US 20080254031A1

(19) **United States**

(12) **Patent Application Publication**

McGowan et al.

(10) **Pub. No.: US 2008/0254031 A1**

(43) **Pub. Date: Oct. 16, 2008**

(54) **TUMOR SPECIFIC GENES AND VARIANT RNAS AND USES THEREOF AS TARGETS FOR CANCER THERAPY AND DIAGNOSIS**

(75) Inventors: **Kevin McGowan**, N. Potomac, MD (US); **Vinayaka Kotraiah**, Germantown, MD (US); **Michael Brenner**, Fredrick, MD (US); **Richard Einstein**, Gaithersburg, MD (US); **Laurent Bracco**, La Garenne Colomba (FR)

Correspondence Address:
OCCHIUTI ROHLICEK & TSAO, LLP
10 FAWCETT STREET
CAMBRIDGE, MA 02138 (US)

(73) Assignee: **EXONHIT THERAPEUTICS SA**, Paris (FR)

(21) Appl. No.: **11/575,032**

(22) PCT Filed: **Sep. 7, 2005**

(86) PCT No.: **PCT/IB05/02934**

§ 371 (c)(1),
(2), (4) Date: **Jan. 10, 2008**

Related U.S. Application Data

(60) Provisional application No. 60/608,076, filed on Sep. 9, 2004.

Publication Classification

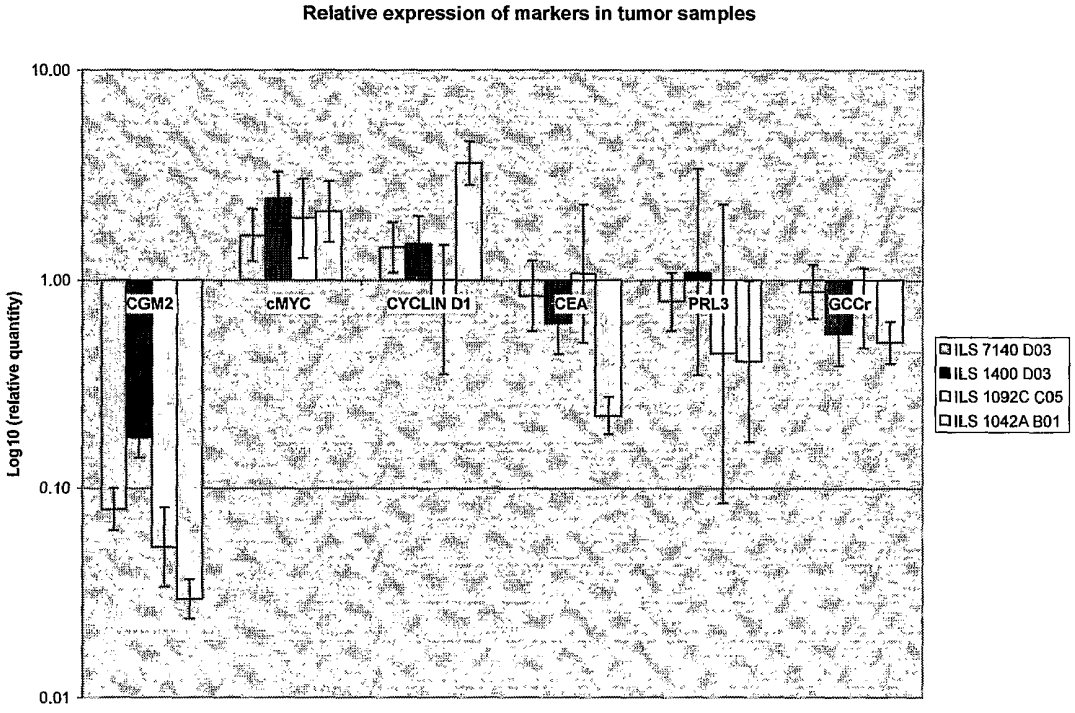
(51) **Int. Cl.**

A61K 39/395	(2006.01)
C07H 21/04	(2006.01)
C07K 14/47	(2006.01)
A61K 39/00	(2006.01)
C12Q 1/68	(2006.01)
G01N 33/574	(2006.01)
A61K 38/02	(2006.01)
A61P 35/00	(2006.01)
G01N 33/566	(2006.01)
G01N 33/53	(2006.01)
C07K 16/32	(2006.01)
C07K 19/00	(2006.01)
A61K 38/47	(2006.01)
A61K 31/7052	(2006.01)

(52) **U.S. Cl.** **424/138.1**; 536/23.5; 536/24.33; 530/300; 530/350; 424/185.1; 435/6; 435/7.23; 435/7.92; 435/7.93; 530/387.7; 530/391.3; 424/94.61; 514/44; 514/2; 436/501

(57) **ABSTRACT**

Genes and variant RNAs that are differentially expressed in human colon tumor tissues compared with normal colon tissue and the corresponding proteins are identified. These genes and the corresponding antigens are suitable targets for the treatment, diagnosis or prophylaxis of colon cancer.



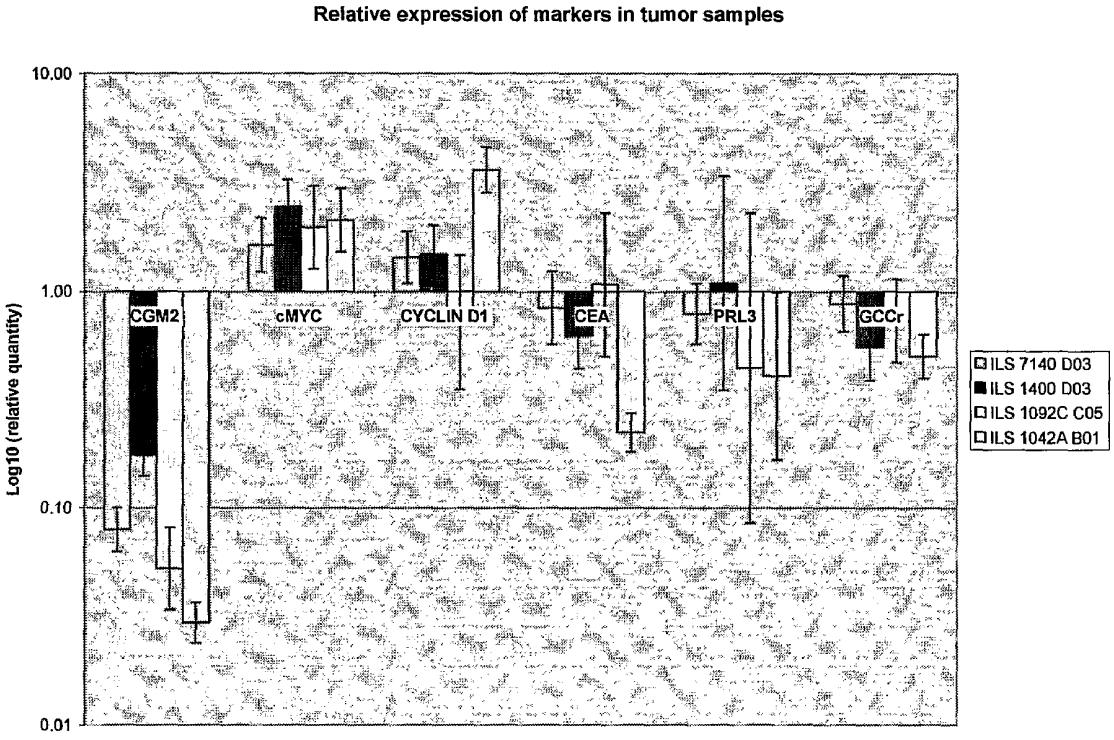

Relative expression of markers in tumor samples
10.00
1.00
0.10
0.01
Log10 (relative quantity)
CGM2
cMYC
CYCLIN D1
CEA
PRL3
GCCr
ILS 7140 D03
ILS 1400 D03
ILS 1092C C05
ILS 1042A B01

Figure 1

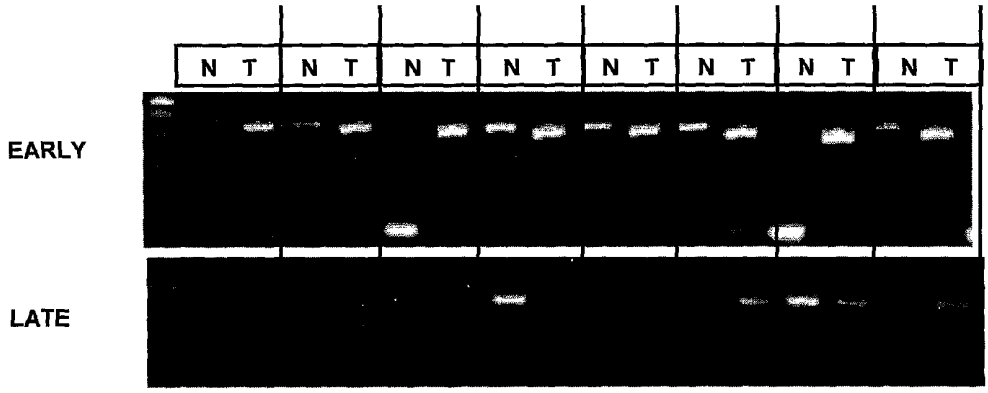


Figure 2

TUMOR SPECIFIC GENES AND VARIANT RNAS AND USES THEREOF AS TARGETS FOR CANCER THERAPY AND DIAGNOSIS

FIELD OF THE INVENTION

[0001] The present invention relates to the identification of nucleic acid sequences that correspond to alternatively spliced events in genes expressed from colon cancer cells. These genes or their corresponding proteins represent novel targets for the treatment, prevention and/or diagnosis of cancers wherein these genes are differentially regulated and/or spliced, particularly in colon cancer. The present invention also relates to compounds that specifically bind or modulate said targets, including antibodies, compositions comprising the same and their uses. The invention also provides novel products or constructs, including primers, probes, cells, chips and the like, for use in diagnostic or pharmacogenomic methods. The invention is suited for use in mammalians, particularly human subjects.

BACKGROUND OF THE INVENTION

[0002] Genetic detection of human disease states is a rapidly developing field (Taparowsky et al., 1982; Slamon et al., 1989; Sidransky et al., 1992; Miki et al., 1994; Dong et al., 1995; Morahan et al., 1996; Lifton, 1996; Barinaga, 1996). However, some problems exist with this approach. A number of known genetic lesions merely predispose an individual to the development of specific disease states. Individuals carrying the genetic lesion may not develop the disease state, while other individuals may develop the disease state without possessing a particular genetic lesion. In human cancers, genetic defects may potentially occur in a large number of known tumor suppresser genes and proto-oncogenes.

[0003] Genetic detection of cancer has a long history. Some of the earliest genetic lesions shown to predispose to cancer were transforming point mutations in the ras oncogenes (Taparowsky et al., 1982). Transforming ras point mutations may be detected in the stool of individuals with benign and malignant colorectal tumors (Sidransky et al., 1992). However, only 50% of such tumors contained a ras mutation (Sidransky et al., 1992). Similar results have been obtained with amplification of HER-2/neu in breast and colon cancer (Slamon et al., 1989), deletion and mutation of p53 in bladder cancer (Sidransky et al., 1991), deletion of DCC in colorectal cancer (Fearon et al., 1990) and mutation of BRCA1 in breast and colon cancer (Miki et al., 1994).

[0004] None of these genetic lesions are capable of predicting a majority of individuals with cancer and most require direct sampling of a suspected tumor, and make screening difficult. Further, none of the markers described above are capable of distinguishing between metastatic and non-metastatic forms of cancer. In effective management of cancer patients, identification of those individuals whose tumors have already metastasized or are likely to metastasize is critical. Because metastatic cancer kills 560,000 people in the U.S. each year (ACS home page), identification of markers for metastatic colon cancer would be an important advance.

[0005] Colon cancer is one of the most prevalent cancers affecting more than 147,500 new patients yearly in the US. Ten million FOBT tests are used as a screening test annually in the US. The other tests for colon cancer are invasive tests with the exception of predisposition tests. A non-invasive test that could detect early-stage CRC would represent a significant improvement from current screening devices. Additionally, a blood diagnostic test could also be utilized as a monitoring test to check for the recurrence of the disease, thereby increasing the number of tests performed yearly.

[0006] Colorectal cancer (CRC) (Midgley and Kerr, 1999) is a leading cause of morbidity and mortality with about 300,000 new cases and 200,000 deaths in Europe and the USA each year. It is the third most common cancer in men and women (American Cancer Society, 2002). In the USA, mortality rates have steadily declined among women since about 1950 and among men since approximately 1985 (American Cancer Society, 2002). The estimated five-year survival rate for patients diagnosed with early stage CRC is nearly 90%. Thus, early detection is a key component for managing the disease. Despite the proven effectiveness and availability of various colorectal cancer screening tests, many adults aged 50 or older are not regularly screened. Prevalence rates are especially low among individuals who are 50-64 years old, have lower incomes little or no health care coverage, and fewer years of education. As a consequence, only 37% of cases are diagnosed when the disease is still localised. Later diagnosis results in a substantially lower 5-year relative survival rate (64.4% for locally spread cancers, 8.3% for metastasised cancers) than would occur if patients were diagnosed when disease is still localized. Additionally, there is increased risk for recurrence of the disease in late-stage patients following surgery, as nearly 50% of patients believed to be cured by surgery will relapse and succumb to the disease.

[0007] Most colorectal cancers arise in the sigmoid colon (the portion just above the rectum). They usually start in the innermost layer and can grow through some or all of the several tissue layers that make up the colon and rectum. Most large bowel cancers arise within pre-existing adenomatous polyps or adenomas. These lesions are common. Necropsies have shown a prevalence of 35% in Europe and the USA with lower rates (10-15%) in Asia and Africa. Adenomas are classified by histological architecture as tubular, tubulovillous or villous. Villous change is associated with a higher malignant potential, as are large (up to 25% of adenomas are >1 cm in diameter) and high-grade epithelial dysplasia (severe dysplasia is found in 5-10% of adenomatous polyps). It is estimated that approximately 5% of adenomatous polyps will become malignant, a transformation that may take 5 to 10 years.

[0008] There is growing recognition that this adenoma-carcinoma sequence results from the interplay of environmental and genetic components. Genetic mutations are either inherited as germline defects or arise in somatic cells, secondary to environmental insults. There are two main inherited predisposition syndromes: Familial Adenomatous Polyposis (FAP) and Hereditary Non Polyposis Colorectal Cancer (HNPCC). Theses inherited predispositions for colorectal cancer share the same random pathway of progression form adenoma to carcinoma with the sporadic form, even if the progression rate and timescale of occurrence differ.

[0009] A multi-step model of progression of sporadic colorectal cancer has been proposed by Vogelstein et al (1988) which hypothesise that a combination of four or five mutations must accumulate in the cell, including activation of oncogenes and inactivation of tumour suppressor genes, to undergo full malignant transformation. This is consistent with the observation that colorectal cancers occur predominantly in the elderly. If one or more defects are present at

birth, less additional mutations will be necessary to occur and the disease will appear earlier.

[0010] The majority of CRCs are treated through surgical removal of the bowel. Traditionally, this has involved open resection using a laparotomy to enable both resection of the primary tumour with sufficient excision margins and an adequate, systematic lymphadenectomy. Additionally, for rectal cancer, a total mesorectal excision is performed to reduce the probability of local recurrence (Vogelstein et al., 1988). Excision of the tumour is the primary treatment for new CRC cases with potential for cure (80%). In the remaining 20%, the disease is too far advanced at presentation (either locally or at distant sites) for any curative intervention. These patients also frequently undergo surgery for palliation, where optimising quality of life is the main objective of treatment. In this setting, chemotherapy has an established role in improving survival and palliating symptoms. In addition, approximately 50% of those patients initially believed to be cured by surgery, subsequently relapse and die of their disease. Adjuvant chemotherapy administered for six months after surgery for Dukes C colon cancer improves absolute survival by 5-10% (Midgley and Kerr, 1999).

Classification	Dukes' A	Dukes' B	Dukes' C	Dukes' D
Cancer Progression	Cancer defined to most superficial cell layers of colon or rectum	Cancer may extend completely through wall of colon or rectum, no lymph node involvement	Cancer may extend completely through wall of colon or rectum, and has spread to lymph nodes	Metastatic disease. The cnacer has spread to distant organs, such as the liver.
Standard Treatment	Surgery	Surgery	Surgery and Chemotherapy (possibly radiation for rectal cancer	Surgery and Chemotherapy (possibly radiation for rectal cancer
Estimated 5-Year survival rate	95%	80%	50%	5%
Percent diagnosed at stage	37%		63%	

Table taken from the American Cancer Society, Cancer facts and figures, 2002

[0011] Assuming correct and early diagnosis, approximately 90% of all colorectal cancer cases are thought to be curable (American Cancer Society, 2002). To improve the likelihood for early detection various screening programs have been recommended for both the general and high-risk populations (Midgley and Kerr, 1999). The recommendation of the American Cancer Society for CRC screening is as follows (Smith et al., 2001):

[0012] 1 low risk patients;

[0013] an annual faecal occult blood test (FOBT) and sigmoidoscopy (FSIG) every five years starting at age 50 or

[0014] colonoscopy and digital rectal examination (DRE) every 10 years or

[0015] double-contrast barium enema and digital rectal examination every 5-10 years

[0016] 2 High risk patients (family history of CRC or polyps or chronic inflammatory bowel disease) screening should begin earlier and be more frequent.

[0017] The specifics of each test are described below. Procedures such as sigmoidoscopy and colonoscopy are utilized as both screening tests and diagnostic tests. For example, colonoscopy is performed routinely on healthy patients over age 50 as a screening test for early detection of colon cancer. However, it is also used for diagnostic purposes when patients present with clear symptoms of a bowel disorder such as blood in stool, positive FOBT test, and/or excessive cramping. It is important to note that positive FOBT can result from a number of factors outside of colon cancer.

[0018] 1 Faecal Occult Blood Test (FOBT). The FOBT detects the presence of blood in stool, derived from colorectal cancer or large (>2 cm) polyps. Individual receive a kit to take home along with dietary instructions. FOBT consists of six small stool samples each taken from three consecutive bowel movements. Upon completing the test, patients return the kit to the physician for evaluation. The test is not specific for colon cancer as other conditions such as ulcers, colitis, hemorrhoids also bleed. A positive FOBT will require further diagnostic work-up such as colonoscopy or double-contrast barium enema. Several studies have proven that the regular use of this screening method saves life and can reduce the incidence of colorectal cancer by diagnosing and removing earlier precancerous lesions (Mandel et al., 1993; Mandel et al., 2000; Kronborg et al., 1996; Hardcastle et al., 1996).

[0019] 2 Flexible Sigmoidoscopy (FSIG). A slender, flexible, hollow, lighted tube is inserted through the rectum into the colon to search for cancer or polyps. The sigmoidoscope is around 2 feet long and, at its maximum of insertion, can only reach about half of the colon. If there is a polyp or tumour present, the patient must be referred for colonoscopy so that the entire colon can be examined. The advantage of FSIG over FOBT is that it allows the examiner to visualize the distal bowel directly and it has higher sensitivity and specificity for both adenocarcinomas and polyps. The limitation of the FSIG is that the instrument cannot visualize the entire length of the colon. Removal of polyps is rare during this procedure as the bowel is not sufficiently prepared for surgery. Additionally, the presence of polyps in the sigmoid colon is usually an indication of polyps or cancer in the proximal bowel.

[0020] 3 Colonoscopy. This procedure allows for direct visual examination of the entire colon as well as surgical removal of the polyps. The procedure is somewhat uncomfortable and may not always be covered under health care plans.

[0021] 4 Barium Enema with air contrast (Double-contrast barium enema). Barium sulphate is introduced into the

colon and allowed to spread to partially fill and open up the colon. The colon is then filled with air so that it can expand and increase the quality of x-rays that are taken. It is superior to FSIG in that it detects small polyps (<9 mm) with accuracy and examines the entire bowel. The test can be uncomfortable especially when done as a double-contrast procedure in which air is introduced into the bowel for contrast.

[0022] There is a need in the art for genetic markers and targets of colon cancers, allowing the design of specific, reliable and sensitive diagnostic and therapeutic approaches of these diseases.

SUMMARY OF THE INVENTION

[0023] The present invention relates to the identification of novel nucleic acid and amino acid sequences that are characteristic of colon cancer cells or tissues, and which represent targets for therapy or diagnosis of such a condition in a subject.

[0024] The invention more specifically discloses 60 specific, isolated nucleic acid molecules that encode expression sequences found to be differentially expressed in colon cancer. Of these, 51 are expressed sequence tags that are differentially spliced and correspond to SEQ ID NOS 1-44, 52, 56, 62, 73, 79, 85, and 91. In addition, 9 specific isoforms of known genes have been identified corresponding to SEQ ID NOS. 47, 53, 57, 65, 70, 76, 82, 88, and 96. These novel sequences were found to be differentially expressed between normal colon and colon cancer. The expressed sequence tag represent novel exons that are alternatively spliced in colon cancer, and as such, directly identify distinct isoforms. These sequences and molecules represent targets and valuable information to develop methods and materials for the detection, diagnosis, and treatment of colon cancer. Furthermore, since deregulations of RNA splicing have been observed in distinct types of cancers, and because said deregulations constitute a mechanism by which response to chemotherapy may be altered, the presently characterized nucleic acids and polypeptides may also represent target molecules suitable for other cancers as well.

[0025] It is thus an object of the invention to provide methods and materials for treatment and diagnosis of cancer, particularly colon cancer.

[0026] In particular, an object of this invention resides in nucleic acids and amino acids, which are differentially regulated in colon cancer cells. More particularly, an object of this invention resides in isolated nucleic acids that are expressed by human cancer cells, particularly colon cancer cells, selected from the group consisting of:

[0027] (i) nucleic acids comprising a sequence contained in SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96;

[0028] (ii) a nucleic acid having a sequence that is at least 70% identical to the sequence of (i) when aligned without allowing for gaps;

[0029] (iii) nucleic acids having a sequence complementary to (i) or (ii); and

[0030] (iv) fragments of (i), (ii) or (iii) having a size of at least 20 nucleotides in length.

[0031] A further object of this invention resides in any polypeptide (or antigen) encoded by a nucleic acid as defined above. More particularly, the invention relates to polypeptides expressed by human cancer cells, selected from the group consisting of:

[0032] (i) the polypeptide encoded by a nucleic acid sequence having at least 90% sequence identity in SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96; or a sequence complementary thereto; and

[0033] (ii) the polypeptide comprising an amino acid sequence having at least 90% sequence identity in SEQ ID NOS. 48, 49, 54, 55, 58, 59, 66, 67, 71, 72, 75, 77, 78, 83, 84, 89, 90, 97, 98 and 99; and

[0034] (iii) an antigenic fragment of (i) or (ii).

[0035] Another object of the invention is to provide novel methods for diagnosis or detection of cancer, particularly colon cancer by using ligands (e.g., monoclonal antibodies, probes, etc.) which specifically bind to a target molecule (i.e., polypeptide or nucleic acid) as defined above. Such methods may be used to detect whether a subject has or is at (increased) risk of developing a cancer, particularly colon cancer or, for instance, whether a treatment regimen is efficient.

[0036] In this respect, a particular object of the invention resides in methods of detecting persons having, or at (increased) risk of developing a cancer, particularly colon cancer, by use of labeled nucleic acid probes that hybridize to a target gene or nucleic acid as defined in the present application.

[0037] According to an other embodiment of the invention, the methods of detecting persons having, or at (increased) risk of developing cancer, particularly colon cancer, use a (labeled) antibody or fragment/derivative thereof that specifically binds a target polypeptide as defined in the present application.

[0038] A further object of this invention relates to diagnostic test kits for the detection of persons having or at (increased) risk of developing cancer, particularly colon cancer, that comprise a ligand that specifically binds to a target molecule as defined above and, optionally, a detectable label, e.g. indicator enzymes, a radiolabels, fluorophores, or paramagnetic particles. In a particular embodiment, the ligand comprises nucleic acid primers or probes specific for target genes or nucleic acids as described above, or an antibody or a derivative thereof, specific for a target polypeptide as described in this application.

[0039] A further aspect of this invention resides in the development of novel therapies for treatment of cancer, particularly colon cancer, involving the administration of an inhibitor of a target molecule as defined in the present application. In a particular embodiment, the method comprises administering an inhibitory nucleic acid (e.g., anti-sense oligonucleotide, ribozyme, iRNA, siRNA or a DNA encoding the same) corresponding to (i.e., complementary and specific for) a target nucleic acid as described herein, thereby inhibiting (e.g., reducing) expression or translation thereof. In an other embodiment, the method comprises administering an antibody that specifically binds a target polypeptide as described herein.

[0040] A further object of this invention relates to methods of treating cancer, particularly colon cancer, in a subject, comprising the administration of a polypeptide antigen as described herein, alone or in combination with adjuvants that elicit an antigen-specific cytotoxic T-cell lymphocyte response against cancer cells that express such antigen.

[0041] It is another object of this invention to provide methods for selecting, identifying, screening, characterizing or optimizing biologically active compounds, comprising a determination of whether a candidate compound binds, pref-

erably selectively, an antigen or a polynucleotide as disclosed in the present application. Such compounds represent drug candidates or leads for treating cancer diseases, particularly colon cancer.

[0042] A further object of this invention resides in a method of producing or selecting ligands that bind a target molecule as described herein, comprising contacting a candidate compound with a target molecule and determining the ability of such compound to bind said target. The method is particularly suited for selecting or producing ligands of an extra-cellular domain of a polypeptide (antigen) encoded by a gene or exon expressed by certain cancers.

[0043] It is another object of the invention to identify genes that are expressed in altered forms in colon cancer cells. These forms represent splice variants of the gene, where the Expressed Sequence Tag either 1) indicates the splice event occurring within the gene, or 2) points to a gene that is actively spliced to produce different gene products. These different splice variants or isoforms can be targets for therapeutic intervention.

LEGEND TO THE FIGURES

[0044] FIG. **1**. Relative expression of colon cancer markers in four colon tumor samples. RNA was purified from the tissue sample and assessed for the level of CGM2, cMyc, cyclin D1, CEA, PRL3, and GCCr. Values are plotted as the fold change from normal samples.

[0045] FIG. **2**. Examples of Expression profile results. Oligonucleotide primers were designed to the novel event identified experimentally. Total RNA was isolated from matched samples for early stage colon tissue (normal and tumor) and late stage cancer. cDNA was prepared and end point RT-PCR analysis was performed to determine the level of expression for the event. cDNA's were normalized for GAPDH prior to the analysis. A) SEQ ID NO: 28; B) SEQ ID NO: 74.

[0046] Table 1. Markers for CRC used to evaluate tissue samples. The table consists of genes that are well known in colon cancer to have differential expression in tumor samples versus normal colon tissue. These genes were chosen to qualify the samples used to generate the DATAS™ libraries.

[0047] Table 2. Expression profiles for Expressed Sequence Tags in Colon Cancer. DATAS™, a differential expression analysis lead to the identification of Expressed Sequence Tags (EST's) that had the potential for serving as biomarkers in colon cancer. Primers were designed to detect the expression level of each EST by RT-PCR in two sets of samples derived from patients with either early or late stage colon cancer. Sequences are identified by the SEQ ID NO, the internal accession number, the GenBank accession number of the gene that is alternatively spliced, and the type of alternative splicing event: novel indicates the sequence suggested a novel exon present in the gene, extension indicates that a known exon from the gene contains additional sequence derived from the intron, leading to an extended exon; amplicon indicates that the sequence was derived by amplifying bioinformatically identified candidates and the detection of a novel splicing event. Expression was scored by the count of samples that were up or down regulated in cancer samples vs normal samples and expressed as a decimal (5.10) where the ones place (5) indicates the number of samples up-regulated, and the decimal (10) indicates the number of samples down regulated.

DETAILED DESCRIPTION OF THE INVENTION

[0048] The present invention relates to novel target molecules suitable for monitoring, treating or developing cancer therapies, particularly colon cancer therapies.

[0049] The deregulation of RNA splicing in human disease is well documented and is supported by an exponentially increasing number of scientific publications. At least ten percent of germline mutations underlying human inherited disease affect RNA splicing, underlining the importance of this process in the development of disease. In addition, it has been estimated that 50% of all human genes undergo alternative splicing (Modrek et al., 2001; Kan et al., 2001). As examples of deregulation, there are isoforms that are specifically expressed in tumours (Obermair et al, 2001; Berggren et al., 2001; Milech et al., 2001; Lucas et al., 2001). In breast cancer, there is significant deregulation of splicing in the estrogen receptor alpha; normal breast tissue primarily expresses only a single variant, while breast tumours have an increased frequency of isoforms with multiple exon deletions (Poola and Speirs, 2001). Furthermore, deregulation of RNA splicing constitutes a mechanism by which response to chemotherapy may be altered. Alternative splicing profoundly affects normal biology, and when altered in a variety of systems can lead to disease.

[0050] DATAS™ (Different Analysis of Transcripts with Alternative Splicing) analyzes structural differences between expressed genes and provides systematic access to alterations in RNA splicing (disclosed in U.S. Pat. No. 6,251,590, the disclosure of which is incorporated by reference in its entirety). Having access to these spliced sequences, which are critical for the cellular homeostasis, represents a useful advance in functional genomics.

[0051] The DATAS™ Technology typically generates two libraries when comparing two samples, such as normal vs. tumor tissue. Each library specifically contains clones of sequences that are present and likely to be more highly expressed in one sample. For example, library A will contain sequences that are present in genes in the normal samples but absent (or expressed at lower levels) in the tumor samples. These sequences are identified as being removed or spliced out from the genes in the tumor samples. In contrast, library B will contain sequences that are present more abundantly and at higher concentrations in the tumor samples as compared to the normal samples. These represent exons/introns that are alternatively spliced into genes expressed predominantly in the tumor samples.

[0052] The present invention is based in part on the identification of exons that are isolated using DATAS™ and then determined to be differentially regulated or expressed in colon tumor samples. Specifically, 51 expressed sequences were identified through DATAS™ and confirmed to be differentially expressed between normal colon tissue and colon tumor tissue. These DATAS™ fragments (DF) are small sections of genes that are selected for inclusion or exclusion in one sample but not the other. These small sections are part of the expressed gene transcript, and can consist of sequences derived from several different regions of the gene, including, but not limited to, portions of single exons, several exons, sequence from introns, and sequences from exons and introns. This alternative usage of exons in different biological

samples produces different gene products from the same gene through a process well known in the art as alternative RNA splicing. In the present application, 60 alternatively spliced isoforms have been identified from the DATAS™ fragment sequences, which produce alternate gene products that fit all the descriptions of target molecules as disclosed below.

[0053] Alternatively spliced mRNA's produced from the same gene contain different ribonucleotide sequence, and therefore translate into proteins with different amino acid sequences. Sequences that are alternatively spliced into or out of the gene products can be inserted or deleted in frame or out of frame from the original gene sequence. This leads to the translation of different proteins from each variant. Differences can include simple sequence deletions, or novel sequence information inserted into the gene product. Sequences inserted out of frame can lead to the production of an early stop codon and produce a truncated form of the protein. Many variations have been identified and produce protein variants that can be agonistic or antagonistic with the original biological activity of the protein.

[0054] The present invention thus identifies genes and proteins which are subject to differential regulation and alternative splicing(s) in colon cancer cells. The present invention thus provides target molecules suitable for diagnosis or therapy of colon cancers, which target molecules comprise all or a portion of genes or RNAs comprising the sequence of a DATAS™ fragment, or of genes or RNA from which the sequence of a DATAS™ fragment derives, as well as corresponding polypeptides or proteins, and variants thereof. These molecules also represent targets for diagnosis or therapy of other types of cancers, particularly those sharing the same type of deregulations as presently identified.

[0055] A first type of target molecule is a target nucleic acid molecule. Preferred target nucleic acid molecules comprise the sequence of a full gene or RNA molecule comprising the sequence of a DATAS™ fragment as disclosed in the present application, or a sequence complementary thereto. Indeed, since DATAS™ identifies genetic deregulations associated with colon tumor, the whole gene or RNA sequence from which said DATAS™ fragment derives can be used as a target of therapeutic intervention or diagnosis.

[0056] Additional target nucleic acid molecules comprise a fragment of a gene or RNA as disclosed above. Indeed, since DATAS™ identifies genes and RNAs that are altered in colon tumor cells, portions of such genes or RNAs, including portions that do not comprise the sequence of a DATAS™ fragment, can be used as a target for therapeutic intervention or diagnosis. Examples of such portions include: DATAS™ fragments, portions thereof, alternative exons or introns of said gene or RNA, junction sequences generated by exon splicing in said RNA, etc. Particular portions comprise a sequence encoding an extra-cellular domain of a polypeptide.

[0057] In this respect, a particular object of this invention resides in a nucleic acid molecule selected from the group consisting of:

[0058] (i) nucleic acids comprising a sequence contained in SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96 or a sequence complementary thereto;

[0059] (ii) a nucleic acid having a sequence that is at least 70% identical to the sequence of (i) when aligned without allowing for gaps; and

[0060] (iii) fragments of (i) or (ii) having a size of at least 20 nucleotides in length.

[0061] Preferred fragments encode alternative exons or introns, junction sequences generated by exon splicing, or an extra-cellular domain of a polypeptide.

[0062] A second type of target molecule is represented by target polypeptides. In this regard, preferred target polypeptides comprise the sequence of a full-length protein comprising the amino acid sequence encoded by a DATAST fragment as disclosed in the present application or the corresponding whole gene or RNA.

[0063] Other target polypeptides of this invention are fragments of a protein as defined above. Such fragments may comprise or not the DATAS™ sequence, and may comprise newly generated amino acid sequence, resulting from a frame shift, the creation of new stop codon, etc.

[0064] A particular object of this invention resides more specifically in a polypeptide (or antigen) selected from the group consisting of:

[0065] (i) the polypeptide encoded by a nucleic acid sequence having at least 90% sequence identity in SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96, or a sequence complementary thereto; and

[0066] (ii) the polypeptide comprising an amino acid sequence having at least 90% sequence identity in SEQ ID NOS. 48, 49, 54, 55, 58, 59, 66, 67, 71, 72, 75, 77, 78, 83, 84, 89, 90, 97, 98 and 99; and

[0067] (iii) an antigenic fragment of (i) or (ii).

[0068] These target molecules (including genes, fragments, proteins and their variants) can serve as diagnostic agents and as targets for the development of therapeutics. For example, these therapeutics may modulate biological processes associated with (colon) tumor formation, viability and/or growth. Agents may also be identified that are associated with the induction of apoptosis (cell death) in colon tumor cells. Other agents can also be developed, such as monoclonal antibodies, that bind to the protein or its variant and alter the biological processes important for cell growth. Alternatively, antibodies can deliver a toxin which can inhibit cell growth and lead to cell death.

[0069] Specifically, the invention provides sequences that are expressed in a variant protein and are colon tumor specific or colon specific. These sequences are portions of proteins identified to be in the plasma membrane of the cell, and the specific sequences of the invention are expressed on the extracellular region of the protein, so that the sequences may be useful in the preparation of (colon) tumor vaccines, including prophylatic and therapeutic vaccines.

[0070] Based thereon, it is anticipated that the disclosed genes that are associated with the differentially expressed sequences and the corresponding variant proteins represent suitable targets for cancer therapy, prevention or diagnosis, e.g. for the development of antibodies, small molecular inhibitors, inhibitory nucleic acids (e.g., anti-sense therapeutics, ribozymes, interfering RNAs, etc.), particularly for colon cancer. The potential therapies are described in greater detail below.

[0071] Inhibitory nucleic acids of this invention include oligonucleotides having sequences in the antisense orientation relative to the subject nucleic acids which appear to be unregulated in colon cancer. Suitable therapeutic inhibitory oligonucleotides typically vary in length from five to several hundred nucleotides, more typically about 20-70 nucleotides in length or shorter. These inhibitory oligonucleotides may be administered as naked nucleic acids or in protected forms,

e.g., encapsulated in liposomes. The use of liposomal or other protected forms may be advantageous as it may enhance in vivo stability and thus facilitate delivery to target sites, e.g., colon tumor cells.

[0072] Also, the subject target genes may be used to design novel ribozymes that target the cleavage of the corresponding mRNAs in tumor cells. Similarly, these ribozymes may be administered in free (naked) form or by the use of delivery systems that enhance stability and/or targeting, e.g., liposomes.

[0073] Also, the subject target genes may be used to design novel siRNAs that can inhibit (e.g., reduce) expression of a target nucleic acid as disclosed in the present application. Similarly, these siRNAs may be administered in free (naked) form or by the use of delivery systems that enhance stability and/or targeting, e.g., liposomes. They may also be administered in the form of their precursors or encoding DNAs.

[0074] Also, the present invention embraces the administration of a ligand of a target molecule of this invention (e.g., a nucleic acid that hybridizes to the novel target nucleic acids identified infra or an antibody that specifically binds a target polypeptide as disclosed above), attached to therapeutic effector moieties, e.g., radiolabels (e.g., ^{90}Y, ^{131}I), cytotoxins, cytotoxic enzymes, and the like in order to selectively target and kill cells that express these targets, e.g., colon tumor cells.

[0075] Also, the present invention embraces the treatment and/or diagnosis of cancer by targeting altered genes or the corresponding altered protein, particularly splice variants that are expressed in altered form in colon tumor cells, as described above. These methods provide for the selective detection of cells and/or eradication of cells that express such altered forms thereby minimizing adverse effects to normal cells.

[0076] Still further, the present invention encompasses other nucleic acid based therapies. For example, the invention encompasses the use of a DNA containing one of the novel cDNAs corresponding to novel antigen identified herein. It is anticipated that the antigens so encoded may be used as therapeutic or prophylactic anti-tumor vaccines. For example, a particular contemplated application of these antigens involves their administration with adjuvants that induce a cytotoxic T lymphocyte response.

[0077] Administration of the subject novel antigens in combination with an adjuvant may result in a humoral immune response against such antigens, thereby delaying or preventing the development of cancer.

[0078] These embodiments of the invention comprise, for instance, administration of one or more of the target polypeptides of this invention, or antigenic fragments thereof, typically in combination with an adjuvant. Such compositions shall be administered in an amount sufficient to be therapeutically or prophylactically effective, e.g. on the order of 50 to 20,000 mg/kg body weight, 100 to 5000 mg/kg body weight. Suitable adjuvants for use in the present invention include PROVAX™, which comprises a microfluidized adjuvant containing Squalene, Tween and Pluronic, ISCOM'S®, DETOX®, SAF, Freund's adjuvant, Alum® and Saponin®, among others.

[0079] Yet another embodiment of the invention comprises the preparation of monoclonal antibodies against a target polypeptide s defined above. Such monoclonal antibodies may be produced by conventional methods and include fragments or derivatives thereof, including, without limitation, human monoclonal antibodies, humanized monoclonal antibodies, chimeric monoclonal antibodies, single chain antibodies, e.g., scFv's and antigen-binding antibody fragments such as Fab and Fab' fragments. Methods for the preparation of monoclonal antibodies are known in the art. In general, the preparation of monoclonal antibodies comprises immunization of an appropriate (non-homologous) host with the subject colon cancer antigens, isolation of immune cells therefrom, use of such immune cells to isolate monoclonal antibodies and screening for monoclonal antibodies that specifically bind to either of such antigens. Antibody fragments may be prepared by known methods, e.g., enzymatic cleavage of monoclonal antibodies.

[0080] These monoclonal antibodies and fragments are useful for passive anti-tumor immunotherapy, or may be attached to therapeutic effector moieties, e.g., radiolabels, cytotoxins, therapeutic enzymes, agents that induce apoptosis, and the like in order to provide for targeted cytotoxicity, i.e., killing of human colon tumor cells. Given the fact that the subject genes are apparently not significantly expressed by many normal tissues this should not result in significant adverse side effects (toxicity to non-target tissues).

[0081] In one embodiment of the present invention, such antibodies or fragments are administered in labeled or unlabeled form, alone or in conjunction with other therapeutics, e.g., chemotherapeutics such as cisplatin, methotrexate, adriamycin, and the like suitable for cancer therapy. The administered composition typically includes a pharmaceutically acceptable carrier, and optionally adjuvants, stabilizers, etc., used in antibody compositions for therapeutic use.

[0082] Preferably, the subject monoclonal antibodies binds the target antigens with high affinity, e.g., possess a binding affinity (Kd) on the order of 10^{-6} to 10^{-12} M.

[0083] The present invention also embraces diagnostic applications that provide for detection of the expression of colon specific splice variants disclosed herein. This comprises detecting the expression of one or more of these genes at the RNA level and/or at the protein level.

[0084] In this respect, a particular object of this invention resides in methods of detecting and/or staging cancer, particularly colon cancer, comprising (i) obtaining a human cell sample, particularly a human colon cell sample; and (ii) determining whether such cell sample expresses a target molecule, wherein said target molecule comprises the sequence of a gene or RNA comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96; a sequence complementary thereto, or of a fragment of said gene or RNA having a size of at least 20 nucleotides in length, or an amino acid sequence encoded by such a nucleic acid. Determination of expression may comprise quantitative and/or qualitative evaluations, e.g., absolute and/or relative measure of such expression levels. Typically, the expression level of said target molecule in said cell sample is compared to a reference expression level, wherein a deviation from said reference expression level is indicative of the presence and/or stage of said cancer in said subject. The reference expression level may be an expression level as determined in a control sample (e.g., from a healthy tissue or subject) or a median expression level from healthy subjects. A "deviation" from said reference expression level designates any significant change, such as an increase or decrease by at least 10%, 20%, or 30%, preferably by at least 40% or 50%, or even more.

[0085] For nucleic acids, expression of the subject genes can be detected by known nucleic acid detection methods,

e.g., Northern blot hybridization, strand displacement amplification (SDA), catalytic hybridization amplification (CHA), and other known nucleic acid detection methods. Preferably, a cDNA library will be made from colon cells obtained from a subject to be tested for colon cancer by PCR using primers corresponding to the novel isoforms disclosed in this application.

[0086] The presence or absence of a cancer can be determined based on whether PCR products are obtained, and the level of expression. The levels of expression of such PCR product may be quantified in order to determine the prognosis of a particular cancer patient, particularly a colon cancer patient (as the levels of expression of the PCR product often will increase or decrease significantly as the disease progresses.) This may provide a method for monitoring the status of a cancer patient.

[0087] Alternatively, the status of a subject to be tested for cancer may be evaluated by testing biological fluids (e.g., blood, urine, lymph), bodily excretions (e.g. fecal matter), exfoliated colonocytes, and the like with an antibody or antibodies or fragment that specifically binds to the novel tumor antigens disclosed herein.

[0088] Methods for using antibodies to detect antigen expression are well known and include ELISA, competitive binding assays, and the like. In general, such assays use an antibody or antibody fragment that specifically binds the target antigen directly or indirectly bound to a label that provides for detection, e.g. indicator enzymes, a radiolabels, fluorophores, or paramagnetic particles.

[0089] Patients which test positive for the enhanced presence of the antigen on cancer cells will be diagnosed as having or being at increased risk of developing cancer. Additionally, the levels of antigen expression may be useful in determining patient status, i.e., how far disease has advanced (stage of cancer).

[0090] As noted, the present invention provides novel splice variants that encode antigens that correlate to human cancer. The present invention also embraces variants thereof. As used herein "variants" means sequences that are at least about 75% identical thereto, more preferably at least about 85% identical, and most preferably at least 90% identical and still more preferably at least about 95-99% identified when these DNA sequences are compared to a nucleic acid sequence encoding the subject DNAs or a fragment thereof having a size of at least about 50 nucleotides. This includes allelic and splice variants of the subject genes. The present invention also encompasses nucleic acid sequences that hybridize to the subject splice variants under high, moderate or low stringency conditions e.g., as described infra.

[0091] Also, the present invention provides for primer pairs that result in the amplification of DNAs encoding the subject novel genes or a portion thereof in an mRNA library obtained from a desired cell source, typically human colon cell or tissue sample. Typically, such primers will be on the order of 12 to 50 nucleotides in length, and will be constructed such that they provide for amplification of the entire or most of the target gene.

[0092] Also, the invention embraces the antigens encoded by the subject DNAs or fragments thereof that bind to or elicits antibodies specific to the full-length antigens. Typically, such fragments will be at least 10 amino acids in length, more typically at least 25 amino acids in length.

[0093] As noted, the subject DNA fragments are expressed in a majority of colon tumor samples tested. The invention further contemplates the identification of other cancers that express such genes and the use thereof to detect and treat such cancers. For example, the subject DNA fragments or variants thereof may be expressed on other cancers, e.g., breast, ovary, pancreas, lung or colon cancers. Essentially, the present invention embraces the detection of any cancer wherein the expression of the subject novel genes or variants thereof correlate to a cancer or an increased likelihood of cancer. To facilitate under-study of the invention, the following definitions are provided.

[0094] "Isolated tumor antigen or tumor protein" refers to any protein that is not in its normal cellular milieu. This includes by way of example compositions comprising recombinant proteins encoded by the genes disclosed infra, pharmaceutical compositions comprising such purified proteins, diagnostic compositions comprising such purified proteins, and isolated protein compositions comprising such proteins. In preferred embodiments, an isolated colon tumor protein according to the invention will comprise a substantially pure protein, in that it is substantially free of other proteins, preferably that is at least 90% pure, that comprises the amino acid sequence contained herein or natural homologues or mutants having essentially the same sequence. A naturally occurring mutant might be found, for instance, in tumor cells expressing a gene encoding a mutated protein according to the invention.

[0095] "Native tumor antigen or tumor protein" refers to a protein that is a non-human primate homologue of the protein having the amino acid sequence contained infra.

[0096] "Isolated colon tumor gene or nucleic acid sequence" refers to a nucleic acid molecule that encodes a tumor antigen according to the invention which is not in its normal human cellular milieu, e.g., is not comprised in the human or non-human primate chromosomal DNA. This includes by way of example vectors that comprise a gene according to the invention, a probe that comprises a gene according to the invention, and a nucleic acid sequence directly or indirectly attached to a detectable moiety, e.g. a fluorescent or radioactive label, or a DNA fusion that comprises a nucleic acid molecule encoding a gene according to the invention fused at its 5' or 3' end to a different DNA, e.g. a promoter or a DNA encoding a detectable marker or effector moiety. Also included are natural homologues or mutants having substantially the same sequence. Naturally occurring homologies that are degenerate would encode the same protein including nucleotide differences that do not change the corresponding amino acid sequence. Naturally occurring mutants might be found in tumor cells, wherein such nucleotide differences may result in a mutant tumor antigen. Naturally occurring homologues containing conservative substitutions are also encompassed.

[0097] "Variant of colon tumor antigen or tumor protein" refers to a protein possessing an amino acid sequence that possess at least 90% sequence identity, more preferably at least 91% sequence identity, even more preferably at least 92% sequence identity, still more preferably at least 93% sequence identity, still more preferably at least 94% sequence identity, even more preferably at least 95% sequence identity, still more preferably at least 96% sequence identity, even more preferably at least 97% sequence identity, still more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity, to the corresponding native tumor antigen wherein sequence identity is as defined infra. Preferably, this variant will possess at least one biological property in common with the native protein.

[0098] "Variant of colon tumor gene or nucleic acid molecule or sequence" refers to a nucleic acid sequence that possesses at least 90% sequence identity, more preferably at least 91%, more preferably at least 92%, even more preferably at least 93%, still more preferably at least 94%, even more preferably at least 95%, still more preferably at least 96%, even more preferably at least 97%, even more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity, to the corresponding native human nucleic acid sequence, wherein "sequence identity" is as defined infra.

[0099] "Fragment of colon antigen encoding nucleic acid molecule or sequence" refers to a nucleic acid sequence corresponding to a portion of the native human gene wherein said portion is at least about 50 nucleotides in length, or 100, more preferably at least 150 nucleotides in length.

[0100] "Antigenic fragments of colon tumor antigen" refer to polypeptides corresponding to a fragment of a colon protein or a variant or homologue thereof that when used itself or attached to an immunogenic carrier elicits antibodies that specifically bind the protein. Typically such antigenic fragments will be at least 8-15 amino acids in length, and may be much longer.

[0101] Sequence identity or percent identity is intended to mean the percentage of the same residues shared between two sequences, referenced to human protein A or protein B or gene A or gene B, when the two sequences are aligned using the Clustal method [Higgins et al, Cabios 8:189-191 (1992)] of multiple sequence alignment in the Lasergene biocomputing software (DNASTAR, INC, Madison, Wis.), or alignment programs available from the Genetics Computer Group (GCG Wisconsin package, Accelrys, San Diego, Calif.). In this method, multiple alignments are carried out in a progressive manner, in which larger and larger alignment groups are assembled using similarity scores calculated from a series of pairwise alignments. Optimal sequence alignments are obtained by finding the maximum alignment score, which is the average of all scores between the separate residues in the alignment, determined from a residue weight table representing the probability of a given amino acid change occurring in two related proteins over a given evolutionary interval. Penalties for opening and lengthening gaps in the alignment contribute to the score. The default parameters used with this program are as follows: gap penalty for multiple alignment=10; gap length penalty for multiple alignment=10; k-tuple value in pairwise alignment=1; gap penalty in pairwise alignment=3; window value in pairwise alignment—5; diagonals saved in pairwise alignment=5. The residue weight table used for the alignment program is PAM25O [Dayhoff et al., in Atlas of Protein Sequence and Structure, Dayhoff, Ed., NDRF, Washington, Vol. 5, suppl. 3, p. 345, (1978)].

[0102] Percent conservation is calculated from the above alignment by adding the percentage of identical residues to the percentage of positions at which the two residues represent a conservative substitution (defined as having a log odds value of greater than or equal to 0.3 in the PAM250 residue weight table). Conservation is referenced to human Gene A or gene B when determining percent conservation with non-human Gene A or gene B, e.g. gene A or gene B, when determining percent conservation. Conservative amino acid changes satisfying this requirement include: R-K; E-D, Y-F, L-M; V-I, Q-H.

Polypeptide Fragments

[0103] The invention provides polypeptide fragments of the disclosed proteins. Polypeptide fragments of the invention can comprise at least 8, more preferably at least 25, still more preferably at least 50 amino acid residues of the protein or an analogue thereof. More particularly such fragment will comprise at least 75, 100, 125, 150, 175, 200, 225, 250, 275 residues of the polypeptide encoded by the corresponding gene. Even more preferably, the protein fragment will comprise the majority of the native protein, e.g. about 100 contiguous residues of the native protein.

Biologically Active Variants

[0104] The invention also encompasses mutants of the novel colon proteins disclosed infra which comprise an amino acid sequence that is at least 80%, more preferably 90%, still more preferably 95-99% similar to the native protein.

[0105] Guidance in determining which amino acid residues can be substituted, inserted, or deleted without abolishing biological or immunological activity can be found using computer programs well known in the art, such as DNASTAR or software from the Genetics Computer Group (GCG). Preferably, amino acid changes in protein variants are conservative amino acid changes, i.e., substitutions of similarly charged or uncharged amino acids. A conservative amino acid change involves substitution of one of a family of amino acids which are related in their side chains. Naturally occurring amino acids are generally divided into four families: acidic (aspartate, glutamate), basic (lysine, arginine, histidine), non-polar (alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), and uncharged polar (glycine, asparagine, glutamine, cystine, serine, threonine, tyrosine) amino acids. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids.

[0106] A subset of mutants, called muteins, is a group of polypeptides in which neutral amino acids, such as serines, are substituted for cysteine residues which do not participate in disulfide bonds. These mutants may be stable over a broader temperature range than native secreted proteins. See Mark et al., U.S. Pat. No. 4,959,314.

[0107] It is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamnate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid will not have a major effect on the biological properties of the resulting secreted protein or polypeptide variant.

[0108] Protein variants include glycosylated forms, aggregative conjugates with other molecules, and covalent conjugates with unrelated chemical moieties. Also, protein variants also include allelic variants, species variants, and muteins. Truncations or deletions of regions which do not affect the differential expression of the gene are also variants. Covalent variants can be prepared by linking functionalities to groups which are found in the amino acid chain or at the N- or C-terminal residue, as is known in the art.

[0109] It will be recognized in the art that some amino acid sequence of the colon proteins of the invention can be varied without significant effect on the structure or function of the

protein. If such differences in sequence are contemplated, it should be remembered that there are critical areas on the protein which determine activity. In general, it is possible to replace residues that form the tertiary structure, provided that residues performing a similar function are used. In other instances, the type of residue may be completely unimportant if the alteration occurs at a non-critical region of the protein. The replacement of amino acids can also change the selectivity of binding to cell surface receptors. Ostade et al., *Nature* 361:266-268 (1993) describes certain mutations resulting in selective binding of TNF-alpha to only one of the two known types of TNF receptors. Thus, the polypeptides of the present invention may include one or more amino acid substitutions, deletions or additions, either from natural mutations or human manipulation.

[0110] The invention further includes variations of the colon proteins disclosed infra which show comparable expression patterns or which include antigenic regions. Such mutants include deletions, insertions, inversions, repeats, and site substitutions. Guidance concerning which amino acid changes are likely to be phenotypically silent can be found in Bowie, J. U., et al., "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science* 247:1306-1310 (1990).

[0111] Of particular interest are substitutions of charged amino acids with another charged amino acid and with neutral or negatively charged amino acids. The latter results in proteins with reduced positive charge to improve the characteristics of the disclosed protein. The prevention of aggregation is highly desirable. Aggregation of proteins not only results in a loss of activity but can also be problematic when preparing pharmaceutical formulations, because they can be immunogenic. (Pinckard et al., *Clin. Exp. Immunol.* 2:331-340 (1967); Robbins et al., *Diabetes* 36:838-845 (1987); Cleland et al., *Crit. Rev. Therapeutic Drug Carrier Systems* 10:307-377 (1993)).

[0112] Amino acids in the polypeptides of the present invention that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, *Science* 244: 1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as binding to a natural or synthetic binding partner. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith et al., *J. Mol. Biol.* 224:899-904 (1992) and de Vos et al. *Science* 255: 306-312 (1992)).

[0113] As indicated, changes are preferably of a minor nature, such as conservative amino acid substitutions that do not significantly affect the folding or activity of the protein. Of course, the number of amino acid substitutions a skilled artisan would make depends on many factors, including those described above. Generally speaking, the number of substitutions for any given polypeptide will not be more than 50, 40, 30, 25, 20, 15, 10, 5 or 3.

Fusion Proteins

[0114] Fusion proteins comprising proteins or polypeptide fragments of the subject colon tumor antigen can also be constructed. Fusion proteins are useful for generating antibodies against amino acid sequences and for use in various assay systems. For example, fusion proteins can be used to identify proteins which interact with a protein of the invention or which interfere with its biological function. Physical methods, such as protein affinity chromatography, or library-based assays for protein-protein interactions, such as the yeast two-hybrid or phage display systems, can also be used for this purpose. Such methods are well known in the art and can also be used as drug screens. Fusion proteins comprising a signal sequence and/or a transmembrane domain of a protein according to the invention or a fragment thereof can be used to target other protein domains to cellular locations in which the domains are not normally found, such as bound to a cellular membrane or secreted extracellularly.

[0115] A fusion protein comprises two protein segments fused together by means of a peptide bond. As noted, these fragments may range in size from about 8 amino acids up to the full length of the protein.

[0116] The second protein segment can be a full-length protein or a polypeptide fragment. Proteins commonly used in fusion protein construction include β-galactosidase, β-glucuronidase, green fluorescent protein (GFP), autofluorescent proteins, including blue fluorescent protein (BFP), glutathione-S-transferase (GST), luciferase, horseradish peroxidase (HRP), and chloramphenicol acetyltransferase (CAT). Additionally, epitope tags can be used in fusion protein constructions, including histidine (His) tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Other fusion constructions can include maltose binding protein (MBP), S-tag, Lex a DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP 16 protein fusions.

[0117] These fusions can be made, for example, by covalently linking two protein segments or by standard procedures in the art of molecular biology. Recombinant DNA methods can be used to prepare fusion proteins, for example, by making a DNA construct which comprises a coding sequence encoding a possible antigen according to the invention or a fragment thereof in proper reading frame with a nucleotide encoding the second protein segment and expressing the DNA construct in a host cell, as is known in the art. Many kits for constructing fusion proteins are available from companies that supply research labs with tools for experiments, including, for example, Promega Corporation (Madison, Wis.), Stratagene (La Jolla, Calif.), Clontech (Mountain View, Calif.), Santa Cruz Biotechnology (Santa Cruz, Calif.), MBL International Corporation (MIC; Watertown, Mass.), and Quantum Biotechnologies (Montreal, Canada; 1-888-DNA-KITS).

[0118] Proteins, fusion proteins, or polypeptides of the invention can be produced by recombinant DNA methods. For production of recombinant proteins, fusion proteins, or polypeptides, a sequence encoding the protein can be expressed in prokaryotic or eukaryotic host cells using expression systems known in the art. These expression systems include bacterial, yeast, insect, and mammalian cells.

[0119] The resulting expressed protein can then be purified from the culture medium or from extracts of the cultured cells using purification procedures known in the art. For example, for proteins fully secreted into the culture medium, cell-free medium can be diluted with sodium acetate and contacted with a cation exchange resin, followed by hydrophobic interaction chromatography. Using this method, the desired pro-

tein or polypeptide is typically greater than 95% pure. Further purification can be undertaken, using, for example, any of the techniques listed above.

[0120] It may be necessary to modify a protein produced in yeast or bacteria, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain a functional protein. Such covalent attachments can be made using known chemical or enzymatic methods.

[0121] A protein or polypeptide of the invention can also be expressed in cultured host cells in a form which will facilitate purification. For example, a protein or polypeptide can be expressed as a fusion protein comprising, for example, maltose binding protein, glutathione-S-transferase, or thioredoxin, and purified using a commercially available kit. Kits for expression and purification of such fusion proteins are available from companies such as New England BioLabs, Pharmacia, and Invitrogen. Proteins, fusion proteins, or polypeptides can also be tagged with an epitope, such as a "Flag" epitope (Kodak), and purified using an antibody which specifically binds to that epitope.

[0122] The coding sequence of the protein variants identified through the sequences disclosed herein can also be used to construct transgenic animals, such as mice, rats, guinea pigs, cows, goats, pigs, or sheep. Female transgenic animals can then produce proteins, polypeptides, or fusion proteins of the invention in their milk. Methods for constructing such animals are known and widely used in the art.

[0123] Alternatively, synthetic chemical methods, such as solid phase peptide synthesis, can be used to synthesize a secreted protein or polypeptide. General means for the production of peptides, analogs or derivatives are outlined in Chemistry and Biochemistry of Amino Acids, Peptides, and Proteins—A Survey of Recent Developments, B. Weinstein, ed. (1983). Substitution of D-amino acids for the normal L-stereoisomer can be carried out to increase the half-life of the molecule.

[0124] Typically, homologous polynucleotide sequences can be confirmed by hybridization under stringent conditions, as is known in the art. For example, using the following wash conditions: 2×SSC (0.3 M NaCl, 0.03 M sodium citrate, pH 7.0), 0.1% SDS, room temperature twice, 30 minutes each; then 2×SSC, 0.1% SDS, 50° C. once, 30 minutes; then 2×SSC, room temperature twice, 10 minutes each, homologous sequences can be identified which contain at most about 25-30% basepair mismatches. More preferably, homologous nucleic acid strands contain 15-25% basepair mismatches, even more preferably 5-15% basepair mismatches.

[0125] The invention also provides polynucleotide probes which can be used to detect complementary nucleotide sequences, for example, in hybridization protocols such as Northern or Southern blotting or in situ hybridizations. Polynucleotide probes of the invention comprise at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, or 40 or more contiguous nucleotides of the nucleic acid sequences provided herein. Polynucleotide probes of the invention can comprise a detectable label, such as a radioisotopic, fluorescent, enzymatic, or chemiluminescent label.

[0126] Isolated genes corresponding to the cDNA sequences disclosed herein are also provided. Standard molecular biology methods can be used to isolate the corresponding genes using the cDNA sequences provided herein. These methods include preparation of probes or primers from the nucleotide sequence disclosed herein for use in identifying or amplifying the genes from mammalian, including human, genomic libraries or other sources of human genomic DNA.

[0127] Polynucleotide molecules of the invention can also be used as primers to obtain additional copies of the polynucleotides, using polynucleotide amplification methods. Polynucleotide molecules can be propagated in vectors and cell lines using techniques well known in the art. Polynucleotide molecules can be on linear or circular molecules. They can be on autonomously replicating molecules or on molecules without replication sequences. They can be regulated by their own or by other regulatory sequences, as is known in the art.

Polynucleotide Constructs

[0128] Polynucleotide molecules comprising the coding sequences of the gene variants identified through the sequences disclosed herein can be used in a polynucleotide construct, such as a DNA or RNA construct. Polynucleotide molecules of the invention can be used, for example, in an expression construct to express all or a portion of a protein, variant, fusion protein, or single-chain antibody in a host cell. An expression construct comprises a promoter which is functional in a chosen host cell. The skilled artisan can readily select an appropriate promoter from the large number of cell type-specific promoters known and used in the art. The expression construct can also contain a transcription terminator which is functional in the host cell. The expression construct comprises a polynucleotide segment which encodes all or a portion of the desired protein. The polynucleotide segment is located downstream from the promoter. Transcription of the polynucleotide segment initiates at the promoter. The expression construct can be linear or circular and can contain sequences, if desired, for autonomous replication.

[0129] Also included are polynucleotide molecules comprising the promoter and UTR sequences of the subject novel genes, operably linked to the associated protein coding sequence and/or other sequences encoding a detectable or selectable marker. Such promoter and/or UTR-based constructs are useful for studying the transcriptional and translational regulation of protein expression, and for identifying activating and/or inhibitory regulatory proteins.

Host Cells

[0130] An expression construct can be introduced into a host cell. The host cell comprising the expression construct can be any suitable prokaryotic or eukaryotic cell. Expression systems in bacteria include those described in Chang et al., *Nature* 275:615 (1978); Goeddel et al., *Nature* 281: 544 (1979); Goeddel et al., *Nucleic Acids Res.* 8:4057 (1980); EP 36,776; U.S. Pat. No. 4,551,433; deBoer et al., *Proc. Natl. Acad. Sci. USA* 80: 21-25 (1983); and Siebenlist et al., *Cell* 20: 269 (1980).

[0131] Expression systems in yeast include those described in Hinnnen et al., *Proc. Natl. Acad. Sci. USA* 75: 1929 (1978); Ito et al., *J Bacteriol* 153: 163 (1983); Kurtz et al., *Mol. Cell. Biol.* 6: 142 (1986); Kunze et al., *J Basic Microbiol.* 25: 141 (1985); Gleeson et al., *J. Gen. Microbiol.* 132: 3459 (1986), Roggenkamp et al., *Mol. Gen. Genet.* 202: 302 (1986)); Das et al., J Bacteriol. 158: 1165 (1984); De Louvencourt et al., *J Bacteriol.* 154:737 (1983), Van den Berg et al., *Bio/Technology* 8: 135 (1990); Kunze et al., *J. Basic Microbiol.* 25: 141 (1985); Cregg et al., *Mol. Cell. Biol.* 5: 3376 (1985); U.S. Pat.

No. 4,837,148; U.S. Pat. No. 4,929,555; Beach and Nurse, *Nature* 300: 706 (1981); Davidow et al., *Curr. Genet.* 10: 380 (1985); Gaillardin et al., Curr. *Genet.* 10: 49 (1985); Ballance et al., *Biochem. Biophys. Res. Commun.* 112: 284-289 (1983); Tilburn et al., *Gene* 26: 205-22 (1983); Yelton et al., *Proc. Natl. Acad, Sci. USA* 81: 1470-1474 (1984); Kelly and Hynes, *EMBO J.* 4: 475479 (1985); EP 244,234; and WO 91/00357.

[0132] Expression of heterologous genes in insects can be accomplished as described in U.S. Pat. No. 4,745,051; Friesen et al. (1986) "The Regulation of Baculovirus Gene Expression" in: THE MOLECULAR BIOLOGY OF BACULOVIRUSES (W. Doerfler, ed.); EP 127,839; EP 155,476; Vlak et al., *J. Gen. Virol.* 69: 765-776 (1988); Miller et al., *Ann. Rev. Microbiol.* 42: 177 (1988); Carbonell et al., *Gene* 73: 409 (1988); Maeda et al., *Nature* 315: 592-594 (1985); Lebacq-Verheyden et al., *Mol. Cell. Biol.* 8: 3129 (1988); Smith et al., *Proc. Natl. Acad. Sci. USA* 82: 8404 (1985); Miyajima et al., *Gene* 58: 273 (1987); and Martin et al., *DNA* 7:99 (1988). Numerous baculoviral strains and variants and corresponding permissive insect host cells from hosts are described in Luckow et al., *Bio/Technology* (1988) δ: 47-55, Miller et al., in GENETIC ENGINEERING (Setlow, J. K. et al. eds.), Vol. 8, pp. 277-279 (Plenum Publishing, 1986); and Maeda et al., *Nature,* 315: 592-594 (1985).

[0133] Mammalian expression can be accomplished as described in Dijkema et al., *EMBO J.* 4: 761 (1985); Gorman et al., *Proc. Natl. Acad. Sci. USA* 79: 6777 (1982b); Boshart et al., *Cell* 41: 521 (1985); and U.S. Pat. No. 4,399,216. Other features of mammalian expression can be facilitated as described in Ham and Wallace, *Meth Enz.* 58: 44 (1979);

[0134] Expression constructs can be introduced into host cells using any technique known in the art. These techniques include transferrin-polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated cellular fusion, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, "gene gun," and calcium phosphate-mediated transfection.

[0135] The invention can also include hybrid and modified forms thereof including fusion proteins, fragments and hybrid and modified forms in which certain amino acids have been deleted or replaced, modifications such as where one or more amino acids have been changed to a modified amino acid or unusual amino acid.

[0136] Also included within the meaning of substantially homologous is any human or non-human primate protein which may be isolated by virtue of cross-reactivity with antibodies to proteins encoded by a gene described herein or whose encoding nucleotide sequences including genomic DNA, mRNA or cDNA may be isolated through hybridization with the complementary sequence of genomic or subgenomic nucleotide sequences or cDNA of a gene herein or fragments thereof. It will also be appreciated by one skilled in the art that degenerate DNA sequences can encode a tumor protein according to the invention and these are also intended to be included within the present invention as are allelic variants of the subject genes.

[0137] Preferred is a colon protein according to the invention prepared by recombinant DNA technology. By "pure form" or "purified form" or "substantially purified form" it is meant that a protein composition is substantially free of other proteins which are not the desired protein.

[0138] The present invention also includes therapeutic or pharmaceutical compositions comprising a protein according to the invention in an effective amount for treating patients with disease, and a method comprising administering a therapeutically effective amount of the protein. These compositions and methods are useful for treating cancers associated with the subject proteins, e.g. colon cancer. One skilled in the art can readily use a variety of assays known in the art to determine whether the protein would be useful in promoting survival or functioning in a particular cell type.

[0139] Anti-Colon Antigen Antibodies

[0140] As noted, the invention includes the preparation and use of anti-colon antigen antibodies and fragments for use as diagnostics and therapeutics. These antibodies may be polyclonal or monoclonal. Polyclonal antibodies can be prepared by immunizing rabbits or other animals by injecting antigen followed by subsequent boosts at appropriate intervals. The animals are bled and sera assayed against purified protein usually by ELISA or by bioassay based upon the ability to block the action of the corresponding gene. When using avian species, e.g., chicken, turkey and the like, the antibody can be isolated from the yolk of the egg. Monoclonal antibodies can be prepared after the method of Milstein and Kohler by fusing splenocytes from immunized mice with continuously replicating tumor cells such as myeloma or lymphoma cells. [Milstein and Kohler, Nature 256:495-497 (1975); Gulfre and Milstein, *Methods in Enzymology: Immunochemical Techniques* 73:1-46, Langone and Banatis eds., Academic Press, (1981) which are incorporated by reference]. The hybridoma cells so formed are then cloned by limiting dilution methods and supernates assayed for antibody production by ELISA, RIA or bioassay.

[0141] The unique ability of antibodies to recognize and specifically bind to target proteins provides an approach for treating an overexpression of the protein. Thus, another aspect of the present invention provides for a method for preventing or treating diseases involving overexpression of the protein by treatment of a patient with specific antibodies to the protein.

[0142] Specific antibodies, either polyclonal or monoclonal, to the protein can be produced by any suitable method known in the art as discussed above. For example, by recombinant methods, preferably in eukaryotic cells murine or human monoclonal antibodies can be produced by hybridoma technology or, alternatively, the protein, or an immunologically active fragment thereof, or an anti-idiotypic antibody, or fragment thereof can be administered to an animal to elicit the production of antibodies capable of recognizing and binding to the protein. Such antibodies can be from any class of antibodies including, but not limited to IgG, IgA, 1 gM, IgD, and IgE or in the case of avian species, IgY and from any subclass of antibodies.

[0143] The availability of isolated protein allows for the identification of small molecules and low molecular weight compounds that inhibit the binding of protein to binding partners, through routine application of high-throughput screening methods (HTS). HTS methods generally refer to technologies that permit the rapid assaying of lead compounds for therapeutic potential. HTS techniques employ robotic handling of test materials, detection of positive signals, and interpretation of data. Lead compounds may be identified via the incorporation of radioactivity or through optical assays that rely on absorbance, fluorescence or luminescence as read-outs. [Gonzalez, J. E. et al, *Curr. Opin. Biotech.* 9:624-63 1 (1998)].

[0144] Model systems are available that can be adapted for use in high throughput screening for compounds that inhibit the interaction of protein with its ligand, for example by competing with protein for ligand binding. Sarubbi et al., *Anal. Biochem.* 237:70-75 (1996) describe cell-free, non-isotopic assays for discovering molecules that compete with natural ligands for binding to the active site of IL-1 receptor. Martens, C. et al., *Anal. Biochem.* 273:20-31 (1999) describe a generic particle-based nonradioactive method in which a labeled ligand binds to its receptor immobilized on a particle; label on the particle decreases in the presence of a molecule that competes with the labeled ligand for receptor binding.

(i) Starting Materials and Methods

[0145] Immunoglobulins (Ig) and certain variants thereof are known and many have been prepared in recombinant cell culture. For example, see U.S. Pat. No. 4,745,055; EP 256, 654; EP 120,694; EP 125,023; EP 255,694; EP 266,663; WO 30 88/03559; Faulkneret al., Nature, 298: 286 (1982); Morrison, J. Immun., 123: 793 (1979); Koehler et al., Proc. Natl. Acad. Sci. USA, 77: 2197 (1980); Raso et al., Cancer Res., 41: 2073 (1981); Morrison et al., Ann. Rev. Immunol., 2: 239 (1984); Morrison, Science, 229: 1202 (1985); and Morrison et al., Proc. Natl. Acad. Sci. USA, 81: 6851 (1984). Reassorted immunoglobulin chains are also known. See, for example, U.S. Pat. No. 4,444,878; WO 88/03565; and EP 68,763 and references cited therein. The immunoglobulin moiety in the chimeras of the present invention may be obtained from IgG-1, IgG-2, IgG-3, or IgG-4 subtypes, IgA, IgE, IgD, or IgM, but preferably from IgG-1 or IgG-3.

(ii) Polyclonal Antibodies

[0146] Polyclonal antibodies to the subject colon antigens are generally raised in animals by multiple subcutaneous (sc) or intraperitoneal (ip) injections of the antigen and an adjuvant. It may be useful to conjugate the antigen or a fragment containing the target amino acid sequence to a protein that is immunogenic in the species to be immunized, e.g., keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, or soybean trypsin inhibitor using a bifunctional or derivatizing agent, for example, maleimidobenzoyl sulfosuccinimide ester (conjugation through cysteine residues), N-hydroxysuccinimide (through lysine residues), glutaraldehyde or succinic anhydride.

[0147] Animals are immunized against the polypeptide or fragment, immunogenic conjugates, or derivatives by combining about 1 mg or 1 μg of the peptide or conjugate (for rabbits or mice, respectively) with 3 volumes of Freund's complete adjuvant and injecting the solution intradermally at multiple sites. One month later the animals are boosted with ⅕ to 1/10 the original amount of peptide or conjugate in Freund's complete adjuvant by subcutaneous injection at multiple sites. Seven to 14 days later the animals are bled and the serum is assayed for antibody titer to the antigen or a fragment thereof. Animals are boosted until the titer plateaus. Preferably, the animal is boosted with the conjugate of the same polypeptide or fragment thereof, but conjugated to a different protein and/or through a different cross-linking reagent. Conjugates also can be made in recombinant cell culture as protein fusions. Also, aggregating agents such as alum are suitably used to enhance the immune response.

(iii) Monoclonal Antibodies

[0148] Monoclonal antibodies are obtained from a population of substantially homogeneous antibodies, i.e., the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor amounts. Thus, the modifier "monoclonal" indicates the character of the antibody as not being a mixture of discrete antibodies.

[0149] For example, monoclonal antibodies using for practicing this invention may be made using the hybridoma method first described by Kohler and Milstein, Nature, 256: 495 (1975), or may be made by recombinant DNA methods (Cabilly et al., supra).

[0150] In the hybridoma method, a mouse or other appropriate host animal, such as a hamster, is immunized as hereinabove described to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the antigen or fragment thereof used for immunization. Alternatively, lymphocytes may be immunized in vitro. Lymphocytes then are fused with myeloma cells using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, pp. 59-103 [Academic Press, 1986]).

[0151] The hybridoma cells thus prepared are seeded and grown in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, parental myeloma cells. For example, if the parental myeloma cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine (HAT medium), which substances prevent the growth of HGPRT-deficient cells.

[0152] Preferred myeloma cells are those that fuse efficiently, support stable high-level production of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. Among these, preferred myeloma cell lines are murine myeloma lines, such as those derived from MOPC-21 and MPC-11 mouse tumors available from the Salk Institute Cell Distribution Center, San Diego, Calif. USA, and SP-2 cells available from the American Type Culture Collection, Rockville, Md. USA.

[0153] Culture medium in which hybridoma cells are growing is assayed for production of monoclonal antibodies directed against the colon antigen. Preferably, the binding specificity of monoclonal antibodies produced by hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA).

[0154] The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107: 220 (1980).

[0155] After hybridoma cells are identified that produce antibodies of the desired specificity, affinity, and/or activity, the clones may be subcloned by limiting dilution procedures and grown by standard methods (Goding, supra). Suitable culture media for this purpose include, for example, D-MEM or RPMI-1640 medium. In addition, the hybridoma cells may be grown in vivo as ascites tumors in an animal.

[0156] The monoclonal antibodies secreted by the subclones are suitably separated from the culture medium, ascites fluid, or serum by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxyapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

[0157] DNA encoding the monoclonal antibodies of the invention is readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma

cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as *E. coli* cells, simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. Review articles on recombinant expression in bacteria of DNA encoding the antibody include Skerra et al., Curr. Opinion in Immunol., 5: 256-262 (1993) and Pluckthun, Immunol. Revs., 130: 151-188 (1992). A preferred expression system is the NEOSPLA™ expression system of IDEC above-referenced.

[0158] The DNA also may be modified, for example, by substituting the coding sequence for human heavy- and light-chain constant domains in place of the homologous murine sequences (Morrison, et al., Proc. Natl. Acad. Sci. USA, 81: 6851 [1984]), or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. In that manner, "chimeric" or "hybrid" antibodies are prepared that have the binding specificity of an anti-colon antigen monoclonal antibody herein.

[0159] Typically such non-immunoglobulin polypeptides are substituted for the constant domains of an antibody of the invention, or they are substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody comprising one antigen-combining site having specificity for colon antigen according to the invention and another antigen-combining site having specificity for a different antigen.

[0160] Chimeric or hybrid antibodies also may be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins may be constructed using a disulfide-exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate.

(iv) Humanized Antibodies

[0161] Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and co-workers (Jones et al., Nature 321, 522-525 [1986]; Riechmann et al., Nature 332, 323-327 [1988]; Verhoeyen et al., Science 239, 1534-1536 [1988]), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (Cabilly et al., supra), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

[0162] The choice of human variable domains, both light and heavy, to be used in making the humanized antibodies is very important to reduce antigenicity. According to the so-called "best-fit" method, the sequence of the variable domain of a rodent antibody is screened against the entire library of known human variable-domain sequences. The human sequence which is closest to that of the rodent is then accepted as the human framework (FR) for the humanized antibody (Sims et al., J. Immunol., 151: 2296 [1993]; Chothia and Lesk, J. Mol. Biol., 196: 901 [1987]). Another method uses a particular framework derived from the consensus sequence of all human antibodies of a particular subgroup of light or heavy chains. The same framework may be used for several different humanized antibodies (Carter et al., Proc. Natl. Acad. Sci. USA, 89: 4285 [1992]; Presta et al., J. Immunol., 151: 2623 [1993]).

[0163] It is further important that antibodies be humanized with retention of high affinity for the antigen and other favorable biological properties. To achieve this goal, according to a preferred method, humanized antibodies are prepared by a process of analysis of the parental sequences and various conceptual humanized products using three-dimensional models of the parental and humanized sequences. Three-dimensional immunoglobulin models are commonly available and are familiar to those skilled in the art. Computer programs are available which illustrate and display probable three-dimensional conformational structures of selected candidate immunoglobulin sequences. Inspection of these displays permits analysis of the likely role of the residues in the functioning of the candidate immunoglobulin sequence, i.e., the analysis of residues that influence the ability of the candidate immunoglobulin to bind its antigen. In this way, FR residues can be selected and combined from the consensus and import sequences so that the desired antibody characteristic, such as increased affinity for the target antigen(s), is achieved. In general, the CDR residues are directly and most substantially involved in influencing antigen binding.

(v) Human Antibodies

[0164] Human monoclonal antibodies can be made by the hybridoma method. Human myeloma and mouse-human heteromyeloma cell lines for the production of human monoclonal antibodies have been described, for example, by Kozbor, J. Immunol. 133, 3001 (1984); Brodeur, et al., Monoclonal Antibody Production Techniques and Applications, pp. 51-63 (Marcel Dekker, Inc., New York, 1987); and Boerner et al., J. Immunol., 147: 86-95 (1991).

[0165] It is now possible to produce transgenic animals (e.g., mice) that are capable, upon immunization, of producing a fall repertoire of human antibodies in the absence of endogenous immunoglobulin production. For example, it has been described that the homozygous deletion of the antibody heavy-chain joining region (JH) gene in chimeric and germ-line mutant mice results in complete inhibition of endogenous antibody production. Transfer of the human germ-line immunoglobulin gene array in such germ-line mutant mice will result in the production of human antibodies upon antigen challenge. See, e.g., Jakobovits et al., Proc. Natl. Acad. Sci. USA, 90: 2551 (1993); Jakobovits et al., Nature, 362: 255-258 (1993); Bruggermann et al., Year in Immuno., 7: 33 (1993).

[0166] Alternatively, the phage display technology (McCafferty et al., Nature, 348: 552-553 [1990]) can be used to produce human antibodies and antibody fragments in vitro, from immunoglobulin variable (V) domain gene repertoires from non-immunized donors. According to this technique, antibody V domain genes are cloned in-frame into either a major or minor coat protein gene of a filamentous bacteriophage, such as M13 or fd, and displayed as functional antibody fragments on the surface of the phage particle. Because the

filamentous particle contains a single-stranded DNA copy of the phage genome, selections based on the functional properties of the antibody also result in selection of the gene encoding the antibody exhibiting those properties. Thus, the phage mimics some of the properties of the B-cell. Phage display can be performed in a variety of formats; for their review see, e.g., Johnson and Chiswell, Curr. Op. Struct. Biol., 3: 564-571 (1993). Several sources of V-gene segments can be used for phage display. Clackson et al., Nature, 352: 624-628 (1991) isolated a diverse array of anti-oxazolone antibodies from a small random combinatorial library of V genes derived from the spleens of immunized mice. A repertoire of V genes from non-immunized human donors can be constructed and antibodies to a diverse array of antigens (including self-antigens) can be isolated essentially following the techniques described by Marks et al., J. Mol. Biol., 222: 581-597 (1991), or Griffith et al., EMBO J., 12: 725-734 (1993).

[0167] In a natural immune response, antibody genes accumulate mutations at a high rate (somatic hypermutation). Some of the changes introduced will confer higher affinity, and B cells displaying high-affinity surface immunoglobulin are preferentially replicated and differentiated during subsequent antigen challenge. This natural process can be mimicked by employing the technique known as "chain shuffling" (Marks et al., Bio/Technology, 10: 779-783 [1992]). In this method, the affinity of "primary" human antibodies obtained by phage display can be improved by sequentially replacing the heavy and light chain V region genes with repertoires of naturally occurring variants (repertoires) of V domain genes obtained from non-immunized donors. This technique allows the production of antibodies and antibody fragments with affinities in the nM range. A strategy for making very large phage antibody repertoires has been described by Waterhouse et al., *Nucl. Acids Res.,* 21: 2265-2266 (1993).

[0168] Gene shuffling can also be used to derive human antibodies from rodent antibodies, where the human antibody has similar affinities and specificities to the starting rodent antibody. According to this method, which is also referred to as "epitope imprinting", the heavy or light chain V domain gene of rodent antibodies obtained by phage display technique is replaced with a repertoire of human V domain genes, creating rodent-human chimeras. Selection on antigen results in isolation of human variable capable of restoring a functional antigen-binding site, i.e., the epitope governs (imprints) the choice of partner. When the process is repeated in order to replace the remaining rodent V domain, a human antibody is obtained (see PCT WO 93/06213, published Apr. 1, 1993). Unlike traditional humanization of rodent antibodies by CDR grafting, this technique provides completely human antibodies, which have no framework or CDR residues of rodent origin.

(vi) Bispecific Antibodies

[0169] Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities will be to a colon antigen according to the invention. Methods for making bispecific antibodies are known in the art.

[0170] Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy chain-light chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305: 537-539 [1983]). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of 10 different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule, which is usually done by affinity chromatography steps, is rather cumbersome, and the product yields are low. Similar procedures are disclosed in WO 93/08829 published May 13, 1993, and in Traunecker et al., EMBO J., 10: 3655-3659 (1991).

[0171] According to a different and more preferred approach, antibody-variable domains with the desired binding specificities (antibody-antigencombining sites) are fused to immunoglobulin constant-domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1), containing the site necessary for light-chain binding, present in at least one of the fusions. DNAs encoding the immunoglobulin heavy chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. This provides for great flexibility in adjusting the mutual proportions of the three polypeptide fragments in embodiments when unequal ratios of the three polypeptide chains used in the construction provide the optimum yields. It is, however, possible to insert the coding sequences for two or all three polypeptide chains in one expression vector when the production of at least two polypeptide chains in equal ratios results in high yields or when the ratios are of no particular significance. In a preferred embodiment of this approach, the bispecific antibodies are composed of a hybrid immunoglobulin heavy chain with a first binding specificity in one arm, and a hybrid immunoglobulin heavy chain-light chain pair (providing a second binding specificity) in the other arm. It was found that this asymmetric structure facilitates the separation of the desired bispecific compound from unwanted immunoglobulin chain combinations, as the presence of an immunoglobulin light chain in only one half of the bispecific molecule provides for a facile way of separation.

[0172] For further details of generating bispecific antibodies, see, for example, Suresh et al., Methods in Enzymology, 121: 210 (1986).

(vii) Heteroconjugate Antibodies

[0173] Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Pat. No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/00373; and EP 03089). Heteroconjugate antibodies may be made using any convenient cross-linking methods. Suitable cross-linking agents are well known in the art, and are disclosed in U.S. Pat. No. 4,676,980, along with a number of cross-linking techniques.

[0174] The polynucleotides and polypeptides of the present invention may be utilized in gene delivery vehicles. The gene delivery vehicle may be of viral or non-viral origin (see generally, Jolly, *Cancer Gene Therapy* 1:51-64 (1994); Kimura, *Human Gene Therapy* 5:845-852 (1994); Connelly, *Human Gene Therapy* 1:185-193 (1995); and Kaplitt, *Nature Genetics* 6:148-153 (1994)). Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic according to the invention can be administered either locally

or systemically. These constructs can utilize viral or non-viral vector approaches. Expression of such coding sequences can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence can be either constitutive or regulated. Preferred vehicles for gene therapy include retroviral and adeno-viral vectors.

[0175] Representative examples of adenoviral vectors include those described by Berkner, *Biotechniques* 6:616-627 (Biotechniques); Rosenfeld et al., *Science* 252:431-434 (1991); WO 93/19191; Kolls et al., *P.N.A.S.* 215-219 (1994); Kass-Bisleret al., P.N.A.S. 90: 11498-11502 (1993); Guzman et al., *Circulation* 88: 2838-2848 (1993); Guzman et al., *Cir. Res.* 73: 1202-1207 (1993); Zabner et al., *Cell* 75: 207-216 (1993); Li et al., *Hum. Gene Ther.* 4: 403-409 (1993); Cailaud et al., *Eur. J. Neurosci.* 5: 1287-1291 (1993); Vincent et al., *Nat. Genet.* 5: 130-134 (1993); Jaffe et al., *Nat. Genet.* 1: 372-378 (1992); and Levrero et al., *Gene* 101: 195-202 (1992). Exemplary adenoviral gene therapy vectors employable in this invention also include those described in WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655. Administration of DNA linked to kill adenovirus as described in Curiel, *Hum. Gene Ther.* 3: 147-154 (1992) may be employed.

[0176] Other gene delivery vehicles and methods may be employed; including polycationic condensed DNA linked or unlinked to kill adenovirus alone, for example Curiel, *Hum. Gene Ther.* 3: 147-154 (1992); ligand-linked DNA, for example see Wu, *J. Biol. Chem.* 264: 16985-16987 (1989); eukaryotic cell delivery vehicles cells, for example see U.S. Ser. No. 08/240,030, filed May 9, 1994, and U.S. Ser. No. 08/404,796; deposition of photopolymerized hydrogel materials; hand-held gene transfer particle gun, as described in U.S. Pat. No. 5,149,655; ionizing radiation as described in U.S. Pat. No. 5,206,152 and in WO 92/11033; nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip, *Mol. Cell. Biol.* 14:2411-2418 (1994), and in Woffendin, *Proc. Natl. Acad. Sci.* 91:1581-1585 (1994).

[0177] Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and U.S. Pat. No. 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm. Liposomes that can act as gene delivery vehicles are described in U.S. Pat. No. 5,422,120, PCT Patent Publication Nos. WO 95/13 796, WO 94/23697, and WO 91/14445, and EP No. 0 524 968.

[0178] Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin et al., *Proc. Natl. Acad. Sci. USA* 91(24): 11581-11585 (1994). Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in U.S. Pat. No. 5,149,655; use of ionizing radiation for activating transferred gene, as described in U.S. Pat. No. 5,206,152 and PCT Patent Publication No. WO 92/11033.

[0179] The subject antibodies or antibody fragments may be conjugated directly or indirectly to effective moieties, e.g., radionuclides, toxins, chemotherapeutic agents, prodrugs, cytostatic agents, enzymes and the like. In a preferred embodiment the antibody or fragment will be attached to a therapeutic or diagnostic radiolabel directly or by use of a chelating agent. Examples of suitable radiolabels are well known and include ^{90}Y, ^{125}I, ^{131}I, ^{111}I, ^{105}Rh, ^{153}Sm, ^{67}Cu, ^{67}Ga, ^{166}Ho, ^{177}Lu, ^{186}Re and ^{188}Re.

[0180] Examples of suitable drugs that my be coupled to antibodies include methotrexate, adriamycine and lymphokines such as interferons, interleukins and the like. Suitable toxins which may be coupled include ricin, cholera and diptheria toxin.

[0181] In a preferred embodiment, the subject antibodies will be attached to a therapeutic radiolabel and used for radioimmunotherapy.

[0182] Inhibitory Oligonucleotides

[0183] In certain circumstances, it may be desirable to modulate or decrease the amount of the protein expressed by a colon cell. Thus, in another aspect of the present invention, inhibitory oligonucleotides can be made and a method utilized for diminishing the level of expression a colon antigen according to the invention by a cell comprising administering one or more inhibitory oligonucleotides, or a precursor thereof or a nucleic acid encoding the same. By inhibitory oligonucleotides reference is made to oligonucleotides that have a nucleotide sequence that interacts through base pairing with a specific complementary nucleic acid sequence involved in the expression of a target molecule, such that the expression of the gene is reduced. Preferably, the specific nucleic acid sequence involved in the expression of the gene is a genomic DNA molecule or mRNA molecule that encodes the gene. This genomic DNA molecule can comprise regulatory regions of the gene, or the coding sequence for the mature gene.

[0184] The term complementary to a nucleotide sequence in the context of inhibitory oligonucleotides and methods therefore means sufficiently complementary to such a sequence as to allow hybridization to that sequence in a cell, i.e., under physiological conditions. Antisense oligonucleotides preferably comprise a sequence containing from about 8 to about 100 nucleotides and more preferably the inhibitory oligonucleotides comprise from about 15 to about 30 nucleotides. They are typically single-stranded, and may be selected from antisense oligonucleotides, ribozymes, siRNAs, etc. Inhibitory oligonucleotides can also contain a variety of modifications that confer resistance to nucleolytic degradation such as, for example, modified internucleoside linages [Uhlmann and Peyman, *Chemical Reviews* 90:543-548 (1990); Schneider and Banner, *Tetrahedron Lett.* 31:335, (1990) which are incorporated by reference], modified nucleic acid bases as disclosed in U.S. Pat. No. 5,958,773 and patents disclosed therein, and/or sugars and the like.

[0185] Any modifications or variations of the inhibitory molecule which are known in the art to be broadly applicable to inhibitory technology are included within the scope of the invention. Such modifications include preparation of phosphorus-containing linkages as disclosed in U.S. Pat. Nos. 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361, 5,625,050 and 5,958,773.

[0186] The inhibitory compounds of the invention can include modified bases. The inhibitory oligonucleotides of the invention can also be modified by chemically linking the oligonucleotide to one or more moieties or conjugates to enhance the activity, cellular distribution, or cellular uptake

of the antisense oligonucleotide. Such moieties or conjugates include lipids such as cholesterol, cholic acid, thioether, aliphatic chains, phospholipids, polyamines, polyethylene glycol (PEG), palmityl moieties, and others as disclosed in, for example, U.S. Pat. Nos. 5,514,758, 5,565,552, 5,567,810, 5,574,142, 5,585,481, 5,587,371, 5,597,696 and 5,958,773.

[0187] Chimeric antisense oligonucleotides are also within the scope of the invention, and can be prepared from the present inventive oligonucleotides using the methods described in, for example, U.S. Pat. Nos. 5,013,830, 5,149, 797, 5,403,711, 5,491,133, 5,565,350, 5,652,355, 5,700,922 and 5,958,773.

[0188] In the antisense art a certain degree of routine experimentation is required to select optimal antisense molecules for particular targets. To be effective, the antisense molecule preferably is targeted to an accessible, or exposed, portion of the target RNA molecule. Although in some cases information is available about the structure of target mRNA molecules, the current approach to inhibition using antisense is via experimentation. mRNA levels in the cell can be measured routinely in treated and control cells by reverse transcription of the mRNA and assaying the cDNA levels. The biological effect can be determined routinely by measuring cell growth or viability as is known in the art.

[0189] Measuring the specificity of antisense activity by assaying and analyzing cDNA levels is an art-recognized method of validating antisense results. It has been suggested that RNA from treated and control cells should be reverse-transcribed and the resulting cDNA populations analyzed. [Branch, A. D., *T.I.B.S.* 23:45-50 (1998)].

[0190] The therapeutic or pharmaceutical compositions of the present invention can be administered by any suitable route known in the art including for example intravenous, subcutaneous, intramuscular, transdermal, intrathecal or intracerebral. Administration can be either rapid as by injection or over a period of time as by slow infusion or administration of slow release formulation.

[0191] Additionally, the subject colon tumor proteins can also be linked or conjugated with agents that provide desirable pharmaceutical or pharmacodynamic properties. For example, the protein can be coupled to any substance known in the art to promote penetration or transport across the blood-brain barrier such as an antibody to the transferrin receptor, and administered by intravenous injection (see, for example, Friden et al., *Science* 259:373-377 (1993) which is incorporated by reference). Furthermore, the subject protein A or protein B can be stably linked to a polymer such as polyethylene glycol to obtain desirable properties of solubility, stability, half-life and other pharmaceutically advantageous properties. [See, for example, Davis et al., *Enzyme Eng.* 4:169-73 (1978); Buruham, *Am. J. Hosp. Pharm.* 51:210-218 (1994) which are incorporated by reference].

[0192] The compositions are usually employed in the form of pharmaceutical preparations. Such preparations are made in a manner well known in the pharmaceutical art. See, e.g. Remington Pharmaceutical Science, 18th Ed., Merck Publishing Co. Eastern PA, (1990). One preferred preparation utilizes a vehicle of physiological saline solution, but it is contemplated that other pharmaceutically acceptable carriers such as physiological concentrations of other non-toxic salts, five percent aqueous glucose solution, sterile water or the like may also be used. It may also be desirable that a suitable buffer be present in the composition. Such solutions can, if desired, be lyophilized and stored in a sterile ampoule ready for reconstitution by the addition of sterile water for ready injection. The primary solvent can be aqueous or alternatively non-aqueous. The subject colon tumor antigens, fragments or variants thereof can also be incorporated into a solid or semi-solid biologically compatible matrix which can be implanted into tissues requiring treatment.

[0193] The carrier can also contain other pharmaceutically-acceptable excipients for modifying or maintaining the pH, osmolarity, viscosity, clarity, color, sterility, stability, rate of dissolution, or odor of the formulation. Similarly, the carrier may contain still other pharmaceutically-acceptable excipients for modifying or maintaining release or absorption or penetration across the blood-brain barrier. Such excipients are those substances usually and customarily employed to formulate dosages for parental administration in either unit dosage or multi-dose form or for direct infusion into the cerebrospinal fluid by continuous or periodic infusion.

[0194] Dose administration can be repeated depending upon the pharmacokinetic parameters of the dosage formulation and the route of administration used.

[0195] It is also contemplated that certain formulations containing the subject antibody or nucleic acid antagonists are to be administered orally. Such formulations are preferably encapsulated and formulated with suitable carriers in solid dosage forms. Some examples of suitable carriers, excipients, and diluents include lactose, dextrose, sucrose, sorbitol, mannitol, starches, gum acacia, calcium phosphate, alginates, calcium silicate, microcrystalline cellulose, polyvinylpyrrolidone, cellulose, gelatin, syrup, methyl cellulose, methyl- and propylhydroxybenzoates, talc, magnesium, stearate, water, mineral oil, and the like. The formulations can additionally include lubricating agents, wetting agents, emulsifying and suspending agents, preserving agents, sweetening agents or flavoring agents. The compositions may be formulated so as to provide rapid, sustained, or delayed release of the active ingredients after administration to the patient by employing procedures well known in the art. The formulations can also contain substances that diminish proteolytic degradation and promote absorption such as, for example, surface active agents.

[0196] The specific dose is calculated according to the approximate body weight or body surface area of the patient or the volume of body space to be occupied. The dose will also be calculated dependent upon the particular route of administration selected. Further refinement of the calculations necessary to determine the appropriate dosage for treatment is routinely made by those of ordinary skill in the art. Such calculations can be made without undue experimentation by one skilled in the art in light of the activity disclosed herein in assay preparations of target cells. Exact dosages are determined in conjunction with standard dose-response studies. It will be understood that the amount of the composition actually administered will be determined by a practitioner, in the light of the relevant circumstances including the condition or conditions to be treated, the choice of composition to be administered, the age, weight, and response of the individual patient, the severity of the patient's symptoms, and the chosen route of administration.

[0197] In one embodiment of this invention, the protein may be therapeutically administered by implanting into patients vectors or cells capable of producing a biologically-active form of the protein or a precursor of protein, i.e., a molecule that can be readily converted to a biological-active form of the protein by the body. In one approach, cells that

secrete the protein may be encapsulated into semipermeable membranes for implantation into a patient. The cells can be cells that normally express the protein or a precursor thereof or the cells can be transformed to express the protein or a precursor thereof. It is preferred that the cell be of human origin and that the protein be a human protein when the patient is human. However, it is anticipated that non-human primate homologues of the protein discussed infra may also be effective.

[0198] Detection of Subject Colon Proteins or Nucleic Acids

[0199] In a number of circumstances it would be desirable to determine the levels of protein or corresponding mRNA in a patient. Evidence disclosed infra suggests the subject colon proteins may be expressed at different levels during some diseases, e.g., cancers, provides the basis for the conclusion that the presence of these proteins serves a normal physiological function related to cell growth and survival. Endogenously produced protein according to the invention may also play a role in certain disease conditions.

[0200] The term "detection" as used herein in the context of detecting the presence of protein in a patient is intended to include the determining of the amount of protein or the ability to express an amount of protein in a patient, the estimation of prognosis in terms of probable outcome of a disease and prospect for recovery, the monitoring of the protein levels over a period of time as a measure of status of the condition, and the monitoring of protein levels for determining a preferred therapeutic regimen for the patient, e.g. one with colon cancer.

[0201] To detect the presence of a colon protein according to the invention in a patient, a sample is obtained from the patient. The sample can be a tissue biopsy sample or a sample of blood, plasma, serum, CSF, urine or the like. It has been found that the subject proteins are expressed at high levels in some cancers. Samples for detecting protein can be taken from colon tissues. When assessing peripheral levels of protein, it is preferred that the sample be a sample of blood, plasma or serum. When assessing the levels of protein in the central nervous system a preferred sample is a sample obtained from cerebrospinal fluid or neural tissue. The sample may be collected by various techniques known per se in the art, including non-invasive techniques, or may be obtained from sample collections.

[0202] In some instances, it is desirable to determine whether the gene is intact in the patient or in a tissue or cell line within the patient. By an intact gene, it is meant that there are no alterations in the gene such as point mutations, deletions, insertions, chromosomal breakage, chromosomal rearrangements and the like wherein such alteration might alter production of the corresponding protein or alter its biological activity, stability or the like to lead to disease processes. Thus, in one embodiment of the present invention a method is provided for detecting and characterizing any alterations in the gene. The method comprises providing an oligonucleotide that contains the gene, genomic DNA or a fragment thereof or a derivative thereof. By a derivative of an oligonucleotide, it is meant that the derived oligonucleotide is substantially the same as the sequence from which it is derived in that the derived sequence has sufficient sequence complementarily to the sequence from which it is derived to hybridize specifically to the gene. The derived nucleotide sequence is not necessarily physically derived from the nucleotide sequence, but may be generated in any manner including for example, chemical synthesis or DNA replication or reverse transcription or transcription.

[0203] Typically, patient genomic DNA is isolated from a cell sample from the patient and digested with one or more restriction endonucleases such as, for example, TaqI and AluI. Using the Southern blot protocol, which is well known in the art, this assay determines whether a patient or a particular tissue in a patient has an intact colon gene according to the invention or a gene abnormality.

[0204] Hybridization to a gene would involve denaturing the chromosomal DNA to obtain a single-stranded DNA; contacting the single-stranded DNA with a gene probe associated with the gene sequence; and identifying the hybridized DNA-probe to detect chromosomal DNA containing at least a portion of a gene.

[0205] The term "probe" as used herein refers to a structure comprised of a polynucleotide that forms a hybrid structure with a target sequence, due to complementarily of probe sequence with a sequence in the target region. Oligomers suitable for use as probes may contain a minimum of about 8-12 contiguous nucleotides which are complementary to the targeted sequence and preferably a minimum of about 20.

[0206] A gene according to the present invention can be DNA or RNA oligonucleotides and can be made by any method known in the art such as, for example, excision, transcription or chemical synthesis. Probes may be labeled with any detectable label known in the art such as, for example, radioactive or fluorescent labels or enzymatic marker. Labeling of the probe can be accomplished by any method known in the art such as by PCR, random priming, end labeling, nick translation or the like. One skilled in the art will also recognize that other methods not employing a labeled probe can be used to determine the hybridization. Examples of methods that can be used for detecting hybridization include Southern blotting, fluorescence in situ hybridization, and single-strand conformation polymorphism with PCR amplification.

[0207] Hybridization is typically carried out at 25°-45° C., more preferably at 32°-40° C. and more preferably at 37°-38° C. The time required for hybridization is from about 0.25 to about 96 hours, more preferably from about one to about 72 hours, and most preferably from about 4 to about 24 hours.

[0208] Gene abnormalities can also be detected by using the PCR method and primers that flank or lie within the gene. The PCR method is well known in the art. Briefly, this method is performed using two oligonucleotide primers which are capable of hybridizing to the nucleic acid sequences flanking a target sequence that lies within a gene and amplifying the target sequence. The terms "oligonucleotide primer" as used herein refers to a short strand of DNA or RNA ranging in length from about 8 to about 30 bases. The upstream and downstream primers are typically from about 20 to about 30 base pairs in length and hybridize to the flanking regions for replication of the nucleotide sequence. The polymerization is catalyzed by a DNA-polymerase in the presence of deoxynucleotide triphosphates or nucleotide analogs to produce double-stranded DNA molecules. The double strands are then separated by any denaturing method including physical, chemical or enzymatic. Commonly, a method of physical denaturation is used involving heating the nucleic acid, typically to temperatures from about 80° C. to 105° C. for times ranging from about 1 to about 10 minutes. The process is repeated for the desired number of cycles.

[0209] The primers are selected to be substantially complementary to the strand of DNA being amplified. Therefore, the primers need not reflect the exact sequence of the template, but must be sufficiently complementary to selectively hybridize with the strand being amplified.

[0210] After PCR amplification, the DNA sequence comprising the gene or a fragment thereof is then directly sequenced and analyzed by comparison of the sequence with the sequences disclosed herein to identify alterations which might change activity or expression levels or the like.

[0211] In another embodiment, a method for detecting a tumor protein according to the invention is provided based upon an analysis of tissue expressing the gene. Certain tissues such as colon tissues have been found to overexpress the subject gene. The method comprises hybridizing a polynucleotide to mRNA from a sample of tissue that normally expresses the gene. The sample is obtained from a patient suspected of having an abnormality in the gene.

[0212] To detect the presence of mRNA encoding the protein, a sample is obtained from a patient. The sample can be from blood or from a tissue biopsy sample. The sample may be treated to extract the nucleic acids contained therein. The resulting nucleic acid from the sample is subjected to gel electrophoresis or other size separation techniques.

[0213] The mRNA of the sample is contacted with a DNA sequence serving as a probe to form hybrid duplexes. The use of labeled probes as discussed above allows detection of the resulting duplex.

[0214] When using the cDNA encoding the protein or a derivative of the cDNA as a probe, high stringency conditions can be used in order to prevent false positives, that is, the hybridization and apparent detection of the gene nucleotide sequence when in fact an intact and functioning gene is not present. When using sequences derived from the gene cDNA, less stringent conditions could be used, however, this would be a less preferred approach because of the likelihood of false positives. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, length of time and concentration of formamide. These factors are outlined in, for example, Sambrook et al. [Sambrook et al. (1989), supra].

[0215] In order to increase the sensitivity of the detection in a sample of mRNA encoding the protein A or protein B, the technique of reverse transcription/polymerization chain reaction (RT/PCR) can be used to amplify cDNA transcribed from mRNA encoding the colon tumor antigen. The method of RT/PCR is well known in the art, and can be performed as follows. Total cellular RNA is isolated by, for example, the standard guanidium isothiocyanate method and the total RNA is reverse transcribed. The reverse transcription method involves synthesis of DNA on a template of RNA using a reverse transcriptase enzyme and a 3' end primer. Typically, the primer contains an oligo(dT) sequence. The cDNA thus produced is then amplified using the PCR method and gene A or gene B specific primers. [Belyavsky et al., *Nucl. Acid Res.* 17:2919-2932 (1989); Krug and Berger, *Methods in Enzymology,* 152:316-325, Academic Press, NY (1987) which are incorporated by reference].

[0216] The polymerase chain reaction method is performed as described above using two oligonucleotide primers that are substantially complementary to the two flanking regions of the DNA segment to be amplified. Following amplification, the PCR product is then electrophoresed and detected by ethidium bromide staining or by phosphoimaging.

[0217] The present invention further provides for methods to detect the presence of the protein in a sample obtained from a patient. Any method known in the art for detecting proteins can be used. Such methods include, but are not limited to immunodiffusion, immunoelectrophoresis, immunochemical methods, binder-ligand assays, immunohistochemical techniques, agglutination and complement assays. [*Basic and Clinical Immunology,* 217-262, Sites and Terr, eds., Appleton & Lange, Norwalk, Conn., (1991), which is incorporated by reference]. Preferred are binder-ligand immunoassay methods including reacting antibodies with an epitope or epitopes of the colon tumor antigen protein and competitively displacing a labeled colon antigen according to the invention or derivative thereof.

[0218] As used herein, a derivative of the subject colon tumor antigen is intended to include a polypeptide in which certain amino acids have been deleted or replaced or changed to modified or unusual amino acids wherein the derivative is biologically equivalent to gene and wherein the polypeptide derivative cross-reacts with antibodies raised against the protein. By cross-reaction it is meant that an antibody reacts with an antigen other than the one that induced its formation.

[0219] Numerous competitive and non-competitive protein binding immunoassays are well known in the art. Antibodies employed in such assays may be unlabeled, for example as used in agglutination tests, or labeled for use in a wide variety of assay methods. Labels that can be used include radionuclides, enzymes, fluorescent tags, chemiluminescers, enzyme substrates or co-factors, enzyme inhibitors, particles, dyes and the like for use in radioimmunoassay (RIA), enzyme immunoassays, e.g., enzyme-linked immunosorbent assay (ELISA), fluorescent immunoassays and the like.

[0220] A further aspect of this invention relates to a method for selecting, identifying, screening, characterizing or optimizing biologically active compounds, comprising a determination of whether a candidate compound binds, preferably selectively, a target molecule as disclosed above. Such target molecules include nucleic acid sequences, polypeptides and fragments thereof, typically colon-specific antigens, even more preferably extracellular portions thereof. Binding may be assessed in vitro or in vivo, typically in vitro, in cell based or accellular systems. Typically, the target molecule is contacted with the candidate compound in any appropriate device, and the formation of a complex is determined. The target molecule and/or the candidate compound may be immobilized on a support. The compounds identified or selected represent drug candidates or leads for treating cancer diseases, particularly colon cancer.

[0221] While the invention has been described supra, including preferred embodiments, the following examples are provided to further illustrate the invention.

EXAMPLE

Tissue Sources

[0222] Appropriate patient samples were obtained with relevant clinical parameters, and patient consent. Histological assessment was performed on all samples and diagnosis by pathology confirmed the presence and/or absence of malignancy within each sample. Clinical data generally included patient history, physiopathology, and parameters relating to colon cancer physiology. The research specimens were

divided into two groups; early-stage CRC (Dukes' stage A or B) and late-stage CRC (Dukes' stage C or D). Eight matched sets containing normal and malignant samples were obtained for each group, resulting in a total of 32 specimens. Two matched pairs from each group were used for the construction of DATAS™ libraries, the remaining samples were used for expression profiling studies by RT-PCR.

Quality Assessment of Tissue Samples

[0223] Six patient samples were purchased from Integrated Laboratory Services (ILS-Bio). Each patient sample contains a matched pair of normal colon tissue and colon tumor that were obtained during surgery. RNA was isolated from each sample using Trizol and was inspected for quality control following isolation using an Agilent 2100 analyzer. One of the late stage samples was degraded, and this sample was removed from consideration and the analysis was performed with the two remaining samples. To maintain continuity between the comparisons of the early and late stage samples, two samples were used to construct the early-stage analysis. The samples were subjected to a patented process, DATAS™ (U.S. Pat. No. 6,251,590), that uses molecular biology techniques to provide information on a alternative RNA Splicing deregulations associated with diseases, colon cancer in this case. Two DATAS™ libraries were constructed, one from the early stage samples, and one from the late stage samples.

[0224] The selection of the two samples was based on qPCR analysis with markers for CRC that were identified in the literature (table 1). As part of this analysis we examined these markers using both end point RT-PCR and qPCR on the tissue RNA samples. Samples 7140 and 1400 were selected based on their qPCR results for marker CGM2 (see FIG. **1**). This gene is a member of the carcinoembryonic antigen family and has been shown to be down-regulated at an early stage in the progression colorectal cancer. The qPCR data for this marker displayed an appropriate trend with the late-stage samples showing a greater extent of down-regulation relative to the early stage. The qPCR data for early-stage sample, #1481, resembled the late-stage samples and was not included in the DATAS™ library construction to maximize the differences between the early- and late-stage libraries (data not shown).

[0225] A comparison of the qPCR data for each of the late- and early-stage samples used in constructing the DATAS™ libraries against this panel of markers is presented in FIG. **1**

Generation of the DATAS™ Library

[0226] Samples were selected based on their expression of tissue markers (normal vs. tumor). Total RNA of 100 ug of each sample was used to construct the DATAS™ libraries as previously disclosed in U.S. Pat. No. 6,251,590, the disclosure of which is incorporated by reference in its entirety. Briefly, total RNA was isolated from the normal and tumor selected samples and mRNA was subsequently purified from the total RNA for each sample. Synthesis of cDNA was performed using a biotinylated oligo (dT) primer. The biotinylated cDNA was hybridized with the mRNA of the opposite sample to form heteroduplexes between the cDNA and the mRNA. For example, the biotinylated cDNA of the normal colon sample was hybridized with colon tumor mRNA. Similarly, colon tumor biotinylated cDNA was hybridized with colon normal RNA to generate the second DATAS™ library. Streptavidin coated beads were used to purify the complexes by binding the biotin present on the cDNA. The heteroduplexes were digested with RNAse H to degrade the RNA that was complementary to the cDNA. All mRNA sequences that were different from the cDNA remained intact. These single stranded RNA fragments or "loops" were subsequently amplified with degenerate primers and cloned into either pGEM-T or pCR II TOPO vector (Invitrogen) to produce the DATAS™ library.

Clone Sequencing and Bioinformatics Analysis:

[0227] *E. coli* was transformed with the DATAS™ library for the isolation of individual clones using standard molecular biology techniques. From these libraries, 9,600 individual clones were isolated and sequenced using an automated Applied Biosystems 3100 sequencer. The nucleotide sequences that were obtained were submitted to ExonHit's proprietary bioinformatics pipeline for analysis. As the DATAS™ library is prepared with PCR amplified DNA, many copies of the same sequence are present in the clones isolated from the libraries. Therefore it is important to reduce the redundancy of the clones to identify the number of unique, nonrepeating sequences that are isolated. From this large set of DATAS™ fragments, 1709 unique, nonredundant sequences were identified and each DATAS™ fragment was annotated with a candidate gene.

[0228] The annotation was performed by aligning the DATAS™ fragment to the human genome sequenceusing proprietary annotation algorithms. Each DATAS™ fragment sequence was annotated with a corresponding gene that overlapped the genomic sequence containing the DATAST fragment. 1467 genes were annotated with either the Refseq accession number, or a hypothetical gene prediction from different algorithms, for example, Genscan, Twinscan, or Fgenesh++, while 242 DATAS™ fragments that matched the genome had no identified overlapping gene. Identified genes were either matched to the sequence of the DATAS™ fragment (in case of exon to fragment match), or overlapped with the DATAS™ fragment (in case of intron to fragment match), and the full length sequence of the gene was identified. These sequences were further analyzed to detect all potential secreted and membrane spanning proteins. Membrane and secreted proteins were predicted through the use of different algorithms commercially available. For example, TMHMM and SignalP (CBS) were used to identify membrane-spanning domains and signal peptide sequences, respectively, present within the amino acid sequence of the candidate gene. DATAS™ fragments were located within the sequence in an attempt to determine whether the spliced event affected intracellular or extracelullar domains for the transmembrane proteins. Genes associated with the sequence were ranked in order to maximize the identification of successful diagnostic and therapeutic targets. The highest priority genes had characteristics where the gene was a known or predicted membrane secreted protein, the function of the gene was known, and the DATAS™ fragment mapped to an intron. In addition, DATAS™ fragments mapping to the extracellular domain of the protein, indicating that the DATAS™ fragment would be presented outside the cell, and secreted proteins were considered the most viable candidates.

[0229] Based on the bioinformatic analysis, clones were prioritized in three groups:

[0230] A) Genes encoding known secreted proteins with DATAS™ fragments located in introns in the secreted proteins.

[0231] B) Genes encoding known and predicted secreted proteins with DATAS™ fragments matching an exon and the neighboring intron, indicating an exon extension.

[0232] C) Genes encoding known secreted proteins with DATAS™ fragments matching known exons.

Fragments were annotated for the gene name, whether the gene had a positive signal for a transmembrane region or a signal secretory sequence, and the type of event that was identified; a novel event that matched a portion of an intron, or a fragment that matched an exon and overlapped with the neighboring intron, suggesting an exon extension. Seventy-three (73) events were found to be completely novel events, and fifty (50) events suggested an exon extension. Reference genes were captured as known gene sequence (SEQ ID NOS. 45, 50, 60, 63, 68, 74, 80, 86, 92, 93, and 95), with the corresponding known protein sequence (SEQ ID NOS. 46, 51, 61, 64, 75, 81, 87, and 94). This information was used later in the process to identify novel exons, exon extensions, and novel epitopes in the protein structure.

[0233] In addition to the candidates identified through DATAS™ experimental methods, a set of candidate genes were examined for novel, alternatively spliced isoforms that would be synthesized in pathological conditions. Potential splice events were first identified through a proprietary bioinformatic method that uses public and private Expressed Sequence Tags (EST's), and aligns them to the gene of interest, looking for differences that would indicate an alternatively spliced isoform exists. Oligonucleotide primers were designed to this event and the subsequent amplicon produced was sequenced to verify the structure of the RNA. The "bioinformatically derived sequence" entries are SEQ ID NOS. 73, 79, 85, and 91. In addition, multiple sets of oligonucleotides were designed to capture novel events that were not indicated through the bioinformatics process. These potentially novel isoforms were treated similarly as the experimentally identified isoforms below.

Expression Monitoring:

[0234] A valid target for colon cancer requires that its expression be differentially expressed in tumor sample compared to the normal tissue. Assessment of the expression profile for each prioritized sequence was performed by RT-PCR, a procedure well known in the art. A protocol known as touchdown PCR was used, described in the user's manual for the GeneAmp PCR system 9700, Applied Biosystems. Briefly, PCR primers were designed to the DATAS™ fragment, or the bioinformatically identified event, and used for end point RT-PCR analysis. Each RT reaction contained 5 ug of total RNA and was performed in a 100 ul volume using Archive RT Kit (Applied Biosystems). The RT reactions were diluted 1:50 with water and 4 ul of the diluted stock was used in a 50 ul PCR reaction consisting of one cycle at 94° C. for 3 min, 5 cycles at 94° C. for 30 seconds, 60° C. for 30 seconds and 72° C. for 45 seconds, with each cycle reducing the annealing temperature by 0.5 degree. This was followed by 30 cycles at 94° C. for 30 seconds, 55° C. for 30 seconds, and 72° C. for 45 seconds. 15 ul was removed from each reaction for analysis and the reactions were allowed to proceed for an additional 10 cycles. This produced reactions for analysis at 30 and 40 cycles, and allowed the detection of differences in expression where the 40 cycle reactions had saturated. Total RNA was used for all reactions. Expression profiles were generated using matched samples for early stage tumors (8 normal and 8 tumor) and late stage tumors (8 normal and 8 tumor). Profiles that showed a differential expression between normal and tumor are summarized in Table 2. An example of the generated expression profiles is illustrated in FIG. **2**, where the differential expression of the novel event is clearly upregulated in tumor samples when compared to the matched control sample. The score for this clone, SEQ ID NO: 28, with the early stage samples is 8.0, while the score for the late stage samples is 3.3. The score reflects the differential expression of this event in tumor samples versus the matching normal colon tissue from the same patient. The early stage set contains samples obtained from patients with confirmed Dukes' A or B stage cancer, and the late stage set contains samples from patients with confirmed Dukes' C. or D stage cancers. The values are reported where the whole number indicates the number of samples that showed an up regulation in the tumor, and the tenths decimal indicates how many samples showed a down regulation with respect to the tumor (i.e. 3.3 would indicate three samples show an upregulation of the splice event in tumor compared to the matched control, and three samples showed a down regulation in tumor compared to matched controls, while an 8.0 indicates eight (8) samples show an upregulation of the splice event in tumor compared to the matched control, and no (zero) samples showed a down regulation in tumor compared to matched controls. Forty four fragments isolated from DATAS™ were found to be differentially expressed in colon tumor samples as compared to normal colon tissue. These sequences are presented in SEQ ID NOS. 1-44. Bioinformatically derived sequence information were either differentially expressed or novel coding sequence was identified to be expressed in tumors (see below).

In addition, while performing these expression profile assays, we identified that SEQ ID NO: 74 (transcobalamin I), corresponding to the wild-type isoform of SEQ ID NO: 73 was differentially expressed between tumor and normal colon tissues with a score of 6.0 for early and 8.0 for late samples (FIG. **2**B).

Verification of RNA Structure:

[0235] DATAS™ identifies sequences that are altered between the experimental samples. However, the exact sequence of the junctions or borders that the DATAS™ fragment represents sometimes needs to be further characterized. The DATAS™ fragment was used, however, to design experiments that refine the sequence of splice event, provide the exact splice sites used, and the sequence of the coding region was identified experimentally. Primers were designed to amplify a region of the gene larger than the proposed DATAS™ fragment sequence. A similar approach was used for the bioinformatically derived gene set to identify the splice event and its junctions.

[0236] These amplicons were subsequently cloned and sequenced for the identification of the exact junctions of all exons and introns in order to identify the splice sites. This required partial cloning of the isoforms from an identified sample to verify the primary structure (sequence) of the isoforms. RNA samples obtained from three individuals were used as the starting material for RT-PCR amplifications. Each reaction was run in duplicate and the products from each reaction were sub-cloned and sequenced. A consensus sequence was obtained by combining the sequencing results from the six separate reactions. Four samples (2 normal, 2

early tumor) were used for the verification of the mRNA structure of the prioritized genes.

[0237] The confirmed structure and sequence of the clones are found in SEQ ID NOS. 52, 56, 62, 73, 79, 85, and 91. Once the event was identified, the novel nucleic acid sequence that was captured in the amplicons was translated to generate the novel protein sequence of the isoform. The novel gene (nucleic acid) sequences are listed in SEQ ID NOS. 47, 53, 57, 65, 70, 76, 82, 88, and 96. These novel protein sequences are listed in SEQ ID NOS. 48, 54, 58, 66, 71, 77, 83, 89, and 97. Comparisons of the novel protein isoform with the known proteins structure (see above) generated significant differences in amino acid content. This difference was chosen as the target for antibody generation for the detection of the novel protein isoform in tissue or serum. These novel epitopes are listed in SEQ ID NOS. 49, 55, 59, 67, 72, 78, 84, 90, 98 and 99.

Isolation of Full-Length Clones of Isoforms:

[0238] Isolation of the full-length clones containing both isoforms was accomplished utilizing the sequence information and DNA fragments generated during the structure validation process. Several methods are applicable to isolation of the full length clone. Where full sequence information regarding the coding sequence is available, gene specific primers were designed from the sequence and used to amplify the coding sequence directly from the total RNA of the tissue samples. An RT-PCR reaction was set up using these gene specific primers. The RT reaction was performed as described infra, using oligo dT to prime for cDNA. Second strand was produced by standard methods to produce double stranded cDNA. PCR amplification of the gene was accomplished using gene specific primers. PCR consisted of 30 cycles at 94° C. for 30 seconds, 55° C. for 30 seconds, and 72° C. for 45 seconds. The reaction products were analyzed on 1% agarose gels and the amplicons were ligated into prepared vectors with A overhangs for amplicon cloning. 1 ul of the ligation mixture was used to transform *E. coli* for cloning and isolation of the amplicon. Once purified, the plasmid containing the amplicon was sequenced on an ABI 3100 automated sequencer.

[0239] Where limited sequence information was available, 3' and 5' RACE was utilized. Briefly, gene-specific oligonucleotides were designed based on the DATAS™ fragment. The oligonucleotides were used for extension using total RNA from normal colon and colon tumor tissue following the procedures of Sambrook et al (1989). The eluted cDNA was converted to double strand plasmid DNA and used to transform *E. coli* cells and the longest cDNA clone was subjected to DNA sequencing. Full length clones were also obtained using sequence specific primers and following the recommended procedures for the First Choice® RLM_RACE kit produced by Ambion, Inc. (Austin, Tex.) using either single patient or pooled RNA samples. Additionally, 3' RACE was performed when additional sequence information was required for designing sequence-specific oligonucleotides. The full-length clones produced in this manner were sequenced in their entirety to verify their nucleic acid sequence and composition.

REFERENCES

[0240] Alcaraz et al., Cancer Res., 55:3998-4002, 1994.
[0241] Allhoff et al., World J. Urol., 7:12-16, 1989.
[0242] American Cancer Society. Cancer Facts and Figures 2002. 2002.
[0243] An et al., Proc. Amer. Assn. Canc. Res., 36:82, 1995.
[0244] An et al., Molec. Urol., 2: 305-309, 1998.
[0245] Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 1988.
[0246] Babian et al., J. Urol., 156:432-437, 1996.
[0247] Badalament et al., J. Urol., 156:1375-1380, 1996.
[0248] Baichwal and Sugden, In: Gene Transfer, Kucherlapati (Ed.), Plenum Press, New York, pp 117-148, 1986.
[0249] Bangharn et al., J. Mol. Biol. 13: 238-252, 1965.
[0250] Barinaga, Science, 271: 1233, 1996.
[0251] Bedzyk et al., J. Biol. Chem., 265:18615, 1990
[0252] Bell et al., "Gynecological and Genitourinary Tumors," In: Diagnostic Immunopathology, Colvin, Bhan and McCluskey (Eds.), 2nd edition, Ch. 31, Raven Press, New York, pp 579-597, 1995.
[0253] Bellus, J. Macromol. Sci. Pure Appl. Chem., A31 (1):1355-1376, 1994.
[0254] Benvenisty and Neshif, Proc. Nat. Acad. Sci. USA, 83:9551-9555, 1986.
[0255] Berggren et al., Arch Biochem Biophys., 389(1): p. 144-9, 2001.
[0256] Bittner et al., Methods in Enzymol, 153:516-544, 1987.
[0257] Bookstein et al., Science, 247:712-715, 1990a.
[0258] Bookstein et al., Proc. Nat'l Acad. Sci. USA, 87:7762-7767, 1990b.
[0259] Bova et al., Cancer Res., 53:3869-3873, 1993
[0260] Brawn et al., The Colon, 28:295-299, 1996.
[0261] Campbell, In: Monoclonal Antibody Technology, Laboratory Techniques in Biochemistry and Molecular Biology, Burden and Von Knippenberg (Eds.), Vol. 13:75-83, Elsevier, Amsterdam, 1984.
[0262] Capaldi et al., Biochem. Biophys. Res. Comm, 76:425, 1977.
[0263] Carter and Coffey, In: Colon Cancer: The Second Tokyo Symposium, J. P. Karr and H. Yamanak (Eds.), Elsevier, New York, pp 19-27, 1989.
[0264] Carter and Coffey, Colon, 16:3948, 1990.
[0265] Carter et al., Proc. Nat'l Acad. Sci. USA, 87:8751-8755, 1990.
[0266] Carter et al., Proc. Nat'l Acad. Sci. USA, 93: 749-753, 1996.
[0267] Carter et al., *J. Urol.,* 157:2206-2209, 1997.
[0268] Cech et al., Cell, 27:487-496, 1981.
[0269] Chang et al., Hepatology, 14: 124A, 1991.
[0270] Chaudhary et al., Proc. Nat'l Acad. Sci., 87:9491, 1990
[0271] Chen and Okayama, MoL Cell. Biol., 7:2745-2752, 1987.
[0272] Chen et al., Clin. Chem., 41:273-282, 1995a.
[0273] Chen et al., Proc. Am. Urol. Assn, 153:267 A, 1995.
[0274] Chinault and Carbon, "Overlap Hybridization Screening: Isolation and Characterization of Overlapping DNA Fragments Surrounding the LEU2 Gene on Yeast Chromosome III," Gene, 5:111-126, 1979.
[0275] Chomczynski and Sacchi, Anal. Biochem., 162:156-159, 1987.
[0276] Christensson et al., J. Urol., 150:100-105, 1993.
[0277] Coffin, In: Virology, Fields et al. (Eds.), Raven Press, New York, pp 1437-1500, 1990.
[0278] Colberre-Garapin et al., J. Mol. Biol., 150:1, 1981.

[0279] Colvin et al., Diagnostic Immunopathology, 2nd edition, Raven Press, New York, 1995.
[0280] Cooner et al., J. Urol., 143:1146-1154, 1990.
[0281] Couch et al., Am. Rev. Resp. Dis., 88:394-403, 1963.
[0282] Coupar et al., Gene, 68:1-10, 1988.
[0283] Culver et al., Science, 256:1550-1552, 1992.
[0284] Davey et al., EPO No. 329 822.
[0285] Deamer and Uster, "Liposome Preparation: Methods and Mechanisms," In: Liposomes, M. Ostro (Ed.), 1983.
[0286] Diamond et al., J. Urol., 128:729-734, 1982.
[0287] Donahue et al., J. Biol. Chem., 269:8604-8609, 1994.
[0288] Dong et al., Science, 268:884-886, 1995.
[0289] Dubensky et al., Proc. Nat. Acad. Sci. USA, 81:7529-7533, 1984.
[0290] Dumont et al., J. Immunol., 152:992-1003, 1994.
[0291] Elledge et al., Cancer Res. 54:3752-3757, 1994
[0292] European Patent Application EPO No. 320 308
[0293] Fearon et al., Science, 247:47-56, 1990.
[0294] Fechheirner et al., Proc. Natl. Acad. Sci. USA, 84:8463-8467, 1987.
[0295] Forster and Symons, Cell, 49:211-220, 1987.
[0296] Fraley et al., Proc. Natl. Acad. Sci. USA, 76:3348-3352, 1979.
[0297] Friedmann, Science, 244:1275-1281, 1989.
[0298] Freifelder, In: Physical Biochemistry Applications to Biochemistry and Molecular Biology, 2nd ed., Wm. Freeman and Co., New York, N.Y., 1982.
[0299] Frohman, In: PCR Protocols: A Guide to Methods and Applications, Academic Press, New York, 1990.
[0300] Gefter et al., Somatic Cell Genet., 3:231-236, 1977.
[0301] Gerlach et al., Nature (London), 328:802-805, 1987.
[0302] Ghosh-Choudhury et al., EMBO J., 6:1733-1739, 1987.
[0303] Gingeras et al., PCT Application WO 88/10315.
[0304] Ghosh and Bachhawat, In: Liver Diseases, Targeted Diagnosis and Therapy Using Specific Receptors and Ligands, Wu et al. (Eds.), Marcel Dekker, New York, pp 87-104, 1991.
[0305] Goding, In: Monoclonal Antibodies: Principles and Practice, 2nd ed., Academic Press, Orlando, Fla., pp 60-61, 65-66, 71-74, 1986.
[0306] Gomez-Foix et al., J. Biol. Chem., 267:25129-25134, 1992.
[0307] Gopal, Mol. Cell. Biol., 5:1188-1190, 1985.
[0308] Graham et al., J. Gen. Virol., 36:59-72, 1977.
[0309] Graham and van der Eb, Virology, 52:456-467, 1973.
[0310] Graham and Prevec, In: Methods in Molecular Biology: Gene Transfer and Expression Protocols 7, E. J. Murray (Ed.), Humana Press, Clifton, N.J., pp 205-225, 1991.
[0311] Gregoriadis (ed.), In: Drug Carriers in Biology and Medicine, pp 287-341, 1979.
[0312] Grunhaus and Horwitz, Sem. Virol., 3:237-252, 1992.
[0313] Hardcastle et al., Lancet; 348(9040):1472-1477, 1996
[0314] Harland and Weintraub, J. Cell Biol., 101: 1094-1099, 1985.
[0315] Harris et al., J. Urol., 157:1740-1743, 1997.
[0316] Heng et al., Proc. Nat. Acad. Sci. USA, 89: 9509-9513, 1992.
[0317] Hermonat and Muzycska, Proc. Nat. Acad. Sci. USA, 81:6466-6470, 1984.
[0318] Hersdorffer et al., DNA Cell Biol., 9:713-723, 1990.
[0319] Herz and Gerard, Proc. Natl. Acad. Sci. USA, 90:2812-2816, 1993.
[0320] Hess et al., J. Adv. Enzyme Reg., 7:149, 1968.
[0321] Hitzeman et al., J. Biol. Chem., 255:2073, 1980.
[0322] Holland et al., Biochemistry, 17:4900, 1978.
[0323] Horoszewicz, Kawinski and Murphy, Anticancer Res., 7:927-936, 1987.
[0324] Horwich, et al., J. Virol., 64:642-650, 1990.
[0325] Huang et al., Colon, 23: 201-212, 1993.
[0326] Innis et al., In: PCR Protocols, Academic Press, Inc., San Diego Calif., 1990.
[0327] Inouye et al., Nucl. Acids Res., 13:3101-3109, 1985.
[0328] Isaacs et al., Cancer Res., 51:4716-4720, 1991.
[0329] Isaacs et al., Sem. Oncol., 21:1-18, 1994.
[0330] Israeli et al., Cancer Res., 54:1807-1811, 1994.
[0331] Jacobson et al., JAMA, 274:1445-1449, 1995.
[0332] Johnson et al., In: Biotechnology and Pharmacy, Pezzuto et al., (Eds.), Chapman and Hall, New York, 1993.
[0333] Jones, Genetics, 85:12, 1977.
[0334] Jones and Shenk, Cell, 13:181-188, 1978.
[0335] Joyce, Nature, 338:217-244, 1989.
[0336] Kan et al., Genome Res 11, p. 889-900, 2001.
[0337] Kaneda et al., Science, 243:375-378, 1989.
[0338] Kato et al., J. Biol. Chem., 266:3361-3364, 1991.
[0339] Kim and Cech, Proc. Natl. Acad. Sci. USA, 84:8788-8792, 1987.
[0340] Kingsman et al., Gene, 7:141, 1979.
[0341] Klein et al., Nature, 327:70-73, 1987.
[0342] Kohler and Milstein, Nature, 256:495-497, 1975.
[0343] Kohler and Milstein, Eur. J. Immunol., 6:511-519, 1976.
[0344] Kronborg et al., Lancet; 348(9040):1467-1471, 1996
[0345] Kwoh et al., Proc. Nat. Acad. Sci. USA, 86:1173, 1989.
[0346] Landis et al., CA Cancer J. Clin., 48: 6-29, 1998.
[0347] Le Gal La Salle et al., Science, 259:988-990, 1993.
[0348] Levrero et al., Gene, 10 1: 195-202, 1991.
[0349] Liang and Pardee, Science, 257:967-971, 1992.
[0350] Liang and Pardee, U.S. Pat. No. 5,262,311, 1993.
[0351] Liang et al., Cancer Res., 52:6966-6968, 1992.
[0352] Lifton, Science, 272:676, 1996.
[0353] Lilja et al., Clin. Chem., 37:1618-1625, 1991.
[0354] Lithrup et al., Cancer, 74:3146-3150, 1994.
[0355] Lowy et al., Cell, 22:817, 1980.
[0356] Lukas et al., Cancer Res., 61(7): p. 3212-9, 2001.
[0357] Macoska et al., Cancer Res., 54:3824-3830, 1994.
[0358] Mandel et al., N Engl J Med; 328(19):1365-1371. 1993
[0359] Mandel et al., N Engl J Med; 343(22):1603-1607. 2000.
[0360] Mann et al., Cell, 33:153-159, 1983.
[0361] Markowitz et al., J. Virol., 62:1120-1124, 1988.
[0362] Marley et al., Urology, 48(6A): 16-22, 1996.
[0363] McCormack et al., Urology, 45:729-744, 1995.
[0364] Michel and Westhof, J. Mol. Biol. 216:585-610, 1990.
[0365] Midgley and Kerr, Lancet; 353(9150):391-3991999

[0366] Miki et al., Science, 266:66-71, 1994.
[0367] Milech et al., Genes Chromosomes Cancer., 32(3): p. 275-80, 2001.
[0368] Miller et al., PCT Application, WO 89/06700.
[0369] Modrek et al., Nucleic Acids Res 29, p. 2850-2859, 2001.
[0370] Mok et al., Gynecol. Oncol., 52:247-252, 1994.
[0371] Morahan et al., Science 272:1811, 1996.
[0372] Mulligan et al., Proc. Nat'l Acad. Sci. USA, 78:2072, 1981.
[0373] Mulligan, Science, 260:926-932, 1993.
[0374] Murphy et al., Cancer, 78: 809-818, 1996.
[0375] Murphy et al., Colon, 26:164-168, 1995.
[0376] Nakamura et al., In: Handbook of Experimental Immunology, (4th Ed.), Weir, E., Herzenberg, L. A., Blackwell, C., Herzenberg, L. (Eds.), Vol. 1, Chapter 27, Blackwell Scientific Publ., Oxford, 1987.
[0377] Nicolas and Rubinstein, In: Vectors: A Survey of Molecular Cloning Vectors and Their Uses, Rodriguez and Denhardt (Eds.), Butterworth, Stoneham, p 494-513, 1988.
[0378] Nicolau and Sene, Biochim. Biophys. Acta, 721: 185-190, 1982.
[0379] Obermair et al., Gynecol Oncol., 83(2): p. 343-7, 2001.
[0380] O'Dowd et al., J. Urol., 158:687-698, 1997.
[0381] O'Hare et al., Proc. Nat'l Acad. Sci. USA, 78:1527, 1981.
[0382] Oesterling et al., J. Urol., 154:1090-1095, 1995.
[0383] Ohara et al., Proc. Nat'l Acad. Sci. USA, 86:5673-5677, 1989.
[0384] Orozco et al., Urology, 51:186-195, 1998.
[0385] Parker et al., CA Cancer J. Clin., 65:5-27, 1996.
[0386] Partin and Oesterling, Urology, 48 (6A):1-3, 1996.
[0387] Partin and Oesterling, J. Urol., 152:1358-1368, 1994.
[0388] Partin and Oesterling (Eds.), Urology, 48(6A) Supplement:1-87, 1996.
[0389] Paskind et al., Virology, 67:242-248, 1975.
[0390] PCT Application No. PCT/US87/00880
[0391] Pettersson et al., Clin. Chem., 41(10):1480-1488, 1995.
[0392] Piironen et al., Clin. Chem. 42:1034-1041, 1996.
[0393] Poola and Speirs, J Steroid Biochem Mol Biol., 78(5): p. 459-69, 2001.
[0394] Potter et al., Proc. Nat. Acad. Sci. USA, 81:7161-7165, 1984.
[0395] Racher et al., Biotechnology Techniques, 9:169-174, 1995.
[0396] Ragot et al., Nature, 361:647-650, 1993.
[0397] Ralph and Veltri, Advanced Laboratory, 6:51-56, 1997.
[0398] Ralph et al., Proc. Natl. Acad. Sci. USA, 90(22): 10710-10714, 1993.
[0399] Reinhold-Hurek and Shub, Nature, 357:173-176, 1992.
[0400] Renan, Radiother. Oncol., 19:197-218, 1990.
[0401] Ribas de Pouplana and Fothergill-Gilmore, Biochemistry, 33:7047-7055, 1994.
[0402] Rich et al., Hum. Gene Ther., 4:461-476, 1993.
[0403] Ridgeway, In: Vectors: A Survey of Molecular Cloning Vectors and Their Uses, Rodriguez R L, Denhardt D T (Eds.), Butterworth, Stoneham, pp 467-492, 1988.
[0404] Rippe et al., Mol. Cell. Biol., 10:689-695, 1990.
[0405] Rosenfeld et al., Science, 252:431-434, 1991.
[0406] Rosenfeld et al., Cell, 68:143-155, 1992.
[0407] Roux et al., Proc. Nat'l Acad. Sci. USA, 86:9079-9083, 1989.
[0408] Sager et al., FASEB J., 7:964-970, 1993.
[0409] Sambrook et al., (ed.), In: Molecular Cloning, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.
[0410] Santerre et al., Gene, 30: 147-156, 1984.
[0411] Sarver, et al., Science, 247:1222-1225, 1990.
[0412] Scanlon et al., Proc Natl Acad Sci USA, 88:10591-10595, 1991.
[0413] Sidransky et al., Science, 252:706-709, 1991.
[0414] Sidransky et al., Cancer Res., 52:2984-2986, 1992.
[0415] Silver et al., Clin. Cancer Res., 3:81-85, 1997.
[0416] Slamon et al., Science, 224:256-262, 1984.
[0417] Slamon et al., Science, 235:177-182, 1987.
[0418] Slamon et al., Science, 244:707-712, 1989.
[0419] Smith, U.S. Pat. No. 4,215,051.
[0420] Smith et al., CA Cancer J Clin, 51(1):38-75, 2001
[0421] Soh et al., J. Urol., 157:2212-2218, 1997.
[0422] Stenman et al., Cancer Res., 51:222-226, 1991.
[0423] Stinchcomb et al., Nature, 282:39, 1979.
[0424] Stratford-Perricaudet and Perricaudet, In: Human Gene Transfer, O. Cohen-Haguenauer et al., (Eds.), John Libbey Eurotext, France, pp 51-61, 1991.
[0425] Stratford-Perricaudet et al., Hum. Gene. Ther., 1:241-256, 1990.
[0426] Sun and Cohen, Gene, 137:127-132, 1993.
[0427] Szoka and Papahadjopoulos, Proc. Nat'l. Acad. Sci. USA, 75: 4194-4198, 1978.
[0428] Szybalska et al., Proc. Nat'l Acad. Sci. USA, 48:2026, 1962.
[0429] Takahashi et al., Cancer Res., 54:3574-3579, 1994.
[0430] Taparowsky et al., Nature, 300:762-764, 1982.
[0431] Temin, In: Gene Transfer, Kucherlapati R. (Ed.), Plenum Press, New York, pp 149-188: 1986.
[0432] Tooze, In: Molecular Biology of DNA Tumor Viruses, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1991.
[0433] Top et al., J. Infect. Dis., 124:155-160, 1971.
[0434] Tschemper et al., Gene, 10:157, 1980.
[0435] Tur-Kaspaet al., Mol. Cell. Biol., 6:716-718, 1986.
[0436] U.S. patent application Ser. No. 08/692,787
[0437] U.S. Pat. No. 4,196,265
[0438] U.S. Pat. No. 4,215,051
[0439] U.S. Pat. No. 4,683,195
[0440] U.S. Pat. No. 4,683,202
[0441] U.S. Pat. No. 4,800,159
[0442] U.S. Pat. No. 4,883,750
[0443] U.S. Pat. No. 5,354,855
[0444] U.S. Pat. No. 5,359,046
[0445] U.S. Pat. No. 6,251,590
[0446] Varmus et al., Cell, 25:23-36, 1981.
[0447] Veltri et al., J. Cell Biochem., 19(suppl):249-258, 1994.
[0448] Veltri et al., Urology, 48: 685-691, 1996.
[0449] Veltri et al., Sem. Urol. Oncol., 16:106-117, 1998.
[0450] Veltri et al., Urology, 53:139-147, 1999.
[0451] Visakorpi et al., Am. J. Pathol., 145:1-7, 1994.
[0452] Vogelstein et al., N Engl J. Med., 319(9):525-532. 1988
[0453] Wagner et al., Science, 260:1510-1513, 1993.
[0454] Walker et al., Proc. Nat'l Acad. Sci. USA, 89:392-396, 1992.

[0455] Watson et al., Cancer Res., 54:4598-4602, 1994.
[0456] Welsh et al., Nucl. Acids Res., 20:4965-4970, 1992.
[0457] Wigler et at, Cell, 11:223, 1977.
[0458] Wigler et al., Proc. Nat'l Acad. Sci. USA, 77:3567, 1980.
[0459] Wingo et al., CA Cancer J. Clin., 47: 239-242, 1997.
[0460] WO 90/07641, filed Dec. 21, 1990.
[0461] Wong et al., Int. J. Oncol., 3:13-17, 1993.
[0462] Wu and Wu, J. Biol. Chem., 262: 4429-4432, 1987.
[0463] Wu and Wu, Biochemistry, 27: 887-892, 1988.
[0464] Wu and Wu, Adv. Drug Delivery Rev., 12: 159-167, 1993.
[0465] Wu et al., Genomics, 4:560, 1989.
[0466] Yang et al., Proc. Natl. Acad. Sci. USA, 87:9568-9572, 1990.
[0467] Yokoda et al., Cancer Res. 52, 3402-3408, 1992.
[0468] Zlotta et al, J. Urol., 157:1315-1321, 1997.

TABLE 1

Markers for CRC used to evaluate tissue samples

Gene	Marker for:	Reference
CGM2	normal colon; down regulated in colon adenocarcinoma	Down-regulation of carcinoembryonic antigen family member 2 expression is an early event in colorectal tumorigenesis Cancer Res. 57 (9), 1776-1784 (1997).

TABLE 1-continued

Markers for CRC used to evaluate tissue samples

Gene	Marker for:	Reference
cMYC	early tumor	Identification of c-MYC as a target of the APC pathway Science 1998 Sep 4; 281(5382): 1509-12
Cyclin D1	early tumor	The APC tumor suppressor controls entry into S-phase through its ability to regulate the cyclinD/RB pathway. Gastroenterology 2002 Sep; 123(3): 751-63
CEA	metastasis, recurrent tumor	Differences in messenger RNA expression of carcinoembryonic antigen in surgical specimens of colorectal carcinoma. Tumour Biol. 1992; 13(5-6): 330-7
GCCr	detection of colon cancer	Guanylyl cyclase C is a selective marker for metastatic colorectal tumors in human extraintestinal tissues. Proc Nat'l Acad. Sci. USA 1996 Dec 10; 93(25): 14827-32
EBAF	detection of colon cancer	Distinct tumor specific expression of TGFB4 (ebaf)*, a novel human gene of the TGF-beta super-family Front Biosci. 1997 Jul 15; 2: a18-25
PRL3	metastasis	A phosphatase associated with metastasis of colorectal cancer. Science 2001 Nov 9; 294(5545): 1343-6

TABLE 2

Expression Values for DATAS ™ fragments.

SEQ ID NO.	NAME	GenBank Accession	Event type	Score Early	Score Late	total score
SEQ ID NO. 1	EXH-DCTB0155-01	NM_000609	DATAS ™ fragment, extension	1.7	0.8	1.15
SEQ ID NO. 2	EXH-DCTB1031-01	NM_000018	DATAS ™ fragment, extension	0.2	0.6	0.8
SEQ ID NO. 3	EXH-DCTB1047-01	NM_000018	DATAS ™ fragment, Novel	0.0	0.7	0.7
SEQ ID NO. 4	EXH-DCTB1107-01	NM_015449	Novel	4.2	0.7	4.9
SEQ ID NO. 5	EXH-DCTB1704-01	NM_024298	Novel	4.4	0.8	4.12
SEQ ID NO. 6	EXH-DCTB1734-01	XM_300663	Novel	4.4	1.7	5.11
SEQ ID NO. 7	EXH-DCTB1839-01	NM_000196	Extension	0.6	0.7	0.13
SEQ ID NO. 8	EXH-DCTB2178-01	NM_005482	Novel	1.0	6.0	7.0
SEQ ID NO. 9	EXH-DCTB2695-01	NM_001132	Novel	4.4	1.6	5.10
SEQ ID NO. 10	EXH-DCTC0602-01	XM_301613	Novel	2.5	0.8	2.13
SEQ ID NO. 11	EXH-DCTC0638-01	NM_024407	Novel	3.4	0.6	3.10
SEQ ID NO. 12	EXH-DCTC0907-01	NM_004448	Novel	4.4	0.6	4.10
SEQ ID NO. 13	EXH-DCTC1082-01	NM_006707	Extension	2.6	0.8	2.14
SEQ ID NO. 14	EXH-DCTC1227-01	NM_025125	Novel	7.0	0.5	7.5
SEQ ID NO. 15	EXH-DCTC1257-01	NM_005532	Novel	7.0	7.0	14.0
SEQ ID NO. 16	EXH-DCTC1743-01	NM_030652	Novel	3.4	0.8	3.12
SEQ ID NO. 17	EXH-DCTC2089-01	NM_032364	Extension	3.3	0.7	3.10
SEQ ID NO. 18	EXH-DCTC2107-01	NM_015940	Extension	4.3	0.6	4.9
SEQ ID NO. 19	EXH-DCTC2167-01	NM_005514	Extension	2.4	0.7	2.11
SEQ ID NO. 20	EXH-DCTD0627-01	NM_004360	Extension	3.5	0.7	3.12
SEQ ID NO. 21	EXH-DCTD0660-01	NM_004360	Novel	7.1	0.3	7.4
SEQ ID NO. 22	EXH-DCTD0784-01	NM_003171	Extension	1.5	0.7	1.12
SEQ ID NO. 23	EXH-DCTD0841-01	NM_006995	Novel	6.0	3.3	9.3
SEQ ID NO. 24	EXH-DCTD0880-01	NM_021910	Novel	0.4	0.8	0.12
SEQ ID NO. 25	EXH-DCTD0991-01	NM_004990	Extension	5.1	0.7	5.8
SEQ ID NO. 26	EXH-DCTD1192-01	NM_018085	Extension	2.3	0.7	2.10
SEQ ID NO. 27	EXH-DCTD1519-01	NM_145169	Extension	5.1	0.8	5.9
SEQ ID NO. 28	EXH-DCTD1536-01	NM_004121	Novel	8.0	3.3	11.3
SEQ ID NO. 29	EXH-DCTD1650-01	NM_002414	Extension	0.7	0.0	0.7
SEQ ID NO. 30	EXH-DCTD1786-01	NM_002089	Extension	6.0	4.1	10.1

TABLE 2-continued

Expression Values for DATAS ™ fragments.

SEQ ID NO.	NAME	GenBank Accession	Event type	Score Early	Score Late	total score
SEQ ID NO. 31	EXH-DCTD1930-01	XM_049733	Novel	2.6	0.8	2.14
SEQ ID NO. 32	EXH-DCTD2188-01	NM_019074	Extension	6.1	0.1	6.2
SEQ ID NO. 33	EXH-DCTD2280-01	XM_305835	Novel	2.6	0.3	2.9
SEQ ID NO. 34	EXH-DCTD2285-01	NM_023068	Extension	1.7	0.1	1.8
SEQ ID NO. 35	EXH-DCTE0390-01	NM_014610	Extension	6.0	2.0	8.0
SEQ ID NO. 36	EXH-DCTE0424-01	NM_032806	Extension	3.0	0.6	3.6
SEQ ID NO. 37	EXH-DCTE0536-01	NM_138373	Novel	5.1	7.0	12.1
SEQ ID NO. 38	EXH-DCTE0648-01	NM_001712	Novel	1.7	0.8	1.15
SEQ ID NO. 39	EXH-DCTE1361-01	NM_014556	Novel	1.6	0.7	1.13
SEQ ID NO. 40	EXH-DCTE1434-01	NM_005588	Novel	6.1	1.5	7.6
SEQ ID NO. 41	EXH-DCTE1480-01	NM_005588	Extension	2.6	0.7	2.13
SEQ ID NO. 42	EXH-DCTE1956-01	XM_290535	Extension	0.4	0.7	0.11
SEQ ID NO. 43	EXH-DCTE2152-01	NM_000958	Extension	0.6	0.6	0.12
SEQ ID NO. 44	EXH-DCTE2366-01	NM_000114	Extension	1.7	0.8	1.15
SEQ ID NO. 74	EXH-DCDA0001-01	NM_001062	Amplicon	6.0	8.0	14.0
SEQ ID NO. 79	EXH-DCDA0015-01	NM_033049	Amplicon	3.0	0.5	3.5
SEQ ID NO. 85	EXH-DCDA0020-01	NM_000358	Amplicon	5.3	8.0	13.3
SEQ ID NO. 91	EXH-DCDA0037-01	NM_002632	Amplicon	6.0	6.0	12.0

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 99

<210> SEQ ID NO 1
<211> LENGTH: 190
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

gatttttacc cgaagctaaa gtggattcag gagtacctgg agaaagcttt aaacaagtaa 60

gcacaacagc caaaaaggac tttccgctag acccactcga ggaaaactaa aaccttgtga 120

gagatgaaag ggcaaagacg tgggggaggg ggccttaacc atgaggacca ggtgtgtgtg 180

tgggggttcc 190

<210> SEQ ID NO 2
<211> LENGTH: 311
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

tcgggagggg aggaatgacg accactagac tattagaggc tacagccacc acttcccctc 60

aggccctgag ggctaaggag acctggctat gtcctatttt gtgttggact tggttgctgg 120

agccaaaact ttagctcagt tgcatacata ttagtcccaa aggccacatt cagtccacct 180

gctttggctt ttgcacacgt ccccagggtg acttgcaaca gccatctgcc caggagccca 240

gtcctgcccc tcagtcctaa gctcccccaa cccaaattca ctcactgggt aatgcccccg 300

aagcccctgt c 311

<210> SEQ ID NO 3
<211> LENGTH: 224
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 3

-continued

attagaggct acagccacca cttcccctca ggccctgagg gctaaggagg cctggctatg 60

tcctattttg tgttggactt ggttgctgga gccaaaactt tagctcagtt gcatacatat 120

tagtcccaaa ggccacattc agtccacctg ctttggcttt tgcacacgtc cccagggtga 180

cttgcaacag ccatctgccc aggagcccag tcctgcccct gcca 224

<210> SEQ ID NO 4
<211> LENGTH: 242
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 4

gggttgcagg tagggaggcc atggcgtccg gcagtaactg gctctccggg gtgaatgtcg 60

tgctggtgat ggcctacggg agcctggtgt ttgtactgct atttatttttt gtgaagaggc 120

aaatcatgcg ctttgcaatg aaatctcgaa ggggacctca tgtccctgtg ggacacaatg 180

cccccaagga cttgaaagag cagattgata ttccactctc cagggttcag gatatcaagt 240

at 242

<210> SEQ ID NO 5
<211> LENGTH: 401
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 5

ggtcgagagg gatgataacc acaacttccc tggcaactgc agtgtccgac aatttagaag 60

ggaccatcct tggcggcttc tctgaatata ctgagtttgg ttgctaaagg actcatagct 120

tagcaaccat agcccttcaa ggcttttcat ggctgtggcg ggccccatta ggtaccaaaa 180

gaagaagaac cccattgtca gtgaactgta ccacccagcc cacccacctt cctaccctac 240

aggcaccctc tgggccaccc tcccttgctg ccctagcaag tctgacagcc agagggccat 300

tgcctggcca ggatcccttc cttagcatcc ggggctggga cactagcagg cgtcgggagg 360

gggcctggct gagctgcatg tctgtccccc accctcacca a 401

<210> SEQ ID NO 6
<211> LENGTH: 371
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 6

ggattggagg gccttgtttt cccagtaatc cacccaccgc cacttcaaga agaaatgata 60

tgaagaagtg ccggttctcc ctcccctctt ccgcactgtc ccgtgatgat gacgcctcca 120

gagaggacga taatctgggt tcctgggaga gatggcttgg tcactattcc caccccttgcc 180

tcgaccactt gtctcaatgt caccacctca cgccctgttc caggtggctg agtccgaatc 240

cagtaaccac cacctcgttt tggttaatct caggctcggg tgttgtagca acattggaaa 300

tgggaggggt ttacgaagtg aacatgaggt caggtgcctg aattcaacag tctacccatt 360

ccccttgcca a 371

<210> SEQ ID NO 7
<211> LENGTH: 273
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

-continued

<400> SEQUENCE: 7

ggtcatggtt ttggttgggg gtgtagtttt ggctgcagac ctggcgcggg ttaaacagtc 60

ctaattggct ttggctcccc tagctgatcc cactctgacc ttgccactcc ttccccagag 120

tcagtgagaa acgtgggtca gtgggaaaag cgcaagcaat tgctgctggc caacctgcct 180

caagagctgc tgcaggccta cggcaaggac tacatcgagc ttgcatgggc agttcctgca 240

ctcgctacgc ctggccatgt ccgacctacc atc 273

<210> SEQ ID NO 8
<211> LENGTH: 312
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 8

gattttgacc cggagaaccc cacgaaatca tgcaaatcaa gaggttccaa tcttcgtgtt 60

cactttaaga acactcgtga aactgctcag gccatcaagg gtatgcatat acgaaaagcc 120

acgaagtatc tgaaagatgt cactttacag aaacagtgtg taccattccg acgttacaat 180

ggtggagttg gcaggtgtgc gcaggccaag caatggggct ggacacaagg tcggtggccc 240

aaaaagagtg ctgaattttt gctgcacatg cttaaaaacg cagagagtaa tgctgaactt 300

aagggtcaac ct 312

<210> SEQ ID NO 9
<211> LENGTH: 499
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (426)..(426)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 9

ggcatgcaag cttgagtatt ctatagtgtc acctatatac caggcggatg cagaaggggg 60

gacttcccct gggatgacga ggatttccgc agtctggccc ttttgggggc aggcgttgcc 120

atgggatttt tctacctcta ttttcgagat cctggaagag aaatcacgtg gaagcacttt 180

gtacagtatt acctggccag agatctggtg gaccggctgg aagtcgtgaa caaacaatct 240

gtgcgtgtta ttcctgcccc tgggacctct tctgaaacgt agctgggtgc gaagaatcaa 300

attggaaatt atggaatttg tgaatttcct gaagaatccc aagcagtatc aggacttagg 360

agcaaaaatt ccaaagggag ccatgctcac tggtcctcct ggtaccggca agacccttct 420

tgccanagca actgcagggg aggccactgc gcccttcatc actgtgaacg ggtctgagtt 480

cctggaaatg tttattggc 499

<210> SEQ ID NO 10
<211> LENGTH: 492
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 10

tacgtgcagg ggccatcatc ggaagggcct gaggaggagg acggagaagg cttctccttc 60

aaatacagcc ccgggaagct gaggggaaac cagtacaaga agatgatgac caaagaggag 120

ctggaggagg agcagagaac tgaagaataa cgaagttatc cttagcgtcc tcctaaaggc 180

-continued

ttttccttttggcatcttaaaagcttgagagataaaacggaaaccccagagaggagtctg 240

ggcaggctcccagggtgcatgctgcctccataaatctgctgagctctagacctcaatca 300

ggacttgtcccttggctagcaggatcctgggaacaccttggccctgcccctgtgtagaga 360

tgttcatgtctgttcctgtgggtcactttgttaagctgaagagttttaaggaggtagagct 420

cagaccctggactgggatttttcttaccactcaaaacttgctatccacacacccttcattc 480

aataacgctt ct 492

<210> SEQ ID NO 11
<211> LENGTH: 458
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 11

tgaaggcagggaggagcacggtccccaaggagggaggagcgcagtcccca gacagggagc 60

acagtccgcagggagggaggagcgcggtccccagggaggggaggagctcag tctgcaagga 120

gggaggagtacagtccacactcgcaacttgccccattggtctgcactgca ggggaagacg 180

ccaggggcatgggcagagcgcctgacccca ggatggacctctctgtgctg tcaagtcaca 240

gggaggcccaggctgccctctccactgccc ccggggtgacctgaaccgtg cagaacgctg 300

aacaaatcaacgccctttccagggaagcgg aatccaaagtcagagcctgt tcctccactt 360

ttgagaggcacaggatgtgcctctcgcttgcccaaacccccagactcgg gactcagggc 420

tgggctctctgcgcatgagctaatgccgca cgcagcac 458

<210> SEQ ID NO 12
<211> LENGTH: 453
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 12

gatgaagggggggcagcgaggggattgccagggacttggcaggatggcga gatgcagtag 60

ggtgtgctatctggtaaaatatccctggagagggctcagcgctcagacct gaacagcaac 120

agagtggcagaaaaggggcctgggggacactggggccctt cagactatga aaaggttcta 180

aggaggtctgtgttggtggctgtgactgtgggctagtggtgagc cctgtgggct 240

caggcgtcagactacctggattcagaccagctcctgcttccaactttgg ttttttattc 300

ctaaaatgggtattgtaataatacctacctgctggggtgtggcaagaat gaaattaaac 360

agggcttggcacagtgaagcacgggaaaggctttctacagagcagtgact gttgttactc 420

gctgttacaccttaggtaatgcgttttcct ctc 453

<210> SEQ ID NO 13
<211> LENGTH: 482
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 13

tgggcaggctgataaaatggggattgaatgtgaaatacaaatgttctgtt gtcagtctga 60

ggacccaataccattgttgggagacaaagtcacattgtt cttccccctg tctacgtcat 120

cccgacacactcccacataccaccctacatttgtcccac gtccacctcc cagtaatgtt 180

tccctgcttggaaaccctgagaagccaccacactcttcct tgtaaatctc ttctcagagt 240

-continued

gaggcacctc ctggggagct tttctatggg ttacagtttt cagatcagaa acgcagagct 300

tcgggtgagc catctctgga tccagagtca cctccactgg ggtgcggtgg gggagagagg 360

aagcatgagt tactgaagac gaaatctggg ctggagcaag aggtcgaccc tcccactgca 420

gactggttcc gggtaggatc cggccagtcc cactatccca cagccatgta ccttctccag 480

gg 482

<210> SEQ ID NO 14
<211> LENGTH: 124
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 14

gggagagtcc ttgtatgcca tagtattgtg caagcaaaat gagctaagaa tacggataaa 60

ggtatgtgaa tgagaacaag ttatgatgac cccaggaaaa cccttctccg ggttacctgc 120

ataa 124

<210> SEQ ID NO 15
<211> LENGTH: 483
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 15

gcggccgagg tcagcttcac attctcagga actctccttc tttgggtctg gctgaagttg 60

aggatctctt actctctagg ccacggaatt aacccgagca ggcatggagg cctctgctct 120

cacctcatca gcagtgacca gtgtggccaa agtggtcagg gtggcctctg gctctgccgt 180

agttttgccc ctggccagga ttgctacagt tgtgattgga ggagttgtgg ccatggcggc 240

tgtgcccatg gtgctcagtg ccatgggctt cactgcggcg ggaatcgcct cgtcctccat 300

agcagccaag atgatgtccg tggcggccat tgccaatggg ggtggagttg cctcgggcag 360

ccttgtggct actctgcagt cactgggagc aactggactc tccggattga ccaagttcat 420

cctgggctcc attgggtctg ccattgcggc tgtcattgcg aggttctact agctccctgc 480

acc 483

<210> SEQ ID NO 16
<211> LENGTH: 447
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 16

gggcaaggcg gaaaaagatg agcgcgctct gaagcaggag attcacgagc tgcgagggcg 60

cctggagcgg ctggacaggt gagccaagcc tgctgggtgg ggcgaggcca gacgtcactg 120

tcaataccct gaggcatctc ttcctttcta gtgggccggt caggctgggg cctgggtcag 180

agcggtgctg cccgtgccgc ctgaagagct gcagccagaa caggtggctg agctgtgggg 240

ccggggtgac cggatcgaat ctctcagcga ccaggtgctg ctgctggagg agaggctagg 300

tgcctgctcc tgtgaggaca acagcctggg cctcggcgtc aatcatcgat aagaagcctc 360

tacagcaccc ctgcccccta atttatacag aaaccggacc cgctaatcct ctgggattgg 420

ccgactgtga gctgcagata aggctat 447

-continued

<210> SEQ ID NO 17
<211> LENGTH: 255
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 17

tgaagagggg gccttcaagg ttttgcgagc agcttgggac attgtcagca atgctgaaaa 60

gcgaaaggag tatgagatgt aagttggaga tgggaagtca tcagataatg gtaaatgaaa 120

aatcctcaat agcagaggca tctggacttg ggggtggagg cttgttgaga tggagagaac 180

tgaagtcact tgtctttctc gctagacagg ggcctcaaga ggccaactga tatgtcttcc 240

tttgtcccta cactt 255

<210> SEQ ID NO 18
<211> LENGTH: 312
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 18

tggttgcagg gggaaatgga gctgaccctg ttgctgtaag tttctgacaa tttctccagt 60

actttctttg gacctgcttc gagcaacagg attttcttgt catgaaagtg aatatcatat 120

ccttaaaaaa aaaaaattta tatcatctta ggaaaatcaa gaatttcctt tcccattacc 180

aagagaaaca acagagtaac cacttaaccc ttaaaatatg aatttcttac agagaacata 240

aacactatac ccatgcagag tcctaagcga taacatatac caccatttta gctatgagaa 300

ccaccctcca ca 312

<210> SEQ ID NO 19
<211> LENGTH: 389
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (274)..(274)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 19

tatggtgagg ggaaggtccc tgctaaggac agaccttagg agagcagttg gtccaggacc 60

cacacttgct ttcctcgtgt ttcctgatcc tgccctgggt ctgtagtcat acttctggaa 120

attccttttg ggtccaagac gaggaggttc ctctaagatc tcatggccct gcttcctccc 180

agtcccctca caggacattt tcttcccaca ggtggaaaag gagggagcta ctctcaggct 240

gcgtgtaagt ggtgggggtg ggagtgtgga gganctcacc caccccataa ttcctcctgt 300

cccacgtctc ctgcgggctc tgaccaggtc ctgtttttgt tctactccag gcagcgacag 360

tgcccaaggc tctgatgtgt ctctcacag 389

<210> SEQ ID NO 20
<211> LENGTH: 374
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 20

gggatgaacc caggaggcgg aggctgcagt gagccgagat tgtaccattg caccccagtg 60

cgcccattgc ctggcctctc cttgtcacat cttctccttg aagcttgctt tcagttagcc 120

aggtgtggtg gtgcatgcct gtggtcccag ctactctgga ggctgaggtg ggaggattgc 180

-continued

ttgagcccag gaggttgagg ctacagtgag ccatgatcgc tcaaatacac tccagcctgg 240

tgacagtgag atcttatctc aaaagaacaa caaaaaaaga ggaatccttt agctccctga 300

gactcagctc tgctagcagt cttggtactt tgtaaatgac acatctcttt gctctgcagt 360

acaagggtca cccc 374

<210> SEQ ID NO 21
<211> LENGTH: 271
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 21

gggatgaacc caggaggcgg aggctgcagt gagccgagat tgtaccattg caccccagtg 60

cgcccattgc ctggcctctc cttgtcacat cttctccttg aagcttgctt tcagttagcc 120

aggtgtggtg gtgcatgcct gtggtcccag ctactctgga ggctgaggtg ggaggattgc 180

ttgagcccag gaggttgagg ctacagagag ccatgatcgc tcaaatacac tccagtctgg 240

tgacagtgag atcttatctc ataagaacaa c 271

<210> SEQ ID NO 22
<211> LENGTH: 197
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 22

tgagtggagg cagaggtcgc agtgagccga gatagtacca ttgcattcca gcctggacaa 60

cagagcaaga ctctgtcttg aaaaaaaaaa aattaaaaat accaaaccat agtttcttaa 120

acagcatata cttacaagga gttgtaacac tgcccatccc aactgtacaa aaaaaatgtg 180

aagcctgttt cccattt 197

<210> SEQ ID NO 23
<211> LENGTH: 443
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 23

aatagggacc atagggaact ggggtcagtt catcaacggt gtcccagcaa tgatgcatca 60

tggctcttct gtcagggaac cccagcagct cttaatgaac cctgcagttt acatccccag 120

gacccttttct cttttggaat aatttaatgc atgattatca aaggtttccc atgcatagta 180

gcctcctaag agttacattt cagaagtcat tgctccttga ttctggatca atttgaaata 240

agataatgct gtcttagcaa ttcatatcct gggaatggta tagactgtgc acagttgaag 300

acctatgtga ggaaaaggcc ataagggtcc ccagtaagcc tgatgcaatg gtgtacaaat 360

aggtggccag aaatacctttt ctgcctctaa gagagcactg atattctgca tgtagtctgc 420

caggggaagg ggtttgggtg agt 443

<210> SEQ ID NO 24
<211> LENGTH: 334
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 24

tcagggaagg gggtatagtg gggccttgca ggccagaggt ggcttggagg agcccctgga 60

-continued

aagaggctta agaggtgaga ctcaacagcc atggcgacag agcatagggc tttaagatga 120

atttgcaggg gttacaggat tacaactgca atgtgggcta ataatagtgc cccctgcatt 180

aagctgcaga gattgagcga gtaagtggga agctgagaaa atgcccccat ggggtagaca 240

ctcaataagc atctgctgtt attaccagga ctcgtatggt catgggtgac agcttcagac 300

cacaggcagt ccacctacaa cctgtgcccc ttgc 334

<210> SEQ ID NO 25
<211> LENGTH: 321
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 25

tatgggcagg taacagtgag gcaagaaaaa gatctattta ggattcagct tgtccagtct 60

cccacagggc ttaagcttca tacttgtttt gtcacttcat ccatcagcgc ttgtatctgc 120

tgtggcttgg ctgttgtaac agtctctaca actgctggct tcggggacgt ttttgcctgg 180

agaacaacaa agttatcacc aacaaccata aatatcccct aacctccagt tttatacagc 240

atctcagagg gaaagtggtt acccttaagt cgaaggtctc ttctagttaa gacaggaaag 300

aaaaactgta agtgaggaag c 321

<210> SEQ ID NO 26
<211> LENGTH: 239
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 26

ttgtcaggca gggggattcc gtgttgcctt atccacttca gatcccgtcg tcgcctcact 60

ggctcaggac atcttcaagg agctgtccca gattgaagcc tgtcagggcc caatgcaaat 120

gaggctgatt cccactctgg tcagcataat gcaggcccca gcagacaaga ttcctgcagg 180

gctttgtgcg acagccattg atatcctgac aacagtagta cgaaatacaa agcctcacc 239

<210> SEQ ID NO 27
<211> LENGTH: 478
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 27

ggtggaggca acagagtggc ggccgctacg gccctgtaac agggccatgg agaagctgcg 60

gcgagtcctg agcggccagg acgacgagga gcagggcctg actgcgcagg tcctggatgc 120

ctcatccctt agtttcaaca ccagattgaa atggtttgcc atctgcttcg tatgtggcgt 180

tttcttttct attcttggaa ctggattgct gtggcttccg ggcggcataa agctttttgc 240

agtgtttttat accctcggca atcttgctgc gttagccagt acatgctttt taatgggacc 300

tgtgaagcaa ctgaagaaaa tgtttgaagc aacaagattg cttgcaataa ttgttatgct 360

tttgtgtttc atatttaccc tgtgtgctgc tctttggtgg cataagaagg gactggccgt 420

gttattctgc atattgcagt tcttgtcaat gacctggtat agcctgcgcc ataacgct 478

<210> SEQ ID NO 28
<211> LENGTH: 337
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

-continued

<400> SEQUENCE: 28

tggccaggct ggtcttgaac tcctgagctc aaaagatcct cctgcgccgg cctcccaaag 60

tgctgggatt acaggtgtga gccaccacgc ccagcccaat ttttgatttt cttttagaga 120

cggtgtcttg ctttgttgcc ctggctgttc tcaaactcct ggcctcaagt gatcctgcca 180

ccttggctcc caagtggcta agactgtagg tatgtgccac cacgcctggc ttttttttat 240

ttcttatttt ttatttttta gaggcaggat cttgctgcat tgcccaggct ggtctcaaac 300

tcctggcctc aggtgatcct ccttcctcgg cctactc 337

<210> SEQ ID NO 29
<211> LENGTH: 402
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 29

gggccaggca gccacaggaa agaaggggaa gaggccgacg ccccaggcgt gatccccggg 60

attgtggggg ctgtcgtggt cgccgtggct ggagccatct ctagcttcat tgcttaccag 120

aaaaagaagc tatgcttcaa agaaaatgca gaacaagggg aggtggacat ggagagccac 180

cggaatgcca acgcagagcc agctggtaag aaggacgggg aacgatggct tgcacacgtg 240

gccagtgttc ccattttatc ttctccatcc tctcccatct tgctgtcctg ctcacattct 300

caaatttggt tgcatggctt tgaatgtctt cctttatgtc tcgttgcttt ggagggatac 360

tttcaaaaga caatgaatgt gtaaacttcc tagggcctac ac 402

<210> SEQ ID NO 30
<211> LENGTH: 251
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 30

ttgttagacc cgcatcgccc atggttaaga aaatcatcga aaagatgctg aaaaagtaag 60

ttataatttc catgtacaca ggcgactgga gctgttggtc agaaatactg gcgtctgccc 120

cctaaaaagt aaatcaggaa aacccagggt tagctgcagg actgaaaaaa ttattatttt 180

cacaaagttg ccattaaggt tattaatctg ttctggtgcc agaggatatt cccagtgccc 240

agggtatccc c 251

<210> SEQ ID NO 31
<211> LENGTH: 397
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 31

tgttgttacc ccggagatgg aggttgcagt gagccaagat cgtgccattg cactccagcc 60

tgggcaacag ggcaagattc cgtctcaaaa acaaacacta ttagaaaatg ccctggaggt 120

ggcggggagt tgttgatttg tgaggacaga ttgaaagcaa ctcccagggt ggccttgtcc 180

acctccccgt cgagaatgtg gctgccggcc tctttgaaga ttgtggtctg gcataaggag 240

aggtgcaggc gcctggttct gagcaccttg gaatttccag ccgcacagca tctggtgccc 300

tcccctccac cctcacaagg agctgccatc ctgtttggat tttctgtttg tggaccagaa 360

acaaacgttt ttccaaagga ttagcaaata gggtacc 397

-continued

<210> SEQ ID NO 32
<211> LENGTH: 381
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 32

tgtaggaagg gcaggaaggg cacaaggggt gaggggcccc ctcacccatc agaggctggc 60

cagcagtgcc gtggctgccc aacccagcac caatacagca ggaaaggacc agaccaggct 120

tgccgccatg gttcctgttt cacaccctta gtggcaatga ggccagcagg gacccaagag 180

tccactgaac ttggcctgcc aaggcagggt tctgacctac ccttcaggaa tggcccatgg 240

gtcaagagtg accaggccac tttccctccc acaaaaggca gaaggaaagg cgagcccagc 300

actcacctcc gtggcaatga cacattcatt cctctcctct gatatcaaac acacagactg 360

gtacatggag tccctacaca t 381

<210> SEQ ID NO 33
<211> LENGTH: 230
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 33

gcctgccagg ttggtctcaa actcatatcc tcaagtgatc caccctcctc gtcctcccaa 60

agtgctggga ttttgccatg agccaccacg ctcagccatg tttagccatt tttaaaaggt 120

gtgaacagat aattaagtct attagcaaca ttagaaaatt taagttaaaa tttaaatgac 180

tctagcaaga attagccttt aggtgccctg ggttcccctc ctcacctccc 230

<210> SEQ ID NO 34
<211> LENGTH: 452
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 34

gccggggagg tggggagggg agaggctctc tgagtccact ccgggccttg cagaaggagg 60

cgtgtttgta agcagagcat gggcgagaat tcggtggaga tggcttttca gaaagagacc 120

acgcaggtat ccattcaaca cacacgggca gatataatac ccagctaacc cacagtgcca 180

actggcatct ggagactgct gtcatttaga ggatggtgcc tgggaccaag gggaggttgg 240

ggagagtgtg cagggaacca ggagcccctc aggagatagg agagccctgt tctgctcggg 300

gcagagggaa tccctgtgtc agctggcggc tggcttagag gactagggga gactttctgg 360

acttggcagc atcccctcca tttccaagac acgtatgttc tgactccccc acctttgttc 420

tgaggcctct tcccactccc agcctgcccg tc 452

<210> SEQ ID NO 35
<211> LENGTH: 462
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 35

aagggtcgaa ctccaaggca gaagaaagaa tatctctcag cactaatgct ccccttccct 60

ccccctaacc cagaacatcc cttgggttat cgcaggtgaa tactccaatc agatgccaca 120

ttgatgccag gcttgcgcag gaccaacaca gaggtctcag ggtcatgctg gaaggacagg 180

-continued

cggctttctg gagatccttt tgtctggagt accacagctg ctggctttcc agcccctatt 240

atcaccaccc gctcaatcca gattggtgtc tcaaagtgtc cttcagggtc tgctgagctg 300

gagacaaggg tgttgccaga gaatgagaat cgacgcagca ggaactcttg gcgagtctgg 360

tagttgaacg tgtgcccatc atccagaaag agctctcctt gagctgtacc ctgagggcta 420

agtgcacaaa gagagtgatg gggtcatcct tcatacattc tg 462

<210> SEQ ID NO 36
<211> LENGTH: 467
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 36

aggaggtggc ttcgtcccgc ggagtccagg cttcaggctt tcaccagttc tcaggatgcc 60

catagggatg ggtgaagcct gcctggcctg tggtgcttcc cagtggccgt catctcatta 120

gggccccaca gtggcattag gatgcacctc tcggcggtgt tcaacgccct cctggtgtcg 180

gtgcaggcag cggtcctgtg gaagcatgtg cggctgcgtg agcatgcagc cacactggag 240

gaggagctgg ccctcagccg acaggccaca gagccagccc cagcactgag gatcgactac 300

ccggaggcac tgcagatcct gatggagggc ggcacacaca tggtgtgcac gggccgcacg 360

cacacagacc gcatctgccg cttcaagtgg ctctgctact ccaacgaggc tgaggagttc 420

atcttcttcc atggcaacac ctctgtcatg ctgcccaacc tctcacc 467

<210> SEQ ID NO 37
<211> LENGTH: 476
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 37

ttgagaggca gtgtctggct ccgggtggag cctcccaggt gctcttacag cctgttccaa 60

gtgtggctta atccgtctcc accaccagat ctttctccgt ggattcctct gctaagaccg 120

ctgccatgcc agtgacggta acccgcacca ccatcacaac caccacgacg tcatcttcgg 180

gcctggggtc ccccatgatt gtggggtccc ctcgggccct gacacagccc ctgggtctcc 240

ttcgcctgct gcagctggtg tctacctgcg tggccttctc gctggtggct agcgtgggcg 300

cctggacggg gtccatgggc aactggtcca tgttcacctg gtgcttctgc ttctccgtga 360

ccctgatcat cctcatcgtg gagctgtgcg ggctccaggc ccgcttcccc ctgtcttggc 420

gcaacttccc catcaccttc gcctgctatg cggccctctt ctgcctctcg gcctca 476

<210> SEQ ID NO 38
<211> LENGTH: 381
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 38

tgagggcacc caggaggtgg agtttgcagt gagccaagat tgcgccattg cgctccagcc 60

tgggcgacag agggaaactc tgtctcaaaa aaaaaagaag ttttcatttt acctttcaa 120

tagagcaaac atacatactg aaaagtgcta aatcataatt atactttcac aatgtatgaa 180

ctttcacaag ttgaacacac ccaggtaagc agcacccaaa tcatgaaata ctacatgacc 240

tcccccttaaa aagagaccct tccagccagg tgcagtggct cacacctgta atcccagcac 300

-continued

tctgggaggc caaggcaggt ggatcactag aagtcaggag ttctaaacca gcctggccaa 360

catagtgaaa ccttgttcta c 381

<210> SEQ ID NO 39
<211> LENGTH: 421
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 39

gggattgagg ggagagagaa gaggatcagc agtgcagccc caggatgcct cgcagacgtg 60

ccttctgcca tgattgtgaa gttccagcca cgtggaactt gtatctctag ctcatggaaa 120

tggaaggaag aaatgtgaag tttaggagat tggagcttcc agtctgaggc tgatccatac 180

aatacaagaa gaaagagagc caggaaaggc ccaacaggag ggaatcaatg caagatttga 240

gcctattttg tggaagcccg cccatcagcc gtcttccctc tcttcgcctc cttgggtctc 300

aggaactcag ccttcagaag tcatgaactg aatctctcct cccacccccа catccaccag 360

tctctaccaa gagtgagtcc tcaggaagca gccaccacca ttcgttatca cgcgggggga 420

c 421

<210> SEQ ID NO 40
<211> LENGTH: 400
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 40

gggaggttaa aacaatgtgt tgtaggagaa tcacatttta ttcaattggt caaggactca 60

catctttatt ttaattgtcc ttgaaagaaa ctgaaatttt agtgtcatct ttttgataat 120

tttatcagag gtaaccgttc actgatttca atttagtttc catcagccta tccatgctaa 180

ttttatgctt cattttgcaa attcaataaa taaaaatgtg actggattac tttaaaataa 240

aagtagctaa tactgtggga atcagctgga ttcaaagcca cttttccatg tcactagggg 300

tgatgggaga tagaagcagc ctggggaact ttcaccagaa gtgggagtta attgtagtca 360

cttactaatc ctggatcttc ctcatgactc tctatttcct 400

<210> SEQ ID NO 41
<211> LENGTH: 416
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 41

gggaggttaa aacaatgtgt tgtaggagaa tcacatttta ttcaattggt caaggactca 60

catctttatt ttaattgtcc ttgaaagaaa ctgaaatttt agtgtcatct ttttgataat 120

tttatcagag gtaaccgttc actgatttca atttagtttc catcagccta tccatgctaa 180

ttttatgctt cattttgcaa attcaataaa taaaaatgtg actggattac tttaaaataa 240

aagtagctaa tactgtggga atcagctgga ttcaaagcca cttttccatg tcactagggg 300

tgatgggaga tagaagcagc ctggggaact ttcaccagaa gtgggagtta attgtagtca 360

gttactaatc ctggatcttc ctcatgactc tctatttcct gaagatatca cccacc 416

<210> SEQ ID NO 42
<211> LENGTH: 450

-continued

<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 42

ggggaggtgg ccaaagtccc aactgtgagc caggccccac attcactggg cctcctccag 60

ggtctgtatg ccatggaacc ctggacatgg ggctatgaag gaaggtgggt gttgctaagc 120

ccaggagcat gggcccctaa ccttggccct gtgccccagg tgaggctggt gccaagttca 180

ttgaggtatc taaagaggcc cggaagcggt tcctgggccc cctgcacccc tccttcaacc 240

tggtaaagat catccgcagt ttcctgctga aggtcctgcc tgctgatagc catgagcatg 300

ccagtgggcg cctgggcatc tccctgaccc gcgtgtcaga cggcgagaat gtcattatat 360

cccacttcaa ctccaaggac gagctcatcc aggccaatgt ctgtagcggt ttcatccccg 420

tgtactgtgg gctcatccct ccctccctcc 450

<210> SEQ ID NO 43
<211> LENGTH: 299
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 43

cttgttcagg ctggtgggag cggcagggct gggcctgccc ctaaggggag ctccctgcaa 60

gtcacatttc ccagtgaaac actgaactta tcagaaaaat gtatataata ggcaaggaaa 120

gaaatacagt actgtttctg gacccttata aaatcctgtg caatagacac atacatgtca 180

catttagctg tgctcagaag ggctatcatc atcctacaac tcacattaga gaacatcctg 240

gcttttgagc acttttcaaa caatcaagtt gactcacgtg ggtcctgagg cctataaac 299

<210> SEQ ID NO 44
<211> LENGTH: 213
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 44

actgccgggc atgagctttg gatggtggag gtctaaagcc tgcttgcttt gttggtcctt 60

gacttcaacc tgtcttgggg aaaggggggca aaagaaaaaa aagaaaggta tgttatgaaa 120

gtgtacctct gtagttctgt gaagctgaaa aagggaacgt cttcccactg cctctccagc 180

ctgttccctg attggaactg tatgtaccct atc 213

<210> SEQ ID NO 45
<211> LENGTH: 2040
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 45

gaagcgcgct cccggggagg tgttgcagcc atggctacgg cagccggcgc gacctacttt 60

cagcgaggca gtctgttctg gttcacagtc atcaccctca gctttggcta ctacacatgg 120

gttgtcttct ggcctcagag tatcccttat cagaaccttg ggcccctggg ccccttcact 180

cagtacttgg tggaccacca tcacaccctc ctgtgcaatg ggtattggct tgcctggctg 240

attcatgtgg gagagtcctt gtatgccata gtattgtgca agcataaagg catcacaagt 300

ggtcgggctc agctactctg gttcctacag actttcttct ttgggatagc gtctctcacc 360

atcttgattg cttacaaacg gaagcgccaa aaacaaactt gaagttgtct gaaagcttgc 420

-continued

tctacacttt tacattcatc ctcacccttt tttttgtggg gtagaggagg tgcagtaatt 480

tactcagtga tctttctact ttctagaaac tgtccttcaa agctctttaa gacccctcg 540

ttagtcagtt ttttctctta tatgctctgg ttgagcttga atagaccagt tgttacttaa 600

gaaagaaaca gagaaagatt ttagcttttc aatcctattt ggcagaggac ttcagctacc 660

ttcttacagt ctttggctgt gttggtaccc tcgtgtgctc tgagctaagc cacatactaa 720

actgactttt tggtttgtat acccttgctc ccgccttctg atgaaaacac cttaccctca 780

caaccaccat ctttcctctc ctttccaaag ctctttccac cttgctgcac taagataaag 840

tgacactttc actatatgtc aattccacac acatttatta ggtacctgtg aggtaggatc 900

ctatcctctc aaacttccat ttctcatgct acagagaaag ataaggaaga tgagcaagtg 960

cctggaatgg ggcaggctga gcagtcacac aggcatagag gcacgctgag aacctggagg 1020

ggagactgca gagtgccttc cctgatgctg cagccggaag tgatccttcc ctccacctgg 1080

cccctgggac actgtgctct gcagtgtgca gggcctgatg gcactgctag attgctcctt 1140

cagctcaggg ccacagctta aacagcttta cctttccccc cagcacctgt cccactatct 1200

tgcacacagg tgctctaacc atgtttattg aacaaaggag ggaaactgat ttcactttca 1260

cttgttcatt atcattccaa tttttatgtg aaaatggcac aacccatttg gggtaccctc 1320

accccaaaat aaaagcccaa gtctaccttt gactggtacc acctttttttg tggtttcgtt 1380

ggtgagaaac ctttatcttt ttcatacctt tctattctca atcacttctc caaaagtgtg 1440

tctttccagc tctgatttat tcaaaacaca agcatttctg tttagagatt ctagcccatg 1500

ggttatctgg ctagttatta cctctcctgt tcacttagtt atactttatt attgctcaca 1560

ggctggggag gcagaatgac tctgtcacca ctaggagcca ttagggcttc ttccctggag 1620

gactgcctgc ttgctttctg gggacactag ccctcatttc ccttctgtgg tacagtgggg 1680

caaattattt gtattaagca aacatttatg ggaaacaacc cgctcccgaa aacggagccc 1740

ccaagtaaag cacaaccctg aaagattatg aactatgaat tgtctctagt agagataaat 1800

ttctgcaaac atatctcagt cttccctctg tttctctggt gattaagaag ttccttttttg 1860

gtaaggaaaa ggatttttaa ccatagagtt aggcatcatg gaaattcaaa ccagatttct 1920

taatacctgg tcttcctcaa agagaaataa taacagtaat agtggtgctg ggaacaatat 1980

ggcagattat tgaatgaaat tgattaactt gaataaaatg ctgtgaattt tctctaaaaa 2040

<210> SEQ ID NO 46
<211> LENGTH: 123
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 46

Met Ala Thr Ala Ala Gly Ala Thr Tyr Phe Gln Arg Gly Ser Leu Phe
1 5 10 15

Trp Phe Thr Val Ile Thr Leu Ser Phe Gly Tyr Tyr Thr Trp Val Val
20 25 30

Phe Trp Pro Gln Ser Ile Pro Tyr Gln Asn Leu Gly Pro Leu Gly Pro
35 40 45

Phe Thr Gln Tyr Leu Val Asp His His His Thr Leu Leu Cys Asn Gly
50 55 60

Tyr Trp Leu Ala Trp Leu Ile His Val Gly Glu Ser Leu Tyr Ala Ile

-continued

65 70 75 80

Val Leu Cys Lys His Lys Gly Ile Thr Ser Gly Arg Ala Gln Leu Leu
85 90 95

Trp Phe Leu Gln Thr Phe Phe Phe Gly Ile Ala Ser Leu Thr Ile Leu
100 105 110

Ile Ala Tyr Lys Arg Lys Arg Gln Lys Gln Thr
115 120

<210> SEQ ID NO 47
<211> LENGTH: 291
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 47

atggctacgg cagccggcgc gacctacttt cagcgaggca gtctgttctg gttcacagtc 60

atcaccctca gctttggcta ctacacatgg gttgtcttct ggcctcagag tatcccttat 120

cagaaccttg ggcccctggg ccccttcact cagtacttgg tggaccacca tcacaccctc 180

ctgtgcaatg ggtattggct tgcctggctg attcatgtgg gagagtcctt gtatgccata 240

gtattgtgca agcaaaatga gctaagaata cggataaagg tatgtgaatg a 291

<210> SEQ ID NO 48
<211> LENGTH: 96
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 48

Met Ala Thr Ala Ala Gly Ala Thr Tyr Phe Gln Arg Gly Ser Leu Phe
1 5 10 15

Trp Phe Thr Val Ile Thr Leu Ser Phe Gly Tyr Tyr Thr Trp Val Val
20 25 30

Phe Trp Pro Gln Ser Ile Pro Tyr Gln Asn Leu Gly Pro Leu Gly Pro
35 40 45

Phe Thr Gln Tyr Leu Val Asp His His His Thr Leu Leu Cys Asn Gly
50 55 60

Tyr Trp Leu Ala Trp Leu Ile His Val Gly Glu Ser Leu Tyr Ala Ile
65 70 75 80

Val Leu Cys Lys Gln Asn Glu Leu Arg Ile Arg Ile Lys Val Cys Glu
85 90 95

<210> SEQ ID NO 49
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 49

Gln Asn Glu Leu Arg Ile Arg Ile Lys Val Cys Glu
1 5 10

<210> SEQ ID NO 50
<211> LENGTH: 2414
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 50

ggggtgaggg cagcagctcg ccacagctgc cagccatctg tccattcacc catctgtcca 60

-continued

tctggcagcc cgctgttcag acctgtctgt ctgtccgccc atctctgtaa gcccatctct 120

gtcccattgt ctatctgacc atctttctct tactgtcctc tttgtctagc tatctggcct 180

atctgtcgat ccatcttcgt gtctgtcttc agcccccacc tgtttttgtc catctgtcca 240

attacctgtg actctgtgca tcttcttgtc cattcatctg cccacccatc cgtccctccg 300

tctgcccacc agccgccccc ctcctcctgg gctgcagagc catggcccgg ggctacgggg 360

ccacggtcag cctagtcctg ctgggtctgg ggctggcgct ggctgtcatt gtgctggctg 420

tggtcctctc tcgacaccag gccccatgtg gcccccaggc ctttgcccac gctgctgttg 480

ccgccgactc caaggtctgc tcggatattg gacgagccat cctccagcag cagggctcac 540

ccgtggatgc caccatcgcg gctctggtct gcaccagcgt cgtcaaccct cagagcatgg 600

gcctgggcgg aggggtcatc ttcaccatct acaatgtgac aacagggaag gtggaggtca 660

tcaatgcccg ggagacggtg ccggccagcc acgccccgag cctgctggac cagtgtgcac 720

aggctctgcc actgggcaca ggggcccagt ggatcggggt gcccggggag ctccgtggct 780

atgccgaggc ccaccgccgc catggccgcc tgccctgggc gcagctgttc cagcccacca 840

tcgcgctgct ccgaggggggg catgtggtgg cccctgtcct cagccgtttc ctgcacaaca 900

gcatcctgcg gccttccttg caggcgtcaa ccctgcgcca gctcttcttc aacgggacag 960

aacccctgag gcctcaggac ccactcccat ggcctgcact ggccaccacc ctggagaccg 1020

tggccacaga gggcgtggag gtcttctaca cggggaggct gggccagatg ctggtggagg 1080

acattgccaa ggaagggagc cagctgacgc tgcaggacct ggccaagttc cagcccgagg 1140

tggtggatgc cctggaggtg cccctggggg actataccct gtactcacca ccgccgcctg 1200

caggggggtgc cattctcagc tttatcctca acgtgctaag agggttcaac ttctcaacag 1260

agtctatggc caggcctgaa gggagggtga acgtgtacca ccaccttgta gagacgctca 1320

agtttgccag ggggcagagg tggaggctgg gggaccctcg aagccacccg aagctccaga 1380

atgcctcccg ggacctgctg ggggagaccc tggcccagct catccgccaa cagatcgatg 1440

gccgggggga ccaccagctc agccactaca gcttggccga ggcctggggc cacgggacag 1500

gcacgtccca tgtgtctgtg ctgggggagg atggcagcgc cgtggctgcc accagcacca 1560

tcaacacacc ctttggagcg atggtgtatt caccacggac aggcatcatc ctcaacaacg 1620

agctcctgga cttatgcgag cgatgcccct ggggttccgg caccaccccc tcacctgtga 1680

gtggagacag ggtgggtgga gctcccggaa ggtgctggcc cccagttcca ggcgagcgtt 1740

ccccatcctc catggtgccc tccatcttga tcaacaaagc ccaggggtcg aagctagtga 1800

ttggcggggc tggcggggag ctcatcatct ctgctgtggc ccaggccatc atgagcaagc 1860

tgtggcttgg ctttgacctg agagcggcca ttgcagcccc catcctgcat gtcaacagca 1920

agggctgtgt ggagtacgag cccaacttca gccaggaggt gcagagggga ctccaagacc 1980

gtggccagaa ccagacccag aggcccttct tcctgaacgt ggtccaggct gtgtcccagg 2040

agggggcctg tgtgtacgcc gtctcggacc tgaggaagag tggggaggcc gcaggctact 2100

aagacactgc tctgcccaga gctgaagtct ggccccacca tgagtcctgt gtccaggccg 2160

gacatggctg ggggaccaac tactctggca ggatctggac ccctggcagg ggagtccagc 2220

tgagagtgga agaggtggcg gggaccagct gggcagatga gaggctgagc ctcatcccta 2280

accccctttc ccagagcccc tggtggtcct gaaccggccc ctctatccct ccgcaggcct 2340

-continued

cttacctggg gccactctcc caccctctcg atctgtatat cctccagtcc aagattaaag 2400

aagaggcgga ctgt 2414

<210> SEQ ID NO 51
<211> LENGTH: 586
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 51

Met Ala Arg Gly Tyr Gly Ala Thr Val Ser Leu Val Leu Leu Gly Leu
1 5 10 15

Gly Leu Ala Leu Ala Val Ile Val Leu Ala Val Val Leu Ser Arg His
20 25 30

Gln Ala Pro Cys Gly Pro Gln Ala Phe Ala His Ala Ala Val Ala Ala
35 40 45

Asp Ser Lys Val Cys Ser Asp Ile Gly Arg Ala Ile Leu Gln Gln Gln
50 55 60

Gly Ser Pro Val Asp Ala Thr Ile Ala Ala Leu Val Cys Thr Ser Val
65 70 75 80

Val Asn Pro Gln Ser Met Gly Leu Gly Gly Gly Val Ile Phe Thr Ile
85 90 95

Tyr Asn Val Thr Thr Gly Lys Val Glu Val Ile Asn Ala Arg Glu Thr
100 105 110

Val Pro Ala Ser His Ala Pro Ser Leu Leu Asp Gln Cys Ala Gln Ala
115 120 125

Leu Pro Leu Gly Thr Gly Ala Gln Trp Ile Gly Val Pro Gly Glu Leu
130 135 140

Arg Gly Tyr Ala Glu Ala His Arg Arg His Gly Arg Leu Pro Trp Ala
145 150 155 160

Gln Leu Phe Gln Pro Thr Ile Ala Leu Leu Arg Gly Gly His Val Val
165 170 175

Ala Pro Val Leu Ser Arg Phe Leu His Asn Ser Ile Leu Arg Pro Ser
180 185 190

Leu Gln Ala Ser Thr Leu Arg Gln Leu Phe Phe Asn Gly Thr Glu Pro
195 200 205

Leu Arg Pro Gln Asp Pro Leu Pro Trp Pro Ala Leu Ala Thr Thr Leu
210 215 220

Glu Thr Val Ala Thr Glu Gly Val Glu Val Phe Tyr Thr Gly Arg Leu
225 230 235 240

Gly Gln Met Leu Val Glu Asp Ile Ala Lys Glu Gly Ser Gln Leu Thr
245 250 255

Leu Gln Asp Leu Ala Lys Phe Gln Pro Glu Val Val Asp Ala Leu Glu
260 265 270

Val Pro Leu Gly Asp Tyr Thr Leu Tyr Ser Pro Pro Pro Pro Ala Gly
275 280 285

Gly Ala Ile Leu Ser Phe Ile Leu Asn Val Leu Arg Gly Phe Asn Phe
290 295 300

Ser Thr Glu Ser Met Ala Arg Pro Glu Gly Arg Val Asn Val Tyr His
305 310 315 320

His Leu Val Glu Thr Leu Lys Phe Ala Arg Gly Gln Arg Trp Arg Leu
325 330 335

Gly Asp Pro Arg Ser His Pro Lys Leu Gln Asn Ala Ser Arg Asp Leu

-continued

```
            340                 345                 350

Leu Gly Glu Thr Leu Ala Gln Leu Ile Arg Gln Gln Ile Asp Gly Arg
        355                 360                 365

Gly Asp His Gln Leu Ser His Tyr Ser Leu Ala Glu Ala Trp Gly His
    370                 375                 380

Gly Thr Gly Thr Ser His Val Ser Val Leu Gly Glu Asp Gly Ser Ala
385                 390                 395                 400

Val Ala Ala Thr Ser Thr Ile Asn Thr Pro Phe Gly Ala Met Val Tyr
                405                 410                 415

Ser Pro Arg Thr Gly Ile Ile Leu Asn Asn Glu Leu Leu Asp Leu Cys
            420                 425                 430

Glu Arg Cys Pro Trp Gly Ser Gly Thr Thr Pro Ser Pro Val Ser Gly
        435                 440                 445

Asp Arg Val Gly Gly Ala Pro Gly Arg Cys Trp Pro Pro Val Pro Gly
    450                 455                 460

Glu Arg Ser Pro Ser Ser Met Val Pro Ser Ile Leu Ile Asn Lys Ala
465                 470                 475                 480

Gln Gly Ser Lys Leu Val Ile Gly Gly Ala Gly Gly Glu Leu Ile Ile
                485                 490                 495

Ser Ala Val Ala Gln Ala Ile Met Ser Lys Leu Trp Leu Gly Phe Asp
            500                 505                 510

Leu Arg Ala Ala Ile Ala Ala Pro Ile Leu His Val Asn Ser Lys Gly
        515                 520                 525

Cys Val Glu Tyr Glu Pro Asn Phe Ser Gln Glu Val Gln Arg Gly Leu
    530                 535                 540

Gln Asp Arg Gly Gln Asn Gln Thr Gln Arg Pro Phe Phe Leu Asn Val
545                 550                 555                 560

Val Gln Ala Val Ser Gln Glu Gly Ala Cys Val Tyr Ala Val Ser Asp
                565                 570                 575

Leu Arg Lys Ser Gly Glu Ala Ala Gly Tyr
            580                 585
```

<210> SEQ ID NO 52
<211> LENGTH: 413
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 52

```
tggtgccctc catcttgatc aacaaagccc aggggtcgaa gctagtgatt ggcggggctg     60

gcggggagct catcatctct gctgtggccc aggccatcat gagcaagctg tggcttggct    120

ttgacctgag agcggccatt gcagccccca tcctgcatgt caacagcaag ggctgtgtgg    180

agtacgagcc caacttcagc caggtgaggc tgaggtccga gctggatgcc tagggcagag    240

cccactcccc aaatccgtgc tgctcaaagc cacctgggag gaactcagtc actgagattc    300

ttaggccaga gacggtgtct tgctttgttg ccctggctgt tctcaaactc ctggcctcaa    360

gtgatcctgc caccttggct cccaagtggc taagactgta ggtatgtgcc acg           413
```

<210> SEQ ID NO 53
<211> LENGTH: 1644
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 53

-continued

```
atggcccggg gctacggggc cacggtcagc ctagtcctgc tgggtctggg gctggcgctg     60

gctgtcattg tgctggctgt ggtcctctct cgacaccagg ccccatgtgg cccccaggcc    120

tttgcccacg ctgctgttgc cgccgactcc aaggtctgct cggatattgg acgagccatc    180

ctccagcagc agggctcacc cgtggatgcc accatcgcgg ctctggtctg caccagcgtc    240

gtcaaccctc agagcatggg cctgggcgga ggggtcatct tcaccatcta caatgtgaca    300

acagggaagg tggaggtcat caatgcccgg gagacggtgc cggccagcca cgccccgagc    360

ctgctggacc agtgtgcaca ggctctgcca ctgggcacag gggcccagtg gatcggggtg    420

cccggggagc tccgtggcta tgccgaggcc caccgccgcc atggccgcct gccctgggcg    480

cagctgttcc agcccaccat cgcgctgctc cgaggggggc atgtggtggc ccctgtcctc    540

agccgtttcc tgcacaacag catcctgcgg ccttccttgc aggcgtcaac cctgcgccag    600

ctcttcttca acgggacaga acccctgagg cctcaggacc cactcccatg gcctgcactg    660

gccaccaccc tggagaccgt ggccacagag ggcgtggagg tcttctacac ggggaggctg    720

ggccagatgc tggtggagga cattgccaag gaagggagcc agctgacgct gcaggacctg    780

gccaagttcc agcccgaggt ggtggatgcc ctggaggtgc ccctggggga ctataccctg    840

tactcaccac cgccgcctgc agggggtgcc attctcagct ttatcctcaa cgtgctaaga    900

gggttcaact tctcaacaga gtctatggcc aggcctgaag ggagggtgaa cgtgtaccac    960

caccttgtag agacgctcaa gtttgccagg gggcagaggt ggaggctggg ggaccctcga   1020

agccacccga agctccagaa tgcctcccgg gacctgctgg gggagaccct ggcccagctc   1080

atccgccaac agatcgatgg ccgggggggac caccagctca gccactacag cttggccgag   1140

gcctggggcc acgggacagg cacgtcccat gtgtctgtgc tgggggagga tggcagcgcc   1200

gtggctgcca ccagcaccat caacacaccc tttggagcga tggtgtattc accacggaca   1260

ggcatcatcc tcaacaacga gctcctggac ttatgcgagc gatgcccctg gggttccggc   1320

accacccccc cacctgtgag tggagacagg gtgggtggag ctcccggaag gtgctggccc   1380

ccagttccag gcgagcgttc cccatcctcc atggtgccct ccatcttgat caacaaagcc   1440

caggggtcga agctagtgat tggcggggct ggcggggagc tcatcatctc tgctgtggcc   1500

caggccatca tgagcaagct gtggcttggc tttgacctga gagcggccat tgcagccccc   1560

atcctgcatg tcaacagcaa gggctgtgtg gagtacgagc ccaacttcag ccaggtgagg   1620

ctgaggtccg agctggatgc ctag                                          1644
```

<210> SEQ ID NO 54
<211> LENGTH: 547
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 54

```
Met Ala Arg Gly Tyr Gly Ala Thr Val Ser Leu Val Leu Leu Gly Leu
1               5                  10                  15

Gly Leu Ala Leu Ala Val Ile Val Leu Ala Val Val Leu Ser Arg His
            20                  25                  30

Gln Ala Pro Cys Gly Pro Gln Ala Phe Ala His Ala Ala Val Ala Ala
        35                  40                  45

Asp Ser Lys Val Cys Ser Asp Ile Gly Arg Ala Ile Leu Gln Gln Gln
    50                  55                  60
```

-continued

```
Gly Ser Pro Val Asp Ala Thr Ile Ala Ala Leu Val Cys Thr Ser Val
65                  70                  75                  80

Val Asn Pro Gln Ser Met Gly Leu Gly Gly Gly Val Ile Phe Thr Ile
                85                  90                  95

Tyr Asn Val Thr Thr Gly Lys Val Glu Val Ile Asn Ala Arg Glu Thr
            100                 105                 110

Val Pro Ala Ser His Ala Pro Ser Leu Leu Asp Gln Cys Ala Gln Ala
        115                 120                 125

Leu Pro Leu Gly Thr Gly Ala Gln Trp Ile Gly Val Pro Gly Glu Leu
    130                 135                 140

Arg Gly Tyr Ala Glu Ala His Arg Arg His Gly Arg Leu Pro Trp Ala
145                 150                 155                 160

Gln Leu Phe Gln Pro Thr Ile Ala Leu Leu Arg Gly Gly His Val Val
                165                 170                 175

Ala Pro Val Leu Ser Arg Phe Leu His Asn Ser Ile Leu Arg Pro Ser
            180                 185                 190

Leu Gln Ala Ser Thr Leu Arg Gln Leu Phe Phe Asn Gly Thr Glu Pro
        195                 200                 205

Leu Arg Pro Gln Asp Pro Leu Pro Trp Pro Ala Leu Ala Thr Thr Leu
    210                 215                 220

Glu Thr Val Ala Thr Glu Gly Val Glu Val Phe Tyr Thr Gly Arg Leu
225                 230                 235                 240

Gly Gln Met Leu Val Glu Asp Ile Ala Lys Glu Gly Ser Gln Leu Thr
                245                 250                 255

Leu Gln Asp Leu Ala Lys Phe Gln Pro Glu Val Val Asp Ala Leu Glu
            260                 265                 270

Val Pro Leu Gly Asp Tyr Thr Leu Tyr Ser Pro Pro Pro Pro Ala Gly
        275                 280                 285

Gly Ala Ile Leu Ser Phe Ile Leu Asn Val Leu Arg Gly Phe Asn Phe
    290                 295                 300

Ser Thr Glu Ser Met Ala Arg Pro Glu Gly Arg Val Asn Val Tyr His
305                 310                 315                 320

His Leu Val Glu Thr Leu Lys Phe Ala Arg Gly Gln Arg Trp Arg Leu
                325                 330                 335

Gly Asp Pro Arg Ser His Pro Lys Leu Gln Asn Ala Ser Arg Asp Leu
            340                 345                 350

Leu Gly Glu Thr Leu Ala Gln Leu Ile Arg Gln Gln Ile Asp Gly Arg
        355                 360                 365

Gly Asp His Gln Leu Ser His Tyr Ser Leu Ala Glu Ala Trp Gly His
    370                 375                 380

Gly Thr Gly Thr Ser His Val Ser Val Leu Gly Glu Asp Gly Ser Ala
385                 390                 395                 400

Val Ala Ala Thr Ser Thr Ile Asn Thr Pro Phe Gly Ala Met Val Tyr
                405                 410                 415

Ser Pro Arg Thr Gly Ile Ile Leu Asn Asn Glu Leu Leu Asp Leu Cys
            420                 425                 430

Glu Arg Cys Pro Trp Gly Ser Gly Thr Thr Pro Ser Pro Val Ser Gly
        435                 440                 445

Asp Arg Val Gly Gly Ala Pro Gly Arg Cys Trp Pro Pro Val Pro Gly
    450                 455                 460
```

-continued

```
Glu Arg Ser Pro Ser Ser Met Val Pro Ser Ile Leu Ile Asn Lys Ala
465                 470                 475                 480

Gln Gly Ser Lys Leu Val Ile Gly Gly Ala Gly Gly Glu Leu Ile Ile
                485                 490                 495

Ser Ala Val Ala Gln Ala Ile Met Ser Lys Leu Trp Leu Gly Phe Asp
            500                 505                 510

Leu Arg Ala Ala Ile Ala Ala Pro Ile Leu His Val Asn Ser Lys Gly
        515                 520                 525

Cys Val Glu Tyr Glu Pro Asn Phe Ser Gln Val Arg Leu Arg Ser Glu
    530                 535                 540

Leu Asp Ala
545
```

<210> SEQ ID NO 55
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 55

```
Val Arg Leu Arg Ser Glu Leu Asp Ala
1               5
```

<210> SEQ ID NO 56
<211> LENGTH: 310
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 56

```
attggtgccc tccatcttga tcaacaaagc ccaggggtcg aagctartga ttggcggggc     60

tggcggggag ctcatcatct ctgctgtggc ccaggccatc atgarcaagc tgtggcttgg    120

ctttgacctg agagcggcca ttgcagcccc catcctgcat gtcaacagca agggctgtgt    180

ggagtacsas cccaacttca sccagagacg gtgtcttgct ttgttgccct ggctgttcat    240

caaactcctg gcctcaagtg atcctgccac cttggctccc aagtggctaa gactgtaggt    300

atgtgccaca                                                           310
```

<210> SEQ ID NO 57
<211> LENGTH: 1698
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 57

```
atggcccggg gctacggggc cacggtcagc ctagtcctgc tgggtctggg gctggcgctg     60

gctgtcattg tgctggctgt ggtcctctct cgacaccagg ccccatgtgg cccccaggcc    120

tttgcccacg ctgctgttgc cgccgactcc aaggtctgct cggatattgg acgagccatc    180

ctccagcagc agggctcacc cgtggatgcc accatcgcgg ctctggtctg caccagcgtc    240

gtcaaccctc agagcatggg cctgggcgga ggggtcatct tcaccatcta caatgtgaca    300

acagggaagg tggaggtcat caatgcccgg gagacggtgc cggccagcca cgccccgagc    360

ctgctggacc agtgtgcaca ggctctgcca ctgggcacag gggcccagtg gatcggggtg    420

cccggggagc tccgtggcta tgccgaggcc caccgccgcc atggccgcct gccctgggcg    480

cagctgttcc agcccaccat cgcgctgctc cgaggggggc atgtggtggc ccctgtcctc    540

agccgtttcc tgcacaacag catcctgcgg ccttccttgc aggcgtcaac cctgcgccag    600
```

-continued

```
ctcttcttca acgggacaga acccctgagg cctcaggacc cactcccatg gcctgcactg    660

gccaccaccc tggagaccgt ggccacagag ggcgtggagg tcttctacac ggggaggctg    720

ggccagatgc tggtggagga cattgccaag gaagggagcc agctgacgct gcaggacctg    780

gccaagttcc agcccgaggt ggtggatgcc ctggaggtgc ccctggggga ctataccctg    840

tactcaccac cgccgcctgc agggggtgcc attctcagct ttatcctcaa cgtgctaaga    900

gggttcaact tctcaacaga gtctatggcc aggcctgaag ggagggtgaa cgtgtaccac    960

caccttgtag agacgctcaa gtttgccagg gggcagaggt ggaggctggg ggaccctcga   1020

agccacccga agctccagaa tgcctcccgg gacctgctgg gggagaccct ggcccagctc   1080

atccgccaac agatcgatgg ccgggggggac caccagctca gccactacag cttggccgag   1140

gcctggggcc acgggacagg cacgtcccat gtgtctgtgc tgggggagga tggcagcgcc   1200

gtggctgcca ccagcaccat caacacaccc tttggagcga tggtgtattc accacggaca   1260

ggcatcatcc tcaacaacga gctcctggac ttatgcgagc gatgcccctg gggttccggc   1320

accacccccct cacctgtgag tggagacagg gtgggtggag ctcccggaag gtgctggccc   1380

ccagttccag gcgagcgttc cccatcctcc atggtgccct ccatcttgat caacaaagcc   1440

caggggtcga agctagtgat tggcggggct ggcggggagc tcatcatctc tgctgtggcc   1500

caggccatca tgagcaagct gtggcttggc tttgacctga gagcggccat tgcagccccc   1560

atcctgcatg tcaacagcaa gggctgtgtg gagtacgagc ccaacttcag ccagagacgg   1620

tgtcttgctt tgttgccctg gctgttctca aactcctggc ctcaagtgat cctgccacct   1680

tggctcccaa gtggctaa                                                 1698
```

<210> SEQ ID NO 58
<211> LENGTH: 560
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 58

```
Met Ala Arg Gly Tyr Gly Ala Thr Val Ser Leu Val Leu Leu Gly Leu
1               5                   10                  15

Gly Leu Ala Leu Ala Val Ile Val Leu Ala Val Val Leu Ser Arg His
            20                  25                  30

Gln Ala Pro Cys Gly Pro Gln Ala Phe Ala His Ala Ala Val Ala Ala
        35                  40                  45

Asp Ser Lys Val Cys Ser Asp Ile Gly Arg Ala Ile Leu Gln Gln Gln
    50                  55                  60

Gly Ser Pro Val Asp Ala Thr Ile Ala Ala Leu Val Cys Thr Ser Val
65                  70                  75                  80

Val Asn Pro Gln Ser Met Gly Leu Gly Gly Gly Val Ile Phe Thr Ile
                85                  90                  95

Tyr Asn Val Thr Thr Gly Lys Val Glu Val Ile Asn Ala Arg Glu Thr
            100                 105                 110

Val Pro Ala Ser His Ala Pro Ser Leu Leu Asp Gln Cys Ala Gln Ala
        115                 120                 125

Leu Pro Leu Gly Thr Gly Ala Gln Trp Ile Gly Val Pro Gly Glu Leu
    130                 135                 140

Arg Gly Tyr Ala Glu Ala His Arg Arg His Gly Arg Leu Pro Trp Ala
145                 150                 155                 160
```

-continued

```
Gln Leu Phe Gln Pro Thr Ile Ala Leu Leu Arg Gly Gly His Val Val
                165                 170                 175

Ala Pro Val Leu Ser Arg Phe Leu His Asn Ser Ile Leu Arg Pro Ser
            180                 185                 190

Leu Gln Ala Ser Thr Leu Arg Gln Leu Phe Phe Asn Gly Thr Glu Pro
        195                 200                 205

Leu Arg Pro Gln Asp Pro Leu Pro Trp Pro Ala Leu Ala Thr Thr Leu
    210                 215                 220

Glu Thr Val Ala Thr Glu Gly Val Glu Val Phe Tyr Thr Gly Arg Leu
225                 230                 235                 240

Gly Gln Met Leu Val Glu Asp Ile Ala Lys Glu Gly Ser Gln Leu Thr
                245                 250                 255

Leu Gln Asp Leu Ala Lys Phe Gln Pro Glu Val Val Asp Ala Leu Glu
            260                 265                 270

Val Pro Leu Gly Asp Tyr Thr Leu Tyr Ser Pro Pro Pro Pro Ala Gly
        275                 280                 285

Gly Ala Ile Leu Ser Phe Ile Leu Asn Val Leu Arg Gly Phe Asn Phe
    290                 295                 300

Ser Thr Glu Ser Met Ala Arg Pro Glu Gly Arg Val Asn Val Tyr His
305                 310                 315                 320

His Leu Val Glu Thr Leu Lys Phe Ala Arg Gly Gln Arg Trp Arg Leu
                325                 330                 335

Gly Asp Pro Arg Ser His Pro Lys Leu Gln Asn Ala Ser Arg Asp Leu
            340                 345                 350

Leu Gly Glu Thr Leu Ala Gln Leu Ile Arg Gln Gln Ile Asp Gly Arg
        355                 360                 365

Gly Asp His Gln Leu Ser His Tyr Ser Leu Ala Glu Ala Trp Gly His
    370                 375                 380

Gly Thr Gly Thr Ser His Val Ser Val Leu Gly Glu Asp Gly Ser Ala
385                 390                 395                 400

Val Ala Ala Thr Ser Thr Ile Asn Thr Pro Phe Gly Ala Met Val Tyr
                405                 410                 415

Ser Pro Arg Thr Gly Ile Ile Leu Asn Asn Glu Leu Leu Asp Leu Cys
            420                 425                 430

Glu Arg Cys Pro Trp Gly Ser Gly Thr Thr Pro Ser Pro Val Ser Gly
        435                 440                 445

Asp Arg Val Gly Gly Ala Pro Gly Arg Cys Trp Pro Pro Val Pro Gly
    450                 455                 460

Glu Arg Ser Pro Ser Ser Met Val Pro Ser Ile Leu Ile Asn Lys Ala
465                 470                 475                 480

Gln Gly Ser Lys Leu Val Ile Gly Gly Ala Gly Gly Glu Leu Ile Ile
                485                 490                 495

Ser Ala Val Ala Gln Ala Ile Met Ser Lys Leu Trp Leu Gly Phe Asp
            500                 505                 510

Leu Arg Ala Ala Ile Ala Ala Pro Ile Leu His Val Asn Ser Lys Gly
        515                 520                 525

Cys Val Glu Tyr Glu Pro Asn Phe Ser Gln Arg Arg Cys Leu Ala Leu
    530                 535                 540

Leu Pro Trp Leu Phe Ser Asn Ser Trp Pro Gln Val Ile Leu Pro Pro
545                 550                 555                 560
```

-continued

<210> SEQ ID NO 59
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 59

Arg Arg Cys Leu Ala Leu Leu Pro Trp Leu Phe Ser Asn Ser Trp Pro
1 5 10 15

Gln Val Ile Leu Pro Pro
20

<210> SEQ ID NO 60
<211> LENGTH: 2428
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 60

gggccactcc gtagtgtgca cttggtgagg gcagcagctc gccacagctg ccagccatct 60

gtccattcac ccatctgtcc gtctggcagc ccgctgttca gacctgtctg tctgtccgcc 120

catctgtaag cccatctctg tcccattgtc tatctgacca tctttctctt actgtcctct 180

ttgtctagct atctggccta tctgtcgatc catcttcgtg tctgtcttca gcccccacct 240

gtttgtccat ctgtccaatt acctgtgact ctgtgcatct tcttgtccat tcatctgccc 300

acccatccgt ccctccgtct gcccaccagc cgcccctctc ctcctgggct gcagagccat 360

ggcccggggc tacggggcca cggtcagcct agtcctgctg ggtctggggc tggcgctggc 420

tgtcattgtg ctggctgtgg tcctctctcg acaccaggcc ccatgtggcc cccaggcctt 480

tgcccacgct gctgttgccg ccgactccaa ggtctgctcg gatattggac gagccatcct 540

ccagcagcag ggctcacccg tggatgccac catcgcggct ctggtctgca ccagcgtcgt 600

caaccctcag agcatgggcc tgggcggagg ggtcatcttc accatctaca atgtgacaac 660

aggggcccag tggatcgggg tgcccgggga gctccgtggc tatgccgagg cccaccgccg 720

ccatggccgc ctgccctggg cgcagctgtt ccagcccacc atcgcgctgc tccgaggggg 780

gcatgtggtg gcccctgtcc tcagccgttt cctgcacaac agcatcctgc ggccttcctt 840

gcaggcgtca accctgcgcc agctcttctt caacgggaca gaaccccctga ggcctcagga 900

cccactccca tggcctgcac tggccaccac cctggagacc gtggccacag agggcgtgga 960

ggtcttctac acggggaggc tgggccagat gctggtggag gacattgcca aggaagggag 1020

ccagctgacg ctgcaggacc tggccaagtt ccagcccgag gtggtggatg ccctggaggt 1080

gcccctgggg gactataccc tgtactcacc accgccgcct gcaggggggtg ccattctcag 1140

ctttatcctc aacgtgctaa gagggttcaa cttctcaaca gagtctatgg ccaggcctga 1200

agggagggtg aacgtgtacc accaccttgt agagacgctc aagtttgcca aggggcagag 1260

gtggaggctg ggggaccctc gaagccaccc gaagctccag aatgcctccc gggacctgct 1320

gggggagacc ctggcccagc tcatccgcca acagatcgat ggccgggggg accaccagct 1380

cagccactac agcttggccg aggcctgggg ccacgggaca ggcacgtccc atgtgtcagt 1440

gctgggggag gatggcagcg ccgtggctgc caccagcacc atctacacac cctttggagc 1500

gatggtgtat tcaccacgga caggcatcat cctcaacaac gagctcctgg acttatgcga 1560

gcgatgcccc cggggttccg gcaccacccc ctcacctgtg agggagacag ggtgggtgga 1620

-continued

gctcccggaa ggtgctggcc cccagttcca ggcgagcgtt ccccatcctc catggtgccc 1680

tccatcttga tcaacaaagc ccaggggtcg aagctagtga ttggcggggc tggcggggag 1740

ctcatcatct ctgctgtggc ccaggccatc atgagcaagc tgtggcttgg ctttgacctg 1800

agagcggcca ttgcagcccc catcctgcat gtcaacagca agggctgtgt ggagtacgag 1860

cccaacttca gccagagacg gtgtcttgct ttgttgccct ggctgttctc aaactcctgg 1920

cctcaagtga tcctgccacc ttggctccca agtggctaag actgtaggag gtgcagaggg 1980

gactccaaga ccgtggccag aaccagaccc agaggccctt cttcctgaac gtggtccagg 2040

ctgtgtccca ggaggggggcc tgtgtgtacg ccgtctcgga cctgaggaag agtggggagg 2100

ccgcaggcta ctaagacact gctctgccca gagctgaagt ctggccccac catgagtcct 2160

gtgtccaggc cggacatggc tgggggacca actactctgg caggatctgg acccctggca 2220

ggggagtcca gctgagagtg gaagaggtgg cggggaccag ctgggcagat gagaggctga 2280

gcctcatccc taacccccctt tcccagagcc cctggtggtc ctgaaccggc ccctctatcc 2340

ctccgcaggc ctcttgcctg gggccactct cccaccctct cgatctgtat atcctccagt 2400

ccaagattaa agaggcggac tatgaaaa 2428

<210> SEQ ID NO 61
<211> LENGTH: 443
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 61

Met Ala Arg Gly Tyr Gly Ala Thr Val Ser Leu Val Leu Leu Gly Leu
1 5 10 15

Gly Leu Ala Leu Ala Val Ile Val Leu Ala Val Val Leu Ser Arg His
20 25 30

Gln Ala Pro Cys Gly Pro Gln Ala Phe Ala His Ala Ala Val Ala Ala
35 40 45

Asp Ser Lys Val Cys Ser Asp Ile Gly Arg Ala Ile Leu Gln Gln Gln
50 55 60

Gly Ser Pro Val Asp Ala Thr Ile Ala Ala Leu Val Cys Thr Ser Val
65 70 75 80

Val Asn Pro Gln Ser Met Gly Leu Gly Gly Gly Val Ile Phe Thr Ile
85 90 95

Tyr Asn Val Thr Thr Gly Ala Gln Trp Ile Gly Val Pro Gly Glu Leu
100 105 110

Arg Gly Tyr Ala Glu Ala His Arg Arg His Gly Arg Leu Pro Trp Ala
115 120 125

Gln Leu Phe Gln Pro Thr Ile Ala Leu Leu Arg Gly Gly His Val Val
130 135 140

Ala Pro Val Leu Ser Arg Phe Leu His Asn Ser Ile Leu Arg Pro Ser
145 150 155 160

Leu Gln Ala Ser Thr Leu Arg Gln Leu Phe Phe Asn Gly Thr Glu Pro
165 170 175

Leu Arg Pro Gln Asp Pro Leu Pro Trp Pro Ala Leu Ala Thr Thr Leu
180 185 190

Glu Thr Val Ala Thr Glu Gly Val Glu Val Phe Tyr Thr Gly Arg Leu
195 200 205

Gly Gln Met Leu Val Glu Asp Ile Ala Lys Glu Gly Ser Gln Leu Thr

-continued

```
    210                 215                 220

Leu Gln Asp Leu Ala Lys Phe Gln Pro Glu Val Val Asp Ala Leu Glu
225                 230                 235                 240

Val Pro Leu Gly Asp Tyr Thr Leu Tyr Ser Pro Pro Pro Pro Ala Gly
                245                 250                 255

Gly Ala Ile Leu Ser Phe Ile Leu Asn Val Leu Arg Gly Phe Asn Phe
            260                 265                 270

Ser Thr Glu Ser Met Ala Arg Pro Glu Gly Arg Val Asn Val Tyr His
        275                 280                 285

His Leu Val Glu Thr Leu Lys Phe Ala Lys Gly Gln Arg Trp Arg Leu
    290                 295                 300

Gly Asp Pro Arg Ser His Pro Lys Leu Gln Asn Ala Ser Arg Asp Leu
305                 310                 315                 320

Leu Gly Glu Thr Leu Ala Gln Leu Ile Arg Gln Gln Ile Asp Gly Arg
                325                 330                 335

Gly Asp His Gln Leu Ser His Tyr Ser Leu Ala Glu Ala Trp Gly His
            340                 345                 350

Gly Thr Gly Thr Ser His Val Ser Val Leu Gly Glu Asp Gly Ser Ala
        355                 360                 365

Val Ala Ala Thr Ser Thr Ile Tyr Thr Pro Phe Gly Ala Met Val Tyr
    370                 375                 380

Ser Pro Arg Thr Gly Ile Ile Leu Asn Asn Glu Leu Leu Asp Leu Cys
385                 390                 395                 400

Glu Arg Cys Pro Arg Gly Ser Gly Thr Thr Pro Ser Pro Val Arg Glu
                405                 410                 415

Thr Gly Trp Val Glu Leu Pro Glu Gly Ala Gly Pro Gln Phe Gln Ala
            420                 425                 430

Ser Val Pro His Pro Pro Trp Cys Pro Pro Ser
        435                 440
```

<210> SEQ ID NO 62
<211> LENGTH: 149
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 62

```
tccctgagtg tcagccccag aatggytcag tgacctgttt tggaccggtg agctgctggc     60

gggytcagag ctgggtggag gggggcagcg agggggattg ccagggactt ggcaggatgg    120

cgagatgcag tagggtgtgc tatctggat                                      149
```

<210> SEQ ID NO 63
<211> LENGTH: 4530
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 63

```
aattctcgag ctcgtcgacc ggtcgacgag ctcgagggtc gacgagctcg agggcgcgcg     60

cccggccccc acccctcgca gcaccccgcg ccccgcgccc tcccagccgg gtccagccgg    120

agccatgggg ccggagccgc agtgagcacc atggagctgg cggccttgtg ccgctggggg    180

ctcctcctcg ccctcttgcc ccccggagcc gcgagcaccc aagtgtgcac cggcacagac    240

atgaagctgc ggctccctgc cagtcccgag acccacctgg acatgctccg ccacctctac    300
```

-continued

```
cagggctgcc aggtggtgca gggaaacctg gaactcacct acctgcccac caatgccagc    360

ctgtccttcc tgcaggatat ccaggaggtg cagggctacg tgctcatcgc tcacaaccaa    420

gtgaggcagg tcccactgca gaggctgcgg attgtgcgag gcacccagct ctttgaggac    480

aactatgccc tggccgtgct agacaatgga gacccgctga acaataccac ccctgtcaca    540

ggggcctccc caggaggcct gcgggagctg cagcttcgaa gcctcacaga gatcttgaaa    600

ggaggggtct tgatccagcg gaaccccccag ctctgctacc aggacacgat tttgtggaag    660

gacatcttcc acaagaacaa ccagctggct ctcacactga tagacaccaa ccgctctcgg    720

gcctgccacc cctgttctcc gatgtgtaag ggctcccgct gctggggaga gagttctgag    780

gattgtcaga gcctgacgcg cactgtctgt gccggtggct gtgcccgctg caaggggcca    840

ctgcccactg actgctgcca tgagcagtgt gctgccggct gcacgggccc caagcactct    900

gactgcctgg cctgcctcca cttcaaccac agtggcatct gtgagctgca ctgcccagcc    960

ctggtcacct acaacacaga cacgtttgag tccatgccca atcccgaggg ccggtataca   1020

ttcggcgcca gctgtgtgac tgcctgtccc tacaactacc tttctacgga cgtgggatcc   1080

tgcaccctcg tctgcccct gcacaaccaa gaggtgacag cagaggatgg aacacagcgg   1140

tgtgagaagt gcagcaagcc ctgtgcccga gtgtgctatg gtctgggcat ggagcacttg   1200

cgagaggtga gggcagttac cagtgccaat atccaggagt ttgctggctg caagaagatc   1260

tttgggagcc tggcatttct gccggagagc tttgatgggg acccagcctc caacactgcc   1320

ccgctccagc cagagcagct ccaagtgttt gagactctgg aagagatcac aggttaccta   1380

tacatctcag catggccgga cagcctgcct gacctcagcg tcttccagaa cctgcaagta   1440

atccggggac gaattctgca caatggcgcc tactcgctga ccctgcaagg gctgggcatc   1500

agctggctgg ggctgcgctc actgagggaa ctgggcagtg gactggccct catccaccat   1560

aacacccacc tctgcttcgt gcacacggtg ccctgggacc agctctttcg gaacccgcac   1620

caagctctgc tccacactgc caaccggcca gaggacgagt gtgtgggcga gggcctggcc   1680

tgccaccagc tgtgcgcccg agggcactgc tggggtccag ggcccaccca gtgtgtcaac   1740

tgcagccagt tccttcgggg ccaggagtgc gtggaggaat gccgagtact gcaggggctc   1800

cccagggagt atgtgaatgc caggcactgt ttgccgtgcc accctgagtg tcagccccag   1860

aatggctcag tgacctgttt tggaccggag gctgaccagt gtgtggcctg tgcccactat   1920

aaggaccctc ccttctgcgt ggcccgctgc cccagcggtg tgaaacctga cctctcctac   1980

atgcccatct ggaagtttcc agatgaggag ggcgcatgcc agccttgccc catcaactgc   2040

acccactcct gtgtggacct ggatgacaag ggctgccccg ccgagcagag agccagccct   2100

ctgacgtcca tcgtctctgc ggtggttggc attctgctgg tcgtggtctt gggggtggtc   2160

tttgggatcc tcatcaagcg acggcagcag aagatccgga agtacacgat gcggagactg   2220

ctgcaggaaa cggagctggt ggagccgctg acacctagcg gagcgatgcc caaccaggcg   2280

cagatgcgga tcctgaaaga gacggagctg aggaaggtga aggtgcttgg atctggcgct   2340

tttggcacag tctacaaggg catctggatc cctgatgggg agaatgtgaa aattccagtg   2400

gccatcaaag tgttgaggga aaacacatcc cccaaagcca acaaagaaat cttagacgaa   2460

gcatacgtga tggctggtgt gggctcccca tatgtctccc gccttctggg catctgcctg   2520

acatccacgg tgcagctggt gacacagctt atgccctatg gctgcctctt agaccatgtc   2580
```

-continued

```
cgggaaaacc gcggacgcct gggctcccag gacctgctga actggtgtat gcagattgcc   2640

aaggggatga gctacctgga ggatgtgcgg ctcgtacaca gggacttggc cgctcggaac   2700

gtgctggtca agagtcccaa ccatgtcaaa attacagact tcgggctggc tcggctgctg   2760

gacattgacg agacagagta ccatgcagat gggggcaagg tgcccatcaa gtggatggcg   2820

ctggagtcca ttctccgccg gcggttcacc caccagagtg atgtgtggag ttatggtgtg   2880

actgtgtggg agctgatgac ttttggggcc aaaccttacg atgggatccc agcccgggag   2940

atccctgacc tgctggaaaa gggggagcgg ctgccccagc cccccatctg caccattgat   3000

gtctacatga tcatggtcaa atgttggatg attgactctg aatgtcggcc aagattccgg   3060

gagttggtgt ctgaattctc ccgcatggcc agggaccccc agcgctttgt ggtcatccag   3120

aatgaggact tgggcccagc cagtcccttg gacagcacct tctaccgctc actgctggag   3180

gacgatgaca tgggggacct ggtggatgct gaggagtatc tggtacccca gcagggcttc   3240

ttctgtccag accctgcccc gggcgctggg ggcatggtcc accacaggca ccgcagctca   3300

tctaccagga gtggcggtgg ggacctgaca ctagggctgg agccctctga agaggaggcc   3360

cccaggtctc cactggcacc ctccgaaggg gctggctccg atgtatttga tggtgacctg   3420

ggaatggggg cagccaaggg gctgcaaagc ctccccacac atgaccccag ccctctacag   3480

cggtacagtg aggaccccac agtacccctg ccctctgaga ctgatggcta cgttgccccc   3540

ctgacctgca gcccccagcc tgaatatgtg aaccagccag atgttcggcc ccagcccccc   3600

tcgccccgag agggccctct gcctgctgcc cgacctgctg gtgccactct ggaaagggcc   3660

aagactctct ccccagggaa gaatggggtc gtcaaagacg tttttgcctt tgggggtgcc   3720

gtggagaacc ccgagtactt gacaccccag ggaggagctg cccctcagcc ccaccctcct   3780

cctgccttca gcccagcctt cgacaacctc tattactggg accaggaccc accagagcgg   3840

ggggctccac ccagcacctt caaagggaca cctacggcag agaacccaga gtacctgggt   3900

ctggacgtgc cagtgtgaac cagaaggcca agtccgcaga agccctgatg tgtcctcagg   3960

gagcagggaa ggcctgactt ctgctggcat caagaggtgg gagggccctc cgaccacttc   4020

caggggaacc tgccatgcca ggaacctgtc ctaaggaacc ttccttcctg cttgagttcc   4080

cagatggctg gaaggggtcc agcctcgttg gaagaggaac agcactgggg agtctttgtg   4140

gattctgagg ccctgcccaa tgagactcta gggtccagtg gatgccacag cccagcttgg   4200

ccctttcctt ccagatcctg ggtactgaaa gccttaggga agctggcctg agaggggaag   4260

cggccctaag ggagtgtcta agaacaaaag cgacccattc agagactgtc cctgaaacct   4320

agtactgccc cccatgagga aggaacagca atggtgtcag tatccaggct ttgtacagag   4380

tgcttttctg tttagttttt acttttttttg ttttgttttt ttaaagacga aataaagacc   4440

caggggagaa tgggtgttgt atggggaggc aagtgtgggg ggtccttctc cacacccact   4500

ttgtccattt gcaaatatat tttggaaaac                                    4530


<210> SEQ ID NO 64
<211> LENGTH: 1255
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 64

Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
1               5                  10                  15
```

-continued

```
Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
            20                  25                  30

Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
        35                  40                  45

Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
    50                  55                  60

Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65                  70                  75                  80

Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
                85                  90                  95

Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
            100                 105                 110

Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
        115                 120                 125

Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
    130                 135                 140

Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145                 150                 155                 160

Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
                165                 170                 175

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

His Pro Cys Ser Pro Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
        195                 200                 205

Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
    210                 215                 220

Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225                 230                 235                 240

Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
                245                 250                 255

His Phe Asn His Ser Gly Ile Cys Glu Leu His Cys Pro Ala Leu Val
            260                 265                 270

Thr Tyr Asn Thr Asp Thr Phe Glu Ser Met Pro Asn Pro Glu Gly Arg
        275                 280                 285

Tyr Thr Phe Gly Ala Ser Cys Val Thr Ala Cys Pro Tyr Asn Tyr Leu
    290                 295                 300

Ser Thr Asp Val Gly Ser Cys Thr Leu Val Cys Pro Leu His Asn Gln
305                 310                 315                 320

Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys Glu Lys Cys Ser Lys
                325                 330                 335

Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met Glu His Leu Arg Glu
            340                 345                 350

Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu Phe Ala Gly Cys Lys
        355                 360                 365

Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu Ser Phe Asp Gly Asp
    370                 375                 380

Pro Ala Ser Asn Thr Ala Pro Leu Gln Pro Glu Gln Leu Gln Val Phe
385                 390                 395                 400

Glu Thr Leu Glu Glu Ile Thr Gly Tyr Leu Tyr Ile Ser Ala Trp Pro
                405                 410                 415
```

-continued

```
Asp Ser Leu Pro Asp Leu Ser Val Phe Gln Asn Leu Gln Val Ile Arg
            420                 425                 430

Gly Arg Ile Leu His Asn Gly Ala Tyr Ser Leu Thr Leu Gln Gly Leu
        435                 440                 445

Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg Glu Leu Gly Ser Gly
    450                 455                 460

Leu Ala Leu Ile His His Asn Thr His Leu Cys Phe Val His Thr Val
465                 470                 475                 480

Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln Ala Leu Leu His Thr
                485                 490                 495

Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu Gly Leu Ala Cys His
            500                 505                 510

Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro Gly Pro Thr Gln Cys
        515                 520                 525

Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu Cys Val Glu Glu Cys
    530                 535                 540

Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val Asn Ala Arg His Cys
545                 550                 555                 560

Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn Gly Ser Val Thr Cys
                565                 570                 575

Phe Gly Pro Glu Ala Asp Gln Cys Val Ala Cys Ala His Tyr Lys Asp
            580                 585                 590

Pro Pro Phe Cys Val Ala Arg Cys Pro Ser Gly Val Lys Pro Asp Leu
        595                 600                 605

Ser Tyr Met Pro Ile Trp Lys Phe Pro Asp Glu Glu Gly Ala Cys Gln
    610                 615                 620

Pro Cys Pro Ile Asn Cys Thr His Ser Cys Val Asp Leu Asp Asp Lys
625                 630                 635                 640

Gly Cys Pro Ala Glu Gln Arg Ala Ser Pro Leu Thr Ser Ile Val Ser
                645                 650                 655

Ala Val Val Gly Ile Leu Leu Val Val Val Leu Gly Val Val Phe Gly
            660                 665                 670

Ile Leu Ile Lys Arg Arg Gln Gln Lys Ile Arg Lys Tyr Thr Met Arg
        675                 680                 685

Arg Leu Leu Gln Glu Thr Glu Leu Val Glu Pro Leu Thr Pro Ser Gly
    690                 695                 700

Ala Met Pro Asn Gln Ala Gln Met Arg Ile Leu Lys Glu Thr Glu Leu
705                 710                 715                 720

Arg Lys Val Lys Val Leu Gly Ser Gly Ala Phe Gly Thr Val Tyr Lys
                725                 730                 735

Gly Ile Trp Ile Pro Asp Gly Glu Asn Val Lys Ile Pro Val Ala Ile
            740                 745                 750

Lys Val Leu Arg Glu Asn Thr Ser Pro Lys Ala Asn Lys Glu Ile Leu
        755                 760                 765

Asp Glu Ala Tyr Val Met Ala Gly Val Gly Ser Pro Tyr Val Ser Arg
    770                 775                 780

Leu Leu Gly Ile Cys Leu Thr Ser Thr Val Gln Leu Val Thr Gln Leu
785                 790                 795                 800

Met Pro Tyr Gly Cys Leu Leu Asp His Val Arg Glu Asn Arg Gly Arg
                805                 810                 815

Leu Gly Ser Gln Asp Leu Leu Asn Trp Cys Met Gln Ile Ala Lys Gly
```

-continued

```
            820                 825                 830

Met Ser Tyr Leu Glu Asp Val Arg Leu Val His Arg Asp Leu Ala Ala
        835                 840                 845

Arg Asn Val Leu Val Lys Ser Pro Asn His Val Lys Ile Thr Asp Phe
    850                 855                 860

Gly Leu Ala Arg Leu Leu Asp Ile Asp Glu Thr Glu Tyr His Ala Asp
865                 870                 875                 880

Gly Gly Lys Val Pro Ile Lys Trp Met Ala Leu Glu Ser Ile Leu Arg
                885                 890                 895

Arg Arg Phe Thr His Gln Ser Asp Val Trp Ser Tyr Gly Val Thr Val
            900                 905                 910

Trp Glu Leu Met Thr Phe Gly Ala Lys Pro Tyr Asp Gly Ile Pro Ala
        915                 920                 925

Arg Glu Ile Pro Asp Leu Leu Glu Lys Gly Glu Arg Leu Pro Gln Pro
    930                 935                 940

Pro Ile Cys Thr Ile Asp Val Tyr Met Ile Met Val Lys Cys Trp Met
945                 950                 955                 960

Ile Asp Ser Glu Cys Arg Pro Arg Phe Arg Glu Leu Val Ser Glu Phe
                965                 970                 975

Ser Arg Met Ala Arg Asp Pro Gln Arg Phe Val Val Ile Gln Asn Glu
            980                 985                 990

Asp Leu Gly Pro Ala Ser Pro Leu  Asp Ser Thr Phe Tyr  Arg Ser Leu
        995                 1000                1005

Leu Glu  Asp Asp Asp Met Gly  Asp Leu Val Asp Ala  Glu Glu Tyr
    1010                1015                1020

Leu Val  Pro Gln Gln Gly Phe  Phe Cys Pro Asp Pro  Ala Pro Gly
    1025                1030                1035

Ala Gly  Gly Met Val His His  Arg His Arg Ser Ser  Ser Thr Arg
    1040                1045                1050

Ser Gly  Gly Gly Asp Leu Thr  Leu Gly Leu Glu Pro  Ser Glu Glu
    1055                1060                1065

Glu Ala  Pro Arg Ser Pro Leu  Ala Pro Ser Glu Gly  Ala Gly Ser
    1070                1075                1080

Asp Val  Phe Asp Gly Asp Leu  Gly Met Gly Ala Ala  Lys Gly Leu
    1085                1090                1095

Gln Ser  Leu Pro Thr His Asp  Pro Ser Pro Leu Gln  Arg Tyr Ser
    1100                1105                1110

Glu Asp  Pro Thr Val Pro Leu  Pro Ser Glu Thr Asp  Gly Tyr Val
    1115                1120                1125

Ala Pro  Leu Thr Cys Ser Pro  Gln Pro Glu Tyr Val  Asn Gln Pro
    1130                1135                1140

Asp Val  Arg Pro Gln Pro Pro  Ser Pro Arg Glu Gly  Pro Leu Pro
    1145                1150                1155

Ala Ala  Arg Pro Ala Gly Ala  Thr Leu Glu Arg Ala  Lys Thr Leu
    1160                1165                1170

Ser Pro  Gly Lys Asn Gly Val  Val Lys Asp Val Phe  Ala Phe Gly
    1175                1180                1185

Gly Ala  Val Glu Asn Pro Glu  Tyr Leu Thr Pro Gln  Gly Gly Ala
    1190                1195                1200

Ala Pro  Gln Pro His Pro Pro  Pro Ala Phe Ser Pro  Ala Phe Asp
    1205                1210                1215
```

-continued

```
Asn Leu  Tyr Tyr Trp Asp Gln  Asp Pro Pro Glu Arg  Gly Ala Pro
    1220                 1225                 1230

Pro Ser  Thr Phe Lys Gly Thr  Pro Thr Ala Glu Asn  Pro Glu Tyr
    1235                 1240                 1245

Leu Gly  Leu Asp Val Pro Val
    1250                 1255
```

<210> SEQ ID NO 65
<211> LENGTH: 1836
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 65

```
atggagctgg cggccttgtg ccgctggggg ctcctcctcg ccctcttgcc ccccggagcc     60

gcgagcaccc aagtgtgcac cggcacagac atgaagctgc ggctccctgc cagtcccgag    120

acccacctgg acatgctccg ccacctctac cagggctgcc aggtggtgca gggaaacctg    180

gaactcacct acctgcccac caatgccagc ctgtccttcc tgcaggatat ccaggaggtg    240

cagggctacg tgctcatcgc tcacaaccaa gtgaggcagg tcccactgca gaggctgcgg    300

attgtgcgag gcacccagct ctttgaggac aactatgccc tggccgtgct agacaatgga    360

gacccgctga acaataccac ccctgtcaca ggggcctccc caggaggcct gcgggagctg    420

cagcttcgaa gcctcacaga gatcttgaaa ggaggggtct tgatccagcg gaaccccccag    480

ctctgctacc aggacacgat tttgtggaag gacatcttcc acaagaacaa ccagctggct    540

ctcacactga tagacaccaa ccgctctcgg gcctgccacc cctgttctcc gatgtgtaag    600

ggctcccgct gctggggaga gagttctgag gattgtcaga gcctgacgcg cactgtctgt    660

gccggtggct gtgcccgctg caaggggcca ctgcccactg actgctgcca tgagcagtgt    720

gctgccggct gcacgggccc caagcactct gactgcctgg cctgcctcca cttcaaccac    780

agtggcatct gtgagctgca ctgcccagcc ctggtcacct acaacacaga cacgtttgag    840

tccatgccca atcccgaggg ccggtataca ttcggcgcca gctgtgtgac tgcctgtccc    900

tacaactacc tttctacgga cgtgggatcc tgcaccctcg tctgcccct gcacaaccaa    960

gaggtgacag cagaggatgg aacacagcgg tgtgagaagt gcagcaagcc ctgtgcccga   1020

gtgtgctatg gtctgggcat ggagcacttg cgagaggtga gggcagttac cagtgccaat   1080

atccaggagt ttgctggctg caagaagatc tttgggagcc tggcatttct gccggagagc   1140

tttgatgggg acccagcctc caacactgcc ccgctccagc cagagcagct ccaagtgttt   1200

gagactctgg aagagatcac aggttaccta tacatctcag catggccgga cagcctgcct   1260

gacctcagcg tcttccagaa cctgcaagta atccggggac gaattctgca caatggcgcc   1320

tactcgctga ccctgcaagg gctgggcatc agctggctgg ggctgcgctc actgagggaa   1380

ctgggcagtg gactggccct catccaccat aacacccacc tctgcttcgt gcacacggtg   1440

ccctgggacc agctctttcg gaacccgcac caagctctgc tccacactgc caaccggcca   1500

gaggacgagt gtgtgggcga gggcctggcc tgccaccagc tgtgcgcccg agggcactgc   1560

tggggtccag ggcccaccca gtgtgtcaac tgcagccagt tccttcgggg ccaggagtgc   1620

gtggaggaat gccgagtact gcaggggctc cccagggagt atgtgaatgc caggcactgt   1680

ttgccgtgcc accctgagtg tcagccccag aatggctcag tgacctgttt tggaccggtg   1740
```

-continued

agctgctggc gggytcagag ctgggtggag gggggcagcg agggggattg ccagggactt 1800

ggcaggatgg cgagatgcag tagggtgtgc tatctg 1836

<210> SEQ ID NO 66
<211> LENGTH: 612
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (585)..(585)
<223> OTHER INFORMATION: Variable amino acid

<400> SEQUENCE: 66

Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
1 5 10 15

Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
20 25 30

Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
35 40 45

Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
50 55 60

Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65 70 75 80

Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
85 90 95

Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
100 105 110

Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
115 120 125

Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
130 135 140

Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145 150 155 160

Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
165 170 175

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

His Pro Cys Ser Pro Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
195 200 205

Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
210 215 220

Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225 230 235 240

Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
245 250 255

His Phe Asn His Ser Gly Ile Cys Glu Leu His Cys Pro Ala Leu Val
260 265 270

Thr Tyr Asn Thr Asp Thr Phe Glu Ser Met Pro Asn Pro Glu Gly Arg
275 280 285

Tyr Thr Phe Gly Ala Ser Cys Val Thr Ala Cys Pro Tyr Asn Tyr Leu
290 295 300

Ser Thr Asp Val Gly Ser Cys Thr Leu Val Cys Pro Leu His Asn Gln
305 310 315 320

-continued

```
Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys Glu Lys Cys Ser Lys
                325                 330                 335

Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met Glu His Leu Arg Glu
            340                 345                 350

Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu Phe Ala Gly Cys Lys
        355                 360                 365

Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu Ser Phe Asp Gly Asp
    370                 375                 380

Pro Ala Ser Asn Thr Ala Pro Leu Gln Pro Glu Gln Leu Gln Val Phe
385                 390                 395                 400

Glu Thr Leu Glu Glu Ile Thr Gly Tyr Leu Tyr Ile Ser Ala Trp Pro
                405                 410                 415

Asp Ser Leu Pro Asp Leu Ser Val Phe Gln Asn Leu Gln Val Ile Arg
            420                 425                 430

Gly Arg Ile Leu His Asn Gly Ala Tyr Ser Leu Thr Leu Gln Gly Leu
        435                 440                 445

Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg Glu Leu Gly Ser Gly
    450                 455                 460

Leu Ala Leu Ile His His Asn Thr His Leu Cys Phe Val His Thr Val
465                 470                 475                 480

Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln Ala Leu Leu His Thr
                485                 490                 495

Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu Gly Leu Ala Cys His
            500                 505                 510

Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro Gly Pro Thr Gln Cys
        515                 520                 525

Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu Cys Val Glu Glu Cys
    530                 535                 540

Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val Asn Ala Arg His Cys
545                 550                 555                 560

Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn Gly Ser Val Thr Cys
                565                 570                 575

Phe Gly Pro Val Ser Cys Trp Arg Xaa Gln Ser Trp Val Glu Gly Gly
            580                 585                 590

Ser Glu Gly Asp Cys Gln Gly Leu Gly Arg Met Ala Arg Cys Ser Arg
        595                 600                 605

Val Cys Tyr Leu
    610


<210> SEQ ID NO 67
<211> LENGTH: 33
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (6)..(6)
<223> OTHER INFORMATION: Variable amino acid

<400> SEQUENCE: 67

Val Ser Cys Trp Arg Xaa Gln Ser Trp Val Glu Gly Gly Ser Glu Gly
1               5                   10                  15

Asp Cys Gln Gly Leu Gly Arg Met Ala Arg Cys Ser Arg Val Cys Tyr
            20                  25                  30

Leu
```

-continued

<210> SEQ ID NO 68
<211> LENGTH: 3383
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 68

gctgcgcgca ggccgggaac acgaggccaa gagccgcagc cccagccgcc ttggtgcagc 60

gtacaccggc actagcccgc ttgcagcccc aggattagac agaagacgcg tcctcggcgc 120

ggtcgccgcc cagccgtagt cacctggatt acctacagcg gcagctgcag cggagccagc 180

gagaaggcca aaggggagca gcgtcccgag aggagcgcct cttttcaggg accccgccgg 240

ctggcggacg cgcgggaaag cggcgtcgcg aacagagcca gattgagggc ccgcgggtgg 300

agagagcgac gcccgagggg atggcggcag cgtcccggag cgcctctggc tgggcgctac 360

tgctgctggt ggcactttgg cagcagcgcg cggccggctc cggcgtcttc cagctgcagc 420

tgcaggagtt catcaacgag cgcggcgtac tggccagtgg gcggccttgc gagcccggct 480

gccggacttt cttccgcgtc tgccttaagc acttccaggc ggtcgtctcg cccggaccct 540

gcaccttcgg gaccgtctcc acgccggtat tgggcaccaa ctccttcgct gtccgggacg 600

acagtagcgg cgggggcgc aaccctctcc aactgccctt caatttcacc tggccgggta 660

ccttctcgct catcatcgaa gcttggcacg cgccaggaga cgacctgcgg ccagaggcct 720

tgccaccaga tgcactcatc agcaagatcg ccatccaggg ctccctagct gtgggtcaga 780

actggttatt ggatgagcaa accagcaccc tcacaaggct gcgctactct taccgggtca 840

tctgcagtga caactactat ggagacaact gctcccgcct gtgcaagaag cgcaatgacc 900

acttcggcca ctatgtgtgc cagccagatg gcaacttgtc ctgcctgccc ggttggactg 960

gggaatattg ccaacagcct atctgtcttt cgggctgtca tgaacagaat ggctactgca 1020

gcaagccagc agagtgcctc tgccgcccag gctggcaggg ccggctgtgt aacgaatgca 1080

tcccccacaa tggctgtcgc cacggcacct gcagcactcc ctggcaatgt acttgtgatg 1140

agggctgggg aggcctgttt tgtgaccaag atctcaacta ctgcacccac cactccccat 1200

gcaagaatgg ggcaacgtgc tccaacagtg ggcagcgaag ctacacctgc acctgtcgcc 1260

caggctacac tggtgtggac tgtgagctgg agctcagcga gtgtgacagc aacccctgtc 1320

gcaatggagg cagctgtaag gaccaggagg atggctacca ctgcctgtgt cctccgggct 1380

actatggcct gcattgtgaa cacagcacct tgagctgcgc cgactccccc tgcttcaatg 1440

ggggctcctg ccgggagcgc aaccaggggg ccaactatgc ttgtgaatgt ccccccaact 1500

tcaccggctc caactgcgag aagaaagtgg acaggtgcac cagcaacccc tgtgccaacg 1560

ggggacagtg cctgaaccga ggtccaagcc gcatgtgccg ctgccgtcct ggattcacgg 1620

gcacctactg tgaactccac gtcagcgact gtgcccgtaa cccttgcgcc cacggtggca 1680

cttgccatga cctggagaat gggctcatgt gcacctgccc tgccggcttc tctggccgac 1740

gctgtgaggt gcggacatcc atcgatgcct gtgcctcgag tccctgcttc aacagggcca 1800

cctgctacac cgacctctcc acagacacct ttgtgtgcaa ctgcccttat ggctttgtgg 1860

gcagccgctg cgagttcccc gtgggcttgc cgcccagctt cccctgggtg gccgtctcgc 1920

tgggtgtggg gctggcagtg ctgctggtac tgctgggcat ggtggcagtg gctgtgcggc 1980

agctgcggct tcgacggccg gacgacggca gcagggaagc catgaacaac ttgtcggact 2040

-continued

tccagaagga caacctgatt cctgccgccc agcttaaaaa cacaaaccag aagaaggagc 2100

tggaagtgga ctgtggcctg gacaagtcca actgtggcaa acagcaaaac cacacattgg 2160

actataatct ggccccaggg cccctggggc gggggaccat gccaggaaag tttccccaca 2220

gtgacaagag cttaggagag aaggcgccac tgcggttaca cagtgaaaag ccagagtgtc 2280

ggatatcagc gatatgctcc cccagggact ccatgtacca gtctgtgtgt ttgatatcag 2340

aggagaggaa tgaatgtgtc attgccacgg aggtataagg caggagccta cctggacatc 2400

cctgctcagc cccgcggctg gaccttcctt ctgcattgtt tacattgcat cctggatggg 2460

acgtttttca tatgcaacgt gctgctctca ggaggaggag ggaatggcag gaaccggaca 2520

gactgtgaac ttgccaagag atgcaatacc cttccacacc tttgggtgtc tgtctggcat 2580

cagattggca gctgcaccaa ccagaggaac agaagagaag agagatgcca ctgggcactg 2640

ccctgccagt agtggccttc agggggctcc ttccggggct ccggcctgtt ttccagagag 2700

agtggcagta gccccatggg gcccggagct gctgtggcct ccactggcat ccgtgtttcc 2760

aaaagtgcct ttggcccagg ctccacggcg acagttgggc ccaaatcaga aaggagagag 2820

ggggccaatg agggcagggc ctcctgtggg ctggaaaacc actgggtgcg tctcttgctg 2880

gggtttgccc tggaggtgag gtgagtgctc gagggagggg agtgctttct gccccatgcc 2940

tccaactact gtatgcaggc ctggctctct ggtctaggcc ctttgggcaa gaatgtccgt 3000

ctacccggct tccaccaccc tctggccctg ggcttctgta agcagacagg cagagggcct 3060

gcccctccca ccagccaagg gtgccaggcc taactggggc actcagggca gtgtgttgga 3120

aattccactg agggggaaat caggtgctgc ggccgcctgg gccctttcct ccctcaagcc 3180

catctccaca acctcgagcc tgggctctgg tccactactg ccccagacca ccctcaaagc 3240

tggtcttcag aaatcaataa tatgagtttt tattttgttt ttttttttt ttttgtagtt 3300

tattttggag tctagtattt caataattta agaatcagaa gcactgacct ttctacattt 3360

tataacatta ttttgtatat aat 3383

<210> SEQ ID NO 69
<211> LENGTH: 685
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 69

Met Ala Ala Ala Ser Arg Ser Ala Ser Gly Trp Ala Leu Leu Leu Leu
1 5 10 15

Val Ala Leu Trp Gln Gln Arg Ala Ala Gly Ser Gly Val Phe Gln Leu
20 25 30

Gln Leu Gln Glu Phe Ile Asn Glu Arg Gly Val Leu Ala Ser Gly Arg
35 40 45

Pro Cys Glu Pro Gly Cys Arg Thr Phe Phe Arg Val Cys Leu Lys His
50 55 60

Phe Gln Ala Val Val Ser Pro Gly Pro Cys Thr Phe Gly Thr Val Ser
65 70 75 80

Thr Pro Val Leu Gly Thr Asn Ser Phe Ala Val Arg Asp Asp Ser Ser
85 90 95

Gly Gly Gly Arg Asn Pro Leu Gln Leu Pro Phe Asn Phe Thr Trp Pro
100 105 110

-continued

```
Gly Thr Phe Ser Leu Ile Ile Glu Ala Trp His Ala Pro Gly Asp Asp
        115                 120                 125

Leu Arg Pro Glu Ala Leu Pro Pro Asp Ala Leu Ile Ser Lys Ile Ala
    130                 135                 140

Ile Gln Gly Ser Leu Ala Val Gly Gln Asn Trp Leu Leu Asp Glu Gln
145                 150                 155                 160

Thr Ser Thr Leu Thr Arg Leu Arg Tyr Ser Tyr Arg Val Ile Cys Ser
                165                 170                 175

Asp Asn Tyr Tyr Gly Asp Asn Cys Ser Arg Leu Cys Lys Lys Arg Asn
            180                 185                 190

Asp His Phe Gly His Tyr Val Cys Gln Pro Asp Gly Asn Leu Ser Cys
        195                 200                 205

Leu Pro Gly Trp Thr Gly Glu Tyr Cys Gln Gln Pro Ile Cys Leu Ser
    210                 215                 220

Gly Cys His Glu Gln Asn Gly Tyr Cys Ser Lys Pro Ala Glu Cys Leu
225                 230                 235                 240

Cys Arg Pro Gly Trp Gln Gly Arg Leu Cys Asn Glu Cys Ile Pro His
                245                 250                 255

Asn Gly Cys Arg His Gly Thr Cys Ser Thr Pro Trp Gln Cys Thr Cys
            260                 265                 270

Asp Glu Gly Trp Gly Gly Leu Phe Cys Asp Gln Asp Leu Asn Tyr Cys
        275                 280                 285

Thr His His Ser Pro Cys Lys Asn Gly Ala Thr Cys Ser Asn Ser Gly
    290                 295                 300

Gln Arg Ser Tyr Thr Cys Thr Cys Arg Pro Gly Tyr Thr Gly Val Asp
305                 310                 315                 320

Cys Glu Leu Glu Leu Ser Glu Cys Asp Ser Asn Pro Cys Arg Asn Gly
                325                 330                 335

Gly Ser Cys Lys Asp Gln Glu Asp Gly Tyr His Cys Leu Cys Pro Pro
            340                 345                 350

Gly Tyr Tyr Gly Leu His Cys Glu His Ser Thr Leu Ser Cys Ala Asp
        355                 360                 365

Ser Pro Cys Phe Asn Gly Gly Ser Cys Arg Glu Arg Asn Gln Gly Ala
    370                 375                 380

Asn Tyr Ala Cys Glu Cys Pro Pro Asn Phe Thr Gly Ser Asn Cys Glu
385                 390                 395                 400

Lys Lys Val Asp Arg Cys Thr Ser Asn Pro Cys Ala Asn Gly Gly Gln
                405                 410                 415

Cys Leu Asn Arg Gly Pro Ser Arg Met Cys Arg Cys Arg Pro Gly Phe
            420                 425                 430

Thr Gly Thr Tyr Cys Glu Leu His Val Ser Asp Cys Ala Arg Asn Pro
        435                 440                 445

Cys Ala His Gly Gly Thr Cys His Asp Leu Glu Asn Gly Leu Met Cys
    450                 455                 460

Thr Cys Pro Ala Gly Phe Ser Gly Arg Arg Cys Glu Val Arg Thr Ser
465                 470                 475                 480

Ile Asp Ala Cys Ala Ser Ser Pro Cys Phe Asn Arg Ala Thr Cys Tyr
                485                 490                 495

Thr Asp Leu Ser Thr Asp Thr Phe Val Cys Asn Cys Pro Tyr Gly Phe
            500                 505                 510

Val Gly Ser Arg Cys Glu Phe Pro Val Gly Leu Pro Pro Ser Phe Pro
```

-continued

```
        515                 520                 525

Trp Val Ala Val Ser Leu Gly Val Gly Leu Ala Val Leu Leu Val Leu
    530                 535                 540

Leu Gly Met Val Ala Val Ala Val Arg Gln Leu Arg Leu Arg Arg Pro
545                 550                 555                 560

Asp Asp Gly Ser Arg Glu Ala Met Asn Asn Leu Ser Asp Phe Gln Lys
                565                 570                 575

Asp Asn Leu Ile Pro Ala Ala Gln Leu Lys Asn Thr Asn Gln Lys Lys
            580                 585                 590

Glu Leu Glu Val Asp Cys Gly Leu Asp Lys Ser Asn Cys Gly Lys Gln
        595                 600                 605

Gln Asn His Thr Leu Asp Tyr Asn Leu Ala Pro Gly Pro Leu Gly Arg
    610                 615                 620

Gly Thr Met Pro Gly Lys Phe Pro His Ser Asp Lys Ser Leu Gly Glu
625                 630                 635                 640

Lys Ala Pro Leu Arg Leu His Ser Glu Lys Pro Glu Cys Arg Ile Ser
                645                 650                 655

Ala Ile Cys Ser Pro Arg Asp Ser Met Tyr Gln Ser Val Cys Leu Ile
            660                 665                 670

Ser Glu Glu Arg Asn Glu Cys Val Ile Ala Thr Glu Val
        675                 680                 685


<210> SEQ ID NO 70
<211> LENGTH: 2142
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 70

atggcggcag cgtcccggag cgcctctggc tgggcgctac tgctgctggt ggcactttgg     60

cagcagcgcg cggccggctc cggcgtcttc cagctgcagc tgcaggagtt catcaacgag    120

cgcggcgtac tggccagtgg gcggccttgc gagcccggct gccggacttt cttccgcgtc    180

tgccttaagc acttccaggc ggtcgtctcg cccggaccct gcaccttcgg gaccgtctcc    240

acgccggtat tgggcaccaa ctccttcgct gtccgggacg acagtagcgg cgggggggcgc    300

aaccctctcc aactgccctt caatttcacc tggccgggta ccttctcgct catcatcgaa    360

gcttggcacg cgccaggaga cgacctgcgg ccagaggcct tgccaccaga tgcactcatc    420

agcaagatcg ccatccaggg ctccctagct gtgggtcaga actggttatt ggatgagcaa    480

accagcaccc tcacaaggct gcgctactct taccgggtca tctgcagtga caactactat    540

ggagacaact gctcccgcct gtgcaagaag cgcaatgacc acttcggcca ctatgtgtgc    600

cagccagatg gcaacttgtc ctgcctgccc ggttggactg gggaatattg ccaacagcct    660

atctgtcttt cgggctgtca tgaacagaat ggctactgca gcaagccagc agagtgcctc    720

tgccgcccag gctggcaggg ccggctgtgt aacgaatgca tcccccacaa tggctgtcgc    780

cacggcacct gcagcactcc ctggcaatgt acttgtgatg agggctgggg aggcctgttt    840

tgtgaccaag atctcaacta ctgcacccac cactccccat gcaagaatgg ggcaacgtgc    900

tccaacagtg ggcagcgaag ctacacctgc acctgtcgcc caggctacac tggtgtggac    960

tgtgagctgg agctcagcga gtgtgacagc aacccctgtc gcaatggagg cagctgtaag   1020

gaccaggagg atggctacca ctgcctgtgt cctccgggct actatggcct gcattgtgaa   1080
```

-continued

```
cacagcacct tgagctgcgc cgactcccc tgcttcaatg ggggctcctg ccgggagcgc   1140

aaccaggggg ccaactatgc ttgtgaatgt ccccccaact tcaccggctc caactgcgag   1200

aagaaagtgg acaggtgcac cagcaacccc tgtgccaacg ggggacagtg cctgaaccga   1260

ggtccaagcc gcatgtgccg ctgccgtcct ggattcacgg gcacctactg tgaactccac   1320

gtcagcgact gtgcccgtaa cccttgcgcc cacggtggca cttgccatga cctggagaat   1380

gggctcatgt gcacctgccc tgccggcttc tctggccgac gctgtgaggt gcggacatcc   1440

atcgatgcct gtgcctcgag tccctgcttc aacagggcca cctgctacac cgacctctcc   1500

acagacacct ttgtgtgcaa ctgcccttat ggctttgtgg gcagccgctg cgagttcccc   1560

gtgggcttgc cgcccagctt cccctgggtg gccgtctcgc tgggtgtggg gctggcagtg   1620

ctgctggtac tgctgggcat ggtggcagtg gctgtgcggc agctgcggct tcgacggccg   1680

gacgacggca gcagggaagc catgaacaac ttgtcggact tccagaagga caacctgatt   1740

cctgccgccc agcttaaaaa cacaaaccag aagaaggagc tggaagtgga ctgtggcctg   1800

gacaagtcca actgtggcaa acagcaaaac cacacattgg actataatct ggccccaggg   1860

cccctggggc gggggaccat gccaggaaag tttccccaca gtgacaagag cttaggagag   1920

aaggcgccac tgcggttaca cagtgaaaag ccagagtgtc ggatatcagc gatatgctcc   1980

cccagggact ccatgtacca gtctgtgtgt ttgatatcag aggagaggaa tgaatgtgtc   2040

attgccacgg aggtgagtgc tgggctcgcc tttccttctg ccttttgtgg gagggaaagt   2100

ggcctggtca ctcttgaccc atgggccatt cctgaagggt ag                      2142
```

<210> SEQ ID NO 71
<211> LENGTH: 713
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 71

```
Met Ala Ala Ala Ser Arg Ser Ala Ser Gly Trp Ala Leu Leu Leu Leu
1               5                   10                  15

Val Ala Leu Trp Gln Gln Arg Ala Ala Gly Ser Gly Val Phe Gln Leu
            20                  25                  30

Gln Leu Gln Glu Phe Ile Asn Glu Arg Gly Val Leu Ala Ser Gly Arg
        35                  40                  45

Pro Cys Glu Pro Gly Cys Arg Thr Phe Phe Arg Val Cys Leu Lys His
    50                  55                  60

Phe Gln Ala Val Val Ser Pro Gly Pro Cys Thr Phe Gly Thr Val Ser
65                  70                  75                  80

Thr Pro Val Leu Gly Thr Asn Ser Phe Ala Val Arg Asp Asp Ser Ser
                85                  90                  95

Gly Gly Gly Arg Asn Pro Leu Gln Leu Pro Phe Asn Phe Thr Trp Pro
            100                 105                 110

Gly Thr Phe Ser Leu Ile Ile Glu Ala Trp His Ala Pro Gly Asp Asp
        115                 120                 125

Leu Arg Pro Glu Ala Leu Pro Pro Asp Ala Leu Ile Ser Lys Ile Ala
    130                 135                 140

Ile Gln Gly Ser Leu Ala Val Gly Gln Asn Trp Leu Leu Asp Glu Gln
145                 150                 155                 160

Thr Ser Thr Leu Thr Arg Leu Arg Tyr Ser Tyr Arg Val Ile Cys Ser
                165                 170                 175
```

-continued

```
Asp Asn Tyr Tyr Gly Asp Asn Cys Ser Arg Leu Cys Lys Lys Arg Asn
            180                 185                 190

Asp His Phe Gly His Tyr Val Cys Gln Pro Asp Gly Asn Leu Ser Cys
        195                 200                 205

Leu Pro Gly Trp Thr Gly Glu Tyr Cys Gln Gln Pro Ile Cys Leu Ser
    210                 215                 220

Gly Cys His Glu Gln Asn Gly Tyr Cys Ser Lys Pro Ala Glu Cys Leu
225                 230                 235                 240

Cys Arg Pro Gly Trp Gln Gly Arg Leu Cys Asn Glu Cys Ile Pro His
                245                 250                 255

Asn Gly Cys Arg His Gly Thr Cys Ser Thr Pro Trp Gln Cys Thr Cys
            260                 265                 270

Asp Glu Gly Trp Gly Gly Leu Phe Cys Asp Gln Asp Leu Asn Tyr Cys
        275                 280                 285

Thr His His Ser Pro Cys Lys Asn Gly Ala Thr Cys Ser Asn Ser Gly
    290                 295                 300

Gln Arg Ser Tyr Thr Cys Thr Cys Arg Pro Gly Tyr Thr Gly Val Asp
305                 310                 315                 320

Cys Glu Leu Glu Leu Ser Glu Cys Asp Ser Asn Pro Cys Arg Asn Gly
                325                 330                 335

Gly Ser Cys Lys Asp Gln Glu Asp Gly Tyr His Cys Leu Cys Pro Pro
            340                 345                 350

Gly Tyr Tyr Gly Leu His Cys Glu His Ser Thr Leu Ser Cys Ala Asp
        355                 360                 365

Ser Pro Cys Phe Asn Gly Gly Ser Cys Arg Glu Arg Asn Gln Gly Ala
    370                 375                 380

Asn Tyr Ala Cys Glu Cys Pro Pro Asn Phe Thr Gly Ser Asn Cys Glu
385                 390                 395                 400

Lys Lys Val Asp Arg Cys Thr Ser Asn Pro Cys Ala Asn Gly Gly Gln
                405                 410                 415

Cys Leu Asn Arg Gly Pro Ser Arg Met Cys Arg Cys Arg Pro Gly Phe
            420                 425                 430

Thr Gly Thr Tyr Cys Glu Leu His Val Ser Asp Cys Ala Arg Asn Pro
        435                 440                 445

Cys Ala His Gly Gly Thr Cys His Asp Leu Glu Asn Gly Leu Met Cys
    450                 455                 460

Thr Cys Pro Ala Gly Phe Ser Gly Arg Arg Cys Glu Val Arg Thr Ser
465                 470                 475                 480

Ile Asp Ala Cys Ala Ser Ser Pro Cys Phe Asn Arg Ala Thr Cys Tyr
                485                 490                 495

Thr Asp Leu Ser Thr Asp Thr Phe Val Cys Asn Cys Pro Tyr Gly Phe
            500                 505                 510

Val Gly Ser Arg Cys Glu Phe Pro Val Gly Leu Pro Pro Ser Phe Pro
        515                 520                 525

Trp Val Ala Val Ser Leu Gly Val Gly Leu Ala Val Leu Leu Val Leu
    530                 535                 540

Leu Gly Met Val Ala Val Ala Val Arg Gln Leu Arg Leu Arg Arg Pro
545                 550                 555                 560

Asp Asp Gly Ser Arg Glu Ala Met Asn Asn Leu Ser Asp Phe Gln Lys
                565                 570                 575
```

-continued

```
Asp Asn Leu Ile Pro Ala Ala Gln Leu Lys Asn Thr Asn Gln Lys Lys
            580                 585                 590

Glu Leu Glu Val Asp Cys Gly Leu Asp Lys Ser Asn Cys Gly Lys Gln
        595                 600                 605

Gln Asn His Thr Leu Asp Tyr Asn Leu Ala Pro Gly Pro Leu Gly Arg
    610                 615                 620

Gly Thr Met Pro Gly Lys Phe Pro His Ser Asp Lys Ser Leu Gly Glu
625                 630                 635                 640

Lys Ala Pro Leu Arg Leu His Ser Glu Lys Pro Glu Cys Arg Ile Ser
                645                 650                 655

Ala Ile Cys Ser Pro Arg Asp Ser Met Tyr Gln Ser Val Cys Leu Ile
            660                 665                 670

Ser Glu Glu Arg Asn Glu Cys Val Ile Ala Thr Glu Val Ser Ala Gly
        675                 680                 685

Leu Ala Phe Pro Ser Ala Phe Cys Gly Arg Glu Ser Gly Leu Val Thr
    690                 695                 700

Leu Asp Pro Trp Ala Ile Pro Glu Gly
705                 710


<210> SEQ ID NO 72
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 72

Ser Ala Gly Leu Ala Phe Pro Ser Ala Phe Cys Gly Arg Glu Ser Gly
1               5                   10                  15

Leu Val Thr Leu Asp Pro Trp Ala Ile Pro Glu Gly
            20                  25


<210> SEQ ID NO 73
<211> LENGTH: 222
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 73

ggccgccagt gtgatggata tctgcagaat tcgcccttcc cagctgcccc gtagtggggc     60

tcttactgtt ttcttttatt ccaagcccac tatgcgagat ttgtgtgtca gatgtaagct    120

cgagagagct tgccttgaag ggcgaattcc agcacactgg gcggccgtta ctagtggatc    180

cgagctcggt accaagcttg atgcatagct tgactattct at                       222


<210> SEQ ID NO 74
<211> LENGTH: 1543
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 74

gctctcatta ccttctgccc atcacttaat aaatagccag ccaattcatc aacattctgg     60

tacactgttg gagagatgag acagtcacac cagctgcccc tagtggggct cttactgttt    120

tcttttattc caagccaact atgcgagatt tgtgaggtaa gtgaagaaaa ctacatccgc    180

ctaaaacctc tgttgaatac aatgatccag tcaaactata acaggggaac cagcgctgtc    240

aatgttgtgt tgtccctcaa acttgttgga atccagatcc aaaccctgat gcaaaagatg    300

atccaacaaa tcaaatacaa tgtgaaaagc agattgtcag atgtaagctc gggagagctt    360
```

-continued

gccttgatta tactggcttt gggagtatgt cgtaacgctg aggaaaactt aatatatgat 420

taccacctga tcgacaagct agaaaataaa ttccaagcag aaattgaaaa tatggaagca 480

cacaatggca ctcccctgac taactactac cagctcagcc tggacgtttt ggccttgtgt 540

ctgttcaatg ggaactactc aaccgccgaa gttgtcaacc acttcactcc tgaaaataaa 600

aactattatt ttggtagcca gttctcagta gatactggtg caatggctgt cctggctctg 660

acctgtgtga agaagagtct aataaatggg cagatcaaag cagatgaagg cagtttaaag 720

aacatcagta tttatacaaa gtcactggta gaaaagattc tgtctgagaa aaaagaaaat 780

ggtctcattg gaaacacatt tagcacagga gaagccatgc aggccctctt tgtatcatca 840

gactattata atgaaaatga ctggaattgc caacaaactc tgaatacagt gctcacggaa 900

atttctcaag gagcattcag caatccaaac gctgcagccc aggtcttacc tgccctgatg 960

ggaaagacct tcttggatat taacaaagac tcttcttgcg tctctgcttc aggtaacttc 1020

aacatctccg ctgatgagcc tataactgtg acacctcctg actcacaatc atatatctcc 1080

gtcaattact ctgtgagaat caatgaaaca tatttcacca atgtcactgt gctaaatggt 1140

tctgtcttcc tcagtgtgat ggagaaagcc cagaaaatga atgatactat atttggtttc 1200

acaatggagg agcgctcatg ggggccctat atcacctgta ttcagggcct atgtgccaac 1260

aataatgaca gaacctactg ggaacttctg agtggaggcg aaccactgag ccaaggagct 1320

ggtagttacg ttgtccgcaa tggagaaaac ttggaggttc gctggagcaa atactaataa 1380

gcccaaactt tcctcagctg cataaaatcc atttgcagtg gagttccatg tttattgtcc 1440

ttatgccttc ttcttcattt atcccagtac gagcaggaga gttaataacc tccccttctc 1500

tctctacatg ttcaataaaa gttgttgaaa gattaacaac tgt 1543

<210> SEQ ID NO 75
<211> LENGTH: 433
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 75

Met Arg Gln Ser His Gln Leu Pro Leu Val Gly Leu Leu Leu Phe Ser
1 5 10 15

Phe Ile Pro Ser Gln Leu Cys Glu Ile Cys Glu Val Ser Glu Glu Asn
20 25 30

Tyr Ile Arg Leu Lys Pro Leu Leu Asn Thr Met Ile Gln Ser Asn Tyr
35 40 45

Asn Arg Gly Thr Ser Ala Val Asn Val Val Leu Ser Leu Lys Leu Val
50 55 60

Gly Ile Gln Ile Gln Thr Leu Met Gln Lys Met Ile Gln Gln Ile Lys
65 70 75 80

Tyr Asn Val Lys Ser Arg Leu Ser Asp Val Ser Ser Gly Glu Leu Ala
85 90 95

Leu Ile Ile Leu Ala Leu Gly Val Cys Arg Asn Ala Glu Glu Asn Leu
100 105 110

Ile Tyr Asp Tyr His Leu Ile Asp Lys Leu Glu Asn Lys Phe Gln Ala
115 120 125

Glu Ile Glu Asn Met Glu Ala His Asn Gly Thr Pro Leu Thr Asn Tyr
130 135 140

-continued

```
Tyr Gln Leu Ser Leu Asp Val Leu Ala Leu Cys Leu Phe Asn Gly Asn
145                 150                 155                 160

Tyr Ser Thr Ala Glu Val Val Asn His Phe Thr Pro Glu Asn Lys Asn
                165                 170                 175

Tyr Tyr Phe Gly Ser Gln Phe Ser Val Asp Thr Gly Ala Met Ala Val
            180                 185                 190

Leu Ala Leu Thr Cys Val Lys Lys Ser Leu Ile Asn Gly Gln Ile Lys
        195                 200                 205

Ala Asp Glu Gly Ser Leu Lys Asn Ile Ser Ile Tyr Thr Lys Ser Leu
    210                 215                 220

Val Glu Lys Ile Leu Ser Glu Lys Lys Glu Asn Gly Leu Ile Gly Asn
225                 230                 235                 240

Thr Phe Ser Thr Gly Glu Ala Met Gln Ala Leu Phe Val Ser Ser Asp
                245                 250                 255

Tyr Tyr Asn Glu Asn Asp Trp Asn Cys Gln Gln Thr Leu Asn Thr Val
            260                 265                 270

Leu Thr Glu Ile Ser Gln Gly Ala Phe Ser Asn Pro Asn Ala Ala Ala
        275                 280                 285

Gln Val Leu Pro Ala Leu Met Gly Lys Thr Phe Leu Asp Ile Asn Lys
    290                 295                 300

Asp Ser Ser Cys Val Ser Ala Ser Gly Asn Phe Asn Ile Ser Ala Asp
305                 310                 315                 320

Glu Pro Ile Thr Val Thr Pro Pro Asp Ser Gln Ser Tyr Ile Ser Val
                325                 330                 335

Asn Tyr Ser Val Arg Ile Asn Glu Thr Tyr Phe Thr Asn Val Thr Val
            340                 345                 350

Leu Asn Gly Ser Val Phe Leu Ser Val Met Glu Lys Ala Gln Lys Met
        355                 360                 365

Asn Asp Thr Ile Phe Gly Phe Thr Met Glu Glu Arg Ser Trp Gly Pro
    370                 375                 380

Tyr Ile Thr Cys Ile Gln Gly Leu Cys Ala Asn Asn Asn Asp Arg Thr
385                 390                 395                 400

Tyr Trp Glu Leu Leu Ser Gly Gly Glu Pro Leu Ser Gln Gly Ala Gly
                405                 410                 415

Ser Tyr Val Val Arg Asn Gly Glu Asn Leu Glu Val Arg Trp Ser Lys
            420                 425                 430

Tyr
```

<210> SEQ ID NO 76
<211> LENGTH: 1122
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 76

```
atgagacagt cacaccagct gcccctagtg gggctcttac tgttttcttt tattccaagc     60

caactatgcg agatttgtgt gtcagatgta agctcgggag agcttgcctt gattatactg    120

gctttgggag tatgtcgtaa cgctgaggaa aacttaatat atgattacca cctgatcgac    180

aagctagaaa ataaattcca agcagaaatt gaaaatatgg aagcacacaa tggcactccc    240

ctgactaact actaccagct cagcctggac gttttggcct tgtgtctgtt caatgggaac    300

tactcaaccg ccgaagttgt caaccacttc actcctgaaa ataaaaacta ttattttggt    360
```

-continued

```
agccagttct cagtagatac tggtgcaatg gctgtcctgg ctctgacctg tgtgaagaag    420

agtctaataa atgggcagat caaagcagat gaaggcagtt taaagaacat cagtatttat    480

acaaagtcac tggtagaaaa gattctgtct gagaaaaaag aaaatggtct cattggaaac    540

acatttagca caggagaagc catgcaggcc ctctttgtat catcagacta ttataatgaa    600

aatgactgga attgccaaca aactctgaat acagtgctca cggaaatttc tcaaggagca    660

ttcagcaatc caaacgctgc agcccaggtc ttacctgccc tgatgggaaa gaccttcttg    720

gatattaaca aagactcttc ttgcgtctct gcttcaggta acttcaacat ctccgctgat    780

gagcctataa ctgtgacacc tcctgactca caatcatata tctccgtcaa ttactctgtg    840

agaatcaatg aaacatattt caccaatgtc actgtgctaa atggttctgt cttcctcagt    900

gtgatggaga aagcccagaa aatgaatgat actatatttg gtttcacaat ggaggagcgc    960

tcatgggggc cctatatcac ctgtattcag ggcctatgtg ccaacaataa tgacagaacc   1020

tactgggaac ttctgagtgg aggcgaacca ctgagccaag gagctggtag ttacgttgtc   1080

cgcaatggag aaaacttgga ggttcgctgg agcaaatact aa                      1122
```

<210> SEQ ID NO 77
<211> LENGTH: 373
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 77

```
Met Arg Gln Ser His Gln Leu Pro Leu Val Gly Leu Leu Leu Phe Ser
1               5                   10                  15

Phe Ile Pro Ser Gln Leu Cys Glu Ile Cys Val Ser Asp Val Ser Ser
            20                  25                  30

Gly Glu Leu Ala Leu Ile Ile Leu Ala Leu Gly Val Cys Arg Asn Ala
        35                  40                  45

Glu Glu Asn Leu Ile Tyr Asp Tyr His Leu Ile Asp Lys Leu Glu Asn
    50                  55                  60

Lys Phe Gln Ala Glu Ile Glu Asn Met Glu Ala His Asn Gly Thr Pro
65                  70                  75                  80

Leu Thr Asn Tyr Tyr Gln Leu Ser Leu Asp Val Leu Ala Leu Cys Leu
                85                  90                  95

Phe Asn Gly Asn Tyr Ser Thr Ala Glu Val Val Asn His Phe Thr Pro
            100                 105                 110

Glu Asn Lys Asn Tyr Tyr Phe Gly Ser Gln Phe Ser Val Asp Thr Gly
        115                 120                 125

Ala Met Ala Val Leu Ala Leu Thr Cys Val Lys Lys Ser Leu Ile Asn
    130                 135                 140

Gly Gln Ile Lys Ala Asp Glu Gly Ser Leu Lys Asn Ile Ser Ile Tyr
145                 150                 155                 160

Thr Lys Ser Leu Val Glu Lys Ile Leu Ser Glu Lys Lys Glu Asn Gly
                165                 170                 175

Leu Ile Gly Asn Thr Phe Ser Thr Gly Glu Ala Met Gln Ala Leu Phe
            180                 185                 190

Val Ser Ser Asp Tyr Tyr Asn Glu Asn Asp Trp Asn Cys Gln Gln Thr
        195                 200                 205

Leu Asn Thr Val Leu Thr Glu Ile Ser Gln Gly Ala Phe Ser Asn Pro
    210                 215                 220
```

-continued

```
Asn Ala Ala Ala Gln Val Leu Pro Ala Leu Met Gly Lys Thr Phe Leu
225                 230                 235                 240

Asp Ile Asn Lys Asp Ser Ser Cys Val Ser Ala Ser Gly Asn Phe Asn
                245                 250                 255

Ile Ser Ala Asp Glu Pro Ile Thr Val Thr Pro Pro Asp Ser Gln Ser
            260                 265                 270

Tyr Ile Ser Val Asn Tyr Ser Val Arg Ile Asn Glu Thr Tyr Phe Thr
        275                 280                 285

Asn Val Thr Val Leu Asn Gly Ser Val Phe Leu Ser Val Met Glu Lys
    290                 295                 300

Ala Gln Lys Met Asn Asp Thr Ile Phe Gly Phe Thr Met Glu Glu Arg
305                 310                 315                 320

Ser Trp Gly Pro Tyr Ile Thr Cys Ile Gln Gly Leu Cys Ala Asn Asn
                325                 330                 335

Asn Asp Arg Thr Tyr Trp Glu Leu Leu Ser Gly Gly Glu Pro Leu Ser
            340                 345                 350

Gln Gly Ala Gly Ser Tyr Val Val Arg Asn Gly Glu Asn Leu Glu Val
        355                 360                 365

Arg Trp Ser Lys Tyr
    370


<210> SEQ ID NO 78
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 78

Gln Leu Cys Glu Ile Cys Val Ser Asp Val Ser Ser Gly
1               5                   10


<210> SEQ ID NO 79
<211> LENGTH: 1876
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 79

ccaccaacca aggcaactca gctgatgctg taacaaccac agaaactgcg actagtggtc     60

ctacagtagc tgcagctgat accactgaaa ctaatttccc tgaaactgct agcaccacag    120

caaatacacc ttctttccca acagctactt cacctgctcc ccccataatt agtacacata    180

gttcctccac aattcctaca cctgctcccc ccataattag tacacatagt tcctccacaa    240

ttcctatacc tactgctgca gacagtgagt caaccacaaa tgtaaattca ttagctacct    300

ctgacataat caccgcttca tctccaaatg atggattaat cacaatggtt ccttctgaaa    360

cacaaagtaa caatgaaatg tcccccacca cagaagacaa tcaatcatca gggcctccca    420

ctggcaccgc tttattggag accagcaccc taaacagcac aggtaaggac aattccctca    480

aagactcccc aaggatggag agtacaattg tggagggagg gaggtgggta tgttctggtg    540

gggtaggaaa atgggtctaa agaggcccaa ggtttctttg gaaatgaact gtgggcttga    600

aatctgtgtc acaatgtagt gacagcagag gccctaaaat cctcagtttc cttctgttct    660

cctttccttc cctcccatcc tacctcaagt tatttgctag ctgtgcagta ttttttttaa    720

tgaacttggg ctttgtgaaa ggtagctcta tagctttcct ctccctaggg gatttttaga    780

aatagtttca gctgctctga gaagaccagg tccaaataag cacagactgt ccagcaccac    840
```

-continued

tggtctagct ctctggtcta cctaaggaag aagtcagttc tacccttgta agatggtggc 900

tgatcacaat gtggaacttg agccaagtca tagaatttga gactgagggg tagaaagcaa 960

aacaacagca gtgttcaaac aatcctaggc ataaccaggg ccattgtcca aaggacagat 1020

gctttgggaa cactaagaga tgacaagtga cctcaagcag aagttaatga aacaggaacc 1080

atttgcttca cttccccaag ctcagcctat aacagaaatc aactacgttt atcagtaaag 1140

gctaaatggc cttgtgggcc atgtggggtc tctgttgtaa ctcctgcact ctactatttt 1200

aacatgaaag cagccacagt taacaaacaa aaggttggtc agatatgact cgtgaaccat 1260

agtttgccag cccttgatct agaagaatcg tcacacttta gagcctaaag aatctagata 1320

aatcctatca ttctcagatg ggaaatctgg gacccagtga gggccatgac tcacccaggg 1380

tcaaagatca agtggatgtc tccctaccct taacttctgg tagcttcctc aatgttcttt 1440

gatagattta agaaatagat ggttaagcaa gtagacccca gaggctgtat ctaagacctc 1500

tttccccaat ctttcatgtt tggaggggcc actctgaagg cgggatccaa tgggacacag 1560

ctgtcctggg atcaggaaag agaggttttc taagccattt ctgcttcgcc aggtgttccc 1620

tcagagtcag gccatcttcc tgtgttctgg ccctaccatg aacaaactgt ggggcatggg 1680

gcaagtcatt tctctcttgg cttcaactta gtgatctgca ataaggcgag actgaactaa 1740

aaagcccccc aaatctcttc tggctgtaac atcctgtgac ttaatcaatt cctggccatg 1800

aaacaagtta atgagtctgt ccttcgttgc tgaagagaaa gcacctcaga gttgtttgtc 1860

tggtgtctcc agaagg 1876

<210> SEQ ID NO 80
<211> LENGTH: 1539
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 80

atgaaagcca tcattcatct tactcttctt gctctccttt ctgtaaacac agccaccaac 60

caaggcaact cagctgatgc tgtaacaacc acagaaactg cgactagtgg tcctacagta 120

gctgcagctg ataccactga aactaatttc cctgaaactg ctagcaccac agcaaataca 180

ccttctttcc caacagctac ttcacctgct ccccccataa ttagtacaca tagttcctcc 240

acaattccta cacctgctcc ccccataatt agtacacata gttcctccac aattcctata 300

cctactgctg cagacagtga gtcaaccaca aatgtaaatt cattagctac ctctgacata 360

atcaccgctt catctccaaa tgatggatta atcacaatgg ttccttctga aacacaaagt 420

aacaatgaaa tgtcccccac cacagaagac aatcaatcat cagggcctcc cactggcacc 480

gctttattgg agaccagcac cctaaacagc acaggtccca gcaatccttg ccaagatgat 540

ccctgtgcag ataattcgtt atgtgttaag ctgcataata caagtttttg cctgtgttta 600

gaagggtatt actacaactc ttctacatgt aagaaaggaa aggtattccc tgggaagatt 660

tcagtgacag tatcagaaac atttgaccca gaagagaaac attccatggc ctatcaagac 720

ttgcatagtg aaattactag cttgtttaaa gatgtatttg gcacatctgt ttatggacag 780

actgtaattc ttactgtaag cacatctctg tcaccaagat ctgaaatgcg tgctgatgac 840

aagtttgtta atgtaacaat agtaacaatt ttggcagaaa ccacaagtga caatgagaag 900

actgtgactg agaaaattaa taaagcaatt agaagtagct caagcaactt tctaaactat 960

-continued

```
gatttgaccc ttcggtgtga ttattatggc tgtaaccaga ctgcggatga ctgcctcaat   1020

ggtttagcat gcgattgcaa atctgacctg caaaggccta acccacagag ccctttctgc   1080

gttgcttcca gtctcaagtg tcctgatgcc tgcaacgcac agcacaagca atgcttaata   1140

aagaagagtg gtggggcccc tgagtgtgcg tgcgtgcccg gctaccagga agatgctaat   1200

gggaactgcc aaaagtgtgc atttggctac agtggactcg actgtaagga caaatttcag   1260

ctgatcctca ctattgtggg caccatcgct ggcattgtca ttctcagcat gataattgca   1320

ttgattgtca cagcaagatc aaataacaaa acgaagcata ttgaagaaga gaacttgatt   1380

gacgaagact ttcaaaatct aaaactgcgg tcgacaggct tcaccaatct tggagcagaa   1440

gggagcgtct ttcctaaggt caggataacg gcctccagag acagccagat gcaaaatccc   1500

tattcaagac acagcagcat gccccgccct gactattag                          1539
```

<210> SEQ ID NO 81
<211> LENGTH: 512
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 81

```
Met Lys Ala Ile Ile His Leu Thr Leu Leu Ala Leu Leu Ser Val Asn
1               5                  10                  15

Thr Ala Thr Asn Gln Gly Asn Ser Ala Asp Ala Val Thr Thr Thr Glu
            20                  25                  30

Thr Ala Thr Ser Gly Pro Thr Val Ala Ala Ala Asp Thr Thr Glu Thr
        35                  40                  45

Asn Phe Pro Glu Thr Ala Ser Thr Thr Ala Asn Thr Pro Ser Phe Pro
    50                  55                  60

Thr Ala Thr Ser Pro Ala Pro Pro Ile Ile Ser Thr His Ser Ser Ser
65                  70                  75                  80

Thr Ile Pro Thr Pro Ala Pro Pro Ile Ile Ser Thr His Ser Ser Ser
                85                  90                  95

Thr Ile Pro Ile Pro Thr Ala Ala Asp Ser Glu Ser Thr Thr Asn Val
            100                 105                 110

Asn Ser Leu Ala Thr Ser Asp Ile Ile Thr Ala Ser Ser Pro Asn Asp
        115                 120                 125

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

Ser Pro Thr Thr Glu Asp Asn Gln Ser Ser Gly Pro Pro Thr Gly Thr
145                 150                 155                 160

Ala Leu Leu Glu Thr Ser Thr Leu Asn Ser Thr Gly Pro Ser Asn Pro
                165                 170                 175

Cys Gln Asp Asp Pro Cys Ala Asp Asn Ser Leu Cys Val Lys Leu His
            180                 185                 190

Asn Thr Ser Phe Cys Leu Cys Leu Glu Gly Tyr Tyr Tyr Asn Ser Ser
        195                 200                 205

Thr Cys Lys Lys Gly Lys Val Phe Pro Gly Lys Ile Ser Val Thr Val
    210                 215                 220

Ser Glu Thr Phe Asp Pro Glu Glu Lys His Ser Met Ala Tyr Gln Asp
225                 230                 235                 240

Leu His Ser Glu Ile Thr Ser Leu Phe Lys Asp Val Phe Gly Thr Ser
                245                 250                 255
```

-continued

```
Val Tyr Gly Gln Thr Val Ile Leu Thr Val Ser Thr Ser Leu Ser Pro
            260                 265                 270

Arg Ser Glu Met Arg Ala Asp Asp Lys Phe Val Asn Val Thr Ile Val
        275                 280                 285

Thr Ile Leu Ala Glu Thr Thr Ser Asp Asn Glu Lys Thr Val Thr Glu
    290                 295                 300

Lys Ile Asn Lys Ala Ile Arg Ser Ser Ser Ser Asn Phe Leu Asn Tyr
305                 310                 315                 320

Asp Leu Thr Leu Arg Cys Asp Tyr Tyr Gly Cys Asn Gln Thr Ala Asp
                325                 330                 335

Asp Cys Leu Asn Gly Leu Ala Cys Asp Cys Lys Ser Asp Leu Gln Arg
            340                 345                 350

Pro Asn Pro Gln Ser Pro Phe Cys Val Ala Ser Ser Leu Lys Cys Pro
        355                 360                 365

Asp Ala Cys Asn Ala Gln His Lys Gln Cys Leu Ile Lys Lys Ser Gly
    370                 375                 380

Gly Ala Pro Glu Cys Ala Cys Val Pro Gly Tyr Gln Glu Asp Ala Asn
385                 390                 395                 400

Gly Asn Cys Gln Lys Cys Ala Phe Gly Tyr Ser Gly Leu Asp Cys Lys
                405                 410                 415

Asp Lys Phe Gln Leu Ile Leu Thr Ile Val Gly Thr Ile Ala Gly Ile
            420                 425                 430

Val Ile Leu Ser Met Ile Ile Ala Leu Ile Val Thr Ala Arg Ser Asn
        435                 440                 445

Asn Lys Thr Lys His Ile Glu Glu Glu Asn Leu Ile Asp Glu Asp Phe
    450                 455                 460

Gln Asn Leu Lys Leu Arg Ser Thr Gly Phe Thr Asn Leu Gly Ala Glu
465                 470                 475                 480

Gly Ser Val Phe Pro Lys Val Arg Ile Thr Ala Ser Arg Asp Ser Gln
                485                 490                 495

Met Gln Asn Pro Tyr Ser Arg His Ser Ser Met Pro Arg Pro Asp Tyr
            500                 505                 510


<210> SEQ ID NO 82
<211> LENGTH: 4284
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 82

ccacgcgtcc gagcaagaac agctaaaatg aaagccatca ttcatcttac tcttcttgct     60

ctcctttctg taaacacagc caccaaccaa ggcaactcag ctgatgctgt aacaaccaca    120

gaaactgcga ctagtggtcc tacagtagct gcagctgata ccactgaaac taatttccct    180

gaaactgcta gcaccacagc aaatacacct tctttcccaa cagctacttc acctgctccc    240

cccataatta gtacacatag ttcctccaca attcctacac ctgctccccc cataattagt    300

acacatagtt cctccacaat tcctatacct actgctgcag acagtgagtc aaccacaaat    360

gtaaattcat tagctacctc tgacataatc accgcttcat ctccaaatga tggattaatc    420

acaatggttc cttctgaaac acaaagtaac aatgaaatgt cccccaccac agaagacaat    480

caatcatcag ggcctcccac tggcaccgct ttattggaga ccagcaccct aaacagcaca    540

ggtaaggaca attccctcaa agactcccca aggatggaga gtacaattgt ggagggaggg    600
```

-continued

aggtgggtat gttctggtgg ggtaggaaaa tgggtctaaa gaggcccaag gtttctttgg 660

aaatgaactg tgggcttgaa atctgtgtca caatgtagtg acagcagagg ccctaaaatc 720

ctcagtttcc ttctgttctc ctttccttcc ctcccatcct acctcaagtt atttgctagc 780

tgtgcagtat tttttttaat gaacttgggc tttgtgaaag gtagctctat agctttcctc 840

tccctagggg atttttagaa atagtttcag ctgctctgag aagaccaggt ccaaataagc 900

acagactgtc cagcaccact ggtctagctc tctggtctac ctaaggaaga agtcagttct 960

acccttgtaa gatggtggct gatcacaatg tggaacttga gccaagtcat agaatttgag 1020

actgaggggt agaaagcaaa acaacagcag tgttcaaaca atcctaggca taaccagggc 1080

cattgtccaa aggacagatg ctttgggaac actaagagat gacaagtgac ctcaagcaga 1140

agttaatgaa acaggaacca tttgcttcac ttccccaagc tcagcctata acagaaatca 1200

actacgttta tcagtaaagg ctaaatggcc ttgtgggcca tgtggggtct ctgttgtaac 1260

tcctgcactc tactatttta acatgaaagc agccacagtt aacaaacaaa aggttggtca 1320

gatatgactc gtgaaccata gtttgccagc ccttgatcta gaagaatcgt cacactttag 1380

agcctaaaga atctagataa atcctatcat tctcagatgg gaaatctggg acccagtgag 1440

ggccatgact cacccagggt caaagatcaa gtggatgtct ccctaccctt aacttctggt 1500

agcttcctca atgttctttg atagatttaa gaaatagatg gttaagcaag tagaccccag 1560

aggctgtatc taagacctct ttccccaatc tttcatgttt ggaggggcca ctctgaaggc 1620

gggatccaat gggacacagc tgtcctggga tcaggaaaga gaggttttct aagccatttc 1680

tgcttcgcca ggtgttccct cagagtcagg ccatcttcct gtgttctggc cctaccatga 1740

acaaactgtg gggcatgggg caagtcattt ctctcttggc ttcaacttag tgatctgcaa 1800

taaggcgaga ctgaactaaa aagcccccca aatctcttct ggctgtaaca tcctgtgact 1860

taatcaattc ctggccatga aacaagttaa tgagtctgtc cttcgttgct gaagagaaag 1920

cacctcagag ttgtttgtct ggtgtctcca gaagggtccc agcaatcctt gccaagatga 1980

tccctgtgca gataattcgt tatgtgttaa gctgcataat acaagttttt gcctgtgttt 2040

agaagggtat tactacaact cttctacatg taagaaagga aaggtattcc ctgggaagat 2100

ttcagtgaca gtatcagaaa catttgaccc agaagagaaa cattccatgg cctatcaaga 2160

cttgcatagt gaaattacta gcttgtttaa agatgtattt ggcacatctg tttatggaca 2220

gactgtaatt cttactgtaa gcacatctct gtcaccaaga tctgaaatgc gtgctgatga 2280

caagtttgtt aatgtaacaa tagtaacaat tttggcagaa accacaagtg acaatgagaa 2340

gactgtgact gagaaaatta ataaagcaat tagaagtagc tcaagcaact ttctaaacta 2400

tgatttgacc cttcggtgtg attattatgg ctgtaaccag actgcggatg actgcctcaa 2460

tggtttagca tgcgattgca aatctgacct gcaaaggcct aacccacaga gccctttctg 2520

cgttgcttcc agtctcaagt gtcctgatgc ctgcaacgca cagcacaagc aatgcttaat 2580

aaagaagagt ggtggggccc ctgagtgtgc gtgcgtgccc ggctaccagg aagatgctaa 2640

tgggaactgc caaaagtgtg catttggcta cagtggactc gactgtaagg acaaatttca 2700

gctgatcctc actattgtgg gcaccatcgc tggcattgtc attctcagca tgataattgc 2760

attgattgtc acagcaagat caaataacaa aacgaagcat attgaagaag agaacttgat 2820

tgacgaagac tttcaaaatc taaaactgcg gtcgacaggc ttcaccaatc ttggagcaga 2880

-continued

agggagcgtc tttcctaagg tcaggataac ggcctccaga gacagccaga tgcaaaatcc 2940

ctattcaaga cacagcagca tgccccgccc tgactattag aatcataaga atgtggaacc 3000

cgccatggcc cccaaccaat gtacaagcta ttatttagag tgtttagaaa gactgatgga 3060

gaagtgagca ccagtaaaga tctggcctcc ggggtttttc ttccatctga catctgccag 3120

cctctctgaa tggaagttgt gaatgtttgc aacgaatcca gctcacttgc taaataagaa 3180

tctatgacat taaatgtagt agatgctatt agcgcttgtc agagaggtgg ttttcttcaa 3240

tcagtacaaa gtactgagac aatggttagg gttgttttct taattctttt cctggtaggg 3300

caacaagaac catttccaat ctagaggaaa gctccccagc attgcttgct cctgggcaaa 3360

cattgctctt gagttaagtg acctaattcc cctgggagac atacgcatca actgtggagg 3420

tccgagggga tgagaaggga tacccaccac ctttcaaggg tcacaagctc actctctgac 3480

aagtcagaat agggacactg cttctatccc tccaatggag agattctggc aacctttgaa 3540

cagcccagag cttgcaacct agcctcaccc aagaagactg gaaagagaca tatctctcag 3600

ctttttcagg aggcgtgcct gggaatccag gaacttttg atgctaatta gaaggcctgg 3660

actaaaaatg tccactatgg ggtgcactct acagtttttg aaatgctagg aggcagaagg 3720

ggcagagagt aaaaaacatg acctggtaga aggaagagag gcaaaggaaa ctgggtgggg 3780

aggatcaatt agagaggagg cacctgggat ccaccttctt ccttaggtcc cctcctccat 3840

cagcaaagga gcacttctct aatcatgccc tcccgaagac tggctgggag aaggtttaaa 3900

aacaaaaaat ccaggagtaa gagccttagg tcagtttgaa attggagaca aactgtctgg 3960

caaagggtgc gagagggagc ttgtgctcag gagtccagcc gtccagcctc ggggtgtagg 4020

tttctgaggt gtgccattgg ggcctcagcc ttctctggtg acagaggctc agctgtggcc 4080

accaacacac aaccacacac acacaaccac acacacaaat gggggcaacc acatccagta 4140

caagctttta caaatgttat tagtgtcctt ttttatttct aatgccttgt cctcttaaaa 4200

gttattttat ttgttattat tatttgttct tgactgttaa ttgtgaatgg taatgcaata 4260

aagtgccttt gttagatggt gaaa 4284

<210> SEQ ID NO 83
<211> LENGTH: 203
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 83

Met Lys Ala Ile Ile His Leu Thr Leu Leu Ala Leu Leu Ser Val Asn
1 5 10 15

Thr Ala Thr Asn Gln Gly Asn Ser Ala Asp Ala Val Thr Thr Thr Glu
20 25 30

Thr Ala Thr Ser Gly Pro Thr Val Ala Ala Ala Asp Thr Thr Glu Thr
35 40 45

Asn Phe Pro Glu Thr Ala Ser Thr Thr Ala Asn Thr Pro Ser Phe Pro
50 55 60

Thr Ala Thr Ser Pro Ala Pro Pro Ile Ile Ser Thr His Ser Ser Ser
65 70 75 80

Thr Ile Pro Thr Pro Ala Pro Pro Ile Ile Ser Thr His Ser Ser Ser
85 90 95

Thr Ile Pro Ile Pro Thr Ala Ala Asp Ser Glu Ser Thr Thr Asn Val

-continued

```
            100                 105                 110

Asn Ser Leu Ala Thr Ser Asp Ile Ile Thr Ala Ser Ser Pro Asn Asp
        115                 120                 125

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

Ser Pro Thr Thr Glu Asp Asn Gln Ser Ser Gly Pro Pro Thr Gly Thr
145                 150                 155                 160

Ala Leu Leu Glu Thr Ser Thr Leu Asn Ser Thr Gly Lys Asp Asn Ser
                165                 170                 175

Leu Lys Asp Ser Pro Arg Met Glu Ser Thr Ile Val Glu Gly Gly Arg
            180                 185                 190

Trp Val Cys Ser Gly Gly Val Gly Lys Trp Val
        195                 200


<210> SEQ ID NO 84
<211> LENGTH: 32
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 84

Gly Lys Asp Asn Ser Leu Lys Asp Ser Pro Arg Met Glu Ser Thr Ile
1               5                   10                  15

Val Glu Gly Gly Arg Trp Val Cys Ser Gly Gly Val Gly Lys Trp Val
            20                  25                  30


<210> SEQ ID NO 85
<211> LENGTH: 171
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 85

gccagagtgc tggaattcgc ccatcctcat cgataaggtc atctccacca tcaccaacaa     60

catccagcag atcattgaga tcgaggacac ctytgagacc cttcgggacg ggggtctgca    120

agaaaatgat gttcccacat agttggcagc acgtgaacag caattgatcc c              171


<210> SEQ ID NO 86
<211> LENGTH: 2691
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 86

gcttgcccgt cggtcgctag ctcgctcggt gcgcgtcgtc ccgctccatg gcgctcttcg     60

tgcggctgct ggctctcgcc ctggctctgg ccctgggccc cgccgcgacc ctggcgggtc    120

ccgccaagtc gccctaccag ctggtgctgc agcacagcag gctccggggc cgccagcacg    180

gccccaacgt gtgtgctgtg cagaaggtta ttggcactaa taggaagtac ttcaccaact    240

gcaagcagtg gtaccaaagg aaaatctgtg gcaaatcaac agtcatcagc tacgagtgct    300

gtcctggata tgaaaaggtc cctggggaga agggctgtcc agcagcccta ccactctcaa    360

acctttacga gaccctggga gtcgttggat ccaccaccac tcagctgtac acggaccgca    420

cggagaagct gaggcctgag atggaggggc ccggcagctt caccatcttc gcccctagca    480

acgaggcctg ggcctccttg ccagctgaag tgctggactc cctggtcagc aatgtcaaca    540

ttgagctgct caatgccctc cgctaccata tggtgggcag gcgagtcctg actgatgagc    600
```

-continued

tgaaacacgg catgaccctc acctctatgt accagaattc caacatccag atccaccact 660

atcctaatgg gattgtaact gtgaactgtg cccggctcct gaaagccgac caccatgcaa 720

ccaacggggt ggtgcacctc atcgataagg tcatctccac catcaccaac aacatccagc 780

agatcattga gatcgaggac acctttgaga cccttcgggc tgctgtggct gcatcagggc 840

tcaacacgat gcttgaaggt aacggccagt acacgctttt ggccccgacc aatgaggcct 900

tcgagaagat ccctagtgag actttgaacc gtatcctggg cgacccagaa gccctgagag 960

acctgctgaa caaccacatc ttgaagtcag ctatgtgtgc tgaagccatc gttgcggggc 1020

tgtctgtaga gaccctggag ggcacgacac tggaggtggg ctgcagcggg gacatgctca 1080

ctatcaacgg gaaggcgatc atctccaata aagacatcct agccaccaac ggggtgatcc 1140

actacattga tgagctactc atcccagact cagccaagac actatttgaa ttggctgcag 1200

agtctgatgt gtccacagcc attgaccttt tcagacaagc cggcctcggc aatcatctct 1260

ctggaagtga gcggttgacc ctcctggctc ccctgaattc tgtattcaaa gatggaaccc 1320

ctccaattga tgcccataca aggaatttgc ttcggaacca cataattaaa gaccagctgg 1380

cctctaagta tctgtaccat ggacagaccc tggaaactct gggcggcaaa aaactgagag 1440

tttttgttta tcgtaatagc ctctgcattg agaacagctg catcgcggcc cacgacaaga 1500

gggggaggta cgggaccctg ttcacgatgg accgggtgct gacccccca atggggactg 1560

tcatggatgt cctgaaggga gacaatcgct ttagcatgct ggtagctgcc atccagtctg 1620

caggactgac ggagaccctc aaccgggaag gagtctacac agtctttgct cccacaaatg 1680

aagccttccg agccctgcca ccaagagaac ggagcagact cttgggagat gccaaggaac 1740

ttgccaacat cctgaaatac cacattggtg atgaaatcct ggttagcgga ggcatcgggg 1800

ccctggtgcg gctaaagtct ctccaaggtg acaagctgga agtcagcttg aaaaacaatg 1860

tggtgagtgt caacaaggag cctgttgccg agcctgacat catggccaca aatggcgtgg 1920

tccatgtcat caccaatgtt ctgcagcctc cagccaacag acctcaggaa agaggggatg 1980

aacttgcaga ctctgcgctt gagatcttca aacaagcatc agcgttttcc agggcttccc 2040

agaggtctgt gcgactagcc cctgtctatc aaaagttatt agagaggatg aagcattagc 2100

ttgaagcact acaggaggaa tgcaccacgg cagctctccg ccaatttctc tcagatttcc 2160

acagagactg tttgaatgtt ttcaaaacca agtatcacac tttaatgtac atgggccgca 2220

ccataatgag atgtgagcct tgtgcatgtg ggggaggagg gagagagatg tacttttaa 2280

atcatgttcc ccctaaacat ggctgttaac ccactgcatg cagaaacttg gatgtcactg 2340

cctgacattc acttccagag aggacctatc ccaaatgtgg aattgactgc ctatgccaag 2400

tccctggaaa aggagcttca gtattgtggg gctcataaaa catgaatcaa gcaatccagc 2460

ctcatgggaa gtcctggcac agtttttgta aagcccttgc acagctggag aaatggcatc 2520

attataagct atgagttgaa atgttctgtc aaatgtgtct cacatctaca cgtggcttgg 2580

aggcttttat ggggccctgt ccaggtagaa aagaaatggt atgtagagct tagatttccc 2640

tattgtgaca gagccatggt gtgtttgtaa taataaaacc aaagaaacat a 2691

<210> SEQ ID NO 87
<211> LENGTH: 683
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

-continued

<400> SEQUENCE: 87

```
Met Ala Leu Phe Val Arg Leu Leu Ala Leu Ala Leu Ala Leu Ala Leu
1               5                  10                  15

Gly Pro Ala Ala Thr Leu Ala Gly Pro Ala Lys Ser Pro Tyr Gln Leu
            20                  25                  30

Val Leu Gln His Ser Arg Leu Arg Gly Arg Gln His Gly Pro Asn Val
        35                  40                  45

Cys Ala Val Gln Lys Val Ile Gly Thr Asn Arg Lys Tyr Phe Thr Asn
    50                  55                  60

Cys Lys Gln Trp Tyr Gln Arg Lys Ile Cys Gly Lys Ser Thr Val Ile
65                  70                  75                  80

Ser Tyr Glu Cys Cys Pro Gly Tyr Glu Lys Val Pro Gly Glu Lys Gly
                85                  90                  95

Cys Pro Ala Ala Leu Pro Leu Ser Asn Leu Tyr Glu Thr Leu Gly Val
            100                 105                 110

Val Gly Ser Thr Thr Thr Gln Leu Tyr Thr Asp Arg Thr Glu Lys Leu
        115                 120                 125

Arg Pro Glu Met Glu Gly Pro Gly Ser Phe Thr Ile Phe Ala Pro Ser
    130                 135                 140

Asn Glu Ala Trp Ala Ser Leu Pro Ala Glu Val Leu Asp Ser Leu Val
145                 150                 155                 160

Ser Asn Val Asn Ile Glu Leu Leu Asn Ala Leu Arg Tyr His Met Val
                165                 170                 175

Gly Arg Arg Val Leu Thr Asp Glu Leu Lys His Gly Met Thr Leu Thr
            180                 185                 190

Ser Met Tyr Gln Asn Ser Asn Ile Gln Ile His His Tyr Pro Asn Gly
        195                 200                 205

Ile Val Thr Val Asn Cys Ala Arg Leu Leu Lys Ala Asp His His Ala
    210                 215                 220

Thr Asn Gly Val Val His Leu Ile Asp Lys Val Ile Ser Thr Ile Thr
225                 230                 235                 240

Asn Asn Ile Gln Gln Ile Ile Glu Ile Glu Asp Thr Phe Glu Thr Leu
                245                 250                 255

Arg Ala Ala Val Ala Ala Ser Gly Leu Asn Thr Met Leu Glu Gly Asn
            260                 265                 270

Gly Gln Tyr Thr Leu Leu Ala Pro Thr Asn Glu Ala Phe Glu Lys Ile
        275                 280                 285

Pro Ser Glu Thr Leu Asn Arg Ile Leu Gly Asp Pro Glu Ala Leu Arg
    290                 295                 300

Asp Leu Leu Asn Asn His Ile Leu Lys Ser Ala Met Cys Ala Glu Ala
305                 310                 315                 320

Ile Val Ala Gly Leu Ser Val Glu Thr Leu Glu Gly Thr Thr Leu Glu
                325                 330                 335

Val Gly Cys Ser Gly Asp Met Leu Thr Ile Asn Gly Lys Ala Ile Ile
            340                 345                 350

Ser Asn Lys Asp Ile Leu Ala Thr Asn Gly Val Ile His Tyr Ile Asp
        355                 360                 365

Glu Leu Leu Ile Pro Asp Ser Ala Lys Thr Leu Phe Glu Leu Ala Ala
    370                 375                 380

Glu Ser Asp Val Ser Thr Ala Ile Asp Leu Phe Arg Gln Ala Gly Leu
385                 390                 395                 400
```

-continued

```
Gly Asn His Leu Ser Gly Ser Glu Arg Leu Thr Leu Leu Ala Pro Leu
                405                 410                 415

Asn Ser Val Phe Lys Asp Gly Thr Pro Pro Ile Asp Ala His Thr Arg
            420                 425                 430

Asn Leu Leu Arg Asn His Ile Ile Lys Asp Gln Leu Ala Ser Lys Tyr
        435                 440                 445

Leu Tyr His Gly Gln Thr Leu Glu Thr Leu Gly Gly Lys Lys Leu Arg
    450                 455                 460

Val Phe Val Tyr Arg Asn Ser Leu Cys Ile Glu Asn Ser Cys Ile Ala
465                 470                 475                 480

Ala His Asp Lys Arg Gly Arg Tyr Gly Thr Leu Phe Thr Met Asp Arg
                485                 490                 495

Val Leu Thr Pro Pro Met Gly Thr Val Met Asp Val Leu Lys Gly Asp
            500                 505                 510

Asn Arg Phe Ser Met Leu Val Ala Ala Ile Gln Ser Ala Gly Leu Thr
        515                 520                 525

Glu Thr Leu Asn Arg Glu Gly Val Tyr Thr Val Phe Ala Pro Thr Asn
    530                 535                 540

Glu Ala Phe Arg Ala Leu Pro Pro Arg Glu Arg Ser Arg Leu Leu Gly
545                 550                 555                 560

Asp Ala Lys Glu Leu Ala Asn Ile Leu Lys Tyr His Ile Gly Asp Glu
                565                 570                 575

Ile Leu Val Ser Gly Gly Ile Gly Ala Leu Val Arg Leu Lys Ser Leu
            580                 585                 590

Gln Gly Asp Lys Leu Glu Val Ser Leu Lys Asn Asn Val Val Ser Val
        595                 600                 605

Asn Lys Glu Pro Val Ala Glu Pro Asp Ile Met Ala Thr Asn Gly Val
    610                 615                 620

Val His Val Ile Thr Asn Val Leu Gln Pro Pro Ala Asn Arg Pro Gln
625                 630                 635                 640

Glu Arg Gly Asp Glu Leu Ala Asp Ser Ala Leu Glu Ile Phe Lys Gln
                645                 650                 655

Ala Ser Ala Phe Ser Arg Ala Ser Gln Arg Ser Val Arg Leu Ala Pro
            660                 665                 670

Val Tyr Gln Lys Leu Leu Glu Arg Met Lys His
        675                 680
```

<210> SEQ ID NO 88
<211> LENGTH: 2596
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 88

```
cttgcccgtc ggtcgctagc tcgctcggtg cgcgtcgtcc cgctccatgg cgctcttcgt     60

gcggctgctg gctctcgccc tggctctggc cctgggcccc gccgcgaccc tggcgggtcc    120

cgccaagtcg ccctaccagc tggtgctgca gcacagcagg ctccggggcc gccagcacgg    180

ccccaacgtg tgtgctgtgc agaaggttat tggcactaat aggaagtact tcaccaactg    240

caagcagtgg taccaaagga aaatctgtgg caaatcaaca gtcatcagct acgagtgctg    300

tcctggatat gaaaaggtcc ctggggagaa gggctgtcca gcagccctac cactctcaaa    360

cctttacgag accctgggag tcgttggatc caccaccact cagctgtaca cggaccgcac    420
```

-continued

ggagaagctg aggcctgaga tggaggggcc cggcagcttc accatcttcg cccctagcaa 480

cgaggcctgg gcctccttgc cagctgaagt gctggactcc ctggtcagca atgtcaacat 540

tgagctgctc aatgccctcc gctaccatat ggtgggcagg cgagtcctga ctgatgagct 600

gaaacacggc atgaccctca cctctatgta ccagaattcc aacatccaga tccaccacta 660

tcctaatggg attgtaactg tgaactgtgc ccggctcctg aaagccgacc accatgcaac 720

caacggggtg gtgcacctca tcgataaggt catctccacc atcaccaaca acatccagca 780

gatcattgag atcgaggaca cctttgagac ccttcgggac gggggtctgc aagaaaatga 840

tgttcccaca tagttggcag cacgtgaaca gcaattgatc cctttgcatc acctcctctt 900

actgtttaga tttggctgct gtggctgcat cagggctcaa cacgatgctt gaaggtaacg 960

gccagtacac gcttttggcc ccgaccaatg aggccttcga gaagatccct agtgagactt 1020

tgaaccgtat cctgggcgac ccagaagccc tgagagacct gctgaacaac cacatcttga 1080

agtcagctat gtgtgctgaa gccatcgttg cggggctgtc tgtagagacc ctggagggca 1140

cgacactgga ggtgggctgc agcggggaca tgctcactat caacgggaag gcgatcatct 1200

ccaataaaga catcctagcc accaacgggg tgatccacta cattgatgag ctactcatcc 1260

cagactcagc caagacacta tttgaattgg ctgcagagtc tgatgtgtcc acagccattg 1320

acctttttcag acaagccggc ctcggcaatc atctctctgg aagtgagcgg ttgaccctcc 1380

tggctcccct gaattctgta ttcaaagatg gaacccctcc aattgatgcc catacaagga 1440

atttgcttcg gaaccacata attaaagacc agctggcctc taagtatctg taccatggac 1500

agaccctgga aactctgggc ggcaaaaaac tgagagtttt tgtttatcgt aatagcctct 1560

gcattgagaa cagctgcatc gcggcccacg acaagagggg gaggtacggg accctgttca 1620

cgatggaccg ggtgctgacc cccccaatgg ggactgtcat ggatgtcctg aagggagaca 1680

atcgctttag catgctggta gctgccatcc agtctgcagg actgacggag accctcaacc 1740

gggaaggagt ctacacagtc tttgctccca caaatgaagc cttccgagcc ctgccaccaa 1800

gagaacggag cagactcttg ggagatgcca aggaacttgc caacatcctg aaataccaca 1860

ttggtgatga aatcctggtt agcggaggca tcggggccct ggtgcggcta aagtctctcc 1920

aaggtgacaa gctggaagtc agcttgaaaa acaatgtggt gagtgtcaac aaggagcctg 1980

ttgccgagcc tgacatcatg gccacaaatg gcgtggtcca tgtcatcacc aatgttctgc 2040

agcctccagc caacagacct caggaaagag gggatgaact tgcagactct gcgcttgaga 2100

tcttcaaaca agcatcagcg ttttccaggg cttcccagag gtctgtgcga ctagcccctg 2160

tctatcaaaa gttattagag aggatgaagc attagcttga agcactacag gaggaatgca 2220

ccacggcagc tctccgccaa tttctctcag atttccacag agactgtttg aatgttttca 2280

aaaccaagta tcacacttta atgtacatgg gccgcaccat aatgagatgt gagccttgtg 2340

catgtggggg aggagggaga gagatgtact ttttaaatca tgttcccccc aaacatggct 2400

gttaacccac tgcatgcaga aacttggatg tcactgcctg acattcactt ccagagagga 2460

cctatcccaa atgtggaatt gactgcctat gccaagtccc tggaaaagga gcttcagtat 2520

tgtggggctc ataaaacatg aatcaagcaa tccagcctca tgggaagtcc tggcacagtt 2580

tttgtaaagc ccttgc 2596

-continued

<210> SEQ ID NO 89
<211> LENGTH: 268
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 89

Met Ala Leu Phe Val Arg Leu Leu Ala Leu Ala Leu Ala Leu Ala Leu
1 5 10 15

Gly Pro Ala Ala Thr Leu Ala Gly Pro Ala Lys Ser Pro Tyr Gln Leu
20 25 30

Val Leu Gln His Ser Arg Leu Arg Gly Arg Gln His Gly Pro Asn Val
35 40 45

Cys Ala Val Gln Lys Val Ile Gly Thr Asn Arg Lys Tyr Phe Thr Asn
50 55 60

Cys Lys Gln Trp Tyr Gln Arg Lys Ile Cys Gly Lys Ser Thr Val Ile
65 70 75 80

Ser Tyr Glu Cys Cys Pro Gly Tyr Glu Lys Val Pro Gly Glu Lys Gly
85 90 95

Cys Pro Ala Ala Leu Pro Leu Ser Asn Leu Tyr Glu Thr Leu Gly Val
100 105 110

Val Gly Ser Thr Thr Thr Gln Leu Tyr Thr Asp Arg Thr Glu Lys Leu
115 120 125

Arg Pro Glu Met Glu Gly Pro Gly Ser Phe Thr Ile Phe Ala Pro Ser
130 135 140

Asn Glu Ala Trp Ala Ser Leu Pro Ala Glu Val Leu Asp Ser Leu Val
145 150 155 160

Ser Asn Val Asn Ile Glu Leu Leu Asn Ala Leu Arg Tyr His Met Val
165 170 175

Gly Arg Arg Val Leu Thr Asp Glu Leu Lys His Gly Met Thr Leu Thr
180 185 190

Ser Met Tyr Gln Asn Ser Asn Ile Gln Ile His His Tyr Pro Asn Gly
195 200 205

Ile Val Thr Val Asn Cys Ala Arg Leu Leu Lys Ala Asp His His Ala
210 215 220

Thr Asn Gly Val Val His Leu Ile Asp Lys Val Ile Ser Thr Ile Thr
225 230 235 240

Asn Asn Ile Gln Gln Ile Ile Glu Ile Glu Asp Thr Phe Glu Thr Leu
245 250 255

Arg Asp Gly Gly Leu Gln Glu Asn Asp Val Pro Thr
260 265

<210> SEQ ID NO 90
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 90

Asp Gly Gly Leu Gln Glu Asn Asp Val Pro Thr
1 5 10

<210> SEQ ID NO 91
<211> LENGTH: 177
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 91

-continued

tagctgacgt tctctcagca cgttcgctgc gaatgccggc ctctgcggga gaagatgaag 60

ccggaaaggt gcggcgatgc tgttccccgg aggtaaccca ccccttggag gagagagacc 120

ccgcacccgg ctcgtgtatt tattaccgtc acactcttca gtgactcctg ctggtac 177

<210> SEQ ID NO 92
<211> LENGTH: 1733
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 92

ctgctgtctg cggaggaaac tgcatcgacg gacggccgcc cagctacggg aggacctgga 60

gtggcactgg gcgcccgacg gaccatcccc gggacccgcc tgcccctcgg cgccccgccc 120

cgccgggccg ctccccgtcg ggttccccag ccacagcctt acctacgggc tcctgactcc 180

gcaaggcttc cagaagatgc tcgaaccacc ggccggggcc tcggggcagc agtgagggag 240

gcgtccagcc ccccactcag ctcttctcct cctgtgccag gggctccccg ggggatgagc 300

atggtggttt tccctcggag ccccctggct cgggacgtct gagaagatgc cggtcatgag 360

gctgttccct tgcttcctgc agctcctggc cgggctggcg ctgcctgctg tgcccccca 420

gcagtgggcc ttgtctgctg ggaacggctc gtcagaggtg gaagtggtac ccttccagga 480

agtgtggggc cgcagctact gccgggcgct ggagaggctg gtggacgtcg tgtccgagta 540

ccccagcgag gtggagcaca tgttcagccc atcctgtgtc tccctgctgc gctgcaccgg 600

ctgctgcggc gatgagaatc tgcactgtgt gccggtggag acggccaatg tcaccatgca 660

gctcctaaag atccgttctg gggaccggcc ctcctacgtg gagctgacgt tctctcagca 720

cgttcgctgc gaatgccggc ctctgcggga gaagatgaag ccggaaagga ggagacccaa 780

gggcaggggg aagaggagga gagagaagca gagacccaca gactgccacc tgtgcggcga 840

tgctgttccc cggaggtaac ccacccctg gaggagagag accccgcacc cggctcgtgt 900

atttattacc gtcacactct tcagtgactc ctgctggtac ctgccctcta tttattagcc 960

aactgtttcc ctgctgaatg cctcgctccc ttcaagacga ggggcaggga aggacaggac 1020

cctcaggaat tcagtgcctt caacaacgtg agagaaagag agaagccagc cacagacccc 1080

tgggagcttc cgctttgaaa gaagcaagac acgtggcctc gtgaggggca agctaggccc 1140

cagaggccct ggaggtctcc aggggcctgc agaaggaaag aaggggccc tgctacctgt 1200

tcttgggcct caggctctgc acagtcaagc agcccttgct ttcggagctc ctgtccaaaa 1260

gtagggatgc ggatcctgct ggggccgcca cggcctggct ggtgggaagg ccggcagcgg 1320

gcggagggga tccagccact tccccctctt cttctgaaga tcagaacatt cagctctgga 1380

gaacagtggt tgcctggggg cttttgccac tccttgtccc ccgtgatctc ccctcacact 1440

ttgccatttg cttgtactgg gacattgttc tttccggcca aggtgccacc accctgcccc 1500

ccctaagaga cacatacaga gtgggccccg ggctggagaa agagctgcct ggatgagaaa 1560

cagctcagcc agtggggatg aggtcaccag gggaggagcc tgtgcgtccc agctgaaggc 1620

agtggcaggg gagcaggttc cccaagggcc ctggcacccc cacaagctgt ccctgcaggg 1680

ccatctgact gccaagccag attctcttga ataaagtatt ctagtgtgga aaa 1733

<210> SEQ ID NO 93
<211> LENGTH: 1645

-continued

<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 93

gggattcggg ccgcccagct acgggaggac ctggagtggc actgggcgcc cgacggacca 60

tccccgggac ccgcctgccc ctcggcgccc cgccccgccg ggccgctccc cgtcgggttc 120

cccagccaca gccttaccta cgggctcctg actccgcaag gcttccagaa gatgctcgaa 180

ccaccggccg gggcctcggg gcagcagtga gggaggcgtc cagcccccca ctcagctctt 240

ctcctcctgt gccaggggct ccccggggga tgagcatggt ggttttccct cggagccccc 300

tggctcggga cgtctgagaa gatgccggtc atgaggctgt tcccttgctt cctgcagctc 360

ctggccgggc tggcgctgcc tgctgtgccc ccccagcagt gggccttgtc tgctgggaac 420

ggctcgtcag aggtggaagt ggtacccttc caggaagtgt ggggccgcag ctactgccgg 480

gcgctggaga ggctggtgga cgtcgtgtcc gagtacccca gcgaggtgga gcacatgttc 540

agcccatcct gtgtctccct gctgcgctgc accggctgct gcggcgatga gaatctgcac 600

tgtgtgccgg tggagacggc caatgtcacc atgcagctcc taaagatccg ttctggggac 660

cggccctcct acgtggagct gacgttctct cagcacgttc gctgcgaatg ccggcctctg 720

cgggagaaga tgaagccgga aaggtgcggc gatgctgttc cccggaggta acccacccct 780

tggaggagag agaccccgca cccggctcgt gtatttatta ccgtcacact cttcagtgac 840

tcctgctggt acctgccctc tatttattag ccaactgttt ccctgctgaa tgcctcgctc 900

ccttcaagac gaggggcagg gaaggacagg accctcagga attcagtgcc ttcaacaacg 960

tgagagaaag agagaagcca gccacagacc cctgggagct tccgctttga aagaagcaag 1020

acacgtggcc tcgtgagggg caagctaggc cccagaggcc ctggaggtct ccaggggcct 1080

gcagaaggaa agaaggggcc cctgctacct gttcttgggc ctcaggctct gcacagacaa 1140

gcagcccttg ctttcggagc tcctgtccaa agtagggatg cggattctgc tggggccgcc 1200

acggcctggt ggtgggaagg ccggcagcgg gcggagggga ttcagccact tccccctctt 1260

cttctgaaga tcagaacatt cagctctgga gaacagtggt tgcctggggg cttttgccac 1320

tccttgtccc ccgtgatctc ccctcacact ttgccatttg cttgtactgg gacattgttc 1380

tttccggccg aggtgccacc accctgcccc cactaagaga cacatacaga gtgggccccg 1440

ggctggagaa agagctgcct ggatgagaaa cagctcagcc agtggggatg aggtcaccag 1500

gggaggagcc tgtgcgtccc agctgaaggc agtggcaggg gagcaggttc cccaagggcc 1560

ctggcacccc cacaagctgt ccctgcaggg ccatctgact gccaagccag attctcttga 1620

ataaagtatt ctagtgtgga aacgc 1645

<210> SEQ ID NO 94
<211> LENGTH: 149
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 94

Met Pro Val Met Arg Leu Phe Pro Cys Phe Leu Gln Leu Leu Ala Gly
1 5 10 15

Leu Ala Leu Pro Ala Val Pro Pro Gln Gln Trp Ala Leu Ser Ala Gly
20 25 30

Asn Gly Ser Ser Glu Val Glu Val Val Pro Phe Gln Glu Val Trp Gly

-continued

```
        35                  40                  45

Arg Ser Tyr Cys Arg Ala Leu Glu Arg Leu Val Asp Val Val Ser Glu
    50                  55                  60

Tyr Pro Ser Glu Val Glu His Met Phe Ser Pro Ser Cys Val Ser Leu
65                  70                  75                  80

Leu Arg Cys Thr Gly Cys Cys Gly Asp Glu Asn Leu His Cys Val Pro
                85                  90                  95

Val Glu Thr Ala Asn Val Thr Met Gln Leu Leu Lys Ile Arg Ser Gly
            100                 105                 110

Asp Arg Pro Ser Tyr Val Glu Leu Thr Phe Ser Gln His Val Arg Cys
        115                 120                 125

Glu Cys Arg Pro Leu Arg Glu Lys Met Lys Pro Glu Arg Cys Gly Asp
    130                 135                 140

Ala Val Pro Arg Arg
145
```

<210> SEQ ID NO 95
<211> LENGTH: 1529
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1358)..(1358)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 95

```
gcacgagttg ggaggtgtag cgcggctctg aacgcgctga gggccgttga gtgtcgcagg     60

cggcgagggc gcgagtgagg agcagaccca ggcatcgcgc gccgagaagg ccgggcgtcc    120

ccacactgaa ggtccggaaa ggcgacttcc gggggctttg gcacctggcg gaccctcccg    180

gagcgtcggc acctgaacgc gaggcgctcc attgcgcgtg cgcgttgagg ggcttcccgc    240

acctgatcgc gagaccccaa cggctggtgg cgtcgcctgc gcgtctcggc tgagctggcc    300

atggcgcagc tgtgcgggct gaggcggagc cgggcgtttc tcgccctgct gggatcgctg    360

ctcctctctg gggtcctggc ggccgaccga gaacgcagca tccacgactt ctgcctggtg    420

tcgaaggtgg tgggcagatg ccgggcctcc atgcctaggt ggtggtacaa tgtcactgac    480

ggatcctgcc agctgtttgt gtatgggggc tgtgacggaa acagcaataa ttacctgacc    540

aaggaggagt gcctcaagaa atgtgccact gtcacagaga atgccacggg tgacctggcc    600

accagcagga atgcagcgga ttcctctgtc ccaagtgctc ccagaaggca ggattctgaa    660

gaccactcca gcgatatgtt caactatgaa gaatactgca ccgccaacgc agtcactggg    720

ccttgccgtg catccttccc acgctggtac tttgacgtgg agaggaactc ctgcaataac    780

ttcatctatg gaggctgccg gggcaataag aacagctacc gctctgagga ggcctgcatg    840

ctccgctgct tccgccagca ggagaatcct cccctgcccc ttggctcaaa ggtggtggtt    900

ctggcggggc tgttcgtgat ggtgttgatc ctcttcctgg gagcctccat ggtctacctg    960

atccgggtgg cacggaggaa ccaggagcgt gccctgcgca ccgtctggag ctccggagat   1020

gacaaggagc agctggtgaa gaacacatat gtcctgtgac cgccctgtcg ccaagaggac   1080

tggggaaggg aggggagact atgtgtgagc ttttttaaa tagagggatt gactcggatt   1140

tgagtgatca ttagggctga ggtctgtttc tctgggaggt aggacggctg cttcctggtc   1200

tggcagggat gggtttgctt tggaaatcct ctaggaggct cctcctcgca tggcctgcag   1260
```

-continued

tctggcagca gccccgagtt gtttcctcgc tgatcgattt ctttcctcca ggtagagttt 1320

tctttgctta tgttgaattc cattgcctcc ttttctcnat cacagaagtg atgttggaat 1380

cgtttctttt gtttgtctga tttatggttt ttttaagtat aaacaaaagt tttttattag 1440

cattctgaaa gaaggaaagt aaaatgtaca agtttaataa aaaggggcct tccccttag 1500

aataaatttc cagcatgttg ctttcaaaa 1529

<210> SEQ ID NO 96
<211> LENGTH: 639
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 96

atggcgcagc tgtgcgggct gaggcggagc cgggcgtttc tcgccctgct gggatcgctg 60

ctcctctctg gggtcctggc ggccgaccga gaacgcagca tccacgactt ctgcctggtg 120

tcgaaggtgg tgggcagatg ccgggcctcc atgcctaggt ggtggtacaa tgtcactgac 180

ggatcctgcc agctgtttgt gtatgggggc tgtgacggaa acagcaataa ttacctgacc 240

aaggaggagt gcctcaagaa atgtgccact gtcacagaga atgccacggg tgacctggcc 300

accagcagga atgcagcgga ttcctctgtc ccaagtgctc ccagaaggca ggattctgaa 360

gaccactcca gcgatatgtt caactatgaa ggtaaaactc caaagaggcc agaatactgc 420

accgcaacgc agtcactggg ccttgccgtg catccttccc acgctggtac tttgacgtgg 480

agaggaactc ctgcaataac ttcatctatg gaggctgccg gggcaataag aacagctacc 540

gctctgagga ggcctgcatg ctccgctgct tccgccagca ggagaatcct cccctgcccc 600

ttggctcaaa ggtggtggtt ctggcggggc tgttcgtga 639

<210> SEQ ID NO 97
<211> LENGTH: 212
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 97

Met Ala Gln Leu Cys Gly Leu Arg Arg Ser Arg Ala Phe Leu Ala Leu
1 5 10 15

Leu Gly Ser Leu Leu Leu Ser Gly Val Leu Ala Ala Asp Arg Glu Arg
20 25 30

Ser Ile His Asp Phe Cys Leu Val Ser Lys Val Val Gly Arg Cys Arg
35 40 45

Ala Ser Met Pro Arg Trp Trp Tyr Asn Val Thr Asp Gly Ser Cys Gln
50 55 60

Leu Phe Val Tyr Gly Gly Cys Asp Gly Asn Ser Asn Asn Tyr Leu Thr
65 70 75 80

Lys Glu Glu Cys Leu Lys Lys Cys Ala Thr Val Thr Glu Asn Ala Thr
85 90 95

Gly Asp Leu Ala Thr Ser Arg Asn Ala Ala Asp Ser Ser Val Pro Ser
100 105 110

Ala Pro Arg Arg Gln Asp Ser Glu Asp His Ser Ser Asp Met Phe Asn
115 120 125

Tyr Glu Gly Lys Thr Pro Lys Arg Pro Glu Tyr Cys Thr Ala Thr Gln
130 135 140

-continued

```
Ser Leu Gly Leu Ala Val His Pro Ser His Ala Gly Thr Leu Thr Trp
145                 150                 155                 160

Arg Gly Thr Pro Ala Ile Thr Ser Ser Met Glu Ala Ala Gly Ala Ile
                165                 170                 175

Arg Thr Ala Thr Ala Leu Arg Arg Pro Ala Cys Ser Ala Ala Ser Ala
            180                 185                 190

Ser Arg Arg Ile Leu Pro Cys Pro Leu Ala Gln Arg Trp Trp Phe Trp
        195                 200                 205

Arg Gly Cys Ser
    210


<210> SEQ ID NO 98
<211> LENGTH: 82
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 98

Gly Lys Thr Pro Lys Arg Pro Glu Tyr Cys Thr Ala Thr Gln Ser Leu
1               5                   10                  15

Gly Leu Ala Val His Pro Ser His Ala Gly Thr Leu Thr Trp Arg Gly
            20                  25                  30

Thr Pro Ala Ile Thr Ser Ser Met Glu Ala Ala Gly Ala Ile Arg Thr
        35                  40                  45

Ala Thr Ala Leu Arg Arg Pro Ala Cys Ser Ala Ala Ser Ala Ser Arg
    50                  55                  60

Arg Ile Leu Pro Cys Pro Leu Ala Gln Arg Trp Trp Phe Trp Arg Gly
65                  70                  75                  80

Cys Ser


<210> SEQ ID NO 99
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 99

Arg Glu Lys Met Lys Pro Glu Arg Cys Gly Asp Ala Val Pro Arg Arg
1               5                   10                  15
```

1. An isolated nucleic acid that is expressed by a human cancer cell, selected from the group consisting of:

i) nucleic acids comprising a sequence contained in SEQ NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96;

ii) a nucleic acid having a sequence that is at least 70% identical to the sequence of (i) when aligned without allowing for gaps;

iii) nucleic acids having a sequence complementary to i) or ii); and

iv) fragments of i), ii) or iii) having a size of at least 20 nucleotides in length.

2. A nucleic acid of claim **1**, consisting of a sequence contained in SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91 or 96, or a sequence complementary thereto.

3. A primer mixture that comprises primers that result in the specific amplification of one of the nucleic acids of claim **1**.

4. A polypeptide expressed by a human cancer cell, that is selected from the group consisting of:

i) the antigen encoded by a nucleic acid sequence having at least 90% sequence identity in SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96, or a sequence complementary thereto,

ii. a polypeptide comprising an amino acid sequence having at least 90% sequence identity in SEQ ID NOS. 48, 49, 54, 55, 58, 59, 66, 67, 71, 72, 75, 77, 78, 83, 84, 89, 90, 97, 98 and 99, and

iii. an antigenic fragment of (i) or (ii).

5. A tumor antigen, comprising an amino acid sequence selected from the group consisting of SEQ ID NOS. 48, 49, 54, 55, 58, 59, 66, 67, 71, 72, 75, 77, 78, 83, 84, 89, 90, 97, 98 and 99 or an antigenic fragment thereof.

6. A method of detecting and/or staging cancer, comprising determining whether a human cell sample, particularly a human colon cell sample, expresses a target nucleic acid molecule, wherein said target nucleic acid molecule com-

prises the sequence of a gene or RNA comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96; a sequence complementary thereto, or of a fragment of said gene or RNA having a size of at least 20 nucleotides in length.

7. The method of claim **6**, wherein said method comprises detecting the expression of said target nucleic acid molecule using a nucleic acid sequence that specifically hybridizes thereto.

8. The method of claim **6**, wherein said method comprises detecting the expression of said target nucleic acid molecule using oligonucleotides that result in the amplification and/or the detection thereof.

9. The method of claim **6**, wherein the expression of said target nucleic acid molecule is detected by assaying for the antigen encoded by said nucleic acid.

10. The method of claim **9**, wherein said assay involves the use of an antibody or a fragment thereof that specifically binds to said antigen.

11. The method of claim **10**, wherein said assay comprises an ELISA or competitive binding assay.

12. The method of claim **10**, wherein said antigen is a polypeptide as defined in claim **4** or **5**.

13. The method of claim **6**, further comprising comparing the expression level of said target molecule in said cell sample to a reference expression level, wherein a deviation from said reference expression level is indicative of the presence and/or stage of said cancer in said subject.

14. The method of claim **13**, wherein said reference expression level is an expression level as determined in a control sample or a median expression level from healthy subjects.

15. An antibody or antigen-binding fragment thereof that specifically binds to a target polypeptide molecule selected from:

i. a polypeptide encoded by a nucleic acid molecule comprising the sequence of a gene or RNA comprising a sequence selected from the group consisting of SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96; a sequence complementary thereto, or by a fragment of said gene or RNA having a size of at least 20 nucleotides in length,

ii. a polypeptide comprising the sequence of a protein comprising a sequence selected from the group consisting of SEQ ID NOS. 48, 49, 54, 55, 58, 59, 66, 67, 71, 72, 75, 77, 78, 83, 84, 89, 90, 97, 98 and 99; or a fragment of said protein having a size of at least 5 amino acids in length.

iii. an antigen according to claim **4** or **5**, and

v. an antigenic fragment of (i), (ii), or (iii).

16. The antibody of claim **15**, which is a monoclonal antibody or an antigen-binding fragment thereof.

17. The antigen of claim **4** which is attached directly or indirectly to a detectable label.

18. The antibody of claim **15** which is attached directly or indirectly to a detectable label.

19. A diagnostic kit for detection and/or staging of cancer, which comprises a DNA according to claim **1** and a detectable label.

20. A diagnostic kit for detection and/or staging of cancer, which comprises primers according to claim **3** and a diagnostically acceptable carrier.

21. A diagnostic kit for detection and/or staging of cancer, which comprises a monoclonal antibody according to claim **15** and a detectable label.

22. A diagnostic kit in the form of a sandwich ELISA in which at least one of the capture of the detection antibodies comprises a monoclonal antibody according to claim **16**.

23. A method for detecting and/or staging cancer using human fluid, in particular whole blood, serum or plasma, as a sample source, with a diagnostic kit described in claim **19**.

24. A method for treating cancer comprises administering to a human subject in need thereof a therapeutically effective amount of a ligand which specifically binds a target molecule selected from

i. a gene or RNA comprising a sequence selected from the group consisting of SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96; a sequence complementary thereto, a variant thereof or a fragment of said gene or RNA having a size of at least 20 nucleotides in length, and

ii. a protein or polypeptide encoded by a gene or RNA comprising a sequence selected from the group consisting of SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96; a sequence complementary thereto, a variant thereof or a fragment of said gene or RNA having a size of at least 20 nucleotides in length; or

iii. A protein or polypeptide comprising a sequence selected from the group consisting of SEQ ID NOS. 48, 49, 54, 55, 58, 59, 66, 67, 71, 72, 75, 78, 83, 84, 89, 90, 97, 98 and 99; a variant thereof or a fragment of said protein having a size of at least 5 amino acids in length.

25. The method of claim **24**, wherein the ligand is a ribozyme or antisense oligonucleotide that inhibits the expression of a gene comprising a DNA sequence selected from the group consisting of SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96, a sequence complementary thereto and a fragment or variant thereof.

26. The method of claim **24**, wherein the ligand is directly or indirectly attached to an effector moiety.

27. The method of claim **26**, wherein said effector moiety is a therapeutic radiolabel, enzyme, cytotoxin, growth factor, or drug.

28. A method for treating cancer, particularly colon cancer, comprising administering to a subject in need thereof a therapeutically effective amount of an antigen according to claim **4**, and optionally an adjuvant that elicits a humoral or cytotoxic T-lymphocyte response to said antigen.

29. A method for treating cancer, particularly colon cancer, comprising administering to a subject in need thereof a therapeutically effective amount of a ligand which specifically binds to a protein encoded by a gene or RNA comprising a sequence selected from the group consisting of SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96; a sequence complementary thereto or a fragment, or variant there, or a protein sequence selected from the group consisting of SEQ ID NOS. 48, 49, 54, 55, 58, 59, 66, 67, 71, 72, 75, 77, 78, 83, 84, 89, 90, 97, 98 and 99; optionally directly or indirectly attached to a therapeutic effector moiety.

30. The method of claim **29**, wherein said effector moiety is a radiolabel, enzyme, cytotoxin, growth factor, or drug.

31. The method of claim **30** wherein the radiolabel is yttrium or indium.

32. The method of claim **29** wherein said ligand is a monoclonal antibody or fragment thereof.

33. The method of claim **29** wherein said ligand is a small molecule.

34. The method of claim **29** wherein said ligand is a peptide.

35. The method of claim **29**, wherein said ligand binds an extracellular domain of said protein.

36. A molecule, selected from:

i. a polypeptide comprising the sequence of an extra-cellular domain of a protein encoded by a gene or RNA comprising a sequence selected from the group consisting of SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96, or a sequence complementary thereto; and

ii. a polypeptide comprising the sequence of an extracellular domain of a protein sequence selected from the group consisting of SEQ ID NOS. 48, 49, 54, 55, 58, 59, 66, 67, 71, 72, 75, 77, 78, 83, 84, 89, 90, 97, 98 and 99; and

iii. a nucleic acid molecule encoding a polypeptide of (i).

37. The molecule of claim **36**, wherein said polypeptide has 8 to 100 amino acids in length.

38. A method for selecting, identifying, screening, characterizing or optimizing biologically active compounds, comprising contacting a candidate compound with a target molecule and determining whether the candidate compound binds said target molecule, wherein said target molecule is selected from

i. a nucleic acid molecule comprising the sequence of a gene or RNA comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOS. 1-44, 47, 52, 53, 56, 57, 62, 65, 70, 73, 74, 76, 79, 82, 85, 88, 91, and 96 or a sequence complementary thereto;

ii. a fragment of said gene or RNA having a size of at least 20 nucleotides in length, and

iii. a polypeptide encoded by (i) or (ii) and (iv) a amino acid molecule comprising the sequence selected from the group consisting of SEQ ID NOS. 48, 49, 54, 55, 58, 59, 66, 67, 71, 72, 75, 77, 78, 83, 84, 89, 90, 97, 98, and 99.

* * * * *

专利名称(译)	肿瘤特异性基因和变异Rnas及其用途作为癌症治疗和诊断的目标		
公开(公告)号	US20080254031A1	公开(公告)日	2008-10-16
申请号	US11/575032	申请日	2005-09-07
申请(专利权)人(译)	EXONHIT THERAPEUTICS SA		
[标]发明人	MCGOWAN KEVIN KOTRAIAH VINAYAKA BRENNER MICHAEL EINSTEIN RICHARD BRACCO LAURENT		
发明人	MCGOWAN, KEVIN KOTRAIAH, VINAYAKA BRENNER, MICHAEL EINSTEIN, RICHARD BRACCO, LAURENT		
IPC分类号	A61K39/395 C07H21/04 C07K14/47 A61K39/00 C12Q1/68 G01N33/574 A61K38/02 A61P35/00 G01N33/566 G01N33/53 C07K16/32 C07K19/00 A61K38/47 A61K31/7052		
CPC分类号	C07K14/4748		
优先权	60/608076 2004-09-09 US		
外部链接	Espacenet USPTO		

摘要(译)

鉴定了与正常结肠组织和相应蛋白质相比在人结肠肿瘤组织中差异表达的基因和变体RNA。这些基因和相应的抗原是治疗，诊断或预防结肠癌的合适靶标。