under hybridizing conditions to form duplexes; and  
(b) detecting any such duplexes that are formed.

As discussed additionally below in connection with assays that may be utilised according to the invention, a nucleic acid molecule as described above may be used as a  
5 hybridization probe for RNA, cDNA or genomic DNA, in order to isolate full-length cDNAs and genomic clones encoding the LBDG2 polypeptide and to isolate cDNA and genomic clones of homologous or orthologous genes that have a high sequence similarity to the gene encoding this polypeptide.

In this regard, the following techniques, among others known in the art, may be utilised  
10 and are discussed below for purposes of illustration. Methods for DNA sequencing and analysis are well known and are generally available in the art and may, indeed, be used to practice many of the embodiments of the invention discussed herein. Such methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7  
15 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proof-reading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, MD). Preferably, the sequencing process may be automated using machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), the Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373  
20 and 377 DNA Sequencers (Perkin Elmer).

One method for isolating a nucleic acid molecule encoding a polypeptide with an equivalent function to that of the LBDG2 polypeptide, particularly with an equivalent function to the LBDG2 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG2 polypeptide, is to probe a genomic or cDNA library with a natural or artificially-  
25 designed probe using standard procedures that are recognised in the art (see, for example, "Current Protocols in Molecular Biology", Ausubel *et al.* (eds). Greene Publishing Association and John Wiley Interscience, New York, 1989,1992). Probes comprising at least 15, preferably at least 30, and more preferably at least 50, contiguous bases that correspond to, or are complementary to, nucleic acid sequences from the appropriate  
30 encoding gene (SEQ ID NO:1), particularly a region from nucleotides 3351-3968 of SEQ

ID NO:1, are particularly useful probes.

Such probes may be labelled with an analytically-detectable reagent to facilitate their identification. Useful reagents include, but are not limited to, radioisotopes, fluorescent dyes and enzymes that are capable of catalysing the formation of a detectable product.

5 Using these probes, the ordinarily skilled artisan will be capable of isolating complementary copies of genomic DNA, cDNA or RNA polynucleotides encoding proteins of interest from human, mammalian or other animal sources and screening such sources for related sequences, for example, for additional members of the family, type and/or subtype.

10 In many cases, isolated cDNA sequences will be incomplete, in that the region encoding the polypeptide will be cut short, normally at the 5' end. Several methods are available to obtain full length cDNAs, or to extend short cDNAs. Such sequences may be extended utilising a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one  
15 method which may be employed is based on the method of Rapid Amplification of cDNA Ends (RACE; see, for example, Frohman *et al.*, Proc. Natl. Acad. Sci. USA (1988) 85: 8998-9002). Recent modifications of this technique, exemplified by the Marathon™ technology (Clontech Laboratories Inc.), for example, have significantly simplified the search for longer cDNAs. A slightly different technique, termed "restriction-site" PCR,  
20 uses universal primers to retrieve unknown nucleic acid sequence adjacent a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Inverse PCR may also be used to amplify or to extend sequences using divergent primers based on a known region (Triglia, T., *et al.* (1988) Nucleic Acids Res. 16:8186). Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent a known  
25 sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. *et al.* (1991) PCR Methods Applic. 1: 111-119). Another method which may be used to retrieve unknown sequences is that of Parker, J.D. *et al.* (1991); Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen  
30 libraries and is useful in finding intron/exon junctions.

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

In one embodiment of the invention, the nucleic acid molecules of the present invention may be used for chromosome localisation. In this technique, a nucleic acid molecule is specifically targeted to, and can hybridize with, a particular location on an individual  
10 human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important step in the confirmatory correlation of those sequences with the gene-associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V.  
15 McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationships between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes). This provides valuable information to investigators searching for disease genes using positional cloning or other gene  
20 discovery techniques. Once the disease or syndrome has been crudely localised by genetic linkage to a particular genomic region, any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleic acid molecule may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

25 The nucleic acid molecules of the present invention are also valuable for tissue localisation. Such techniques allow the determination of expression patterns of the polypeptide in tissues by detection of the mRNAs that encode them. These techniques include in situ hybridization techniques and nucleotide amplification techniques, such as PCR. Results from these studies provide an indication of the normal functions of the  
30 polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by a mutant gene provide valuable

insights into the role of mutant polypeptides in disease. Such inappropriate expression may be of a temporal, spatial or quantitative nature.

The vectors of the present invention comprise nucleic acid molecules of the invention and may be cloning or expression vectors. The host cells of the invention, which may be  
5 transformed, transfected or transduced with the vectors of the invention may be prokaryotic or eukaryotic.

The polypeptides of the invention may be prepared in recombinant form by expression of their encoding nucleic acid molecules in vectors contained within a host cell. Such expression methods are well known to those of skill in the art and many are described in  
10 detail by Sambrook *et al.* (*supra*) and Fernandez & Hoeffler (1998, eds. "Gene expression systems. Using nature for the art of expression". Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto).

Generally, any system or vector that is suitable to maintain, propagate or express nucleic acid molecules to produce a polypeptide in the required host may be used. The  
15 appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those described in Sambrook *et al.*, (*supra*). Generally, the encoding gene can be placed under the control of a control element such as a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator, so that the DNA sequence encoding the desired polypeptide  
20 is transcribed into RNA in the transformed host cell.

Examples of suitable expression systems include, for example, chromosomal, episomal and virus-derived systems, including, for example, vectors derived from: bacterial plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast  
25 chromosomal elements, viruses such as baculoviruses, papova viruses such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, or combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, including cosmids and phagemids. Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid.

30 Particularly suitable expression systems include microorganisms such as bacteria

transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (for example, baculovirus); plant cell systems transformed with virus expression vectors (for example, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (for example, Ti or pBR322 plasmids); or animal cell systems. Cell-free translation systems can also be employed to produce the polypeptides of the invention.

Introduction of nucleic acid molecules encoding a polypeptide of the present invention into host cells can be effected by methods described in many standard laboratory manuals, such as Davis *et al.*, Basic Methods in Molecular Biology (1986) and Sambrook *et al.*, (*supra*). Particularly suitable methods include calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection (see Sambrook *et al.*, 1989 [*supra*]; Ausubel *et al.*, 1991 [*supra*]; Spector, Goldman & Leinwald, 1998). In eukaryotic cells, expression systems may either be transient (for example, episomal) or permanent (chromosomal integration) according to the needs of the system.

The encoding nucleic acid molecule may or may not include a sequence encoding a control sequence, such as a signal peptide or leader sequence, as desired, for example, for secretion of the translated polypeptide into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals. Leader sequences can be removed by the bacterial host in post-translational processing.

In addition to control sequences, it may be desirable to add regulatory sequences that allow for regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory sequences are those which cause the expression of a gene to be increased or decreased in response to a chemical or physical stimulus, including the presence of a regulatory compound or to various temperature or metabolic conditions. Regulatory sequences are those non-translated regions of the vector, such as enhancers, promoters and 5' and 3' untranslated regions. These interact with host cellular proteins to

carry out transcription and translation. Such regulatory sequences may vary in their strength and specificity. Depending on the vector system and host utilised, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript phagemid (Stratagene, LaJolla, CA) or pSport1<sup>TM</sup> plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (for example, heat shock, RUBISCO and storage protein genes) or from plant viruses (for example, viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

An expression vector is constructed so that the particular nucleic acid coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the regulatory sequences being such that the coding sequence is transcribed under the "control" of the regulatory sequences, i.e., RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence. In some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; i.e., to maintain the reading frame.

The control sequences and other regulatory sequences may be ligated to the nucleic acid coding sequence prior to insertion into a vector. Alternatively, the coding sequence can be cloned directly into an expression vector that already contains the control sequences and an appropriate restriction site.

For long-term, high-yield production of a recombinant polypeptide, stable expression is preferred. For example, cell lines which stably express the polypeptide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow

for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells that successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalised cell lines available from the American Type Culture Collection (ATCC) including, but not limited to, Chinese hamster ovary (CHO), HeLa, baby hamster kidney (BHK), monkey kidney (COS), C127, 3T3, BHK, HEK 293, Bowes melanoma and human hepatocellular carcinoma (for example Hep G2) cells and a number of other cell lines.

In the baculovirus system, the materials for baculovirus/insect cell expression systems are commercially available in kit form from, inter alia, Invitrogen, San Diego CA (the "MaxBac" kit). These techniques are generally known to those skilled in the art and are described fully in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987). Particularly suitable host cells for use in this system include insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells.

There are many plant cell culture and whole plant genetic expression systems known in the art. Examples of suitable plant cellular genetic expression systems include those described in US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, (1991) *Phytochemistry* 30, 3861-3863.

In particular, all plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be utilised, so that whole plants are recovered which contain the transferred gene. Practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugar cane, sugar beet, cotton, fruit and other trees, legumes and vegetables.

Examples of particularly preferred bacterial host cells include streptococci, staphylococci, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells.

Examples of particularly suitable host cells for fungal expression include yeast cells (for example, *S. cerevisiae*) and *Aspergillus* cells.

Any number of selection systems are known in the art that may be used to recover transformed cell lines. Examples include the herpes simplex virus thymidine kinase (Wigler, M. *et al.* (1977) *Cell* 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. *et al.* (1980) *Cell* 22:817-23) genes that can be employed in tk- or aprt± cells, respectively.

Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dihydrofolate reductase (DHFR) that confers resistance to methotrexate (Wigler, M. *et al.* (1980) *Proc. Natl. Acad. Sci.* 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. *et al.* (1981) *J. Mol. Biol.* 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. Additional selectable genes have been described, examples of which will be clear to those of skill in the art.

Although the presence or absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the relevant sequence is inserted within a marker gene sequence, transformed cells containing the appropriate sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding a polypeptide of the invention under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain a nucleic acid sequence encoding a polypeptide of the invention and which express said polypeptide may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassays, for example, fluorescence activated cell sorting (FACS) or immunoassay techniques (such as the enzyme-linked immunosorbent assay [ELISA] and radioimmunoassay [RIA]), that include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein (see Hampton, R. *et al.* (1990) *Serological*

Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. *et al.* (1983) J. Exp. Med, 158, 1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing  
5 labelled hybridization or PCR probes for detecting sequences related to nucleic acid molecules encoding polypeptides of the present invention include oligolabelling, nick translation, end-labelling or PCR amplification using a labelled polynucleotide. Alternatively, the sequences encoding the polypeptide of the invention may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art,  
10 are commercially available, and may be used to synthesise RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labelled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp., Cleveland, OH)).

15 Suitable reporter molecules or labels, which may be used for ease of detection, include radionuclides, enzymes and fluorescent, chemiluminescent or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Nucleic acid molecules according to the present invention may also be used to create transgenic animals, particularly rodent animals. Such transgenic animals form a further  
20 aspect of the present invention. This may be done locally by modification of somatic cells, or by germ line therapy to incorporate heritable modifications. Such transgenic animals may be particularly useful in the generation of animal models for drug molecules effective as modulators of the polypeptides of the present invention.

The polypeptide can be recovered and purified from recombinant cell cultures by well-  
25 known methods including ammonium sulphate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. High performance liquid chromatography is particularly useful for purification. Well known techniques for refolding proteins may be  
30 employed to regenerate an active conformation when the polypeptide is denatured during

isolation and or purification.

Specialised vector constructions may also be used to facilitate purification of proteins, as desired, by joining sequences encoding the polypeptides of the invention to a nucleotide sequence encoding a polypeptide domain that will facilitate purification of soluble proteins. Examples of such purification-facilitating domains include metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilised metals, protein A domains that allow purification on immobilised immunoglobulin, and the domain utilised in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and the polypeptide of the invention may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing the polypeptide of the invention fused to several histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification by IMAC (immobilised metal ion affinity chromatography as described in Porath, J. *et al.* (1992) Prot. Exp. Purif. 3: 263-281) while the thioredoxin or enterokinase cleavage site provides a means for purifying the polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. *et al.* (DNA Cell Biol. 199312:441-453).

If the polypeptide is to be expressed for use in screening assays, generally it is preferred that it be produced at the surface of the host cell in which it is expressed. In this event, the host cells may be harvested prior to use in the screening assay, for example using techniques such as fluorescence activated cell sorting (FACS) or immunoaffinity techniques. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the expressed polypeptide. If polypeptide is produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

The polypeptide of the invention can be used to screen libraries of compounds in any of a variety of drug screening techniques. Such compounds may activate (agonise) or inhibit (antagonise) the level of expression of the gene or the activity of the polypeptide of the invention and form a further aspect of the present invention. Preferred compounds are

effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

Agonist or antagonist compounds may be isolated from, for example, cells, cell-free  
5 preparations, chemical libraries or natural product mixtures. These agonists or antagonists may be natural or modified substrates, ligands, enzymes, receptors or structural or functional mimetics. For a suitable review of such screening techniques, see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991).

Compounds that are most likely to be good antagonists are molecules that bind to the  
10 polypeptide of the invention without inducing the biological effects of the polypeptide upon binding to it. Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to the polypeptide of the invention and thereby inhibit or extinguish its activity. In this fashion, binding of the polypeptide to normal cellular binding molecules may be inhibited, such that the normal biological activity of  
15 the polypeptide is prevented.

The polypeptide of the invention that is employed in such a screening technique may be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. In general, such screening procedures may involve using appropriate cells or cell membranes that express the polypeptide that are contacted with a test compound to  
20 observe binding, or stimulation or inhibition of a functional response. The functional response of the cells contacted with the test compound is then compared with control cells that were not contacted with the test compound. Such an assay may assess whether the test compound results in a signal generated by activation of the polypeptide, using an appropriate detection system. Inhibitors of activation are generally assayed in the  
25 presence of a known agonist and the effect on activation by the agonist in the presence of the test compound is observed.

Alternatively, simple binding assays may be used, in which the adherence of a test compound to a surface bearing the polypeptide is detected by means of a label directly or indirectly associated with the test compound or in an assay involving competition with a  
30 labelled competitor. In another embodiment, competitive drug screening assays may be

used, in which neutralising antibodies that are capable of binding the polypeptide specifically compete with a test compound for binding. In this manner, the antibodies can be used to detect the presence of any test compound that possesses specific binding affinity for the polypeptide.

- 5 Assays may also be designed to detect the effect of added test compounds on the production of mRNA encoding the polypeptide in cells. For example, an ELISA may be constructed that measures secreted or cell-associated levels of polypeptide using monoclonal or polyclonal antibodies by standard methods known in the art, and this can be used to search for compounds that may inhibit or enhance the production of the
- 10 polypeptide from suitably manipulated cells or tissues. The formation of binding complexes between the polypeptide and the compound being tested may then be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the polypeptide of interest

15 (see International patent application WO84/03564). In this method, large numbers of different small test compounds are synthesised on a solid substrate, which may then be reacted with the polypeptide of the invention and washed. One way of immobilising the polypeptide is to use non-neutralising antibodies. Bound polypeptide may then be detected using methods that are well known in the art. Purified polypeptide can also be

20 coated directly onto plates for use in the aforementioned drug screening techniques.

The polypeptide of the invention may be used to identify membrane-bound or soluble receptors, through standard receptor binding techniques that are known in the art, such as ligand binding and crosslinking assays in which the polypeptide is labelled with a radioactive isotope, is chemically modified, or is fused to a peptide sequence that

25 facilitates its detection or purification, and incubated with a source of the putative receptor (for example, a composition of cells, cell membranes, cell supernatants, tissue extracts, or bodily fluids). The efficacy of binding may be measured using biophysical techniques such as surface plasmon resonance and spectroscopy. Binding assays may be used for the purification and cloning of the receptor, but may also identify agonists and

30 antagonists of the polypeptide, that compete with the binding of the polypeptide to its

receptor. Standard methods for conducting screening assays are well understood in the art.

The invention also includes a screening kit useful in the methods for identifying agonists, antagonists, ligands, receptors, substrates, and enzymes that are described above.

- 5 The invention includes the agonists, antagonists, ligands, receptors, substrates and enzymes, and other compounds which modulate the activity or antigenicity of the polypeptide of the invention discovered by the methods that are described above.

The invention also provides pharmaceutical compositions comprising a polypeptide, nucleic acid, ligand or compound of the invention in combination with a suitable  
10 pharmaceutical carrier. These compositions may be suitable as therapeutic or diagnostic reagents, as vaccines, or as other immunogenic compositions, as outlined in detail below.

According to the terminology used herein, a composition containing a polypeptide, nucleic acid, ligand or compound [X] is "substantially free of" impurities [herein, Y] when at least 85% by weight of the total X+Y in the composition is X. Preferably, X  
15 comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95%, 98% or even 99% by weight.

The pharmaceutical compositions should preferably comprise a therapeutically effective amount of the polypeptide, nucleic acid molecule, ligand, or compound of the invention. The term "therapeutically effective amount" as used herein refers to an amount of a  
20 therapeutic agent needed to treat, ameliorate, or prevent a targetted disease or condition, or to exhibit a detectable therapeutic or preventative effect. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, for example, of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and  
25 route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

The precise effective amount for a human subject will depend upon the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and

tolerance/response to therapy. This amount can be determined by routine experimentation and is within the judgement of the clinician. Generally, an effective dose will be from 0.01 mg/kg to 50 mg/kg, preferably 0.05 mg/kg to 10 mg/kg. Compositions may be administered individually to a patient or may be administered in combination with other agents, drugs or hormones.

A pharmaceutical composition may also contain a pharmaceutically acceptable carrier, for administration of a therapeutic agent. Such carriers include antibodies and other polypeptides, genes and other therapeutic agents such as liposomes, provided that the carrier does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolised macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers and inactive virus particles.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulphates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable carriers is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may additionally contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such compositions. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

The pharmaceutical compositions utilised in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal or transcutaneous

applications (for example, see WO98/20734), subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, intravaginal or rectal means. Gene guns or hyposprays may also be used to administer the pharmaceutical compositions of the invention. Typically, the therapeutic compositions may be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared.

Direct delivery of the compositions will generally be accomplished by injection, subcutaneously, intraperitoneally, intravenously or intramuscularly, or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion.

10 Dosage treatment may be a single dose schedule or a multiple dose schedule.

If the activity of the polypeptide of the invention is in excess in a particular disease state, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as described above, along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the polypeptide, such as by blocking the binding of ligands, substrates, enzymes, receptors, or by inhibiting a second signal, and thereby alleviating the abnormal condition. Preferably, such antagonists are antibodies. Most preferably, such antibodies are chimeric and/or humanised to minimise their immunogenicity, as described previously.

In another approach, soluble forms of the polypeptide that retain binding affinity for the ligand, substrate, enzyme, receptor, in question, may be administered. Typically, the polypeptide may be administered in the form of fragments that retain the relevant portions.

In an alternative approach, expression of the gene encoding the polypeptide can be inhibited using expression blocking techniques, such as the use of antisense nucleic acid molecules (as described above), either internally generated or separately administered. Modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions (signal sequence, promoters, enhancers and introns) of the gene encoding the polypeptide. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of

the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. *et al.* (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The  
5 complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Such oligonucleotides may be administered or may be generated in situ from expression *in vivo*.

In addition, expression of the polypeptide of the invention may be prevented by using ribozymes specific to its encoding mRNA sequence. Ribozymes are catalytically active  
10 RNAs that can be natural or synthetic (see for example Usman, N, *et al.*, Curr. Opin. Struct. Biol (1996) 6(4), 527-33). Synthetic ribozymes can be designed to specifically cleave mRNAs at selected positions thereby preventing translation of the mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate backbone and natural bases, as normally found in RNA molecules. Alternatively the  
15 ribozymes may be synthesised with non-natural backbones, for example, 2'-O-methyl RNA, to provide protection from ribonuclease degradation and may contain modified bases.

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

For treating abnormal conditions related to an under-expression of the polypeptide of the invention and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound that activates the polypeptide, i.e., an agonist as described above, to alleviate the abnormal condition.  
30 Alternatively, a therapeutic amount of the polypeptide in combination with a suitable

pharmaceutical carrier may be administered to restore the relevant physiological balance of polypeptide.

Gene therapy may be employed to effect the endogenous production of the polypeptide by the relevant cells in the subject. Gene therapy is used to treat permanently the  
5 inappropriate production of the polypeptide by replacing a defective gene with a corrected therapeutic gene.

Gene therapy of the present invention can occur *in vivo* or *ex vivo*. *Ex vivo* gene therapy requires the isolation and purification of patient cells, the introduction of a therapeutic gene and introduction of the genetically altered cells back into the patient. In contrast, *in*  
10 *vivo* gene therapy does not require isolation and purification of a patient's cells.

The therapeutic gene is typically "packaged" for administration to a patient. Gene delivery vehicles may be non-viral, such as liposomes, or replication-deficient viruses, such as adenovirus as described by Berkner, K.L., in *Curr. Top. Microbiol. Immunol.*, 158, 39-66 (1992) or adeno-associated virus (AAV) vectors as described by Muzyczka,  
15 N., in *Curr. Top. Microbiol. Immunol.*, 158, 97-129 (1992) and U.S. Patent No. 5,252,479. For example, a nucleic acid molecule encoding a polypeptide of the invention may be engineered for expression in a replication-defective retroviral vector. This expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding the polypeptide,  
20 such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo* (see Chapter 20, *Gene Therapy and other Molecular Genetic-based Therapeutic Approaches*, (and references cited therein) in *Human Molecular Genetics* (1996), T Strachan and A P Read, BIOS Scientific Publishers  
25 Ltd).

Another approach is the administration of "naked DNA" in which the therapeutic gene is directly injected into the bloodstream or muscle tissue.

In situations in which the polypeptides or nucleic acid molecules of the invention are disease-causing agents, the invention provides that they can be used in vaccines to raise  
30 antibodies against the disease causing agent.

Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection). Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with pharmaceutically-acceptable carriers as described above, which include  
5 any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, and other pathogens.

10 Since polypeptides may be broken down in the stomach, vaccines comprising polypeptides are preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation  
15 isotonic with the blood of the recipient, and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents.

The vaccine formulations of the invention may be presented in unit-dose or multi-dose containers. For example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use.

20 The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

This invention also relates to the use of nucleic acid molecules according to the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the nucleic acid molecules of the invention which is associated with a dysfunction will  
25 provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acid molecules for diagnosis may be obtained from a subject's cells, such as from  
30 blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used

directly for detection or may be amplified enzymatically by using PCR, ligase chain reaction (LCR), strand displacement amplification (SDA), or other amplification techniques (see Saiki *et al.*, Nature, 324, 163-166 (1986); Bej, *et al.*, Crit. Rev. Biochem. Molec. Biol., 26, 301-334 (1991); Birkenmeyer *et al.*, J. Virol. Meth., 35, 117-126 (1991); Van Brunt, J., Bio/Technology, 8, 291-294 (1990)) prior to analysis.

In one embodiment, this aspect of the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to the invention and comparing said level of expression to a control level, wherein a level that is different to said control level is indicative of disease. The method may comprise the steps of:

- a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule of the invention and the probe;
- b) contacting a control sample with said probe under the same conditions used in step a);
- 15 c) and detecting the presence of hybrid complexes in said samples;

wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.

A further aspect of the invention comprises a diagnostic method comprising the steps of:

- a) obtaining a tissue sample from a patient being tested for disease;
- 20 b) isolating a nucleic acid molecule according to the invention from said tissue sample; and,
- c) diagnosing the patient for disease by detecting the presence of a mutation in the nucleic acid molecule which is associated with disease.

To aid the detection of nucleic acid molecules in the above-described methods, an amplification step, for example using PCR, may be included.

Deletions and insertions can be detected by a change in the size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labelled RNA of the invention or alternatively, labelled antisense DNA

sequences of the invention. Perfectly-matched sequences can be distinguished from mismatched duplexes by RNase digestion or by assessing differences in melting temperatures. The presence or absence of the mutation in the patient may be detected by contacting DNA with a nucleic acid probe that hybridises to the DNA under stringent  
5 conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation in the corresponding portion of the DNA strand.

10 Such diagnostics are particularly useful for prenatal and even neonatal testing.

Point mutations and other sequence differences between the reference gene and "mutant" genes can be identified by other well-known techniques, such as direct DNA sequencing or single-strand conformational polymorphism, (see Orita *et al.*, *Genomics*, 5, 874-879 (1989)). For example, a sequencing primer may be used with double-stranded PCR  
15 product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabelled nucleotides or by automatic sequencing procedures with fluorescent-tags. Cloned DNA segments may also be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. Further, point mutations and  
20 other sequence variations, such as polymorphisms, can be detected as described above, for example, through the use of allele-specific oligonucleotides for PCR amplification of sequences that differ by single nucleotides.

DNA sequence differences may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA  
25 sequencing (for example, Myers *et al.*, *Science* (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, *Proc. Natl. Acad. Sci. USA* (1985) 85: 4397-4401).

In addition to conventional gel electrophoresis and DNA sequencing, mutations such as  
30 microdeletions, aneuploidies, translocations, inversions, can also be detected by in situ

analysis (see, for example, Keller *et al.*, DNA Probes, 2nd Ed., Stockton Press, New York, N.Y., USA (1993)), that is, DNA or RNA sequences in cells can be analysed for mutations without need for their isolation and/or immobilisation onto a membrane. Fluorescence in situ hybridization (FISH) is presently the most commonly applied method and numerous reviews of FISH have appeared (see, for example, Trachuck *et al.*, Science, 250: 559-562 (1990), and Trask *et al.*, Trends, Genet. 7:149-154 (1991)).

In another embodiment of the invention, an array of oligonucleotide probes comprising a nucleic acid molecule according to the invention can be constructed to conduct efficient screening of genetic variants, mutations and polymorphisms. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee *et al.*, Science (1996) 274: 610-613).

In one embodiment, the array is prepared and used according to the methods described in PCT application WO95/11995 (Chee *et al.*); Lockhart, D. J. *et al.* (1996) Nat. Biotech. 14: 1675-1680); and Schena, M. *et al.* (1996) Proc. Natl. Acad. Sci. 93: 10614-10619). Oligonucleotide pairs may range from two to over one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support. In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application WO95/251116 (Baldeschweiler *et al.*). In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other number between two and over one million which lends itself to the efficient use of commercially-available instrumentation.

In addition to the methods discussed above, diseases may be diagnosed by methods

comprising determining, from a sample derived from a subject, an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

Assay techniques that can be used to determine levels of a polypeptide of the present invention in a sample derived from a host are well-known to those of skill in the art and are discussed in some detail above (including radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays). This aspect of the invention provides a diagnostic method which comprises the steps of: (a) contacting a ligand as described above with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

Protocols such as ELISA, RIA, and FACS for measuring polypeptide levels may additionally provide a basis for diagnosing altered or abnormal levels of polypeptide expression. Normal or standard values for polypeptide expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably humans, with antibody to the polypeptide under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, such as by photometric means.

Antibodies which specifically bind to a polypeptide of the invention may be used for the diagnosis of conditions or diseases characterised by expression of the polypeptide, or in assays to monitor patients being treated with the polypeptides, nucleic acid molecules, ligands and other compounds of the invention. Antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for the polypeptide include methods that utilise the antibody and a label to detect the polypeptide in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labelled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules known in the art may be used, several of which are described above.

Quantities of polypeptide expressed in subject, control and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease. Diagnostic assays may be used to distinguish between absence, presence, and excess expression of polypeptide and to  
5 monitor regulation of polypeptide levels during therapeutic intervention. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials or in monitoring the treatment of an individual patient.

A diagnostic kit of the present invention may comprise:

- (a) a nucleic acid molecule of the present invention;
- 10 (b) a polypeptide of the present invention; or
- (c) a ligand of the present invention.

In one aspect of the invention, a diagnostic kit may comprise a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to the invention; a second container containing primers useful for  
15 amplifying the nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease. The kit may further comprise a third container holding an agent for digesting unhybridised RNA.

In an alternative aspect of the invention, a diagnostic kit may comprise an array of nucleic acid molecules, at least one of which may be a nucleic acid molecule according to  
20 the invention.

To detect polypeptide according to the invention, a diagnostic kit may comprise one or more antibodies that bind to a polypeptide according to the invention; and a reagent useful for the detection of a binding reaction between the antibody and the polypeptide.

Such kits will be of use in diagnosing a disease or susceptibility to disease, particularly  
25 cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposis' sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation,

asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hyperparathyroidism, hypercalcemia, hypercholesterolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, dermatological disorders, including, acne, eczema, and wound healing, negative effects of aging, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated.

Various aspects and embodiments of the present invention will now be described in more detail by way of example, with particular reference to the LBDG2 polypeptide.

It will be appreciated that modification of detail may be made without departing from the scope of the invention.

### **Brief description of the Figures**

Figure 1: This is the front end of the Biopendium Target Mining Interface. A search of the database is initiated using the PDB code "1BSX:A".

Figure 2A: A selection is shown of the Inpharmatica Genome Threader results for the search using 1BSX:A. The arrow indicates *Homo Sapiens* Thyroid Hormone Receptor Beta-1, which has a typical Nuclear Hormone Receptor Ligand Binding Domain.

Figure 2B: A selection is shown of the Inpharmatica Genome Threader results for the search using 1BSX:A. The arrow indicates BAA22563.1 (LBDG2).

Figure 2C: Full list of forward PSI-BLAST results for the search using 1BSX:A. BAA22563.1 (LBDG2) is not identified.

Figure 3: The Redundant Sequence Display results page for BAA22563.1 (LBDG2).

Figure 4: InterPro PFAM search results for BAA22563.1 (LBDG2), see arrow ①.

Figure 5: NCBI Protein Report for BAA22563.1 (LBDG2).

Figure 6A: This is the front end of the Biopendium database. A search of the database is initiated using BAA22563.1 (LBDG2), as the query sequence.

Figure 6B: A selection of the Inpharmatica Genome Threader results of search using  
5 BAA22563.1 (LBDG2), as the query sequence. The arrow points to 1BSX:A.

Figure 6C: A selection of the reverse-maximised PSI-BLAST results obtained using BAA22563.1 (LBDG2), as the query sequence. The arrows numbered ① to ④ point to homologues of BAA22563.1 (LBDG2).

Figure 7: AIEye sequence alignment of BAA22563.1 (LBDG2) and 1BSX:A.

10 Figure 8A: LigEye for 1BSX:A that illustrates the sites of interaction of 3,5,3'-Triiodothyronine with the Ligand Binding Domain of *Homo sapiens* Thyroid Hormone Receptor Beta, 1BSX:A.

Figure 8B: iRasMol view of 1BSX:A, the Ligand Binding Domain of *Homo sapiens* Thyroid Hormone Receptor Beta.

15 Figure 9: AIEye sequence alignment of BAA22563.1 (LBDG2; *Homo sapiens* PRP8) with 4 homologues; AAF58573.1 (*Drosophila melanogaster* ①), P34369 (*Caenorhabditis elegans* ②), BAA78744.1 (*Oryza sativa* ③) and CAB80541.1 (*Arabidopsis thaliana* ④).

20 Figure 10: The linear dynamic range for target BAA22563.1 (LBDG2) reactions on colon cDNA.

Figure 11: The linear dynamic range for internal control 18s rRNA reactions on colon cDNA.

Figure 12: The linear dynamic range for internal control human ribosomal protein mRNA reactions on IM9 cell cDNA.

25 Figure 13: Normalised expression of BAA22563.1 (LBDG2) in 18 normal human tissues.

Figure 14: Normalised expression of BAA22563.1 (LBDG2) in a number of cell lines.

**Example: 1**

In order to initiate a search for novel, distantly related Nuclear Hormone Receptor Ligand Binding Domains, an archetypal family member is chosen, the Ligand Binding Domain of *Homo sapiens* Thyroid Hormone Receptor Beta. More specifically, the search is initiated using a structure from the Protein Data Bank (PDB) which is operated by the Research Collaboratory for Structural Bioinformatics.

The structure chosen is the Ligand Binding Domain of *Homo sapiens* Thyroid Hormone Receptor Beta (PDB code 1BSX:A; see Figure 1).

A search of the Biopendium (using the Target Mining Interface) for relatives of 1BSX:A takes place and returns 4096 Genome Threader results. The 4096 Genome Threader results include examples of typical Nuclear Hormone Receptor Ligand Binding Domains, such as that found between residues 211-461 of the *Homo sapiens* Thyroid Hormone Receptor Beta (see arrow in Figure 2A).

Among the proteins known to contain a Nuclear Hormone Receptor Ligand Binding Domain appears a protein which is not annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain, BAA22563.1 (LBDG2; see arrow in Figure 2B). The Inpharmatica Genome Threader has identified a region of the sequence BAA22563.1 (LBDG2), between residues 1104-1309, as having a structure similar to the Ligand Binding Domain of *Homo sapiens* Thyroid Hormone Receptor Beta. The possession of a structure similar to the Ligand Binding Domain of *Homo sapiens* Thyroid Hormone Receptor Beta suggests that residues 1104-1309 of BAA22563.1 (LBDG2) function as a Nuclear Hormone Receptor Ligand Binding Domain. The Genome Threader identifies this with 86% confidence.

The search of the Biopendium (using the Target Mining Interface) for relatives of 1BSX:A also returns 852 Forward PSI-Blast results. Forward PSI-Blast (see Figure 2C) is unable to identify this relationship; only the Inpharmatica Genome Threader is able to identify BAA22563.1 (LBDG2) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

In order to assess what is known in the public domain databases about BAA22563.1 (LBDG2) the Redundant Sequence Display Page (Figure 3) is viewed. There are no PROSITE or PRINTS hits which identify BAA22563.1 (LBDG2) as containing a Nuclear Hormone Receptor Ligand Binding Domain. PROSITE and PRINTS are databases that help to describe proteins of similar families. Returning no Nuclear Hormone Receptor Ligand Binding Domain hits from both databases means that BAA22563.1 (LBDG2) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using PROSITE or PRINTS.

In order to identify if any other public domain annotation vehicle is able to annotate BAA22563.1 (LBDG2) as containing a Nuclear Hormone Receptor Ligand Binding Domain, the BAA22563.1 (LBDG2) protein sequence is searched against the PFAM database (Protein Family Database of Alignment and hidden Markov models) at the InterPro website (see Figure 4 arrow ①). A PFAM-A match is found to PF00527/IPR000148, which is diagnostic of relatedness to Papillomavirus E7 protein. The Papillomavirus E7 protein match is located between residues 1561-1570 of BAA22563.1 (LBDG2). However, there are no PFAM-A matches annotating BAA22563.1 (LBDG2) as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus PFAM does not identify BAA22563.1 (LBDG2) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

Interestingly, PROSITE PFscan (see Figure 4 arrow ②) identifies a bipartite nuclear localisation signal in BAA22563.1 (LBDG2) at residues 35-52. A typical (although non-diagnostic) feature of Nuclear Hormone Receptors is the possession of a bipartite nuclear localisation signal.

The National Center for Biotechnology Information (NCBI) Genebank protein database is then viewed to examine if there is any further information that is known in the public domain relating to BAA22563.1 (LBDG2). This is the US public domain database for protein and gene sequence deposition (Figure 5). BAA22563.1 (LBDG2) is a *Homo sapiens* sequence, its Genebank protein ID is BAA22563.1 and it is 2335 amino acids in length. BAA22563.1 (LBDG2) is called a *Homo sapiens* homologue of *Saccharomyces cerevisiae* PRP8, a Pre-mRNA splicing factor. BAA22563.1 (LBDG2) was cloned by a

group of scientists at the Otuka GEN Research Institute; Kagasuno, Kawauchi-cho, Tokushima, Japan. The public domain information for this gene does not annotate it as containing a Nuclear Hormone Receptor Ligand Binding Domain.

Therefore, it can be concluded that using all public domain annotation tools,  
5 BAA22563.1 (LBDG2) may not be annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain. Only the Inpharmatica Genome Threader is able to annotate this protein as containing a Nuclear Hormone Receptor Ligand Binding Domain.

The reverse search is now carried out. BAA22563.1 (LBDG2) is now used as the query sequence in the Biopendium (see Figure 6A). The Inpharmatica Genome Threader  
10 identifies residues 1104-1309 of BAA22563.1 (LBDG2) as having a structure that is the same as the Ligand Binding Domain of *Homo sapiens* Thyroid Hormone Receptor Beta with 86% confidence (see arrow in Figure 6B). The Ligand Binding Domain of *Homo sapiens* Thyroid Hormone Receptor Beta (1BSX:A) was the original query sequence. Positive iterations of PSI-Blast do not return this result (Figure 6C). It is only the  
15 Inpharmatica Genome Threader that is able to identify this relationship.

The sequence of the *Homo sapiens* Thyroid Hormone Receptor Beta Ligand Binding Domain is chosen against which to view the sequence alignment of BAA22563.1 (LBDG2). Viewing the AlEye alignment (Figure 7) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology.

20 The *Homo sapiens* Thyroid Hormone Receptor Beta Ligand Binding Domain contains an “LBD motif” which has been found in all annotated Nuclear Hormone Receptor Ligand Binding Domains to date. The “LBD motif” is involved in recruiting Nuclear Hormone Receptor Co-Activators and Co-Repressors. The 6 residues; PHE293, LEU296, ASP300, GLN301, LEU304 and LEU305 constitute this motif in the *Homo sapiens* Thyroid  
25 Hormone Receptor Beta Ligand Binding Domain (see square boxes Figure 7). 4 residues (PHE1174, ASP1181, LEU1185, LEU1186) in BAA22563.1 (LBDG2) precisely match 4 (PHE293, ASP300, LEU304, LEU305) out of the 6 “LBD motif” residues in the *Homo sapiens* Thyroid Hormone Receptor Beta Ligand Binding Domain. Furthermore VAL1177 and ASN1182 in BAA22563.1 (LBDG2) conservatively substitute for the  
30 remaining 2 residues LEU296 and GLN301 in the “LBD motif” of *Homo sapiens*

Thyroid Hormone Receptor Beta Ligand Binding Domain. This indicates that BAA22563.1 (LBDG2) contains a Nuclear Hormone Receptor Ligand Binding Domain similar to The *Homo sapiens* Thyroid Hormone Receptor Beta Ligand Binding Domain.

In order to ensure that the protein identified is in fact a relative of the query sequence, the visualization programs "LigEye" (Figure 8A) and "iRasmol" (Figure 8B) are used. These visualization tools identify the active site of known protein structures by indicating the amino acids with which known small molecule inhibitors interact at the active site. These interactions are either through a direct hydrogen bond or through hydrophobic interactions. In this manner, one can see if the active site fold/structure is conserved between the identified homologue and the chosen protein of known structure. The LigEye view of the *Homo sapiens* Thyroid Hormone Receptor Beta Ligand Binding Domain reveals 5 residues which bind 3,5,3'-Triiodothyronine (circled in Figure 7). However, only 4 (ILE276, LEU330, ASN331 and LEU346) of these 5 residues lie within the Genome Threader alignment. Thus only these 4 residues can be used to consolidate the Genome Threader annotation of BAA22563.1 (LBDG2) as containing a Nuclear Hormone Receptor Ligand Binding Domain. Of these 4 residues there are 3 hydrophobic residues which line the pocket of the *Homo sapiens* Thyroid Hormone Receptor Beta Ligand Binding Domain; ILE276, LEU330 and LEU346. LEU330 and LEU346 of *Homo sapiens* Thyroid Hormone Receptor Beta Ligand Binding Domain are perfectly conserved in BAA22563.1 (LBDG2); LEU1216 and LEU1230 (circled in Figure 7). ILE276 of the *Homo sapiens* Thyroid Hormone Receptor Beta Ligand Binding Domain is conservatively substituted by LEU1161 in BAA22563.1 (LBDG2): (broken circle in Figure 7). This conservation of hydrophobicity in 3 out of the 3 hydrophobic residues (within the region of Genome Threader alignment) which line the binding pocket indicates that BAA22563.1 (LBDG2) will bind a hydrophobic steroid-like ligand.

ASN331 of the *Homo sapiens* Thyroid Hormone Receptor Beta Ligand Binding Domain is conservatively substituted by GLN1217 in BAA22563.1 (LBDG2): (broken circle in Figure 7). This indicates that indeed as predicted by the Inpharmatica Genome Threader, BAA22563.1 (LBDG2) folds in a similar manner to the *Homo sapiens* Thyroid Hormone Receptor Beta Ligand Binding Domain and as such is identified as containing a Nuclear Hormone Receptor Ligand Binding Domain.

Reverse-maximised PSI-BLAST of BAA22563.1 (LBDG2) identifies a *Drosophila melanogaster* homologue (AAF58573.1, see Figure 6C arrow ①), a *Caenorhabditis elegans* homologue (P34369, see Figure 6C arrow ②), an *Oryza sativa* homologue (BAA78744.1, see Figure 6C arrow ③), and an *Arabidopsis thaliana* homologue (CAB80541.1, see Figure 6C arrow ④). BAA22563.1 (LBDG2), AAF58573.1 (*Drosophila melanogaster* homologue), P34369 (*Caenorhabditis elegans* homologue), BAA78744.1 (*Oryza sativa* homologue) and CAB80541.1 (*Arabidopsis thaliana* homologue) are aligned and viewed in AIEye (Figure 9). AIEye reveals that the 4 predicted ligand binding residues (within the Genome Threader alignment; LEU1161, LEU1216, GLN1217 and LEU1230) of BAA22563.1 (LBDG2) are all precisely conserved in AAF58573.1 (*Drosophila melanogaster* homologue), P34369 (*Caenorhabditis elegans* homologue), BAA78744.1 (*Oryza sativa* homologue) and CAB80541.1 (*Arabidopsis thaliana* homologue). Furthermore all of the predicted “LBD motif” residues in BAA22563.1 (LBDG2; PHE1174, VAL1177, ASP1181, ASN1182, LEU1185 and LEU1186) are precisely conserved in AAF58573.1 (*Drosophila melanogaster* homologue), P34369 (*Caenorhabditis elegans* homologue), BAA78744.1 (*Oryza sativa* homologue) and CAB80541.1 (*Arabidopsis thaliana* homologue). The only exception is P34369 (*Caenorhabditis elegans* homologue) in which LEU1185 is conservatively substituted with a MET. Residues which are essential for the function of a protein will be conserved in homologues of that protein. Thus the conservation of residues which would be essential for the function of the predicted BAA22563.1 (LBDG2) Nuclear Hormone Receptor Ligand Binding Domain in AAF58573.1 (*Drosophila melanogaster* homologue), P34369 (*Caenorhabditis elegans* homologue), BAA78744.1 (*Oryza sativa* homologue) and CAB80541.1 (*Arabidopsis thaliana* homologue) strongly supports the annotation of BAA22563.1 (LBDG2) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

### **Example: 2**

In order to determine the tissue expression of the proposed LBD, Taqman RT-PCR quantitation was used. The TaqMan 3'- 5' exonuclease assay signals the formation of PCR amplicons by a process involving the nucleolytic degradation of a double-labeled fluorogenic probe that hybridises to the target template at a site between the two primer

recognition sequences (cf. U. S. Patent 5,876,930). The ABI Prism 7000 automates the detection and quantitative measurement of these signals, which are stoichiometrically related to the quantities of amplicons produced, during each cycle of amplification. In addition to providing substantial reductions in the time and labour requirements for PCR analyses, this technology permits simplified and potentially highly accurate quantification of target sequences in the reactions.

Human RNA prepared from non-diseased organs was purchased from either Ambion Europe (Huntingdon, UK) or Clontech (BD, Franklin Lakes, NJ). Oligonucleotide primers and probes were designed using Primer Express software (Applied Biosystems, Foster City CA) with a GC-content of 40-60%, no G-nucleotide at the 5'-end of the probe, and no more than 4 contiguous Gs. Each primer and probe was then analysed using BLAST<sup>®</sup> (Basic Local Alignment Search Tool, Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.: J Mol Biol 1990 Oct 5;215(3):403-10). Results confirmed that each oligonucleotide recognised the target sequence with a specificity >3 bp when compared to other known cDNA's or genomic sequence represented in the Unigene and GoldenPath publicly available databases. The sequence of the primers and probes were as follows:

BAA22563.1 (LBDG2) Forward primer: CAG ACA GGC CGC TGA CAT T

BAA22563.1 (LBDG2) Reverse primer: GCC ATC AGG AGG TCA ACA ACA

BAA22563.1 (LBDG2) Probe: AGT TTG GCC TCT TTC CCT CTG TCT GTG C

18s and human ribosomal protein pre-optimised primers and probe were purchased from Applied Biosystems, Foster City, CA. Probes were covalently conjugated with a fluorescent reporter dye (e.g. 6carboxy-fluorescein [FAM];  $\lambda_{em} = 518\text{nm}$ ) and a fluorescent quencher dye (6carboxytetram-ethyl-rhodamine [TAMRA];  $\lambda_{em} = 582\text{nm}$ ) at the most 5' and most 3' base, respectively. All primers and probes were obtained from Applied Biosystems, Germany. Primer/probe concentrations were titrated in the range of 50nM to 900nM and optimal concentrations for efficient PCR reactions were determined. Optimal primer and probe concentrations vary in between 100nM and 900nM depending on the target gene that was amplified. cDNA is prepared using components from Applied Biosystems, Foster City CA. 50 $\mu\text{l}$  reactions are prepared in 0.5ml RNase free tubes. Reactions contain 500ng total RNA; 1x reverse transcriptase buffer; 5.5mM MgCl<sub>2</sub>;

1mM dNTP's; 2.5µl random hexamers; 20U RNase inhibitor; and 62.5U reverse transcriptase. 25µl reactions were prepared in 0.5 ml thin-walled, optical grade PCR 96 well plates (Applied Biosystems, Foster City CA). Reactions contained: 1x final concentration of TaqMan Universal Master Mix (a proprietary mixture of AmpliTaq Gold DNA polymerase, AmpEraseX UNG, dNTPs with UTP, passive reference dye and optimised buffer components, Applied Biosystems, Foster City CA); 100nM Taqman probe; 300nM forward primer; 900nM reverse primer and 15ng of cDNA template.

Standard procedures for the operation of the ABI Prism 7000 or similar detection system were used. This included, for example with the ABI Prism 7000, use of all default program settings with the exception of reaction volume which was changed from 50 to 25 ul. Thermal cycling conditions consisted of two min at 50 C, 10 min at 95 C, followed by 40 cycles of 15 sec at 95 C and 1 min at 60 C. Cycle threshold (Ct) determinations, (i.e. non-integer calculations of the number of cycles required for reporter dye fluorescence resulting from the synthesis of PCR products to become significantly higher than background fluorescence levels), were automatically performed by the instrument for each reaction using default parameters. Assays for target sequences and ribosomal 18s (reference) sequences in the same cDNA samples were performed in separate reaction tubes. Within each experiment, a standard curve was carried out of a typical tissue sample, from 50ng to 0.78ng of cDNA template. From this standard curve, the amount of actual starting target or 18s cDNA in each test sample is determined.

The levels of target cDNA in each sample were normalised to the level of expression of target in a comparative sample. The levels of internal control cDNA in each sample were normalised to the level of expression of internal control in a comparative sample. The data was then represented as fold expression of normalised target sequence relative to the level of expression in the comparative sample, which is set arbitrarily to 1. Taqman RT-PCR was carried out on 2-fold dilutions of colon cDNA using primers/probes specific for BAA22563.1 (LBDG2) as described above. Figure 10 shows the Ct values plotted vs. the log input cDNA and illustrates that a linear relationship was seen over this range of input cDNA concentrations. Linear regression analysis of the standard curve was used to calculate the starting amount of cDNA from test Ct values.

Taqman RT-PCR was carried out on 2-fold dilutions of colon cDNA using primers/probes specific for 18s rRNA as described above. Figure 11 shows the Ct values plotted vs. the log input cDNA value, and illustrates that a linear relationship was seen over this range of input cDNA concentrations. Linear regression analysis of the standard curve was used to calculate the starting amount of cDNA from test Ct values.

Taqman RT-PCR was carried out on 2-fold dilutions of IM9 cDNA using primers/probes specific for human ribosomal protein mRNA as described above. Figure 12 shows the Ct values plotted vs. the log input cDNA value, and illustrates that a linear relationship was seen over this range of input cDNA concentrations. Linear regression analysis of the standard curve was used to calculate the starting amount of cDNA from test Ct values.

Taqman RT-PCR was carried out using 15ng of the indicated cDNA using primers/probes specific for BAA22563.1 (LBDG2) and 18s rRNA as described above. A standard curve for target and internal control was also carried out, using between 50ng to 0.78ng of cDNA template of a typical tissue sample. Using linear regression analysis of the standard curves, the Ct values were used to calculate the amount of actual starting target or 18s cDNA in each test sample.

The levels of target cDNA in each sample were normalised to the level of expression of target in a comparative sample, in this case, stomach. The levels of 18s cDNA in each sample were also normalised to the level of expression of 18s in stomach. The expression levels of BAA22563.1 (LBDG2) were then normalised to the expression levels of 18s. Figure 13 represents the fold expression of normalised target sequence relative to the level of expression in stomach cDNA, which is set arbitrarily to 1. Each sample was quantitated in between 2-4 individual experiments. Figure 13 shows the mean  $\pm$  SEM for the multiple experiments.

Taqman RT-PCR was carried out using 15ng of the indicated cDNA using primers/probes specific for BAA22563.1 (LBDG2) and human ribosomal protein mRNA as described in the detailed description. A standard curve for target and internal control was also carried out using a typical cell line sample, using between 50ng to 0.78ng of cDNA template. Using linear regression analysis of the standard curves, the Ct values

were used to calculate the amount of actual starting target or human ribosomal protein cDNA in each test sample.

The levels of target cDNA in each sample were normalised to the level of expression of target in a comparative sample, in this case LAK cells. The levels of human ribosomal protein cDNA in each sample were also normalised to the level of expression of human  
5 ribosomal protein cDNA in LAK cells. The expression levels of BAA22563.1 (LBDG2) were then normalised to the expression levels of human ribosomal protein. Figure 14 represents the fold expression of normalised target sequence relative to the level of expression in LAK cDNA, which is set arbitrarily to 1. Figure 14 shows the mean  $\pm$  SEM  
10 for duplicate measurements of each sample.

The mRNA for LBDG2 has been found in extracts from a variety of human tissues (Figure 13). The finding of high levels of the transcript in the human spleen is consistent with a role of LBDG2 in the immune system and in particular in lymphocyte development and function and in particular in B cell development and function.  
15 Development of agonists and antagonists for LBDG2 may therefore have a role in the therapeutic intervention in various human diseases of the immune system including autoimmunity, allergies and diseases associated with immunoglobulin dysfunction. These diseases include type I diabetes mellitus, rheumatoid arthritis, multiple sclerosis, psoriasis, renal failure arising from glomerulopathies, scleroderma, inflammatory bowel  
20 disease (both Crohns disease and ulcerative colitis), transplant rejection, asthma, atopic dermatitis, eczema, myelomas and in infectious diseases that require production of antibodies e.g. intracellular pathogen such as virus infected cells, tuberculosis, listeria.

The finding of mRNA for LBDG2 in human B cell lines such as Daudi, IM9 and Raji cells (Figure 14) is consistent with the finding of mRNA in the spleen. Finding high  
25 levels of expression of the mRNA in U937 cells suggests a role for LBDG2 in monocyte/macrophage functions and as such agonists or antagonists may be valuable in treating inflammatory diseases including chronic obstructive pulmonary disease (COPD), osteoarthritis, rheumatoid arthritis, inflammatory bowel disease, fibrosis such as liver fibrosis (cirrhosis) and skin fibrosis (scarring), atherosclerosis, dementia, multiple  
30 sclerosis, inflammatory pain.

In addition, the finding of significant levels of LBDG2 in adrenal, ovary, prostate and testis indicates that development of agonists and antagonists to LBDG2 may be of value in diseases such as benign prostatic hypertrophy, prostatic cancer, ovarian cancer, testicular cancer. In addition, agonists or antagonists for LBDG2 may be developed for  
5 treatment of diseases including but not exclusive to hypertension, responses to stress including stress of infectious diseases, regulation of salt and water homeostasis, control of fertility through regulation of ovulation (infertility and contraception), regulation of implantation (infertility and contraception) and regulation of spermatogenesis (infertility and contraception).

10 The finding that the mRNA for LBDG2 is expressed at significant levels in the human brain is noteworthy as this provides a potential link to human disease states and development of agonists and antagonists for the ligand binding domain of LBDG2 offers the potential for therapeutic intervention in various human diseases including cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas,  
15 head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposi's sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, oedema, angina,  
20 atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hyperparathyroidism, hypercalcemia,  
25 hypercholestrolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, dermatological disorders, including, acne, eczema, and wound healing, negative effects of aging, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are  
30 implicated.

The finding of a “non-classical” nuclear hormone receptor such as LBDG2 which contains a ligand binding domain in the absence of a DNA binding domain is consistent with the known literature which has consistently reported widespread effects of steroids in the brain (known as neurosteroids) and that these effects, in general, are mediated not through the known classic steroid hormone nuclear receptors which requires transcriptional activation. For instance, neurosteroids have been shown to influence neurotransmission particularly in the field of receptors such as those for GABA and NMDA and Sigma receptors. Neurosteroids have been shown to play a neuroprotective role. Therapeutic intervention through the development of agonists (or antagonists) to LBDG2 may therefore have a role in treatment of neurodegenerative conditions such as dementia, Parkinson’s disease and neurodegeneration following cerebrovascular disease such as infarction or haemorrhage (stroke) and trauma to the central nervous system and spinal cord. In addition, neurosteroids have been shown to influence cognitive processing, spatial learning and memory, anxiety and behaviours such as craving which leads to addictive behaviour patterns. Development of agonists and antagonists to LBDG2 may therefore lead to therapeutic intervention to treat dementias, learning difficulties, anxiety, addictive behaviours such as but not exclusively alcoholism, eating disorders and drug addiction.

**CLAIMS**

1. A polypeptide, which polypeptide:
  - (i) comprises or consists of the amino acid sequence as recited in SEQ ID NO:2;
  - (ii) is a fragment thereof having Nuclear Hormone Receptor Ligand Binding Domain activity or having an antigenic determinant in common with the polypeptide of (i); or
  - (iii) is a functional equivalent of (i) or (ii).
2. A polypeptide which is a fragment according to claim 1(ii), which includes the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG2 polypeptide, said Nuclear Hormone Receptor Ligand Binding Domain region being defined as including residues 1104 to 1309 inclusive, of the amino acid sequence recited in SEQ ID NO:2, wherein said fragment possesses the "LBD motif" residues PHE1174, VAL1177, ASP1181, ASN1182, LEU1185 and LEU1186, or equivalent residues, and possesses Nuclear Hormone Receptor Ligand Binding Domain activity.
3. A polypeptide which is a functional equivalent according to claim 1(iii), is homologous to the amino acid sequence as recited in SEQ ID NO:2, possesses the catalytic residues PHE1174, VAL1177, ASP1181, ASN1182, LEU1185 and LEU1186, or equivalent residues, and has Nuclear Hormone Receptor Ligand Binding Domain activity.
4. A polypeptide according to claim 3, wherein said functional equivalent is homologous to the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG2 polypeptide.
5. A fragment or functional equivalent according to any one of claims 1-4, which has greater than 80% sequence identity with an amino acid sequence as recited in SEQ ID NO:2, or with a fragment thereof that possesses Nuclear Hormone Receptor Ligand Binding Domain activity, preferably greater than 85%, 90%, 95%, 98% or 99% sequence identity, as determined using BLAST version 2.1.3 using the default

parameters specified by the NCBI (the National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].

- 5 6. A functional equivalent according to any one of claims 1-5, which exhibits significant structural homology with a polypeptide having the amino acid sequence given in any one of SEQ ID NO:2, or with a fragment thereof that possesses Nuclear Hormone Receptor Ligand Binding Domain activity.
7. A fragment as recited in claim 1, 2, or 5, having an antigenic determinant in common with the polypeptide of claim 1(i), which consists of 7 or more (for example, 8, 10,  
10 12, 14, 16, 18, 20 or more) amino acid residues from the sequence of SEQ ID NO:2.
8. A purified nucleic acid molecule which encodes a polypeptide according to any one of the preceding claims.
9. A purified nucleic acid molecule according to claim 8, which has the nucleic acid  
15 sequence as recited in SEQ ID NO:1, or is a redundant equivalent or fragment thereof.
10. A fragment of a purified nucleic acid molecule according to claim 8 or claim 9, which comprises nucleotides 3351 to 3968 of SEQ ID NO:1, or is a redundant equivalent thereof.
11. A purified nucleic acid molecule which hybridizes under high stringency conditions  
20 with a nucleic acid molecule according to any one of claims 8-10.
12. A vector comprising a nucleic acid molecule as recited in any one of claims 8-11.
13. A host cell transformed with a vector according to claim 12.
14. A ligand which binds specifically to, and which preferably inhibits the Nuclear  
Hormone Receptor Ligand Binding Domain activity of, a polypeptide according to  
25 any one of claims 1-7.
15. A ligand according to claim 14, which is an antibody.
16. A compound that either increases or decreases the level of expression or activity of a polypeptide according to any one of claims 1-7.

17. A compound according to claim 16 that binds to a polypeptide according to any one of claims 1-7 without inducing any of the biological effects of the polypeptide.
18. A compound according to claim 16 or claim 17, which is a natural or modified substrate, ligand, enzyme, receptor or structural or functional mimetic.
- 5 19. A polypeptide according to any one of claim 1-7, a nucleic acid molecule according to any one of claims 8-11, a vector according to claim 12, a ligand according to claim 14 or 15, or a compound according to any one of claims 16-18, for use in therapy or diagnosis of disease.
20. A method of diagnosing a disease in a patient, comprising assessing the level of  
10 expression of a natural gene encoding a polypeptide according to any one of claim 1-7, or assessing the activity of a polypeptide according to any one of claim 1-7, in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease.
21. A method according to claim 20 that is carried out *in vitro*.
- 15 22. A method according to claim 20 or claim 21, which comprises the steps of: (a) contacting a ligand according to claim 14 or claim 15 with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.
23. A method according to claim 20 or claim 21, comprising the steps of:
- 20 a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 8-11 and the probe;
- b) contacting a control sample with said probe under the same conditions used in step a);  
and
- 25 c) detecting the presence of hybrid complexes in said samples;  
wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.
24. A method according to claim 20 or claim 21, comprising:

a) contacting a sample of nucleic acid from tissue of the patient with a nucleic acid primer under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 8-11 and the primer;

5 b) contacting a control sample with said primer under the same conditions used in step a); and

c) amplifying the sampled nucleic acid; and

d) detecting the level of amplified nucleic acid from both patient and control samples;

10 wherein detection of levels of the amplified nucleic acid in the patient sample that differ significantly from levels of the amplified nucleic acid in the control sample is indicative of disease.

25. A method according to claim 20 or claim 21 comprising:

a) obtaining a tissue sample from a patient being tested for disease;

15 b) isolating a nucleic acid molecule according to any one of claims 8-11 from said tissue sample; and

c) diagnosing the patient for disease by detecting the presence of a mutation which is associated with disease in the nucleic acid molecule as an indication of the disease.

20 26. The method of claim 25, further comprising amplifying the nucleic acid molecule to form an amplified product and detecting the presence or absence of a mutation in the amplified product.

25 27. The method of either claim 25 or 26, wherein the presence or absence of the mutation in the patient is detected by contacting said nucleic acid molecule with a nucleic acid probe that hybridises to said nucleic acid molecule under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and

detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation.

28. A method according to any one of claims 20-27, wherein said disease is selected from cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposis' sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hyperparathyroidism, hypercalcemia, hypercholestrolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, dermatological disorders, including, acne, eczema, and wound healing, negative effects of aging, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated..
29. Use of a polypeptide according to any one of claims 1-7 as a Nuclear Hormone Receptor Ligand Binding Domain.
30. Use of a nucleic acid molecule according to any one of claims 8-11 to express a protein that possesses Nuclear Hormone Receptor Ligand Binding Domain activity.
31. A method for effecting cell-cell adhesion, utilising a polypeptide according to any one of claims 1-7.
32. A pharmaceutical composition comprising a polypeptide according to any one of claims 1-7, a nucleic acid molecule according to any one of claims 8-11, a vector according to claim 12, a ligand according to claim 14 or 15, or a compound according to any one of claims 16-18.
33. A vaccine composition comprising a polypeptide according to any one of claims 1-7 or a nucleic acid molecule according to any one of claims 8-11.

34. A polypeptide according to any one of claims 1-7, a nucleic acid molecule according to any one of claims 8-11, a vector according to claim 12, a ligand according to claim 14 or 15, a compound according to any one of claims 16-18, or a pharmaceutical composition according to claim 32 for use in the manufacture of a medicament for the treatment of cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposis' sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hyperparathyroidism, hypercalcemia, hypercholestrolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, dermatological disorders, including, acne, eczema, and wound healing, negative effects of aging, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated..
35. A method of treating a disease in a patient, comprising administering to the patient a polypeptide according to any one of claims 1-7, a nucleic acid molecule according to any one of claims 8-11, a vector according to claim 12, a ligand according to claim 14 or 15, a compound according to any one of claims 16-18, or a pharmaceutical composition according to claim 32.
36. A method according to claim 35, wherein, for diseases in which the expression of the natural gene or the activity of the polypeptide is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide,

nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an agonist.

37. A method according to claim 35, wherein, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an antagonist.
38. A method of monitoring the therapeutic treatment of disease in a patient, comprising monitoring over a period of time the level of expression or activity of a polypeptide according to any one of claims 1-7, or the level of expression of a nucleic acid molecule according to any one of claims 8-11 in tissue from said patient, wherein altering said level of expression or activity over the period of time towards a control level is indicative of regression of said disease.
39. A method for the identification of a compound that is effective in the treatment and/or diagnosis of disease, comprising contacting a polypeptide according to any one of claims 1-7, a nucleic acid molecule according to any one of claims 8-11, or a host cell according to claim 13 with one or more compounds suspected of possessing binding affinity for said polypeptide or nucleic acid molecule, and selecting a compound that binds specifically to said nucleic acid molecule or polypeptide.
40. A kit useful for diagnosing disease comprising a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to any one of claims 8-11; a second container containing primers useful for amplifying said nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease.
41. The kit of claim 40, further comprising a third container holding an agent for digesting unhybridised RNA.
42. A kit comprising an array of nucleic acid molecules, at least one of which is a nucleic acid molecule according to any one of claims 8-11.

43. A kit comprising one or more antibodies that bind to a polypeptide as recited in any one of claims 1-7; and a reagent useful for the detection of a binding reaction between said antibody and said polypeptide.
44. A transgenic or knockout non-human animal that has been transformed to express  
5 higher, lower or absent levels of a polypeptide according to any one of claims 1-7.
45. A method for screening for a compound effective to treat disease, by contacting a non-human transgenic animal according to claim 44 with a candidate compound and determining the effect of the compound on the disease of the animal.

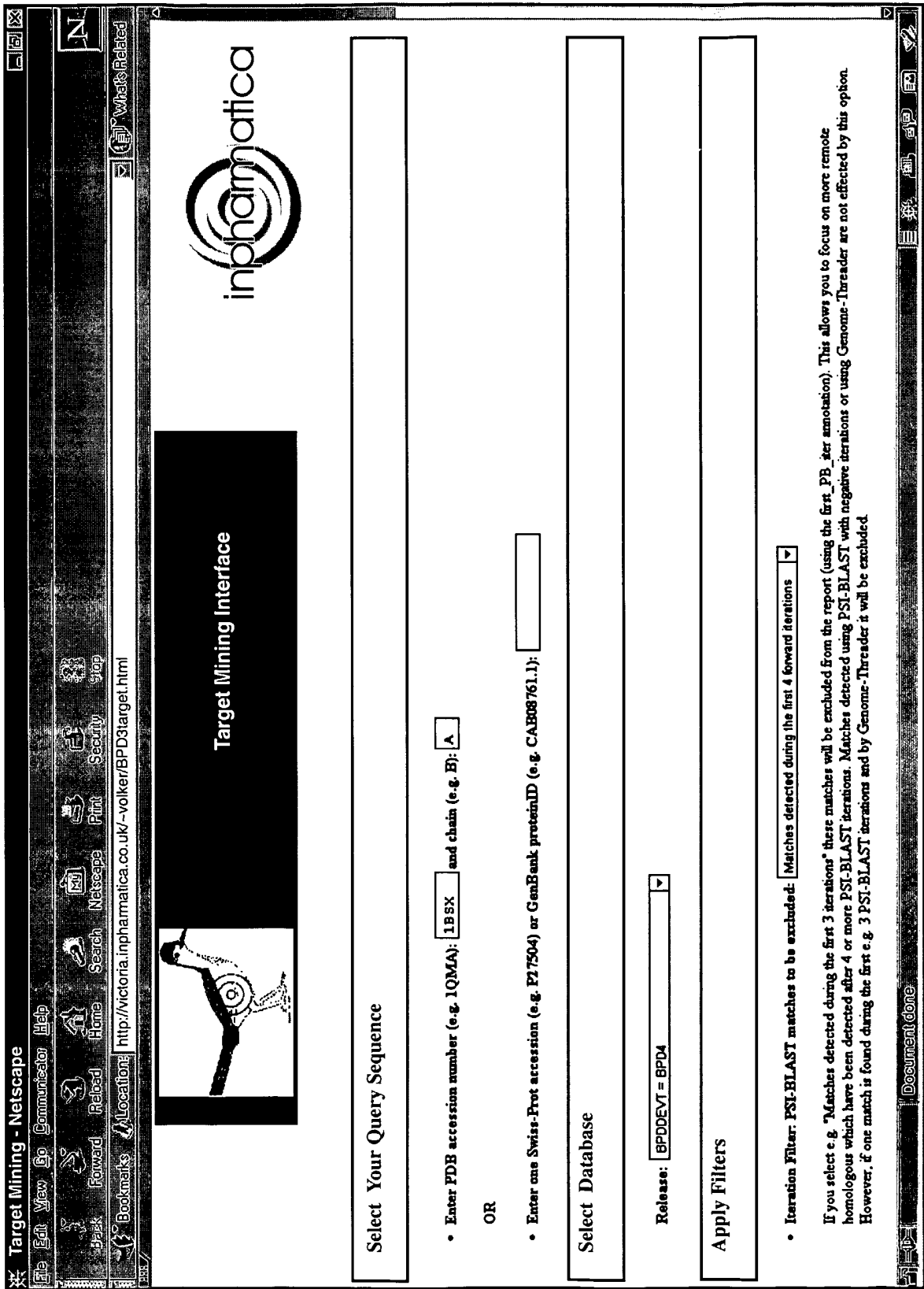


FIG. 1

Add/Link	BPD links	WWW link	Title	Organism	Div	%ID(G/PSI)	Query rgn (GT, PSI)	Target rgn (GT, PSI)	Align score (GT)	Conf (GT)	1st Iter (PSI)	Best Iter (PSI)	Best E-value (PSI)
<input type="checkbox"/>	<a href="#">E1082E</a> dell through Top-SIBlastHit Red-Seg View	<a href="#">E1082E</a>	THYROID HORMONE RECEPTOR BETA-1.	Homo sapiens (Human).	PRU	96.83% 96% unmaskedSV	1-243, 1-243	211-461, 211-461	570	100%	1	1	0
<input type="checkbox"/>	<a href="#">E17243</a> dell through Top-SIBlastHit Red-Seg View	<a href="#">E17243</a>	THYROID HORMONE RECEPTOR BETA-2.	Homo sapiens (Human).	PRU	96% 96% unmaskedSV	1-243, 1-243	226-476, 226-476	568	100%	1	1	0
<input type="checkbox"/>	<a href="#">E21203</a> dell through Top-SIBlastHit Red-Seg View	<a href="#">E21203</a>	THYROID HORMONE RECEPTOR ALPHA-1 (C-ERBA-ALPHA-1) (EAR-7-1) (EAR7).	Homo sapiens (Human).	PRU	83.3% 83% unmaskedSV	1-243, 1-243	157-407, 157-407	567	100%	1	1	1E-121
<input type="checkbox"/>	<a href="#">E19793</a> dell through Top-SIBlastHit Red-Seg View	<a href="#">E19793</a>	RETINOIC ACID RECEPTOR RXR-ALPHA.	Homo sapiens (Human).	PRU	21.6% 21% unmaskedSV	1-242, 1-242	201-456, 201-456	567	100%	1	2	2E-73
<input type="checkbox"/>	<a href="#">AAE001121</a> dell through Top-SIBlastHit Red-Seg View	<a href="#">AAE001121</a>	microphosamin-retinoic acid receptor alpha fusion protein NPM-RAR long form	Homo sapiens	PRU	24.9% 31% unmaskedSV	1-243, 1-243	273-517, 273-517	551	100%	1	2	6E-78
<input type="checkbox"/>	<a href="#">AAE196072</a> dell through Top-SIBlastHit Red-Seg View	<a href="#">AAE196072</a>	retinoic acid receptor alpha	Homo sapiens	PRU	24.9% 31% unmaskedSV	1-243, 1-243	125-370, 125-370	551	100%	1	2	6E-78
<input type="checkbox"/>	<a href="#">AAE001131</a> dell through Top-SIBlastHit Red-Seg View	<a href="#">AAE001131</a>	microphosamin-retinoic acid receptor alpha fusion protein NPM-RAR short form	Homo sapiens	PRU	24.9% 31% unmaskedSV	1-243, 1-243	229-474, 229-474	551	100%	1	2	6E-78
<input type="checkbox"/>	<a href="#">AAA601261</a> dell through Top-SIBlastHit Red-Seg View	<a href="#">AAA601261</a>	PML-RAR protein	Homo sapiens	PRU	24.9% 31% unmaskedSV	1-243, 1-243	506-751, 506-751	551	100%	1	2	6E-78
<input type="checkbox"/>	<a href="#">AAD052211</a> dell through Top-SIBlastHit Red-Seg View	<a href="#">AAD052211</a>	retinoic acid receptor alpha	Homo sapiens	PRU	31.2% 36% unmaskedSV	1-243, 1-243	171-416, 171-416	550	100%	1	2	1E-77
<input type="checkbox"/>	<a href="#">E12631</a> dell through Top-SIBlastHit Red-Seg View	<a href="#">E12631</a>	RETINOIC ACID RECEPTOR GAMMA-1 (RAR-GAMMA-1).	Homo sapiens (Human).	PRU	34% 34% unmaskedSV	3-243, 3-243	181-418, 181-418	549	100%	1	2	2E-77
<input type="checkbox"/>	<a href="#">E22922</a> dell through Top-SIBlastHit Red-Seg View	<a href="#">E22922</a>	RETINOIC ACID RECEPTOR GAMMA-2 (RAR-GAMMA-2).	Homo sapiens (Human).	PRU	24% 34% unmaskedSV	3-243, 3-243	170-407, 170-407	549	100%	1	2	2E-77
<input type="checkbox"/>	<a href="#">CAC013021</a> dell through Top-SIBlastHit Red-Seg View	<a href="#">CAC013021</a>	d1013A22.1 (hepatic nuclear factor 4, alpha)	Homo sapiens	PRU	21.6% 21% unmaskedSV	1-241, 1-241	126-374, 126-374	520	100%	1	3	7E-55



FIG. 2A

Target Mining Results - Netscape  
 Life Edit View Go Communicator Help  
 Back Forward Reload Home Search Netscape Print Security Stop  
 Locations: <http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl> What's Related

<input type="checkbox"/>	Red Seq View	CAB4127.1 deft through Top 50BlastHits	CAB4127.1	hypothetical protein	Homo sapiens	PRI	11.2%	16-238	204-399	62	24.31%
<input type="checkbox"/>	Red Seq View	AA04662.1 deft through Top 50BlastHits	AA04662.1	U59	Human herpesvirus 6B	VRL	8.1%	83-231	199-332	75	23.82%
<input type="checkbox"/>	Red Seq View	AA07821.1 deft through Top 50BlastHits	AA07821.1	alpha-crabbin	Homo sapiens	PRI	11.9%	81-222	387-715	72	23.87%
<input type="checkbox"/>	Red Seq View	BA03283.1 deft through Top 50BlastHits	BA03283.1	KIAA1033 protein	Homo sapiens	PRI	11.7%	9-243	415-646	63	23.82%
<input type="checkbox"/>	Red Seq View	BA091816.1 deft through Top 50BlastHits	BA091816.1	Not given	Homo sapiens	PRI	12.8%	9-243	392-673	63	22.76%
<input type="checkbox"/>	Red Seq View	CAB4124.1 deft through Top 50BlastHits	CAB4124.1	hypothetical protein	Homo sapiens	PRI	10.1%	4-241	191-398	57	22.26%
<input type="checkbox"/>	Red Seq View	CA03350.1 deft through Top 50BlastHits	CA03350.1	Not given	Human rotavirus	VRL	11.7%	86-230	130-279	70	20.57%
<input type="checkbox"/>	Red Seq View	BA069479.1 deft through Top 50BlastHits	BA069479.1	Not given	Homo sapiens	PRI	11.2%	20-242	54-287	61	28.17%
<input type="checkbox"/>	Red Seq View	AA025679.1 deft through Top 50BlastHits	AA025679.1	cyclic AMP-specific phosphodiesterase HSPDEA1A	Homo sapiens	PRI	9.7%	1-277	76-321	55	27.73%
<input type="checkbox"/>	Red Seq View	BA022563.1 deft through Top 50BlastHits	BA022563.1	PRP8 protein	Homo sapiens	PRI	12.2%	6-209	1104-1309	85	26.63%
<input type="checkbox"/>	Red Seq View	AA030438.1 deft through Top 50BlastHits	AA030438.1	3',5'-cyclic AMP phosphodiesterase HPDEA6	Homo sapiens	PRI	9.7%	1-277	115-360	55	26.63%
<input type="checkbox"/>	Red Seq View	AA03178.1 deft through Top 50BlastHits	AA03178.1	F-box protein FBL2	Homo sapiens	PRI	1.6%	54-221	194-367	45	26.63%
<input type="checkbox"/>	Red Seq View	AA032011.1 deft through Top 50BlastHits	AA032011.1	putative novel cAMP specific phosphodiesterase 4A variant	Homo sapiens	PRI	9.7%	1-277	288-534	55	23.57%
<input type="checkbox"/>	Red Seq View	AA032012.1 deft through Top 50BlastHits	AA032012.1	chromodomain-helicase-DNA-binding protein	Homo sapiens	PRI	13.1%	82-242	33-177	58	24.43%
<input type="checkbox"/>	Red Seq View	AA032013.1 deft through Top 50BlastHits	AA032013.1	cAMP specific phosphodiesterase 4A variant pd-46	Homo sapiens	PRI	9.7%	1-277	315-560	55	24.43%
<input type="checkbox"/>	Red Seq View	BA091845.1 deft through Top 50BlastHits	BA091845.1	Not given	Homo sapiens	PRI	14.6%	81-229	40-193	66	23.26%
<input type="checkbox"/>	Red Seq View	EA0784 deft through Top 50BlastHits	EA0784	CAPSID ASSEMBLY PROTEIN UL47	Human cytomegalovirus (strain AD169)	VRL	11.2%	54-222	200-387	52	23.26%

Document Done

FIG. 2B



FIG. 3

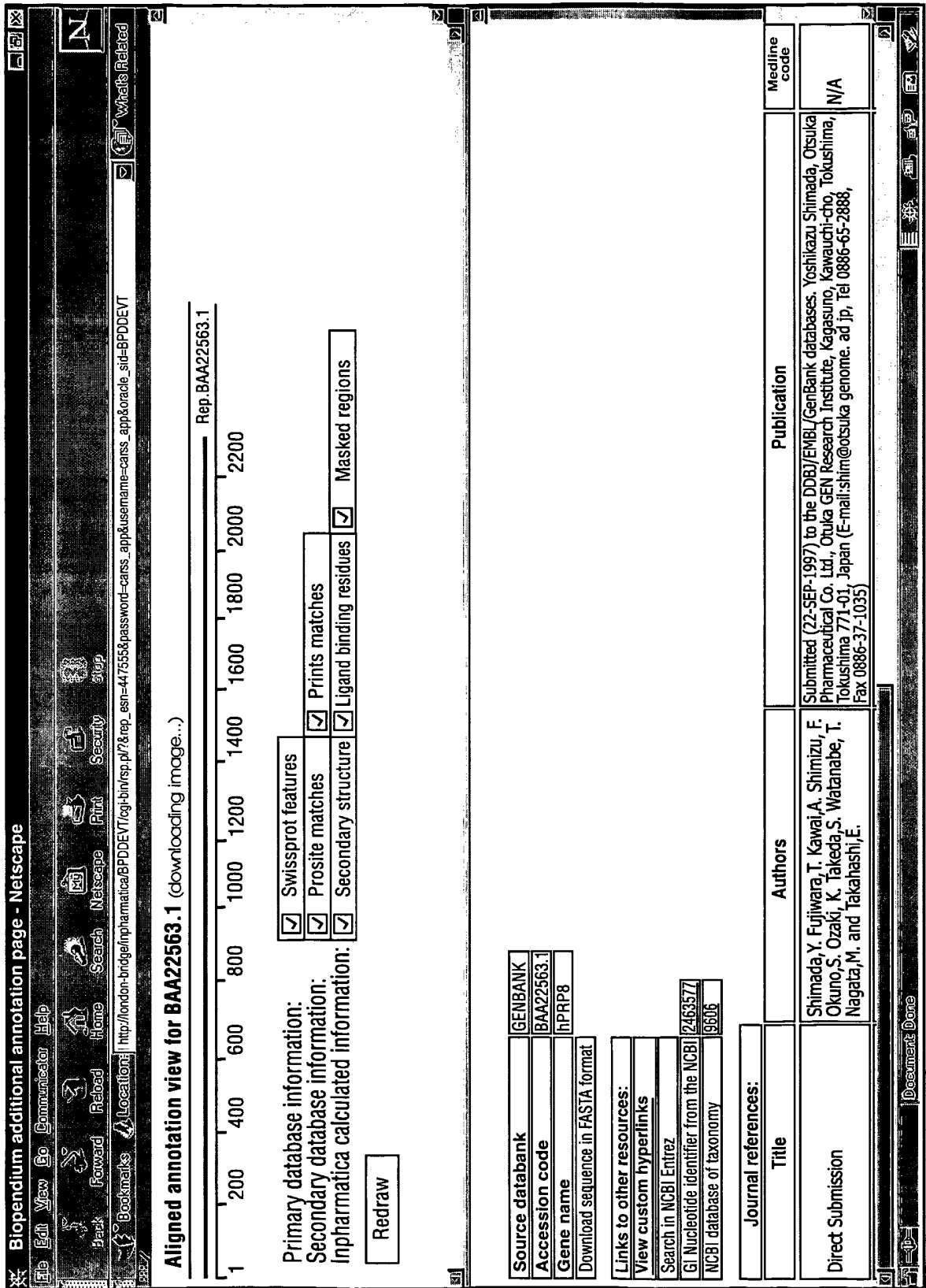


FIG. 4

InterPro search Results.

1 Query Sequence submitted Length 2335aa.

InterPro	Results of PPsearch against PROSITE	Results of PFScan against PROSITE	Results of FingerPRINTScan against PRINTS	Results of HMMDecypher against PFAM-A
IPR000148 Papillomavirus E7 protein				PF00527 [1561-1570]
IPR0001472 Bipartite nuclear localization signal		PS50079 [35-52]		

IPR000148 PF00527  
IPR000172 PF00579

XML / TXT formatted

Applications furnished normally.

FIG. 5

NCBI Sequence Viewer - Netscape

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=Protein&list\_uids=2463577&dopt=GenPept

BLINK, Related Sequences, Nucleotide, Taxonomy, LinkOut

**1. BAA22563** **PRP8 protein [Homo sapiens]** **PRI** **03-OCT-1997**

LOCUS BAA22563 2335 aa  
 DEFINITION PRP8 Protein [Homo sapiens].  
 ACCESSION BAA22563  
 PII 92463577  
 VERSION BAA22563.1 GI:2463577  
 DBSOURCE Locust AB007510 accession AB007510.1  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (residues 1 to 2335)  
 AUTHORS Shimada, Y., Fujiwara, T., Kawai, A., Shimizu, F., Okuno, S., Ozaki, K.,  
 Takeda, S., Watanabe, T., Nagata, M. and Takahashi, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-SEP-1997) Yoshikazu Shimada, Otsuka Pharmaceutical  
 Co. Ltd., Otsuka GEN Research Institute; Kagasuno, Kawauchi-cho,  
 Tokushima, Tokushima 771-01, Japan  
 (E-mail: shim@otsuka.genome.ad.jp, Tel:0886-65-2888,  
 Fax:0886-37-1035)  
 REFERENCE 2 (residues 1 to 2335)  
 AUTHORS Shimada, Y., Fujiwara, T., Kawai, A., Shimizu, F., Okuno, S., Ozaki, K.,  
 Takeda, S., Watanabe, T., Nagata, M. and Takahashi, E.  
 TITLE Human homologue of Saccharomyces cerevisiae PRP8, Pre-mRNA splicing  
 factor  
 JOURNAL Published Only in Database (1997) In press  
 FEATURES  
 source  
 location/qualifiers  
 1..2335  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17p13.3"  
 /tissue\_lib="fetal brain"  
 Protein 1..2335  
 /product="PRP8 protein"  
 CDS 1..2335  
 /gene="hPRP8"  
 /coded\_by="AB007510.1:42..7049"  
 ORIGIN  
 1 magvfpytgp gnpvpglep lpdymseekl cekarkwqql cekryaekrk fgfvdaqked  
 61 mpehvreli rdhgdtnrk fthdkrvylg alkymphavl kllenmpmpw eqirdvpvly  
 121 hitgaistfvn eipwievpy isqwgsmwim mrekdrrh fkrmrtpffd deppldyad  
 181 nilnveplea iqleldpeed apvlwfydh qplrdskryv ngstyqrwqf tlpmmatlyr  
 241 lanqltclv dnnfyfled kafftskain maipgpkfe plvrainlqd edwnefndin  
 301 kiirqpirt eykiafply mlphhvhtl wyhtpvnvfi ktedpdpaf yfdplinpis  
 361 hrhsvksqep lpdadeetel pefveprfkd tplytdantan giailwairp fnlrsgrtrr

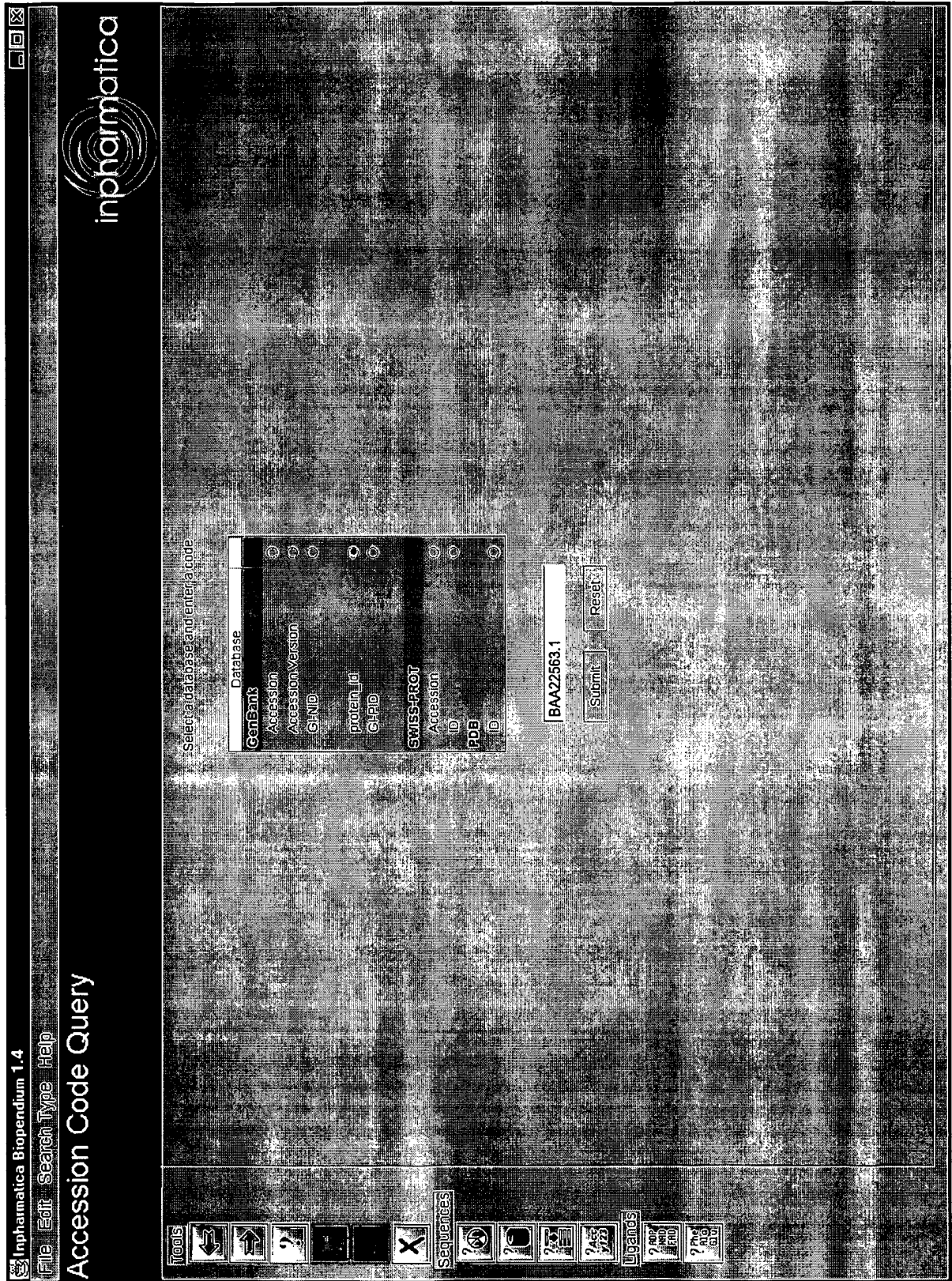


FIG. 6A

FIG. 6B

Inpharmatica Biopendium 1.4  
 File Edit Search Type Help  
**Aligned Sequence Display**  
 Query details:  
 BAA22563.1 PRP8 protein

inpharmatica

Total selected: 0. [View alignment](#)

Inpharmatica Genome Threader results

Page 1 of 2    Total hits: 53    Selected: 0

[Select all](#)    [Deselect all](#)

Cluster Detail	Code	Title	Organism	%ID	Query Len.	Target Len.	Aln. score	Method	Confidence
1007.A		STRUCTURE OF ASPERGILLUS NIGER	<i>A. niger</i>	7.3	872 - 1,158	27 - 324	62	Local	Low(91%)
1007.B		STRUCTURE OF ASPERGILLUS NIGER	<i>A. niger</i>	7.3	872 - 1,158	27 - 324	62	Local	Low(91%)
185X.A		STRUCTURE AND SPECIFICITY OF NU	<i>H. sapiens</i>	12.5	1,104 - 1,309	6 - 209	86	Local	Low(86%)
185X.B		STRUCTURE AND SPECIFICITY OF NU	<i>H. sapiens</i>	12.5	1,104 - 1,309	6 - 209	86	Local	Low(86%)
1A40		PIG ALDOSE REDUCTASE COMPLEXE	<i>S. scrofa</i>	16.2	465 - 705	57 - 296	74	Local	Low(80%)
1A45		ALDOSE REDUCTASE COMPLEXED W.	<i>S. scrofa</i>	16.2	465 - 705	56 - 295	69	Local	Low(64%)
1EK0A		PIG ALDOSE REDUCTASE COMPLEXE	<i>S. scrofa</i>	16.2	465 - 705	56 - 295	69	Local	Low(64%)
1A44		PIG ALDOSE REDUCTASE HOLOFORM	<i>S. scrofa</i>	16.2	465 - 705	56 - 295	69	Local	Low(64%)
1A0S		ALDOSE REDUCTASE (E.C.1.1.1.21) C	<i>H. sapiens</i>	15.2	477 - 705	68 - 295	69	Local	Low(64%)
1MAR		ALDOSE REDUCTASE (E.C.1.1.1.21) C	<i>H. sapiens</i>	15.2	477 - 705	68 - 295	69	Local	Low(64%)
2ACO		ALDOSE REDUCTASE (E.C.1.1.1.21) W	<i>H. sapiens</i>	15.2	477 - 705	68 - 295	69	Local	Low(64%)
2ACR		ALDOSE REDUCTASE (E.C.1.1.1.21) W	<i>H. sapiens</i>	15.2	477 - 705	68 - 295	69	Local	Low(64%)
1EL3A		HUMAN ALDOSE REDUCTASE COMPL	<i>H. sapiens</i>	15.2	477 - 705	69 - 296	69	Local	Low(63%)
2ACS		ALDOSE REDUCTASE (E.C.1.1.1.21) W	<i>H. sapiens</i>	15.2	477 - 705	67 - 284	69	Local	Low(59%)
1LW1A		3-ALPHA-HYDROXYSTEROID/DIHYDRO	<i>R. norvegicus</i>	12.9	510 - 705	97 - 297	62	Local	Low(44%)
2UCZ		UBIQUITIN CONJUGATING ENZYME [UB	<i>S. cerevisiae</i>	19.4	746 - 1,693	7 - 123	61	Local	Low(34%)
1AF5A		RECOMBINANT RAT LIVER 3-ALPHA-H	<i>R. norvegicus</i>	13.8	510 - 705	100 - 303	58	Local	Low(33%)
1AF5B		RECOMBINANT RAT LIVER 3-ALPHA-H	<i>R. norvegicus</i>	13.8	510 - 705	100 - 303	58	Local	Low(33%)
1REV.A		HIV-1 REVERSE TRANSCRIPTASE	<i>H. immunodeficiency</i>	14.0	1,420 - 1,742	204 - 519	57	Local	Low(29%)
1C91B		PEPTIDE-IN-GROOVE INTERACTIONS	<i>R. norvegicus</i>	20.2	1,721 - 1,888	55 - 207	75	Local	Low(25%)
1C91A		PEPTIDE-IN-GROOVE INTERACTIONS	<i>R. norvegicus</i>	20.2	1,721 - 1,888	56 - 208	75	Local	Low(25%)
1C91B		PEPTIDE-IN-GROOVE INTERACTIONS	<i>R. norvegicus</i>	20.2	1,721 - 1,888	56 - 208	75	Local	Low(25%)
1C91A		PEPTIDE-IN-GROOVE INTERACTIONS	<i>R. norvegicus</i>	20.2	1,721 - 1,888	56 - 208	75	Local	Low(25%)
2CLJ		HOMOLOGY-BUILT MODEL OF HUMAN	N/A	13.9	1,692 - 1,955	96 - 376	47	Local	Low(24%)
1UIP		ADENOSINE DEAMINASE (HIS 238 GLU	<i>M. musculus</i>	15.8	793 - 876	82 - 171	65	Local	Low(22%)
1AZ1		ALRESTATIN BOUND TO C298A/W219	<i>H. sapiens</i>	15.5	474 - 705	64 - 294	60	Local	Low(21%)
1AZ2		CITRATE BOUND, C298A/W219 MUTA	<i>H. sapiens</i>	15.5	474 - 705	64 - 294	60	Local	Low(21%)
1EX.A		ENANTIOMER DISCRIMINATION ILLUST	<i>H. sapiens</i>	8.3	1,105 - 1,299	4 - 195	73	Local	Low(21%)
1EX.A		ENANTIOMER DISCRIMINATION ILLUST	<i>H. sapiens</i>	8.3	1,105 - 1,299	4 - 195	73	Local	Low(21%)

GenBank SWISS-PROT PDB

FIG. 6C

Inpharmatica Biopendium 1.4  
 File Edit Search Type Help  
**Aligned Sequence Display**  
 Query details:  
 BAA22563.1 PRP8 protein

inpharmatica

Reverse Maximised PSI-BLAST Results

Total selected: 0  
 Total hits: 32  
 Selected: 0

View alignment  
 Select all  
 Deselect all

Cluster/Detail	Code	Title	Organism	%ID	Query.com.	Target.com.	Filter	E-value
o	AAC61776.1	splicing factor Prp8	<i>H. sapiens</i>	99.0	1 - 2,335	1 - 2,335	1	0.00E+00
o	AAE58573.1	Not given	<i>D. melanogaster</i>	90.0	8 - 2,335	70 - 2,996	1	0.00E+00
o	P34369	HYPOTHETICAL 272.0 KDA PROTEIN C50C3.6 IN	<i>C. elegans</i>	86.0	11 - 2,335	6 - 2,329	1	0.00E+00
o	BAA18744.1	Not given	<i>O. sativa</i>	83.0	20 - 2,334	43 - 2,360	1	0.00E+00
o	AAD55467.1	Not given	<i>A. thaliana</i>	80.0	10 - 2,334	36 - 2,358	1	0.00E+00
o	CAB80541.1	splicing factor-like protein	<i>A. thaliana</i>	78.0	7 - 2,335	31 - 2,352	1	0.00E+00
o	CAB11062.1	probable pre-mRNA splicing factor	<i>S. pombe</i>	73.0	27 - 2,335	48 - 2,363	1	0.00E+00
o	AA457042.1	Prp8p RNA splicing factor	<i>S. cerevisiae</i>	61.0	33 - 2,335	110 - 2,413	1	0.00E+00
o	AAB68011.1	pre-mRNA processing 8 protein homolog PRPB	<i>S. cerevisiae</i>	61.0	33 - 2,335	110 - 2,413	1	0.00E+00
o	AA029088.1	pre-mRNA processing 8 protein homolog PRPB	<i>S. cerevisiae</i>	55.0	30 - 2,329	19 - 2,314	1	0.00E+00
o	CAA13186.1	PRP8 protein homolog	<i>T. brucei</i>	40.0	36 - 2,317	37 - 2,303	1	0.00E+00
o	CAC00558.1	similarity to human amsh protein	<i>S. pombe</i>	21.0	250 - 369	2,068 - 2,234	2	1.00E-04
o	P40085	HYPOTHETICAL 64.8 KDA PROTEIN IN GD1-COX	<i>S. cerevisiae</i>	29.0	466 - 526	1,959 - 2,020	3	1.00E-12
o	BAA92611.1	KIAA1373 protein	<i>H. sapiens</i>	17.0	187 - 461	2,022 - 2,303	4	9.00E-82
o	AAF57051.1	Not given	<i>D. melanogaster</i>	20.0	251 - 417	2,105 - 2,274	4	8.00E-44
o	CA235570.1	Not given	<i>K. lactis</i>	17.0	225 - 392	176 - 341	4	6.00E-40
o	AAB82533.1	unknown	<i>D. discoideum</i>	25.0	309 - 409	2,144 - 2,248	4	6.00E-11
o	AAB87991.1	Not given	<i>C. elegans</i>	20.0	62 - 215	2,110 - 2,255	4	3.00E-04
o	AAD05037.1	AMSH	<i>H. sapiens</i>	17.0	214 - 414	2,067 - 2,265	5	3.00E-54
o	P411883	HYPOTHETICAL 35.8 KDA PROTEIN F37A4.5 IN C.	<i>C. elegans</i>	16.0	7 - 207	2,062 - 2,274	5	5.00E-44
o	BA488335.1	Not given	<i>O. sativa</i>	18.0	24 - 178	2,099 - 2,247	5	3.00E-33
o	AAC51866.1	26S proteasome-associated pad1 homolog	<i>H. sapiens</i>	15.0	27 - 204	2,100 - 2,274	5	2.00E-32
o	CA473514.1	26S proteasome non-ATPase subunit	<i>M. musculus</i>	15.0	26 - 703	2,100 - 2,274	5	2.00E-32
o	CAA56098.1	Not given	<i>S. cerevisiae</i>	18.0	23 - 163	2,100 - 2,235	5	1.00E-28
o	AAD46836.1	BcDNA GM14618	<i>D. melanogaster</i>	14.0	2 - 140	2,109 - 2,249	5	6.00E-13
o	AAB67823.1	sk31 multidrug resistance protein homolog	<i>D. discoideum</i>	13.0	20 - 201	2,093 - 2,274	6	3.00E-37
o	AAF08394.1	26S proteasome regulatory complex subunit p37B	<i>D. melanogaster</i>	15.0	25 - 199	2,100 - 2,257	6	2.00E-31
o	AAC33765.1	jab1 protein	<i>O. sativa</i>	16.0	61 - 217	2,105 - 2,253	6	2.00E-29
o	BAA20419.1	reverse transcriptase	<i>M. musculus</i>	11.0	86 - 1,107	1,290 - 1,524	6	2.00E-04
o	BA397246.1	Not given	<i>A. thaliana</i>	17.0	3 - 166	2,074 - 2,235	7	2.00E-37

GenBank SWISS-PROT PDB

FIG. 7-1

AIEye output (January 18, 2001 5:45 PM)

BAA22563.1    10    20    30    40    50    60    70  
1BSX:A    . . . . . MAGVFPYRGGNPPGPLAPLPDYMSSEKLEKARKWQQLQAKRYAEKRKFGFVDAQKEDMPPEHVREIIRDHGDMTNR  
. . . . . 80    90    100    110    120    130    140    150  
BAA22563.1    KFRHDKRVLGALKYMPHAVLKLEENMPWEQIRDVPVLYHITGAI SFVNEIPWVIEPVYISQWGSMWIMMRREKRDR  
1BSX:A    . . . . . 160    170    180    190    200    210    220    230  
BAA22563.1    RHFKRMRFPFDEEPPLDYADNILNVEPLEAIQLELDPEEDAPVLDWFYDQPLRDSRKYVNGSTYQRWQFTLPMMST  
1BSX:A    . . . . . 240    250    260    270    280    290    300    310  
BAA22563.1    LYRLANQLTDLVDDNYFYLFDLKAFFTSKALNMAIPGGPKFEPLVRDINLQDEDWNEFNDINKIIRQPIRTEYKIAF  
1BSX:A    . . . . . 320    330    340    350    360    370    380    390  
BAA22563.1    PLYNLPHHVHLT WYHTPNVVFIKTEDPDLPAFYFDPLINPISHRHSVKSQEPLPDDDEEFELPEFVEPFLKDTPLYT  
1BSX:A    . . . . . 400    410    420    430    440    450    460    470  
BAA22563.1    DNTANGIALLWAPRPFNLRSGRTRRALDIPLVKNWYREHCPAGQPVKVRVSYQKLLKYYVLNALKHRPPKAKKRYLFR  
1BSX:A    . . . . . 480    490    500    510    520    530    540    550  
BAA22563.1    SFKATKFFQSTKLDWVEGWLQVCRQGYNMLNLLIHRKNLNYLHLDYNFNLPVKTLLTKERKSRFGNAFHLCREVLR  
1BSX:A    . . . . . 560    570    580    590    600    610    620    630  
BAA22563.1    TKLVVD SHVQYRLGNVDAFQLADGLQYIFAHVQGLTGMRYKYKLMRQIRVCKDLKHLIYRFRNTGTPVGKGGCGFWAA  
1BSX:A    . . . . . 640    650    660    670    680    690    700    710  
BAA22563.1    GWRVWLFFMRGITPLLERWLNLLARQFEGRHSKGVAKTVTQKRVESHFDLELRAAVMHDILDMMPGKIKONKARTILQ  
1BSX:A    . . . . .

AI Eye output (January 18, 2001 5:45 PM)

FIG. 7-2

BAA22563.1 H L S E A W R C W K A N I P W K V P G L P T P I E N M I L R Y V K A K A D W W T N T A H Y N R E R I R R G A T V D K T V C K K N L G R L T R L Y L K A E Q E R 720 730 740 750 760 770 780 790  
1BSX:A

BAA22563.1 Q H N Y L K D G P Y I T A E E T V A V Y T T T V H W L E S R R F S P I P F P P L S Y K H D T K L L I A L E R L K E A Y S V K S R L N Q S R E E L G L I E Q 800 810 820 830 840 850 860  
1BSX:A

BAA22563.1 A Y D N L H E A L S R I K R H L L T Q R A F K E V G I E F M D L Y S H L V P V Y D V E P L E K I T D A Y L D Q Q Y L W Y E A D K R R L F P P W I K P A D T E P P 870 880 890 900 910 920 930 940  
1BSX:A

BAA22563.1 P L L V Y K W C Q G I N N L Q D V W E T S E G E C N V M L E S R F E K M Y E K I D L T L L N R L V R L I V D H N I A D Y M T A K N N V V I N Y K D M N H T N S 950 960 970 980 990 1000 1010 1020  
1BSX:A

BAA22563.1 Y G I R G L Q F A S F I V Q Y Y G L V M D L L V L G L H R A S E M A G P P Q M P N D F L S F Q D I A T E A A H P I R L F C R Y I D R I H I F F R F T A D E A 1030 1040 1050 1060 1070 1080 1090 1100  
1BSX:A

BAA22563.1 R D L I Q R Y L T E H P D P N N E N I V G Y N N K K C W P R D A R M R L M K H D V N L G R A V F W D I K N R L P R S V - T T V Q W - - - E N S F V S V Y S K D 1110 1120 1130 1140 1150 1160 1170 1180 1181  
1BSX:A W E L I K T V T E A H V A T N A Q G S H W K Q R K F L P E D I G - - - - Q A P K V D L E A F S H F T K I P A I T R V V D F A K K L P M F C E L P C E D 1174 1177 283 296 300

BAA22563.1 N P N L F N M C G F E C R I L P K C R T S Y E E F T H K D G V W N Q N - - - E V T K E R T A Q C F R V D D E S M Q R F H N R V R Q I L M A S G S T T F T K 1182 1185 1186 1200 1210 1220 1230 1240 1250  
1BSX:A Q I I L L K G C C M E I M S L R A A V R Y D P E - - - - S E T L T N G E M A V T R G Q L K N G G G V V S D A I F D L G M S L S S F N L D D T E V A L L Q 1216 1217 330 331 346 1320 1330 1340  
301 304 1270 1280 1290 1300 1310 1320 1330 1340

BAA22563.1 I V N K W N T A L I G L - - - - M T Y F R E A V N T Q E L L D L L V K C E H K I Q T R I K I G L N S K M P S R F P P V V F Y T P K E L G G L G M - - - - 1350 1360 1370 1380 1390 1400 1410 1420  
1BSX:A A V L L M S S D R P G L A C V E R I E K Y Q D S F - - L L A F E H Y I N Y R K H H V T H F W P - K L L M K V T D L - - - - - R M I G A C H A S R F L H 1400 435 1420

BAA22563.1 - - - L S M G H V L I P Q - - - - - S D L R W S K Q T D V G I T H F R S G M S H E E D Q L I P N L Y R Y I Q P W E S E F I D S Q R V W A E Y S L K R Q E A I A 1350 1360 1370 1380 1390 1400 1410 1420  
1BSX:A M K V E C P T E L F P P L F L E V F E D

FIG. 7-3

AIEye output (January 18, 2001 5:45 PM)

BAA22563.1 1430 1440 1450 1460 1470 1480 1490 1500  
 1BSX:A QNRRLTLEDLEDSDRGIPRINTLFQKDRHTLAYDKGWRVRTDFKQYQVLKQNPFFWWTHORHDGKLVNLLNNYRTDMIQA

BAA22563.1 1510 1520 1530 1540 1550 1560 1570 1580  
 1BSX:A LGGVEGILEHTLFFKGTYPFTWEGLFWEKASGFEESMKWKKLTNAQRSGLNQIPNRRFTLWWSPTINRANVYVGFQVQLD

BAA22563.1 1590 1600 1610 1620 1630 1640 1650  
 1BSX:A LTGIFMHGKIPTLKI SLIQIFRAHLWQKIHESI VMDLCQVFDQELD ALEIETVQKETIHPRKSYKMNSSCADILLFASY

BAA22563.1 1660 1670 1680 1690 1700 1710 1720 1730  
 1BSX:A KWNVSRPSSLADSKDVMDSTTTQKYWIDIQLRWGDYDSDHIERYARAKFLDYTTDNMSIYPSPTGVLI AIDLAYNLHSA

BAA22563.1 1740 1750 1760 1770 1780 1790 1800 1810  
 1BSX:A YGNWFFPGSKPLIQAMAKIMKANPALYVLRERIRKGLQLYSSEPT EPYLSSQNYGELFSNQI IWFVDDTNVYRVTIHKKT

BAA22563.1 1820 1830 1840 1850 1860 1870 1880 1890  
 1BSX:A FEGNLTTPINGAIFIFNPRGTGQLFLKIIHTSVWAGQKRLGQLAKWKTAEEVAALIRSLPVEEQPKQIIVTRKMDLDP

BAA22563.1 1900 1910 1920 1930 1940 1950 1960 1970  
 1BSX:A EVHLDFPNIVIKGSELQLPFQACLKVEKFGDLILKATEPQMVLFNLYDDWLKTISSYTAFSRLILLRALHVNNDRAK

BAA22563.1 1980 1990 2000 2010 2020 2030 2040 2050  
 1BSX:A VILKPKDKTITTEPHHIWPTLTDEEWIKVEVQLKDLILADYGKKNVNVASLTQSEIRDII LGMEISAPSQRQQAIEIE

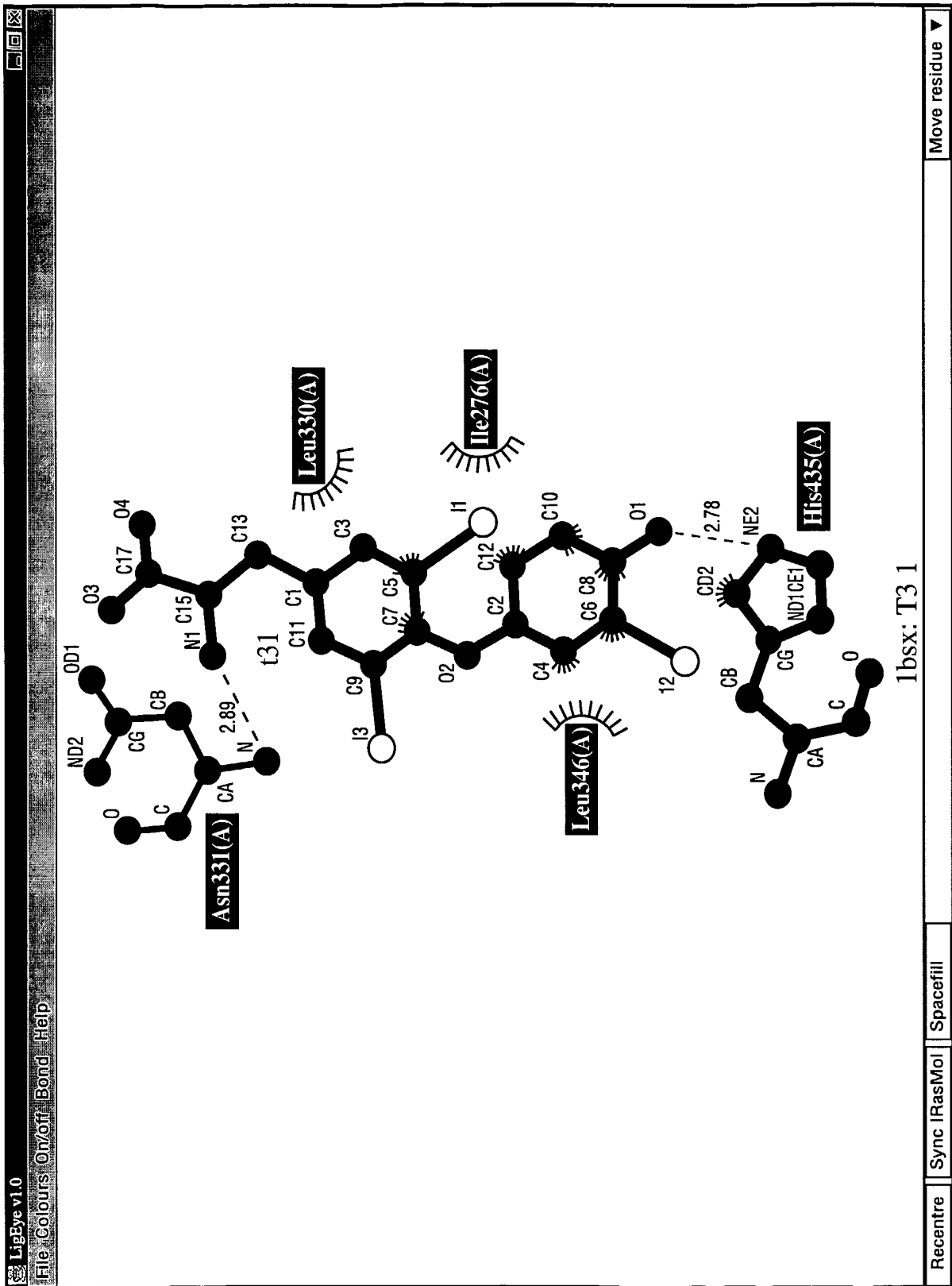
BAA22563.1 2060 2070 2080 2090 2100 2110 2120 2130  
 1BSX:A KQTKEQSQLTATQTRTVNKHGDEIITSTTSNYETQTFSSKTEWRVRAISAANLHLRTNHIIYSSDDIKETGYTYILPKN

# FIG. 7-4

AIEye output (January 18, 2001 5:45 PM)

BAA22563.1	2140	2150	2160	2170	2180	2190	2200	2210
1BSX:A	V L K K F I C I S D L R A Q I A G Y L Y G V S P P D N P Q V K E I R C I V M V P Q W G T H Q T V H L P G Q L P Q H E Y L K E M E P L G W I H T O P N E S P Q L							
BAA22563.1	2220	2230	2240	2250	2260	2270	2280	2290
1BSX:A	S P Q D V T T H A K I M A D N P S W D G E K T I I I T C S F T P G S C T L T A Y K L T P S G Y E W G R Q N T D K G N N P K G Y L P S H Y E R V Q M L L S D R F							
BAA22563.1	2300	2310	2320	2330	2340	2350		
1BSX:A	L G F F M V P A Q S S W N Y N F M G V R H D P N M K Y E L Q L A N P K E F Y H E V H R P S H F L N F A L L Q E G E V Y S A D R E D L Y A							

FIG. 8A



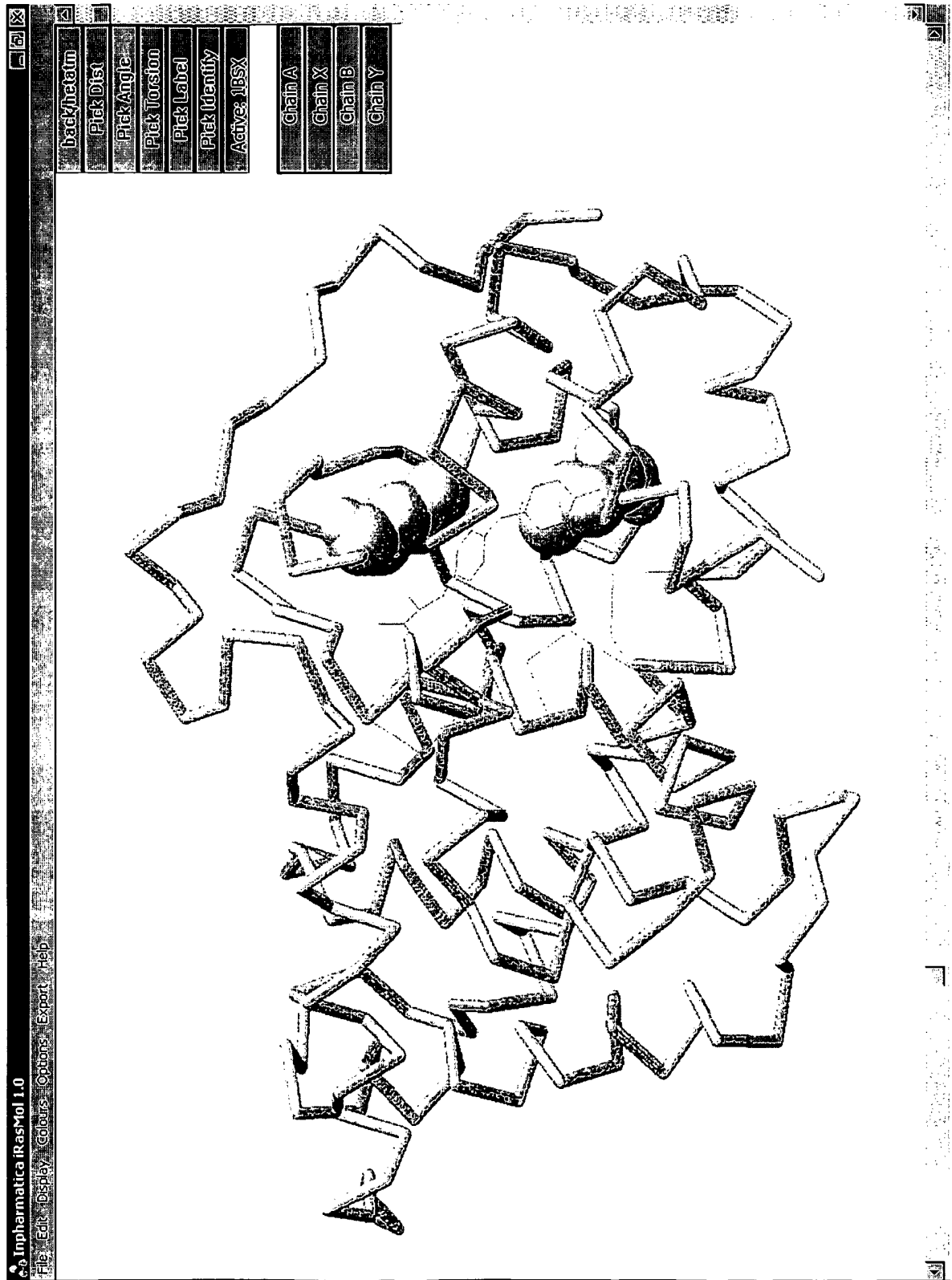


FIG. 8B

FIG. 9-1 AIEye output (February 26, 2001 9:53 AM)

```

BAA22563.1  . . . . . 10 20 30 40 50 60 70
AAF58573.1  M S I P P Y M I P O N A W A A Q L M A Q Q A Y A A A H A Q Q A Q L H A Q Q M A N Q I Q Q I P P P G A P L P P A G G H T N G I P I P V G G . . . . . M A G V F P Y R G P G . N
P34369      . . . . . 20 30 40 50 60 70
BAA78744.1  . . . . . 20 30 40 50 60 70
CAB80541.1  . . . . . 20 30 40 50 60 70

BAA22563.1  . . . . . 90 100 110 120 130 140 150
AAF58573.1  P V P G P L A . . . . . P L P D Y M S E E K L Q E K A R K W Q Q L Q A K R Y A E K R K F G F V D A Q K E D M P P E H V R E I I R D H G D M T N R K
P34369      Q I P . . . . . T . . . . . P K P D I L T E E K L Q E K A L K W H L Q S K R F A E K R K F G F V D T Q K E D M P P E H I R K I I R D H G D M T S R K
BAA78744.1  P Q T E P H A . . . . . I P D S I . . . . . L E E K S R K W K Q L Q G K R Y S E K K K F G M S D T Q K E E M P P E H V R K V I R D H G D M T S R K
CAB80541.1  P P P P A A V G P P G G V G G K P L T A A E L E A Q L V E K A R K W H Q L N S K R Y G D K R K F G F V E A Q K E D M P P E H V R K I I R D H G D M S S K K
N T P . . . . . T V P T P E D . . . . . A E A K L E K K A R T W M Q L N S K R Y G D K R K F G F V E T Q K E D M P P E H V R K I I R . . . . . R K

BAA22563.1  . . . . . 170 180 190 200 210 220 230
AAF58573.1  F R H D K R V Y L G A L K Y M P H A V L K L L E N M P M P W E Q I R D V P V L Y H I T G A I S F V N E I P W V I E P V Y I S Q W G S M W I M M R R E K R D R R
P34369      Y R H D K R V Y L G A L K Y M P H A V L K L L E N M P M P W E Q I R D V Q V L Y H I T G A I T F V N E I P W V I E P V Y I A Q W G T M W I M M R R E K R D R R
BAA78744.1  Y R H D K R V Y L G A L K Y M P H A V L K L L E N M P M P W E Q I R D V K V L Y H I T G A I T F V N D I P R V I E P V Y M A Q W G T M W I M M R R E K R D R R
CAB80541.1  Y R H D K R V Y L G A L K F V P H A V Y K L L E N M P M P W E Q V R H V K I L Y H I T G A I T F V N E I P W V V E P I Y L A Q W G T M W I M M R R E K R D R R
H R L D K R V Y L G A L K F V P H A V F K L L E N M P M P W E Q V I I H L V L Y H I T G A I T F V N E V R V V V E P I Y M A Q W G S M W I M M R R E K R D R R

BAA22563.1  . . . . . 240 250 260 270 280 290 300 310
AAF58573.1  H F K R M R F P P F D D E E P P L D Y A D N I L N V E P L E A I Q L E L D P E E D A P V L D W F Y D H Q P L R D S R K Y V N G S T Y Q R W Q F T L P M M S T L
P34369      H F K R M R F P P F D D E E P P L D Y A D N V L D V E P L E A I Q I E L D N D E D N A V Y K W F Y D H R P L V D T - Q F V N G T T Y R K W N L S L P Q L A T L
BAA78744.1  H F K R M R F P P F D D E E P P L D Y A D N I L D V E P L E P I Q M E L D P E E D G A V A E W F Y D H K P L A T T R - F V N G P T Y R K W A F S I P Q M S T L
CAB80541.1  H F K R M R F P P F D D E E P P L D Y A D N L L D V E P L E A I Q L E L D E E E D S A V H E W F Y D H K P L V K T - K L I N G P S Y R K W H L S L P I M A T L
H F K R M R F P P F D D E E P P L D Y A D N L L D V D P L E A I Q L E L D E E E D S A V Y S W F Y D H K P L V K T - K M I N G P S Y Q T W N L S L P I M S T L

BAA22563.1  . . . . . 320 330 340 350 360 370 380 390
AAF58573.1  Y R L A N Q L L T D L V D D N Y F Y L F D L K A F F T S K A L N M A I P G G P K F E P L V R D I N L Q E D E D W N E F N D I N K I I R Q P I R T E Y K I A F P
P34369      Y R L A N Q L L T D L V D N N F F Y L F D P K S F F T A K A L N M A I P G G P K F E P L I K D H N V G D E D W N E F N D I N K V I R Q P I R T E Y R I A F P
BAA78744.1  Y R L A N Q L L T D L V D D N Y F Y L F D M K S F F T A K A L N V A I P G G P K F E P L V K D L H T - D E D W N E F N D I N K V I I R A P I R T E Y R I A F P
CAB80541.1  Y R L A G Q L L S D L I D R N Y F Y L F D M E S F F T A K A L N M C I P G G P K F E P L Y R D M E K G D E D W N E F N D I N K L I R Q P L R T E Y R I A F P
H R L A A Q L L S D L V D R N Y F Y L F D M P S F F T A K A L N M C I P G G P K F E P L H R D M E K G D E D W N E F N D I N K L I I R S P L R T E Y K V A F P

```

FIG. 9-2

AIEye output (February 26, 2001 9:53 AM)

BAA22563.1 Y L Y N N L P H - - H V H L T W Y H T P N V V F I K T E D P D L P A F Y F D P L I N P I S H R H S V K S - Q E P L P - - D D D - E E F E L P E F V E P F L K D  
AAF58573.1 Y L Y N N M P H - - F V H L S W Y H T P N V V Y I K T E D P D L P A F Y F D P L I N P I S H R N A N S K I Q E P L P - - D D D - E D F T L P D D V Q P F L Q D  
P34369 F M Y N N L I S S L P V Q V S W Y H T P S V V F I K T E D P D L P A F Y Y D P L I N P I V L S N - L K A T E E N L P E G E E E - D E W E L P E D V R P I F E D  
BAA78744.1 H L Y N N R P R - - K V R L G V Y H T P M I M Y I K T E D P D L P A F Y Y D P L I N P I T S T N K V D R R E R R T T E E D E D - E D F R L P D G V E P L L K G  
CAB80541.1 H L Y N N R P R - - K V K L C V Y H T P M V M Y I K T E D P D L P A F Y Y D P L I H P I S N S N N T N K E Q R K S N G Y D D D G D D F V L P E G L E P L L N N

BAA22563.1 T P L Y T D N T A N G I A L L W A P R P F N L R S G R T R R A L D I P L V K N W Y R E H C P A G Q P V K V R V S Y Q K L L K Y Y V L N A L K H R P P K A Q K K  
AAF58573.1 T P L Y T D N T A N G I A L L W A P R P F N M R S G R S R R A I D V P L V K C W Y K E H C P P G H P V K V R V S Y Q K L L K Y Y V L N A L K H R K P K P Q K K  
P34369 V P L Y T D N T A N G L A L L W A P R P F N L R S G R T R R A V D V P L V K S W Y R E H C P A G M P V K V R V S Y Q K L L K V F V L N A L K H R P P K P Q K R  
BAA78744.1 T E L Y T D T T A A G I S L L F A P K P F N M R S G R T R R A E D I P L V S E W Y K E H C P P A Y P V K V R V S Y Q K L L K C Y V L N E L H H R P P K A Q K K  
CAB80541.1 S P L Y T D T T A P G I S L L F A P R P F N M R S G R T R R A E D I P L V A E W F K E H C P P A Y P V K V R V S Y Q K L L K C Y L L N E L H H R P P K A Q K K

BAA22563.1 R Y L F R S F K A T K F F Q S T K L D W V E G L Q V C R Q G Y N M L N L I H R K N L N Y L H L D Y N F N L K P V K T L T T K E R K K S R F G N A F H L C R  
AAF58573.1 R Y L F R S F K A T K F F Q T T L D W V E A G L Q V C R Q G Y N M L N L I H R K N L N Y L H L D Y N F N L K P V K T L T T K E R K K S R F G N A F H L C R  
P34369 R Y L F R S F K A T K F F Q T T L D W V E A G L Q V L R Q G Y N M L N L I H R K N L N Y L H L D Y N F N L K P V K T L T T K E R K K S R F G N A F H L C R  
BAA78744.1 K H L F R S L O A T K F F Q T T E L D W A E A G L Q V C K Q G Y N M L N L I H R K N L N Y L H L D Y N F N L K P V K T L T T K E R K K S R F G N A F H L C R  
CAB80541.1 K H L F R S L A A T K F F Q S T E L D W V E V G L Q V C R Q G Y N M L N L I H R K N L N Y L H L D Y N F N L K P V K T L T T K E R K K S R F G N A F H L C R

BAA22563.1 E V L R L T K L V V D S H V Q Y R L G N V D A F Q L A D G L Q Y I F A H V G Q L T G M Y R Y K Y K L M R Q I R V C K D L K H L I Y Y R F N T G P V G K G P G C  
AAF58573.1 E I L R L T K L I I D S H V Q Y R L N N V D A F Q L A D G L Q Y I F A H V G Q L T G M Y R Y K Y K L M R Q I R M C K D L K H L I Y Y R F N T G P V G K G P G C  
P34369 E I L R L T K L V D A H V Q Y R L N N V D A Y Q L A D G L Q Y I F A H V G Q L T G M Y R Y K Y K L M R Q V R M C K D L K H L I Y Y R F N T G P V G K G P G C  
BAA78744.1 E I L R L T K L V D A N I Q F R L G N V D A F Q L A D G L Q Y I F S H V G Q L T G M Y R Y K Y R L M R Q I R M C K D L K H L I Y Y R F N T G P V G K G P G C  
CAB80541.1 E I L R L T K L V D A N V Q F R L G N V D A F Q L A D G L Q Y I F S H V G Q L T G M Y R Y K Y R L M R Q I R M C K D L K H L I Y Y R F N T G P V G K G P G C

BAA22563.1 G F W A A G W R V W L F F M R G I T P L L E R W L G N L L A R Q F E G R H S K G V A K T V T K Q R V E S H F D L E L R A A V M H D I L D M M P E G I K Q N K A  
AAF58573.1 G F W A P G W R V W L F F M R G I T P L L E R W L G N L L S R Q F E G R H S K G V A K T V T K Q R V E S H F D L E L R A S V M H D I V D M M P E G I K Q N K A  
P34369 G F W A P G W R V W L F F L R G I T P L L E R W L G N L L S R Q F E G R H S K G V A K T V T K Q R V E S H F D L E L R A A V M H D I L D M M P D G I K Q N K A  
BAA78744.1 G F W A P M W R V W L F F L R G I V P L L E R W L G N L L A R Q F E G R H S K G V A K T V T K Q R V E S H F D L E L R A A V M H D V L D A M P E G I K Q N K A  
CAB80541.1 G F W A P M W R V W L F F L R G I V P L L E R W L G N L L A R Q F E G R H S K G V A K T V T K Q R V E S H F D L E L R A A V M H D V V D A M P E G I K Q N K A

FIG. 9-3

AIEye output (February 26, 2001 9:53 AM)

	800	810	820	830	840	850	860	
BAA22563.1	R T I L Q H L S E A W R C W K A N I P W K V P G L P T P I E N M I L R Y V K A K A D W W T N T A H Y N R E R I R R G A T V D K T V C K K N L G R L T R L Y L K							
AAF58573.1	R T I L Q H L S E A W R C W K A N I P W K V P G L P I P I E N M I L R Y V K M K A D W W T N T A H Y N R E R I R R G A T V D K T V C K K N L G R L T R L Y L K							
P34369	R V I L Q H L S E A W R C W K A N I P W K V P G L P T P V E N M I L R Y V K A K A D W W T N S A H Y N R E R V R R G A T V D K T V C K K N L G R L T R L Y L K							
BAA78744.1	R T I L Q H L S E A W R C W K A N I P W K V P G L P V P I E N M I L R Y V K S K A D W W T N V A H Y N R E R I R R G A T V D K T V C R K N L G R L T R L W L K							
CAB80541.1	R T I L Q H L S E A W R C W K A N I P W K V P G L P V A I E N M I L R Y V K S K A D W W T N V A H Y N R E R I R R G A T V D K T V C R K N L G R L T R L W L K							
	870	880	890	900	910	920	930	940
BAA22563.1	A E Q E R Q H N Y L K D G P Y I T A E E T V A V Y T T T V H W L E S R R F S P I P F P P L S Y K H D T K L L I L A L E R L K E A Y S V K S R L N Q S Q R E E L							
AAF58573.1	A E Q E R Q H N Y L K D G P Y I S P E E A V A I Y T T T V H W L E S R R F A P I P F P P L S Y K H D T K L L I L A L E R L K E A Y S V K S R L N Q S Q R E E L							
P34369	S E Q E R Q H N Y L K D G P Y I S A E E A V A I Y T T T V H W L E S R R F S P I P F P P L S Y K H D T K L L I L A L E R L K E S Y S V A V R L N Q L Q R E E L							
BAA78744.1	A E Q E R Q H N Y L K D G P Y V T P E E A V A I Y T T T V H W L E S R K F S P I P F P P L S Y K H D T K L L I L A L E R L K E S Y S V A V R L N Q L Q R E E L							
CAB80541.1	A E Q E R Q H N F Q K D G P Y V T A D E G I A I Y S T T V N W L E S R K F S A I P F P P L S Y K H D T K L L I L A L E R L K E S Y S A A V K L N Q Q R E E L							
	950	960	970	980	990	1000	1010	1020
BAA22563.1	G L I E Q A Y D N L H E A L S R I K R H L L T Q R A F K E V G I E F M D L Y S H L P V Y D V E P L E K I T D A Y L D Q Y L W Y E A D K R R L F P P W I K P A							
AAF58573.1	G L I E Q A Y D N P H E A L S R I K R H L L T Q R A F K E V G I E F M D L Y S H L I P V Y E V E P L E K I T D A Y L D Q Y L W Y E A D K R R L F P P W I K P S							
P34369	A L I E Q A Y D N P H E A L S R I K R H M L T Q R A F K E V G I E F M D L Y T H L I P V Y D I E P L E K V T D A Y L D Q Y L W Y E A D K R R L F P A W V K P G							
BAA78744.1	G L I E Q A Y D N P H E A L S R I K R H L L T Q R A F K E V G I E F M D L Y S Y L I P V Y E I E P L E K I T D A Y L D Q Y L W Y E G D K R H L F P N W V K P A							
CAB80541.1	G L I E Q A Y D N P H E A L M R I K R H L L T Q H S F K E V G I E F M D L Y S H L I P V Y Q I D P L E K I T D A Y L D Q Y L W Y E G D K R H L F P N W I K P A							
	1030	1040	1050	1060	1070	1080	1090	1100
BAA22563.1	D T E P P L L V Y K W C Q G I N N L Q D V W E T S E G E C N V M L E S R F E K M Y E K I D L T L L N R L V R L I V D H N I A D Y M T A K N N V V I N Y K D M							
AAF58573.1	D T E P P L L A Y K W C Q G I N N L Q D V W D V G E G E C N V L E S R F E K L Y E K I D L T L L N R L L R L I V D H N I A D Y M T A K N N V V I N Y K D M							
P34369	D T E P P L L T Y K W C Q G L N N L Q D V W E T S E G E C N V I M E T K L E K I A E K M D L T L L N R L L R L I V D H N I A D Y M T S K N N V L I N Y K D M							
BAA78744.1	D S E P P L L V Y K W C Q G I N N L Q D V W D T S D G Q C V V M L Q T K F E K F F E K I D L T L L N R L L R L V L D H N I A D Y V T A K N N V L S Y K D M							
CAB80541.1	D S E P P L L V Y K W C Q G I N N L Q G I W D T S D G Q C V V M L Q T K F E K L F E K I D L T V L N S L L R L V L D P K L A N Y V T G K N N V V L S Y K D M							
	1110	1120	1130	1140	1150	1160	1170	1180
BAA22563.1	N H T N S Y G I I R G L Q F A S F I V Q Y Y G L V M D L L V L G L H R A S E M A G P P Q M P N D F L S F Q D I A T E A A H P I R L F C R Y I D R I H I F F R F							
AAF58573.1	N H T N S Y G I I R G L Q F S S F I T Q Y Y G L V L D L L V L G L H R S S E M A G P P Q M P N D F L T F Q D T V T E T A H P I R L Y C R Y V D R I H L F F R F							
P34369	N H T N S F G I I R G L Q F A S F I V Q F Y G L V L D L L V L G L R R A S E I A G P P Q C P N E F L Q F Q D V A T E I G H P I R L Y C R Y I D R V W I M F R F							
BAA78744.1	S H T N S Y G L I R G L Q F A S F V V Q Y Y G L V L D L L L G L T R A S E I A G P P T M P N E F L T Y A D T K V E T R H P I R L Y S R Y I D K V H I M F R F							
CAB80541.1	S Y T N T Y G L I R G L Q F A S F V V Q F Y G L V L D L L L G L T R A S E I A G P P Q R P N E F M T Y W D T K V E T R H P I R L Y S R Y I D K V H I M F K F							

FIG. 9-4

AIEye output (February 26, 2001 9:53 AM)

1181 1185 1270 1186  
1182 1183 1184 1177  
1174 1177  
1260 1250 1240 1230 1220 1210 1200 1190  
BAA22563.1 T A D E A R D L I Q R Y L T E H P D P N N E N I V G Y N N K K C W P R D A R M R L M K H D V N L G R A V F W D I K N R L P R S V T T V Q W E N S F V S V Y S K  
AAF58573.1 S A E E A R D L I Q R Y L T E H P D P N N E N I V G Y N N K K C W P R D A R M R L M K H D V N L G R A V F W D I K N R L P R S V T T I G W E S T F V S V Y S K  
P34369 S A D E A R D L I Q R Y L T E H P D P N N E N I V G Y N N K K C W P R D A R M R L M K H D V N L G R A V F W D I K N R L P R S I T T V E W E N S F V S V Y S K  
BAA78744.1 T H E E A R D L I Q R Y L T E H P D P N N E N M V G Y N N K K C W P R D A R M R L M K H D V N L G R S V F W D M K N R L P R S I T T L E W E N S F V S V Y S K  
CAB80541.1 T H E E A R D L I Q R H L T E R P D P N N E N M V G Y N N K K C W P R D A R M R L M K H D V N L G R S V F W D M K N R L P R S I T T L E W E N G F V S V Y S K

1310 1300 1216 1217  
1230 1230  
1330 1340  
BAA22563.1 D I N P N L L F N M C G F E C R I L P K C R T S Y E E F T H - K D G V W N L Q N E V T K E R T A Q C F L R V D D E S M Q R F H N R V R Q I L M A S G S T T F T K  
AAF58573.1 D I N P N L L F N M S G F E C R I L P K C R T Q N E E F T H - R D G V W N L Q N E I T K E R T A Q C F L R V D D E S L G R F H N R V R Q I L M A S G S T T F T K  
P34369 D I N P N M L F D M S G F E C R I L P K C R T A N E E F V H - R D G V W N L Q N E V T K E R T A Q C F L R V D E E S L S K F H N R I R Q I L M S S G S T T F T K  
BAA78744.1 D I N P N L L F S M C G F E V R I L P K I R M T Q E A F S N T K D G V W N L Q N E Q T K E R T A I A F L R V D D E H M K V F E N R V R Q I L M S S G S T T F T K  
CAB80541.1 D I N P N L L F S M C G F E V R V L P K I R M G Q E A F S T R D G V W N L Q N E Q T K E R T A V A F L R A D D E H M K V F E N R V R Q I L M S S G S T T F T K

1350 1360 1370 1380 1390 1400 1410 1420  
BAA22563.1 I V N K W N T A L I G L M T Y F R E A V N T Q E L L D L L V K C E H K I Q T R I K I G L N S K M P S R F P P V V F Y T P K E L G G L G M L S M G H V L I P Q  
AAF58573.1 I V N K W N T A L I G L M T Y F R E A V N T Q E L L D L L V K C E N K I Q T R I K I G L N S K M P S R F P P V V F Y T P K E L G G L G M L S M G H V L I P Q  
P34369 I V N K W N T A L I G L M T Y F R E A V N T Q E L L D L L V K C E N K I Q T R I K I G L N S K M P S R F P P V V F Y T P K E I G G L G M L S M G H V L I P Q  
BAA78744.1 I V N K W N T A L I G L M T Y F R E A T V H T Q E L L D L L V K C E N K I Q T R I K I G L N S K M P S R F P P V I F Y T P K E I G G L G M L S M G H I L I P Q  
CAB80541.1 I V N K W N T A L I G L M T Y F R E A T V H T Q E L L D L L V K C E N K I Q T R V K I G L N S K M P S R F P P V I F Y T P K E I G G L G M L S M G H I L I P Q

1430 1440 1450 1460 1470 1480 1490 1500  
BAA22563.1 S D L R W S K Q T D V - G I T H F R S G M S H E E D Q L I P N L Y R Y I Q P W E S E F I D S Q R V W A E Y S L K R Q E A I A Q N R R L T L E D L E D S W D R G  
AAF58573.1 S D L R W S K Q T D V - G I T H F R S G M S H D E D Q L I P N L Y R Y I Q P W E S E F I D S Q R V W A E Y A L K R Q E A N A Q N R R L T L E D L E D S W D R G  
P34369 S D L R W M Q Q T E A G G V T H F R S G M S H D E D Q L I P N L Y R Y I Q P W E A E F V D S V R V W A E Y A L K R Q E A N A Q N R R L T L E D L D D S W D R G  
BAA78744.1 S D L R Y S K Q T D V - G V T H F R S G M S H E E D Q L I P N L Y R Y I Q P W E S E F I D S Q R V W A E Y A L K R Q E A Q S Q N R R L T L E D L E D S W D R G  
CAB80541.1 S D L R Y S N Q T D V - G V S H F R S G M S H E E D Q L I P N L Y R Y I Q P W E S E F I D S Q R V W A E Y A L K R Q E A Q A Q N R R L T L E D L E D S W D R G

1510 1520 1530 1540 1550 1560 1570 1580  
BAA22563.1 I P R I N T L F Q K D R H T L A Y D K G W R V R T D F K Q Y Q V L K Q N P F W W T H Q R H D G K L W N L N N Y R T D M I Q A L G G V E G I L E H T L F K G T Y  
AAF58573.1 I P R I N T L F Q K D R H T L A Y D K G W R I R T E F K Q Y Q V L K Q N P F W W T H Q R H D G K L W N L N N Y R T D M I Q A L G G V E G I L E H T L F K G T Y  
P34369 I P R I N T L F Q K D R H T L A Y D K G W R V R T E F K A Y Q I L K Q N P F W W T H Q R H D G K L W N L N N Y R T D M I Q A L G G V E G I L E H T L F R G T Y  
BAA78744.1 I P R I N T L F Q K D R H T L A Y D K G W R V R T D F K Q Y Q V L K Q N P F W W T H Q R H D G K L W N L N N Y R T D V I Q A L G G V E G I L E H T L F K G T Y  
CAB80541.1 I P R I N T L F Q K D R H T L A Y D K G W R V R T D F K Q Y Q A L K Q N P F W W T H Q R H D G K L W N L N N Y R T D V I Q A L G G V E G I L E H T L F K G T Y

AIEye output (February 26, 2001 9:53 AM)

FIG. 9-5

	1590	1600	1610	1620	1630	1640	1650	
BAA22563.1	FPTW	EGLF	WEKA	SGFE	ESMK	KWK	KL	TNAQ
AAF58573.1	FPTW	EGLF	WEKA	SGFE	ESMK	KWK	KL	TNAQ
P34369	FPTW	EGLF	WEKA	SGFE	ESMK	KWK	KL	TNAQ
BAA78744.1	FPTW	EGLF	WEKA	SGFE	ESMK	KWK	KL	TNAQ
CAB80541.1	FPTW	EGLF	WEKA	SGFE	ESMK	KWK	KL	TNAQ
	1660	1670	1680	1690	1700	1710	1720	1730
BAA22563.1	IQI	FRAH	LWQK	I	HES	I	VMDL	CQV
AAF58573.1	IQI	FRAH	LWQK	I	HES	I	VMDL	CQV
P34369	IQI	FRAH	LWQK	I	HES	I	VMDL	CQV
BAA78744.1	IQI	FRAH	LWQK	I	HES	I	VMDL	CQV
CAB80541.1	IQI	FRAH	LWQK	I	HES	I	VMDL	CQV
	1740	1750	1760	1770	1780	1790	1800	1810
BAA22563.1	DST	TTQK	YWID	I	QLRW	G	YD	SHD
AAF58573.1	DNT	TTQK	YWLD	I	QLRW	G	YD	SHD
P34369	DNT	TTQK	YWLD	I	QLRW	G	YD	SHD
BAA78744.1	DQK	ASNK	YWID	I	VQLR	W	G	YD
CAB80541.1	DQK	ASNK	YWID	I	VQLR	W	G	YD
	1820	1830	1840	1850	1860	1870	1880	1890
BAA22563.1	KIM	KAN	PAL	Y	VL	R	E	R
AAF58573.1	KIM	KAN	PAL	Y	VL	R	E	R
P34369	KIM	KAN	PAL	Y	VL	R	E	R
BAA78744.1	KIM	KAN	PAL	Y	VL	R	E	R
CAB80541.1	KIM	KAN	PAL	Y	VL	R	E	R
	1900	1910	1920	1930	1940	1950	1960	1970
BAA22563.1	NPR	TG	Q	L	F	L	K	I
AAF58573.1	NPR	TG	Q	L	F	L	K	I
P34369	NPR	TG	Q	L	F	L	K	I
BAA78744.1	NPR	TG	Q	L	F	L	K	I
CAB80541.1	NPR	TG	Q	L	F	L	K	I

AIEye output (February 26, 2001 9:53 AM)

FIG. 9-6

	1980	1990	2000	2010	2020	2030	2040	2050
BAA22563.1	QLPFQACLKVEKFGDL	ILKATEPQMVLFNLYDDWLKTI	SSYTAFSRLILIRALHVNNDRAK	VILKPKDKTT	ITEPHHIW			
AAF58573.1	QLPFQACLKVEKFGDL	ILKATEPQMVLFNLYDDWLKTI	SSYTAFSRLILIRALHVNNDRAK	VILKPKDKTT	ITEPHHIW			
P34369	MLPFQAIMKVEKFGDL	ILKATEPQMVLFNLYDDWLKTI	SSYTAFSRVVIMRGMHINPD	KTKVILKPKDKTT	ITEPHHIW			
BAA78744.1	QLPFQACLKIEKFGDL	ILKATEPQMVLYNIYDDWLKSI	SSYTAFSRIVLIRALHVNNEKAK	MLLKPKDKTIVTEPHHIW				
CAB80541.1	QLPFQACLKIEKFGDL	ILKATEPQMALFNIYDDWLM	TVSSYTAQRLLILRALHVNNEKAK	MLLKPKDMSVVTEPHHIW				
	2060	2070	2080	2090	2100	2110	2120	2130
BAA22563.1	PTLTDEEVIKVEVQLKDL	ILADYGKKNVNVASLTQSEIRD	ILGMEISAPSQORQ	IAEIEKQTK	KEQSQ	LTATQTRTV		
AAF58573.1	PTLTDEEVIKVEVQLKDL	ILADYGKKNVNVASLTQSEIRD	ILGMEISAPSAQRQ	IAEIEKQTK	KEQNQLT	ATTTRTT		
P34369	PTLSDDDWIKVELAKDM	ILADYGKKNVNVASLTQSEVRD	ILGMEISAPSQORQ	IAEIEKQTK	KEQSQ	VTATTRTV		
BAA78744.1	PTLTDEQWLKVECALRD	LILSDYAKKNVNTSAL	TQSEIRDILGAEI	APPSQORQ	IAEIEKQSR	ETTQLTAVTTRTT		
CAB80541.1	PSLTDDQWMKVEVALRD	LILSDYAKKNKVNTSAL	TQSEIRDILGAEI	PPSQORQ	IAEIEKQAK	EASQ	LTAVTTRTT	
	2140	2150	2160	2170	2180	2190	2200	2210
BAA22563.1	NKHGDEIITSTTSNYETQ	TFSSKTEWRVRAIS	ANLHLR	TNHIYVSSDD	IKETGYTYILPKN	VLKKFICISDLRAQ	IAAG	
AAF58573.1	NKHGDEIITSTTSNYETQ	TFSSKTEWRVRAIS	ATNLHLR	TNHIYVSSDD	IKETGYTYILPKN	ILKKFVTSDLRAQ	IAAG	
P34369	NKHGDEIITATTSNYET	ASFASRTEWRVRAIS	STNLHLR	TQHIYVNSDD	VKDTGYTYILPKN	ILKKFITISDLRTQ	IAAG	
BAA78744.1	NVHGDEIITTSPEEQ	AFASKTDWRVRAIS	ATNLYLR	VNHIYVNSDD	IKETGYTYIMP	KNILKKFICISDLRTQ	IAAG	
CAB80541.1	NVHGDELISTTISPEEQ	SFAFGSKTDWRVRAIS	ATNLYLR	VNHIYVNSDD	IKETGYTYIMP	KNILKKFICISDLRTQ	IAAG	
	2220	2230	2240	2250	2260	2270	2280	2290
BAA22563.1	YLYGVSPDDNPQVKEIR	CIVMVPQWGTHQTVHL	PGQLPQHEYLKEMEPL	GWHTQPNESPQLSPQDV	TTHAKIMADNPS			
AAF58573.1	YLYGVSPDDNPQVKEIR	CIVMPPQWGTHQTNL	PNTLPTHQYLKMEPL	GWHTQPNELPQLSPQDI	TTHAKIMQENSN			
P34369	FMYGVSPDDNPQVKEIR	CIVLVPQTGSHQQVNL	PTQLPDHELLRDFEPL	GWHTQPNELPQLSPQDV	TTHAKLLTDNIS			
BAA78744.1	FLYGLSPDDNPQVKEIR	CIAPPHGTHQMVTL	PANLPEHEFLNDLEPL	GWHTQPNELPQLSPQDL	TSHAKILENNKQ			
CAB80541.1	YLYGISPPDDNPQVKEIR	CVVMVPCCGNHQQVQL	PSSLPEHQFLDDLEPL	GWHTQPNELPQLSPQDV	TFHTRVLENNKQ			
	2300	2310	2320	2330	2340	2350	2360	2370
BAA22563.1	WDGKTIITCSFTPGSC	TLTAYKLT	PSGYEWGRQNTDKGN	NPKGYLPSHYER	VQMLLSDRFL	GLFFMVP	PAQSSWN	YNFM
AAF58573.1	WDGKTIITCSFTPGSC	SLTAYKLT	PSGFEGWGSKNTDKGN	NPKGYLPSHYER	VQMLLSN	KFLGLFFMVP	PAQSSWN	YNFM
P34369	WDGKTVMTCSFTPGSV	SLTAYKLT	PSGYEWGKANTDKGN	NPKGYMPTHYEK	VQMLLSDRFL	GLGYFMV	PSNGVWN	YNFQ
BAA78744.1	WDGKCIILCSFTPGSC	SLTAYKLT	PSGYEWGRSNKDT	GSNPHGYLPTHYEK	VQMLLSDRFL	GLGYMVP	PDNT	TPWN
CAB80541.1	WDAEKCIILCSFTPGSC	SLTSYKLT	QAGYEWGR	LNKDTGSNPHGYLPTHYEK	VQMLLSDRFL	GLGYMVP	PENGPWN	YNFM



FIG. 10

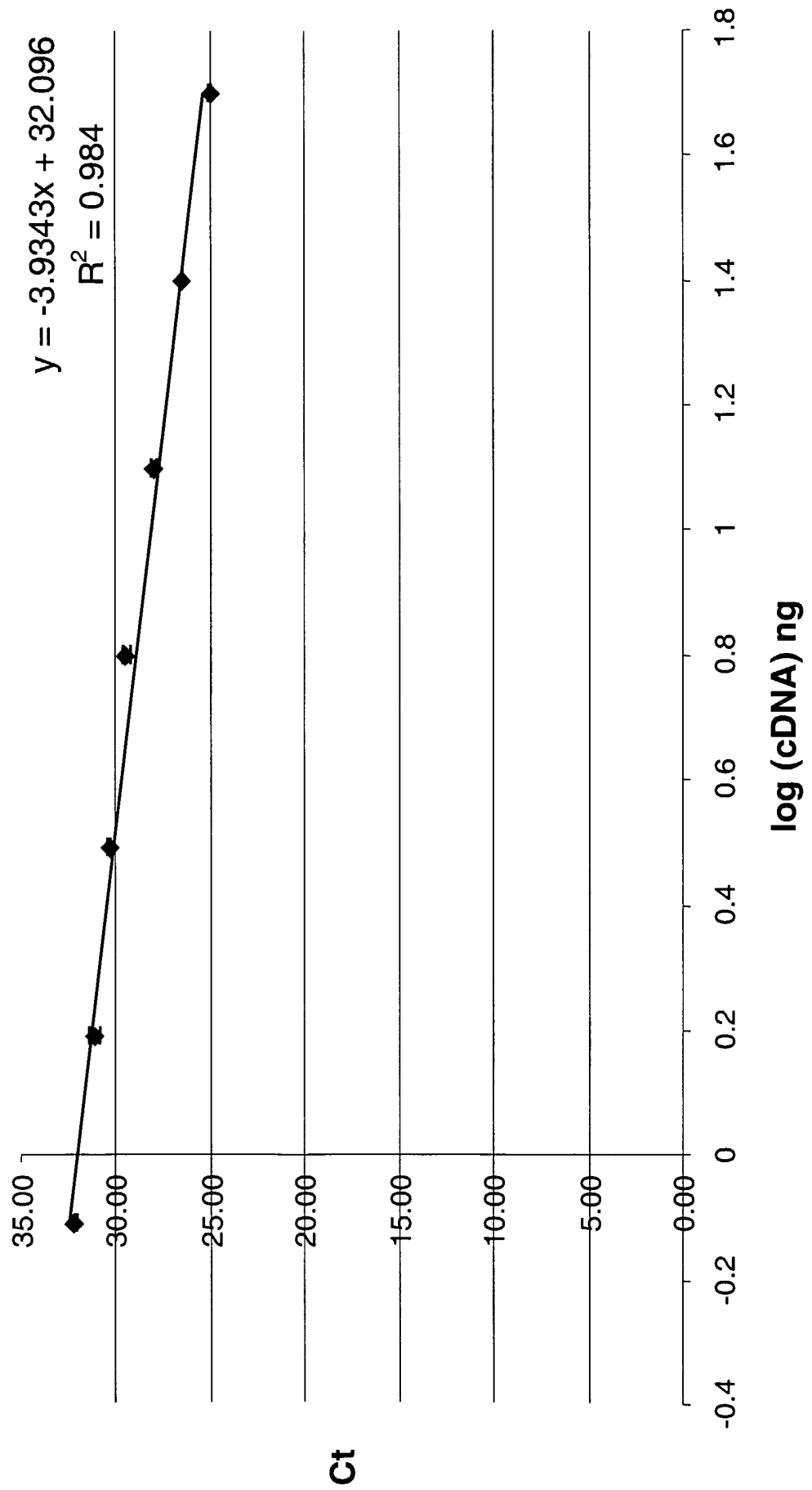


FIG. 11

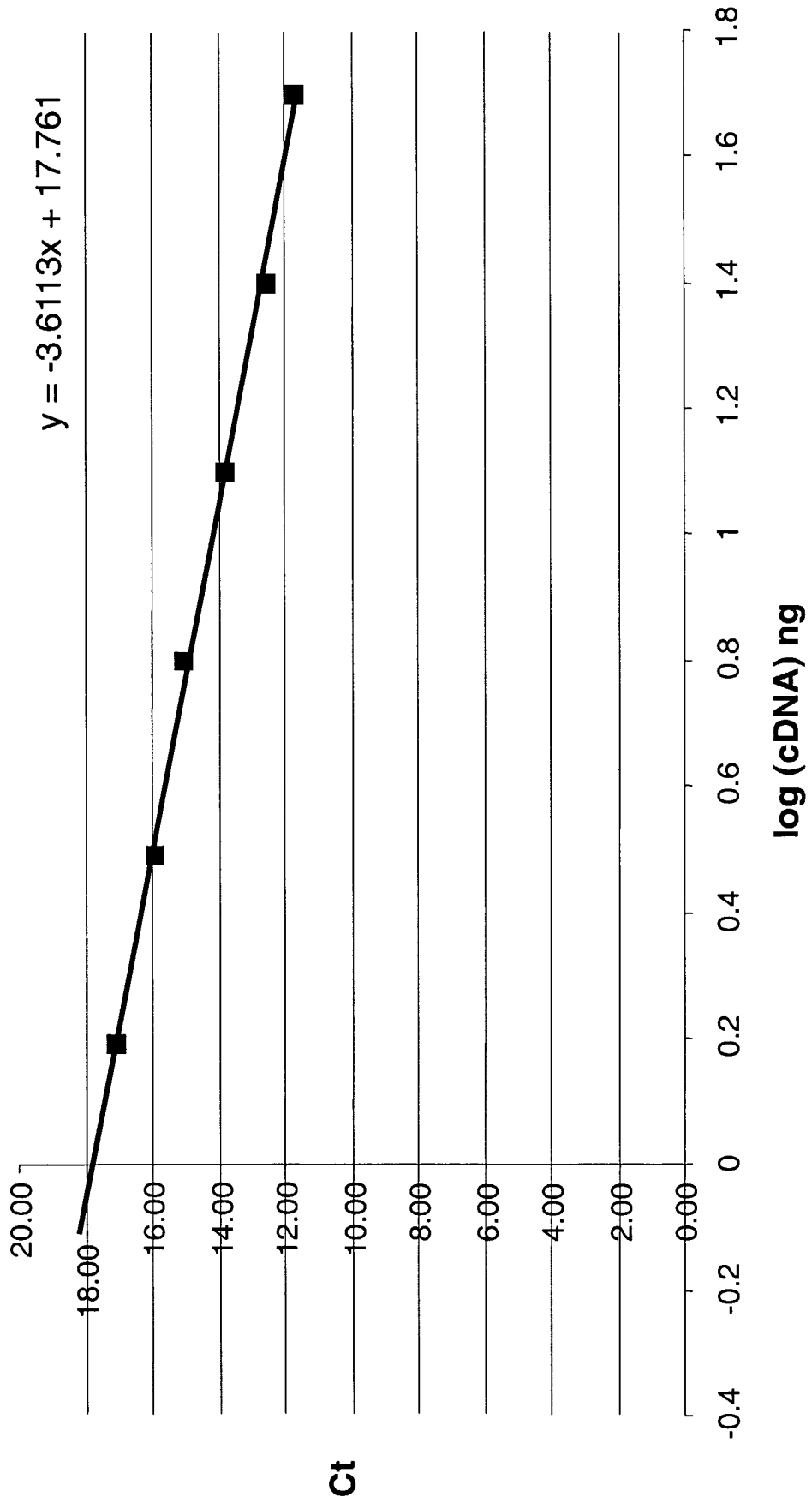


FIG. 12

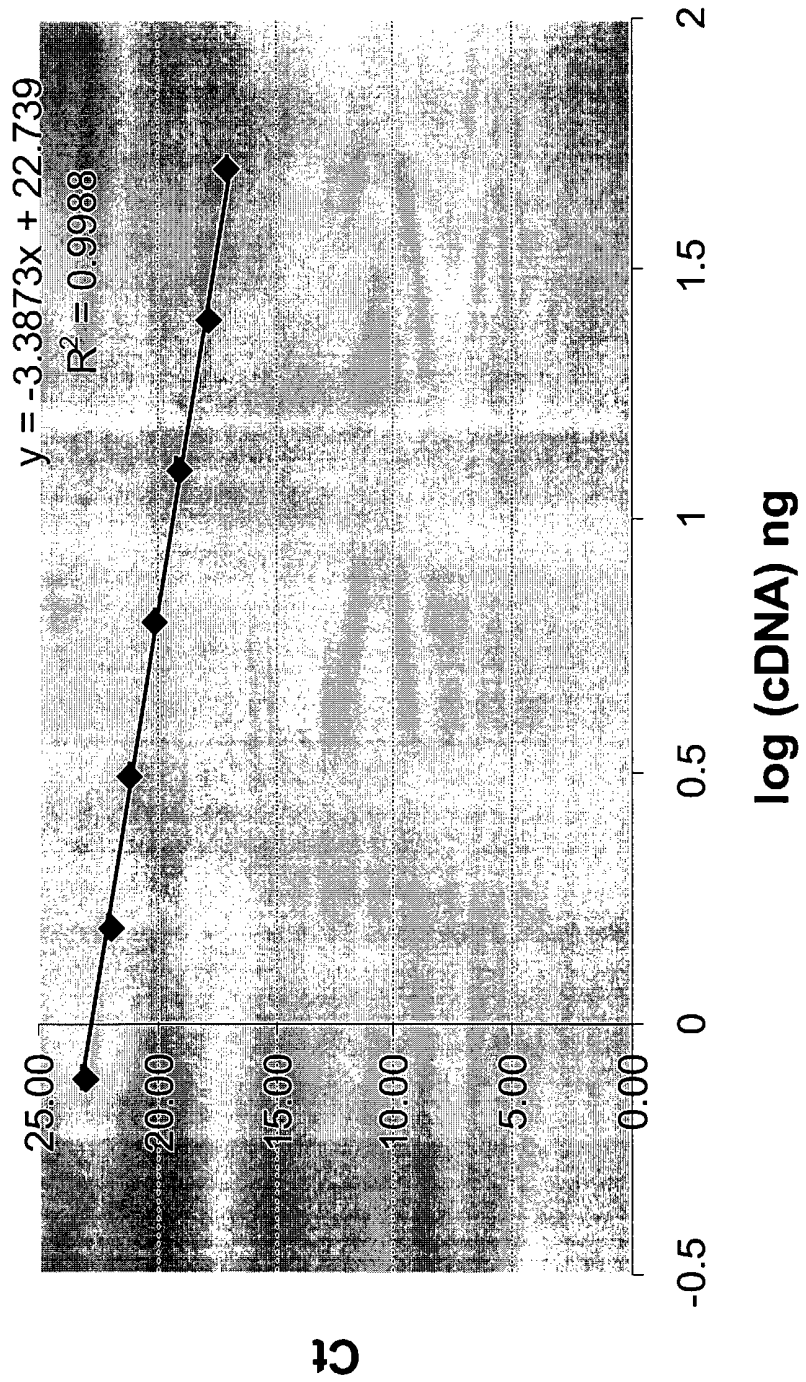


FIG. 13

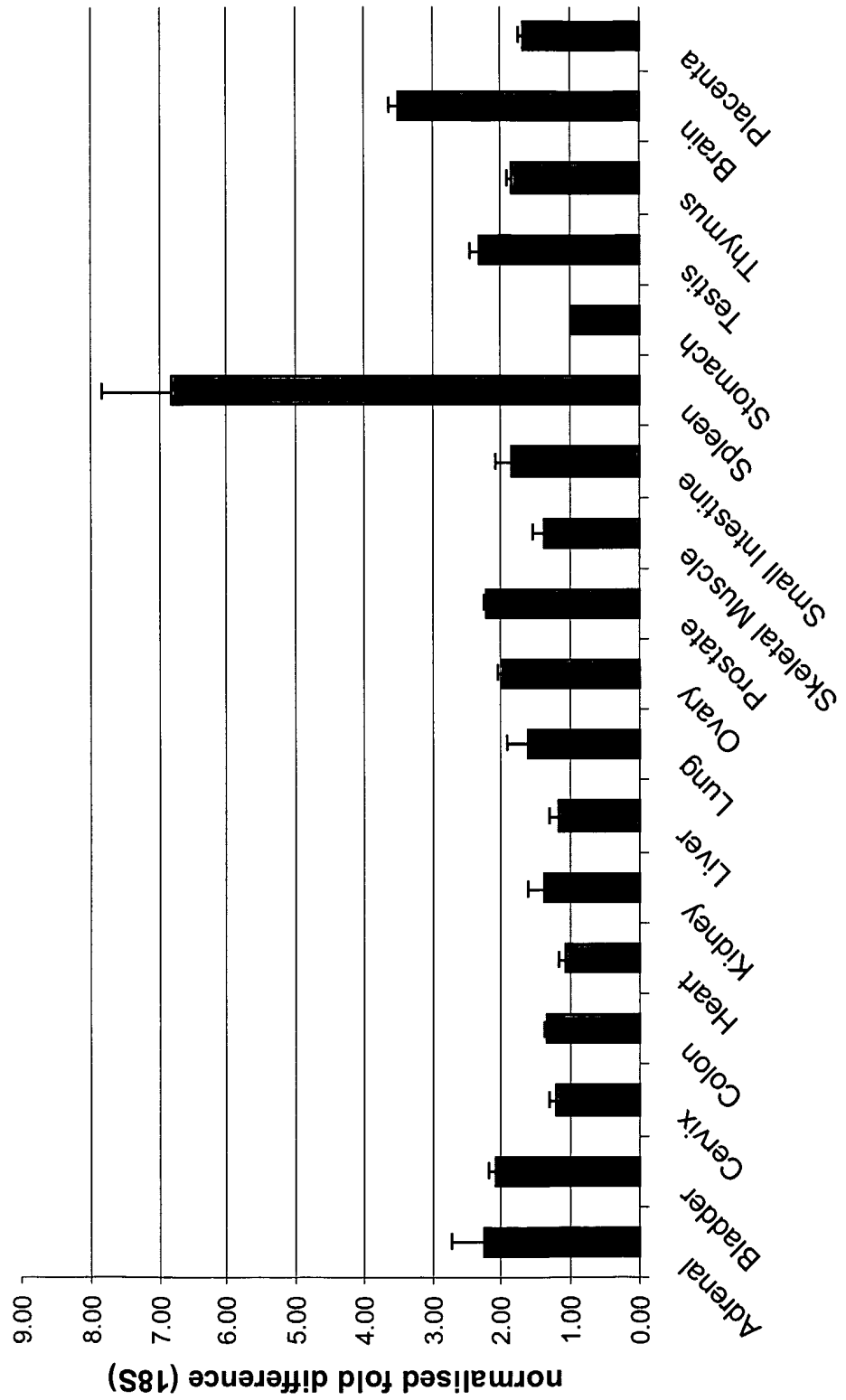
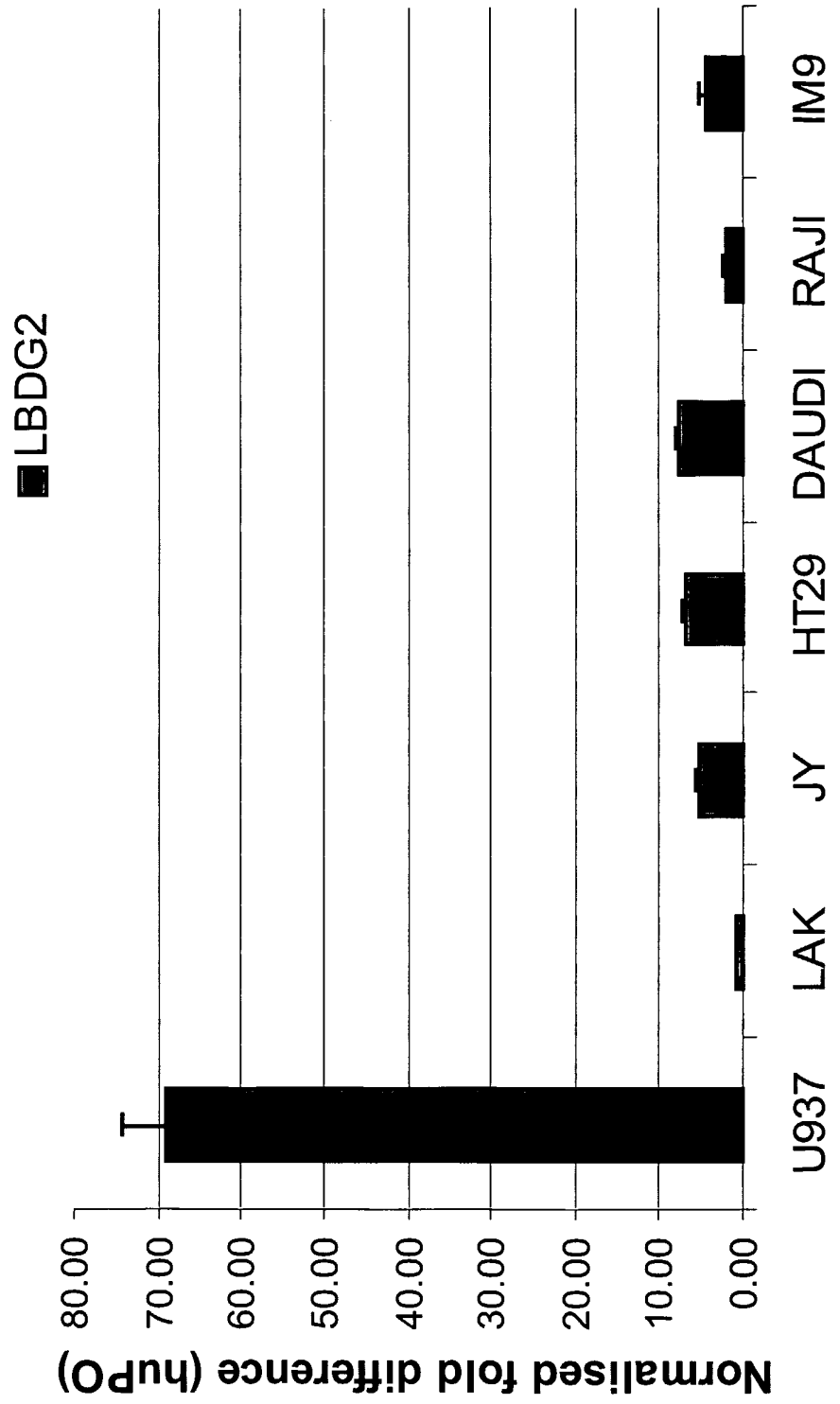


FIG. 14



## Sequence Listing

SEQ ID NO: 1 Nucleotide coding sequence for BAA22563.1 (LBDG2) protein

```
1  cgggcccgcct cttgtgtgag ggcctgtggg attctccgga tatggccgga gtgtttcctt
5  61  atcgagggcc gggtaaccgc gtgectggcc ctctagcccc gctaccggac tacatgtcgg
121 aggagaagct gcaggagaaa gctcgaaaat ggcagcaatt gcaggccaag cgctatgcag
181 aaaagcggaa gtttgggttt gtggatgccc agaaggaaga catgccccca gaacatgtca
241 gggagatcat tcgagaccat ggagacatga ccaacaggaa gttccgccat gacaaaaggg
301 tttacttggg tgccctaaag tacatgcccc acgcagtcct caaactcctg gagaacatgc
10 361 ctatgccttg ggagcagatt cgggatgtgc ccgtgctgta ccacatcact ggagccattt
421 ctttcgtcaa tgagattccc tgggtcattg aaactgtcta catctcccag tgggggtcaa
481 tgtggattat gatgcgccga gaaaaagag ataggaggca tttcaagaga atgcgttttc
541 ccccttttga tgatgaggag ccgcccttgg actatgctga caacatccta aatggtgagc
601 cactggaggc cattcagcta gagctggacc ctgaggagga cgcccctgtg ttggactggt
15 661 tctatgacca ccagccgttg agggacagca ggaagtatgt aaatggctcc acttaccagc
721 gctggcagtt cacactacct atgatgtcaa ctctctaccg cctggcta at cagctcctga
781 cagacttggg ggatgacaac tacttctacc tgtttgattt gaaggccttc tttacgtcca
841 aggcactcaa tatggccatt cctggaggcc ccaaatttga acctcttgtt cgagacatca
901 acctacagga tgaagactgg aatgaattca atgatattaa caagattatc atccggcagc
20 961 ctatccggac tgagtacaag attgcttttc cttacttgta caacaatctt ccacaccatg
1021 tccacctcac ctggtaccat actcccaatg ttgtattcat caaaactgaa gatcctgact
1081 tgccagcttt ctactttgac cctttgatca acccaatctc ccataggcac tcagtcaaga
1141 gccaggaacc attgccggat gatgatgagg aatttgagct cccggagttt gtggagccct
1201 tcctgaagga cacaccctc tatacagaca atacagccaa tggcattgcc ctgctctggg
25 1261 ccccgcggcc cttcaacctc cgctctggtc gcaccgctcg ggcctggac atacccttg
1321 tcaagaactg gtatcgggag cattgtcctg ccgggcagcc tgtgaaagtg agggctctct
1381 accagaagct gcttaagtac tatgtgctga atgccctgaa gcatcgcccc cctaaggctc
1441 aaaagaagag gtatttgttc cgctccttca aagccaccaa attctttcag tcacaaagc
1501 tggactgggt ggagggttgg ctccaggttt gccgccaggg ctacaacatg ctcaaccttc
30 1561 tcattcaccg caaaaacctc aactacctgc acctggacta caacttcaac ctcaagcctg
1621 tgaaaacgct caccaccaag gaaagaaaga aatctcgttt tgggaatgct ttccacctgt
```

1681 gtcgggaagt tctgcgtttg actaagctgg tggatgatag tcacgtgcag tatcggctgg  
1741 gcaatgtgga tgccttccag ctggcagatg gattgcagta tatatttggc catggttgggc  
1801 agttgacggg catgtatcga tacaataca agctgatgcg acagattcgc gtgtgcaagg  
1861 acctgaagca tctcatctat tatcgtttca acacaggccc tgtaggaag ggtcctggct  
5 1921 gtggcttctg ggctgccggg tggcgagtct ggctcttttt catgctggc attaccctt  
1981 tattagagcg atggcttggc aacctcctgg cccggcagtt tgaagtcga cactcaaagg  
2041 ggggtggcaa gacagtaaca aagcagcag tggagtcaca ttttgacctt gagctgctgg  
2101 cagctgtgat gcatgatatt ctggacatga tgctgaggg gatcaagcag aacaaggccc  
2161 ggacaatcct gcagcacctc agtgaagcct ggcgctgctg gaaagccaac attcctgga  
10 2221 aggtccctgg gctgccgacg cccatagaga atatgatcct tcgatacgtg aaggccaagg  
2281 ctgactgggt gaccaacact gccactaca accgagaacg gatccgccga ggggccactg  
2341 tggacaagac tgtttgtaaa aagaatctgg gccgcctcac cgggtctat ctgaaggcag  
2401 aacaggagcg gcagcacaac tacctgaagg acgggcctta catcacagcg gaggaacag  
2461 tggcagtata taccaccaca gtgcattggt tggaaagccg caggttttca cccatcccat  
15 2521 tccccccact ctctataag catgacacca agttgctcat cttggcattg gagcggctca  
2581 aggaagctta tagtgtgaag tctcggttga accagtctca gagggaggag ctagggtctga  
2641 tcgagcagcg ctacgataac ctccacgagg cgctgtcccg cataaagcgt cacctcctca  
2701 cacagagagc cttcaaagag gtgggcattg agttcatgga tctgtatagc cacctcgttc  
2761 cagtatatga tgttgagccc ctggagaaga taactgatgc ttacctggac cagtacctgt  
20 2821 ggtatgaagc cgacaagcgc cgctgttcc caccctggat taagcctgca gacacagaac  
2881 cacctccact gcttgtttac aagtgtgtc aaggcatcaa taacctgcag gacgtgtggg  
2941 agacgagtga aggcgagtgc aatgtcatgc tggaatcccg ctttgagaag atgtatgaga  
3001 agatcgactt gactctgctc aacaggctcg tgcgcctcat cgtggaccac aacatagccg  
3061 actacatgac agccaagaac aacgtcgtca tcaactataa ggacatgaac catacgaatt  
25 3121 catatgggat catcagaggc ctgcagtttg cctcattcat agtgcagtat tatggcctgg  
3181 tgatggattt gcttgtattg ggattgcacc gggccagtga gatggctggg cccctcaga  
3241 tgccaaatga ctttctcagt ttccaggaca tagccactga ggctgcccac cccatccgtc  
3301 tcttctgcag atacattgat cgcattcata ttttttctcag gttcacagca gatgaggctc  
3361 gggacctgat tcaacgttac ctgacagagc accctgacct caataatgaa aacatcgttg  
30 3421 gctataataa caagaagtgc tggccccgag atgcccgcag gcgcctcatg aaacatgatg  
3481 ttaacttagg ccgggcggtg ttctgggaca tcaagaaccg cttgccacgg tcagtgacta

3541 cagttcagtg ggagaacagc ttcgtgtctg tgtacagtaa ggacaacccc aacctgctgt  
3601 tcaacatgtg tggcttcgag tgccgcatcc tgcctaagtg cgcaccagc tatgaggagt  
3661 tcaccacaaa ggacggggtc tggaacctgc agaatgaggt tactaaggag cgcacagctc  
3721 agtgtttctt gcgtgtggac gatgagtcaa tgcagcgctt ccacaaccgc gtgcgtcaga  
5 3781 ttctcatggc ctctgggtcc accaccttca ccaagattgt gaataagtgg aatacagctc  
3841 tcattggcct tatgacatac tttcgggagg ctgtggtgaa cacccaagag ctcttgact  
3901 tactggtgaa gtgtgagcac aaaatccaga cacgtatcaa gattggactc aactccaaga  
3961 tgccaagtcg gttccccccg gttgtgttct acaccctaa ggagttgggt ggactcggca  
4021 tgctctcaat gggccatgtg ctcatcccc aatccgacct caggtggtcc aaacagacag  
10 4081 atgtaggtat cacacacttt cgttcaggaa tgagccatga agaagaccag ctcattecca  
4141 acttgtaccg ctacatacag ccatgggaga gcgagtcat tgattctcag cgggtctggg  
4201 ctgagtactc actcaagaga caagaggcca ttgctcagaa cagacgcctg actttaagaag  
4261 acctagaaga ttcattggat cgtggcattc ctgcaatcaa taccctcttc cagaaggacc  
4321 ggcacacact ggcttatgat aagggtggc gtgtcagaac tgactttaag cagtatcagg  
15 4381 ttttgaagca gaatccgttc tgggtggacac accagcggca tgatgggaag ctctggaacc  
4441 tgaacaacta ccgtacagac atgatccagg ccctgggagg tgtggaaggc attctggaac  
4501 acacactctt taagggcact tacttcccta cctgggaggg gcttttctgg gagaaggcca  
4561 gtggctttga ggaatctatg aagtggaaga agctaactaa tgctcagcga tcaggactga  
4621 accagattcc caatcgtaga ttcaccctct ggtggtcccc gaccattaat cgagccaatg  
20 4681 tatatgtagg ctttcagggtg cagctagacc tgacgggtat ctcatgcac ggcaagatcc  
4741 ccacgctgaa gatctctctc atccagatct tccgagctca cttgtggcag aagatccatg  
4801 agagcattgt tatggactta tgtcaggtgt ttgaccagga acttgatgca ctggaaattg  
4861 agacagtaca aaaggagaca atccatcccc gaaagtcata taagatgaac tcttcctgtg  
4921 cagatattct gctctttgcc tcctataagt ggaatgtctc ccggccctca ttgctggctg  
25 4981 actccaagga tgtgatggac agcaccacca occagaaata ctggattgac atccagttgc  
5041 gctgggggga ctatgattcc cacgacattg agcgctacgc ccgggccaag ttcttgact  
5101 acaccaccga caacatgagt atctaccctt cgccacagg tgtactcatc gccattgacc  
5161 tggcctataa cttgcacagt gcctatggaa actggttccc aggagcaag cctctcatac  
5221 aacaggccat ggccaagatc atgaaggcaa accctgccct gtatgtgta cgtgaacgga  
30 5281 tccgcaaggg gctacagctc tattcatctg aaccactga gccttatttg tcttctcaga  
5341 actatggtga gctcttctcc aaccagatta tctggtttgt ggatgacacc aacgtctaca

5401 gagtgactat tcacaagacc tttgaaggga acttgacaac caagcccatc aacggagcca  
5461 tcttcatctt caaccacgc acagggcagc tgttctcaa gataatccac acgtccgtgt  
5521 gggcgggaca gaagcgtttg gggcagttgg ctaagtggaa gacagctgag gaggtggccg  
5581 ccctgatccg atctctgcct gtggaggagc agcccaagca gatcattgtc accaggaagg  
5 5641 acatgctgga cccactggag gtgcacttac tggacttccc caatattgtc atcaaaggat  
5701 cggagctcca actccctttc caggcgtgtc tcaaggtgga aaaattcggg gatctcatcc  
5761 ttaaagccac tgagccccag atggttctct tcaacctcta tgacgactgg ctcaagacta  
5821 tttcatctta cacggccttc tcccgtctca toctgattct gcgtgcccta catgtgaaca  
5881 acgatcgggc aaaagtgatc ctgaagccag acaagactac tattacagaa ccacaccaca  
10 5941 tctggcccac tctgactgac gaagaatgga tcaaggtcga ggtgcagctc aaggatctga  
6001 tcttggtgta ctacggcaag aaaaacaatg tgaacgtggc atcactgaca caatcagaaa  
6061 ttcgagacat catcctgggt atggagatct cggcacctgc acagcagcgg cagcagatcg  
6121 ctgagatcga gaagcagacc aaggaacaat cgcagctgac ggcaacacag actcgcactg  
6181 tcaacaagca tggcgatgag atcatcacct ccaccaccag caactatgag acccagactt  
15 6241 tctcatccaa gactgagtgg agggtcaggg ccatctctgc tgccaacctg cacctaagga  
6301 ccaatcacat ctatgtttca tctgacgaca tcaaggagac tggctacacc tacatccttc  
6361 ccaagaatgt gcttaagaag ttcattctgca tatctgacct tcgggcccac attgcaggat  
6421 acctatatgg ggtgagccca ccagataacc ccaggtgaa ggagatccgc tgcattgtga  
6481 tgggtgccga gtggggcact caccagaccg tgcacctgcc tggccagctg cccagcatg  
20 6541 agtacctcaa ggagatggaa cccttaggtt ggatccacac tcagcccaat gagtccccgc  
6601 agttatcacc ccaggatgtc accaccatg ccaagatcat ggctgacaac ccatcttggg  
6661 atggcgagaa gaccattatc atcacatgca gcttcacgcc aggtcctgt aactgacgg  
6721 cctacaagct gaccctagt ggctacgaat ggggccgcca gaacacagac aagggcaaca  
6781 accccaaggg ctacctgcct tcacactatg agaggggtgca gatgctgctg tcggaccgtt  
25 6841 tccttggett ctatcatggtc cctgcccagt cctcgtggaa ctacaacttc atgggtgttc  
6901 ggcattgacc caacatgaaa tatgagctac agctggcgaa ccccaaagag ttctaccacg  
6961 aggtgcacag gccctctcac ttctcaact ttgctctcct gcaggagggg gaggtttact  
7021 ctgcgatcga ggaggacctg tatgcctgac cgtttcctg cctcctgctt cagcctcccg  
7081 aggccgaagc ctcagccctt ccagacaggc cgctgacatt cagcagtttg gcctctttcc  
30 7141 ctctgtctgt gcttgtgttg ttgacctcct gatggcttgt catcctgaat aaaatataat  
7201 aataaatttt gtataaatag g

## SEQ ID NO: 2 Protein sequence for BAA22563.1 (LBDG2)

1 magvfpyrgp gnpvpgplap lpdymseekl qekarkwqql qakryaekrk fgfvdaqked  
 5 61 mppehvreii rdhgdmtnrk frhdkrvylg alkymphavl kllenmpmpw eqirdvpvly  
 121 hitgaisfvn eipwviepvy isqwgsmwim mrrekrdrh fkrmr fppfd deeppldyad  
 181 nilnveplea iqleldpeed apvldwfydh qplrdsrkyv ngstyqrwqf tlpmmstlyr  
 241 lanqlltdlv ddnyfylfdl kafftskaln maipggpkfe plvr dinlqd edwnefn din  
 301 kiiirqpi rt eykiafpyly nnlphvhlt wyhtpnv vfi ktedpdlpaf yfdplinpis  
 10 361 hrhsvksqep lpdddeefel pefvepflkd tplytdntan giallwaprp fnlrsgrtrr  
 421 aldiplvknw yrehcpagqp vkvrsvyqkl lkyyvlnalk hrppkaqkkr ylf rfskatk  
 481 ffqstkldwv egwlqvcrqg ynmlnllhr knlnylhldy nfnlkpvktl ttkerksrf  
 541 gnafhlcrev lrltklvvds hvqyrlgnvd afqladglqy ifahvgqltg myrykyklmr  
 601 qirvckdlkh liyyrfntgp vqkgpgcgfw aagwrwv lff mrgitpller wlg nllarqf  
 15 661 egrhskgvak tvtkqrvesh fdlelraavm hdildmmp eg ikqnkartil qhlseawrcw  
 721 kanipkwkpg lptpienmil ryvkakadwv tntahynrer irrgatv dkt vckknlgrlt  
 781 rlylkae qer qhnylkdgpy itaetvavy tttvhwlesr rfs pipfppl sykhdtklli  
 841 lalerlkeay svksrlnqsq reelglieqa ydnhealsr ikrhl ltqra fkevgiefmd  
 901 lyshlvpyvd veplekitda yldqylwea dkrrlfppwi kpadtepppl lvykwccgin  
 20 961 nlqdvwetse gecnvmlesr fekmyekidl tllnrlvrli vdhniady mt aknnvinyk  
 1021 dmnhtnsygi irglqfasfi vqyyglvmdl lvlglhrase magppqmpnd flsfqdi ate  
 1081 aahpir lfc r yidrihiffr ftadeardli qryltehpdp nnenivgynn kkcwpr darm  
 1141 rlmkhdvnl g ravfwdiknr lprsvttvqw ensfsvvysk dnpnllfnmc gfecrilpkc  
 1201 rtsyee fthk dgwvnlqnev tkertaqcfl rvddesmqrf hnrvrqilma sgsttftkiv  
 25 1261 nkwntaligl mtyfreavn tqelldllvk cekiqtrik iglnskmpsr fppvvyf tpk  
 1321 elggglmsm ghvlpqsdl rwsqtdv gi thfrsgmshe edqlipnlyr yiqpwesefi  
 1381 dsqrwaeys lkrqeaiaqn rrltledled swdrqiprin tlfqkdrhtl aydkgwrvrt  
 1441 dfkqyqvlkq npfwthqrh dgklwnlnny rtdmiqalgg vegilehtlf kgtyfptweg  
 1501 lfwekasgfe esmkwkltn aqrsglnqip nrrftlwwsp tinranvyvg fqvqldltgi  
 30 1561 fmhgkiptlk isliqifrah lwqkihesiv mdlcqvf dqe ldaleietvq ketihprksy  
 1621 kmnsscadi l lfasykwnvs rpslladskd vmdstttqky widiqlrwgd ydshdierya  
 1681 rakfldy ttd nmsiypsptg vliaidlayn lhsaygnwfp gskpliqqam akimkanpal  
 1741 yvlrerirk g lqlyssepte pylssqnyge lfsnqiifv ddtvyrvti hktfegnltt  
 1801 kpingaifif nprtqqflk iihtsvwagq krlgqlakwk taeeva alir slpveeqpkq  
 35 1861 iivtrk dml d plevhlldfp nivikgselq lpfqaclkve kfgdlilkat epqmv lfnly  
 1921 ddwlktissy tafsr lili l ralhvndra kvilkpdktt itephhiwpt ltdeewikve  
 1981 vqlkdlilad ygkknvnva sltqseirdi ilgmeisaps qqrqqiaeie kqtkeqsqt  
 2041 atqtrtvnkh gdeitstts nyetqtfssk tewrvraisa anlhlrt nhi yvssddiket  
 2101 gytyilpknv lkkficsdl rraqiagylyg vsppdnqvk eircivmvpq wgtqtvhlp  
 40 2161 gqlpqheyk emeplgwiht qpnespqlsp qdvtthakim adnpswdgek tiiitcsftp  
 2221 gsctltaykl tpsgyewgrq ntdkg nnpkg ylpshyervq mllsdrflgf fmvpaqsswn  
 2281 ynfmgvrhdp nmkyelqlan pkefyhevhr pshflnfall qegevysadr edlya